THE EMPLOYMENT OF WIVES AND THE INEQUALITY OF FAMILY INCOME*

James A. Sweet, Center for Demography and Ecology, and Institute for Research on Poverty
University of Wisconsin

In his 1960 Census Monograph, Income Distribution in the United States, Herman Miller reports that the incomes of families in which the wife is in the labor force are more evenly distributed than those of families in which the wife is not in the labor force.

	Wife Works	Wife Does Not Work
Percent of Aggregate Money Income Received by Highest:		
5% of families	13%	19%
20% of families	37%	43%
Gini ratio	.29	.38

Miller suggests that since the proportion of wives who are working has increased considerably in recent years, the effect has been to reduce family income inequality.1/ In discussion of the income inequality within urban areas, Wilber Thompson asserts, "The existence of jobs for women acts to reduce inequality (of family income) in that working wives come more proportionately from the lower income groups."2/

In this paper we will describe the effects of the employment of wives on the distribution of family income in the United States, and the possible reasons for the observed effects. We will then look at the trend in family income inequality in relation to the trend in labor force participation of wives, and finally at the effect of the employment of wives on the inequality of income between blacks and whites.

As a statement of logical necessity, Miller's argument supposes that the dispersion of two combined samples is some sort of a weighted average of their separate dispersions. Thus as more wives enter the labor force, their relative weight increases and the dispersion tends to move toward the within class dispersion of families of working wives -- i.e., the dispersion tends to become less. A simple example will demonstrate that this is clearly not necessarily the case. If one combines two samples each with a different mean and zero dispersion, the dispersion of the combined sample is clearly non-zero and may be considerable depending on the difference in means.

Although Miller does not assert that the lower income inequality of families in which the wife is employed is caused by the employment of the wife, it seems to be implicit in his discussion. It is equally plausible to hypothesize that among families in which the wife is employed there is initially less inequality in husband's income, and for that reason, less inequality in family income. Comparisons of the inequality of family income by themselves tell us nothing about the effect of wife's employment on the degree of family income inequality, unless we can demonstrate that there is a comparable degree of inequality to start with before adding in wife's income.

Since the employment of wives is related to a great variety of factors including age, education, color, husband's income, type of residence, and family composition, it is difficult to determine a priori how husband's income inequality would be related to wife's employment. To the extent that the employment of wives is inversely correlated with husband's income, we would expect that husband's income inequality would be less for working wives, simply on the basis of the systematic underrepresentation of families with high income husbands.

On the other hand, the association of wife's employment with the absence of young children in the family has an effect working in the opposite direction. Wives of young men are more likely to have young children, and are thus less likely to work than are wives of older men. Young men are more likely to have low incomes, and less dispersion in income because they consist disproportionately both of men in dead-end jobs and of men at the beginning of careers. As these latter men age, they will experience relatively rapid income increases, while the men in dead-end jobs will have smaller income increases. Thus dispersion in husband's income ought to be related positively to age. Data from the 1960 census indicate that older married men have greater income dispersion than younger men (Table 1).

In Table 2 we present Gini coefficients computed on each of three income measures: family income, husband's income, and family income minus wife's earnings. Our sample consists of 32,521 nonfarm married couples in which the wife is under the age of 60. It was drawn from the 1/1000 sample of the 1960 United States Census. Data are shown separately by color, and by presence or absence of children under 18. In each case the degree of inequality of non-Negro families is very little different for husband's income and family income minus wife's earnings. For Negro families, the contribution of family members other than the wife and husband tends to increase income inequality over the inequality of husband's income alone. This pattern is particularly pronounced for families with no children under 18. For non-Negro families, the effect of the employment of wives is to decrease income inequality somewhat (from 32.9 to 30.9). For the Negro population, the employment of wives has almost no effect on income inequality. The effect of wife's employment on income inequality is greater for families with no children under 18 than for families with children. In the case of the Negro population, the employment of wives with no children under 18 reduces the degree of income inequality from 39.0 to 37.9, while the employment of Negro mothers raises the degree of inequality. Clearly the effect of employment of wives on income inequality is rather small and not invariant in direction.

In Table 3 we show Gini coefficients for the same three income variables computed separately for families in which the wife earned income in 1959 and for those in which she did not. Again the population is disaggregated by color and child status. Among non-Negro families there is less inequality in husband's income and family income minus wife's earnings in families in which the wife is employed (had income in 1959) than in those in which she is not. For the Negro population, the reverse tends to be true -- the husband's income inequality tends to be greater for families in which the wife is employed. Miller's comparisons, then, are clearly distorted by systematic differences in dispersion of husband's income between men whose wives are in and those whose wives are not in the labor force.

Thompson argues that income inequality is reduced when the wife works because wives of low income husbands are more likely to work -- i.e., the bottom end of the distribution of families in terms of husband's income are more likely to have their incomes incremented than the upper end of the distribution. Thompson's argument would be logically valid if there were no variation among working wives in the amount earned. However, working wives of high income husbands tend to receive more income than working wives of low income husbands (see Table 4). The combination of a strong negative relationship between employment and husband's income and positive relationship between the earnings of employed women and husband's income results in only a small differential in the average amount of income per family (irrespective of whether or not the wife is employed) among various levels of husband's income. This means that while families toward the lower end of the husband's income distribution are being disproportionately moved upward in the distribution, the amount by which they move is relatively less than the movements achieved by families with employed wives in the middle and upper end of the husband's income distribution.

Trends

Figure 1 plots the time series of labor force participation rates of married women and Gini coefficients for various income measures as published in the U.S. Census Report, "Trends in the Income of Families and Persons in the United States, 1947-1964."3/ The labor force participation rate has risen rapidly and regularly by almost one percentage point per year. The income inequality measures show little evidence of trend. To the extent that there is a trend of decline in the inequality of family income, it appears to be matched by a similar decline in the inequality of husband's income. It does not appear that the increase in labor force participation of wives that has been occurring over the past two or three decades has had any impact on the level of income inequality.

The effect on family income inequality of any increase in the employment of wives depends on at least three things: (1) The pattern of change in labor force participation of wives in relation to husband's income; (2) The pattern of change in both the mean and the dispersion of wife's earnings in relation to husband's income; and, (3) The change in the shape of the distribution of husbands among income levels. Without attempting to specify exactly what has happened in recent history to each of these relationships, it does appear that the outcome of these changes has been neutral with respect to income inequality.

Income Inequality Between Black and White Families

Negro wives have considerably higher rates of employment than do white wives. A Negro husband-wife family is considerably more likely to have its total income result from the contributions of more than one earner than a white family, but the size of the Negro wife's contribution is on average considerably smaller than that of the white wife's. What effect do these differences have on the inequality of family income between the races?

To summarize the overall effect of differential employment and differential earnings of wives on the inequality or dissimilarity of the income distributions of the Negro and white populations, we have again used the Gini coefficient. Here, rather than comparing the cumulative distributions of families and money income, we are comparing the cumulative distributions of black and white families ordered with respect to incomes. In these comparisons a value of $35,000 was used for the category $25,000 and over. Other reasonable values were tried and produced no major change in results. Again, three separate income measures were used: (1) husband's income, (2) family income minus wife's earnings, and (3) total family income. Differences between 2 and 3 reflect the effect of differential contributions of other income recipients (and wife's non-earnings income) on income inequality between Negroes and non-Negroes.

Overall, the Gini coefficient for husband's income is 54.3 in comparison to a coefficient of 47.3 for total family income (bottom row, Table 5). Quite clearly, then, family income is less inequitably distributed than is husband's income. When family income is compared with family income minus wife's earnings, the differential is very small, 47.3 vs. 47.4, indicating that the effect of differential employment and earnings patterns of wives makes an insignificant difference to the inequality of distribution of income.

The differential in inequality between the total family income and husband's income results from the much greater incidence of earnings of adult family members other than the husband and wife, and may have nothing to do with family economic welfare. There are more earners, who may or may not pool their resources with those of other family members, and there are more adult consumers.

The employment rates of black wives are especially high relative to those of whites in the case of women with young children. Black mothers of children under six are 66 percent more likely to be working than their white counterparts. For mothers with children 6-11 and 12-17, the differentials are 45 and 5 percent respectively. Married Negro women without children have employment rates that are not much greater than those of white women, and in the case of women 14-29 with no children, the black employment rate is considerably lower than the white rate.

If we disaggregate the population into two categories, those couples with children and those with none, and examine the separate Gini coefficients, we discover that the aggregate pattern presented above results from differential patterns within these two groups. For both categories income inequality is substantially reduced by virtue of the greater contribution of other family members. For childless couples (i.e., those with no children present), the racial inequality increases as a result of contributions by wives to family income, while for couples with children, inequality decreases somewhat. The effect of income of other relatives is greater, however, than that of income of wives.

Further disaggregation of couples in relation to age of youngest own child reveals that the wife's contribution in the case of couples with youngest own child aged 12-17 tend to slightly increase inequality, just as it does for childless couples. Thus in those groups with very much higher employment rates, the degree of racial inequality is slightly reduced by virtue of the income of wives. In groups where the black employment rates are only slightly higher, the degree of racial inequality of income is unaffected, or in some cases increased.

The effect of wife's earnings on the inequality of family income between the races is small because despite the higher rate of employment of Negro wives, their earnings are on the average considerably lower. Thus, a higher proportion of Negro families move up in the income distribution from where they would be in the absence of wife's income. The distance which they move in the distribution, however, is smaller, on average, than is the distance moved by white families with employed wives.

FOOTNOTES

*The research reported here was supported by funds granted to the Institute for Research on Poverty at the University of Wisconsin by the Office of Economic Opportunity persuant to the provisions of the Economic Opportunity Act of 1964. The conclusions are the sole responsibility of the author. Certain data used in this analysis were derived by the author from a computer tape file furnished under the joint project sponsored by the U.S. Bureau of the Census and the Population Council and containing selected 1960 Census information for a 0.1 percent sample of the population of the United States. Neither the Census Bureau nor the Population Council assumes any responsibility for the validity of any of the figures or interpretations of the figures presented herein or based on this material.

1/ Herman Miller, Income Distribution in the United States, U.S. Bureau of the Census, 1960 Census Monograph (Washington, D.C.: U.S. Government Printing Office, 1966), p. 22.

2/ Wilber Thompson, "Internal and External Factors in the Development of Urban Economics" in Harvey Perloff and Lowden Wingo, Jr. (eds.), Issues in Urban Economics (Baltimore: Johns Hopkins University Press, 1968). See also, Wilber Thompson, A Preface to Urban Economics (Baltimore: Johns Hopkins University Press, 1965), pp. 106 ff.

3/ U.S. Bureau of the Census, "Trends in the Income of Families and Persons in the United States: 1947-1964," Technical Paper #17, 1967, Tables 2 and 37; U.S. Bureau of Labor Statistics, Statistics on Manpower: A Supplement to the Manpower Report of the President (1969), Table B-1.

TABLE 1

Income Inequality of Married Men Living in Urbanized Areas, by Age (1959 Income as Enumerated in 1960 Census)

Age of Husband	Gini Coefficients Total	Nonwhite
Less than 18	47.2	44.7
18 - 24	28.8	31.1
25 - 34	26.6	27.4
35 - 44	28.2	29.1
45 - 54	30.9	30.9
55 - 64	35.0	35.6
65 - 74	45.8	44.3
75 and over	53.0	47.7
Total	33.1	32.3

SOURCE: Derived from data in 1960 United States Census, "Persons by Family Characteristics," Subject Report PC(2) 4B, Table 10b.

TABLE 2

Gini Coefficients on Three Income Measures: Husband-Wife Families, Wife Under age 60, Nonfarm

	Husband's Income	Family Income Minus Wife's Earnings	Family Income
Total	35.6	35.8	33.7
Negro	33.5	35.3	35.1
Non-Negro	34.7	35.1	32.9
Families with No Children	39.2	39.7	36.0
Families with Children	33.6	33.8	32.4
Families with Children Under 18			
Negro	31.9	32.5	32.9
Non-Negro	32.8	33.1	31.7
Families with No Children Under 18			
Negro	36.2	39.5	38.3
Non-Negro	38.5	39.0	35.1

Incomes of $25,000 and over are coded as $44,000.
SOURCE: 1/1000 Sample

TABLE 3

Gini Coefficients Computed on Three Income Measures: Husband-Wife Families, Wife Under Age 60, Nonfarm, by Color, Family Status, and Whether or Not Wife Received Income

	Husband's Income		Family Income Minus Wife's Earnings		Family Income	
	Wife with Income	Wife without Income	Wife with Income	Wife without Income	Wife with Income	Wife without Income
Total	32.6	36.6	33.3	36.0	30.2	36.0
Negro	34.1	32.8	36.3	34.0	34.6	34.0
Non-Negro	31.6	35.8	32.4	35.3	29.1	35.3

Incomes of $25,000 and over are coded as $44,000.

SOURCE: 1/1000 Sample

TABLE 4

Wife's Contribution to Family Income by Husband's Income and Age

Husband's Income	N	Percent	Proportion of Wives with Income	Wife's Average Income per Recipient	Wife's Average Income per Family
None	614	1.6	44.8	$ 2424	$ 1085
Less than $1000	1,941	5.2	51.0	1542	787
$1000 - $1999	2,736	7.3	50.3	1625	818
$2000 - $2999	3,463	9.3	49.5	1752	868
$3000 - $3999	4,493	12.0	48.7	1994	972
$4000 - $4999	5,471	14.6	47.3	2192	1036
$5000 - $5999	5,895	15.8	44.3	2312	1024
$6000 - $6999	4,192	11.2	40.8	2356	960
$7000 - $9999	5,350	14.3	35.9	2365	850
$10,000 - $14,999	2,006	5.4	31.2	2412	754
$15,000 +	1,245	3.3	32.1	2613	838
Total	37,406	100.0	43.8	2112	925

SOURCE: 1960 U.S. Census, Sources and Structure of Family Income, Table 17.

TABLE 5

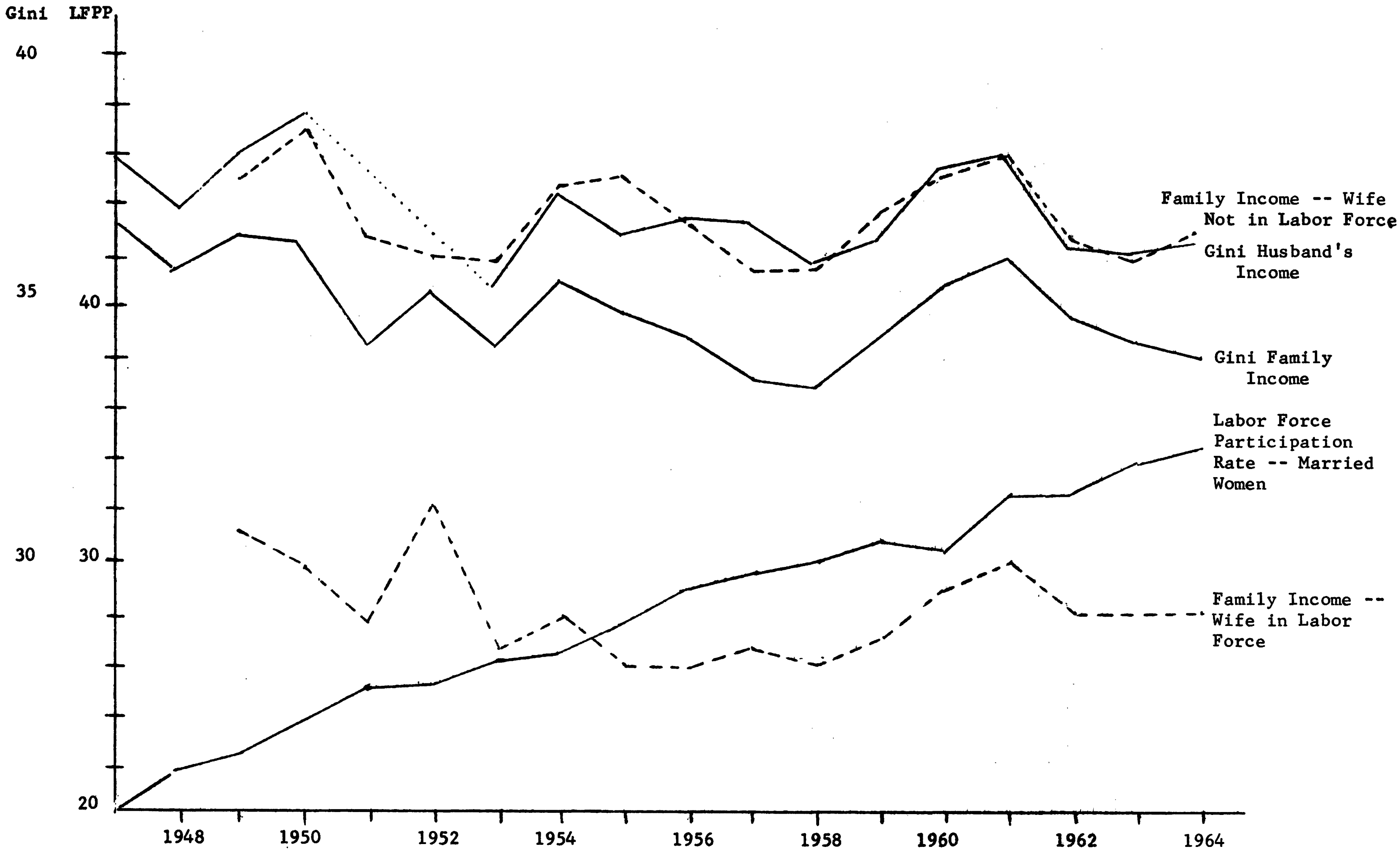
Negro-White Income Inequality by Family Status for Three Income Measures (Gini Coefficients)

Family Status	Husband's Income	Family Income Minus Wife's Earnings	Family Income
Husband-Wife Families with one or more Children Under 18			
Youngest 0 - 2	56.2	48.8	45.9
3 - 5	61.6	55.7	50.2
6 - 11	60.4	56.1	51.4
12 - 17	54.0	49.2	50.9
Total	58.2	52.5	49.6
Husband-Wife Families with No Children Under 18			
Wife 14 - 29	38.1	26.2	36.9
30 - 44	49.2	41.0	44.8
45 - 59	50.0	46.2	47.6
Total	46.0	40.5	43.5
Husband-Wife Families, Wife Under Age 60	54.3	47.4	47.3

SOURCE: 1/1000 Tabulations

FIGURE 1

Recent Trend in Income Inequality and Labor Force Participation Rates of Married Women



SOURCE: U.S. Bureau of the Census, Trends in the Income of Families and Persons in the United States: 1947-1964. Technical Paper #17, 1967. Tables 2 and 37; U.S. Bureau of Labor Statistics, Statistics on Manpower: A Supplement to the Manpower Report of the President (1969) Table B-1.

COMPONENTS OF NEGRO-WHITE INCOME DIFFERENCES*

H. H. Winsborough and Peter Dickinson
Center for Demography and Ecology
University of Wisconsin

Introduction

This paper investigates black-white income differences for urban males in the labor force in 1960. Specifically, it addresses the question, "How much of the observed difference can be accounted for by differences in the educational level of the two populations?" In so doing, we have generalized components-of-a difference techniques somewhat.1/ That generalization is the most interesting methodological or statistical aspect of our paper.

We begin by discussing the method. Then we apply the technique to the analysis of Negro-white income differences. Finally, we will extend the generalization of our method to cover composit functions.

Components of a Difference

Suppose we have two functions on the real plane:

$$Y = f_1\ (X_1)$$
$$Y = f_2\ (X_2) \qquad (1)$$

Let D_1 and D_2 denote the domains of f_1 and f_2 respectively, and R_1 and R_2 the co-domains. For any Y_1 an element of R_1, and Y_2 an element of R_2, there exists an a which is an element of D_1 and a b, an element of D_2, such that $Y_1 = f_1(a)$ and $Y_2 = f_2(b)$. The difference $Y_1 - Y_2$ can be written as $f_1(a) - f_2(b)$.

Let:

$$\delta_1 = f_2(a) - f_2(b)$$
$$\delta_2 = f_1(b) - f_2(b) = \Delta f(b) \qquad (2)$$
$$\delta_3 = \Delta f(a) - \Delta f(b)$$

Clearly,

$$\sum_{i=1}^{3} \delta_i = [f_2(a) - f_2(b)] + [f_1(b) - f_2(b)] + [f_1(a) - f_2(a) - f_1(b) + f_2(b)] \qquad (3)$$
$$= f_1(a) - f_2(b) = Y_1 - Y_2.$$

We see then that the difference between two values, Y_1 and Y_2, can be expressed in terms of three additive components: a change in the argument of the function, a change in the function, and an interaction term, the result of a simultaneous change in both argument and function.

Application to Negro-white Income Differences

Suppose now that we let population one be urban white males in the labor force in 1960 while population two are comparably defined black males. Suppose, further, that we believe income is a linear function of education. In this example the Y_1 and X_1 of the above section are white income and education means respectively while Y_2 and X_2 are comparable Negro means. The function f_1 is the linear equation to predict income from education for whites while f_2 is the equation for blacks.

The component δ_1, then, is given by:

$$\delta_1 = (\alpha_N + \beta_N \bar{X}_w) - (\alpha_N + \beta_N \bar{X}_N) = \beta_N \bar{X}_w - \beta_N \bar{X}_N \qquad (4)$$

where the subscripts indicate the population of reference. This component can be interpreted as the gain in income which would result from an improvement of mean Negro education to the white level but with no change in the parameters of the Negro function.

The component δ_2 is given by:

$$\delta_2 = (\alpha_w + \beta_w \bar{X}_N) - (\alpha_N + \beta_N \bar{X}_N) = (\alpha_w - \alpha_N) + (\beta_w - \beta_N) \bar{X}_N \qquad (5)$$

This component may be interpreted as the improvement which would result if Negro education were translated into income by the white rule but the level of Negro education were unchanged.

The component δ_3 is given by:

$$\delta_3 = [(\alpha_w + \beta_w \bar{X}_w) - (\alpha_N + \beta_N \bar{X}_w)] - [(\alpha_w + \beta_w \bar{X}_N) - (\alpha_N + \beta_N \bar{X}_N)] = (\beta_w - \beta_N)(\bar{X}_w - \bar{X}_N) \qquad (6)$$

In interpreting the preceding components we have used the rearranging of means or functions by some policy procedure as a kind of model to lend meaning to our components. If we continue using the policy model for interpretation, we might tell ourselves the following story about this component. Suppose the political or financial situation were such that it was possible to deal with only one aspect of the income discrimination problem at a time. Suppose a policy maker chose to improve the Negro mapping of education into income first. If that policy worked, the subsequent value of an improvement in Negro educational levels would no longer be simply the value of δ_1 but the sum of δ_1 and δ_3. Conversely, if we chose to improve Negro education levels first, the subsequent value of improving the Negro function would be the sum of δ_2 and δ_3.

Component three, then, is the increment (or decrement) in effect due to modifying both aspects of the situation simultaneously or in series over the effect of changing each singly.

The Data and Some Results

Let us now estimate these components using data from the one-in-a-thousand sample of the 1960 Census. We will estimate equations and components separately for three age groups, 25-34, 35-44, and 45-54. The dependent variable is total individual income scored at the mid-point of thousand dollar intervals. Educational attainment is measured by the highest grade of school completed scored as follows:

Years	Score
0-4	0
5-7	1
8	2
9-11	3
12	4
13-15	5
16	6
17 and above	7

Some preliminary analysis showed that these intervals yield the best approximation to linearity for the relationship between income and education. As we shall see presently, specification of a believable zero-point for educational attainment permits further analysis of component two -- the component for a change in functions. Since there is some reason to believe that functional literacy is generally obtained in the 5-7 interval, we have chosen to score 4 and fewer years of school as zero because attainment of functional literacy seems such a critical component of education.

Table 1 presents estimates of the six regression equations along with mean levels of education and income. All estimates are several times their standard deviation. Table 2 presents the components of the difference in mean income. Over all, whites make between $2359 and $3411 more money than do Negroes of the comparable ages and also have between 1.11 and 1.44 more units of education.

Table 1: Parameters for the regression of income on education by age group for Negro and for white urban males in the labor force, 1960

	Age and Race					
	25-34		35-44		45-54	
Coefficient	White	Negro	White	Negro	White	Negro
Mean education	3.98	2.87	3.75	2.41	3.30	1.86
Mean income	4630	2270	5880	2670	5780	1821
bIE	572	219	1012	354	1128	295
aIE	2353	1641	2085	1817	2058	1821

Table 2: Components of Negro-white income difference by age

	Age		
Component	25-34	35-44	45-54
δ_1	243	474	425
δ_2	1725	1854	1786
δ_3	391	882	1200

Recall that the component δ_1 is the improvement in black income which would accrue if blacks had the same level of education as whites. From Table 2 we see that this improvement would be $243 for the youngest age group, $474 for the middle one and $425 for the oldest. These quantities are between ten and fifteen percent of the total Negro-white income difference for the age group.

The component δ_2 is the improvement which would result if blacks had their own educational level but that education translated into income by the white equation. Estimating this component from our data we find the improvement in Negro income for the age groups would be $1725 for the youngest, $1854 for the middle and $1786 for the oldest. These amounts range from 52 percent to 73 percent of the total difference.

Under the linear model it is possible to divide this component into two parts as shown in the second line of equation (5). One part has to do with differences in α and represents a fixed amount by which Negroes of <u>any</u> educational level are deprived of income in comparison to whites. The other part represents the difference in slope; the payoff of a unit of education. This separation of the component into two parts, however, depends on believing that one has set the zero-point of the education scale at the proper place. Had we set the zero-point of the educational scale at another place, the parts would divide differently but their sum would remain constant. With our assumption that 5 years of education represents functional literacy and all fewer years represent no education at all, we find change in the slope part of this component to be worth $1013, $1586, and $1549 for the age group respectively. Leaving $712, $268, and $237 attributable to intercept differences.

Returning to our main components, we find that δ_3, the interaction component has the value of $391, $882, and $1200 for the age group. These amounts are about 17, 28, and 35 percent of the total difference.

From these data, then, it seems clear that δ_1, difference in educational level alone, is only a modest contribution to income differences. The component δ_2, differences in the equation mapping education into income, is by far the more important factor. Further, if one believes our setting of the zero-point for educational attainment, the villain can be narrowed down to differences in the return per unit of education. Perhaps this conclusion should have been apparent simply by an inspection of Table 1 where we find the differences in slopes ranging from \$833 to \$353. Even in this rather simple problem, we found some security of mind in working through the components to arrive at the conclusion. In more complex problems we have found the method invaluable. For that reason, we will conclude by extending our method somewhat.

Extensions of the Method

An important aspect of the components method presented above is that it holds not only for the class of continuous function but other kinds as well. Thus, the Kitagawa three component method is a special case of the method presented here.

In this regard it is worth noting that the Kitagawa two-component method has the virtue of a symmetry not obvious in our method. In the above example it has seemed natural to think of improving the Negro income to the white level. In that sense we have been using the white population as a standard. In another problem, it might not be clear which population should be standard. Thus, it is worth considering what would happen if we were to imagine reducing the white income level to that of blacks. Viewed from that perspective our components would be:

$$\begin{aligned} \delta_1' &= f_1(b) - f_1(a) \\ \delta_2' &= f_2(a) - f_1(a) = \Delta' f(a) \\ \delta_3' &= \Delta' f(b) - \Delta' f(a) \end{aligned} \qquad (7)$$

Writing out δ_3' we see that

$$\delta_3' = f_2(b) - f_1(b) - f_2(a) + f_1(a) = \Delta f(a) - \Delta f(b) = \delta_3 \qquad (8)$$

Thus the interaction component is identical in the two approaches. Further, it is easy to show that:

$$- \delta_1' = \delta_1 + \delta_3 \quad \text{and} \quad - \delta_2' = \delta_2 + \delta_3 \qquad (9)$$

Thus, in spite of the lack of symmetry in the three component model with changes in choice of the standard population, it is easy to move from one perspective to the other. Further, it is easy to show that the Kitagawa two-component method achieves its symmetry by simply dividing δ_3 among the other two components. In our notation her combined IJ effect is:

$$\frac{\delta_1 - \delta_1'}{2} = \frac{\delta_1 + \delta_1 + \delta_3}{2} = \delta_1 + \frac{\delta_3}{2} \qquad (10)$$

while her residual IJ effect is

$$\frac{\delta_2 - \delta_2'}{2} \quad \frac{\delta_2 + \delta_2 + \delta_3}{2} \quad \delta_2 + \frac{\delta_3}{2} \qquad (11)$$

The above methods are easily generalizable to composit functions. Suppose the Y-values are obtained from the composit of two functions,

$$\begin{aligned} Y &= f_1 \cdot g_1 (x) \\ Y &= f_2 \cdot g_2 (x) \end{aligned} \qquad (12)$$

Letting $Y_1 = f_1g_1(a)$ and $Y_2 = f_2g_2(b)$ we can decompose the differences $Y_1 - Y_2$ as follows. Let

$$\begin{aligned} \delta_1 &= f_2g_2(a) - f_2g_2(b) \\ \delta_2 &= f_2g_2(b) - f_2g_2(b) \\ \delta_3 &= \Delta f g_2(b) \\ \delta_4 &= \Delta f \Delta g(b) \\ \delta_5 &= f_2\Delta g(a) - f_2\Delta g(b) \\ \delta_6 &= \Delta g(a) - \Delta g(b) \\ \delta_7 &= \Delta f\Delta g(a) - \Delta f\Delta g(b) \end{aligned} \qquad (13)$$

It is easy to verify that:

$$\sum_{i=1}^{7} \delta_i = f_1g_1(a) - f_2g_2(b) \qquad (14)$$

We have used these components to decompose income differences for a recursive system.

---

* The research reported here was supported by funds granted to the Institute for Research on Poverty at the University of Wisconsin by the Office of Economic Opportunity persuant to the provisions of the Economic Opportunity Act of 1964. The conclusions are the sole responsibility of the author. Certain data used in this analysis were derived by the authors from a computer tape file furnished under the joint project sponsored by the U. S. Bureau of the Census and the Population Council and containing selected 1960 Census information for a 0.1 percent sample of the population of the United States. Neither the Census Bureau nor the Population Council assumes any responsibility for the validity of any of the figures or interpretations of the figures presented herein or based on this material.

During the preparation of this report, Mr. Dickinson was supported as a trainee under NIH Training Grant #GM-01190 from the National Institute of General Medical Sciences.

1/ Evelyn M. Kitagawa, "Components of a Difference Between Two Rates," JASA, Vol. 50 (December 1955), pp. 1168-1194.

ADVANCING STANDARDS IN AN ADVANCING FIELD

Paul B. Sheatsley, National Opinion Research Center

As the title of this small paper indicates, I mean to talk briefly about the problem of standards in survey research. I suppose standards can be defined or viewed in at least three ways. First, standards in the sense of standardization or comparability. Certainly it would be desirable if surveys could better be compared with one another, if we all asked certain basic questions such as occupation or income in an agreed upon way, if we all calculated our response rates in exactly the same manner. Second, there are what we might call ethical standards, such as respect for respondent anonymity and willingness to disclose full details of methodology. And third, there are technical standards which might cover such matters as sample size, questionnaire construction, level of interviewer trainong and supervision, and similar steps in conducting the usual survey.

A very brief history of survey research may be instructive. We must remember that it started in the business world. It was not until after World War II that the campuses took much notice of it. Businessmen and advertising agencies were questioning samples of readers and consumers almost fifty years ago. The Gallup Poll, the first of its kind, was begun by George Gllup, himself a market researcher, and the other preeminent pollsters, such as Elmo Roper, Archibald Crossley and Louis Harris, have also been marketing researchers. That early research 30 or 40 years ago was remarkably simple, even simplistic, by modern standards. They selected their samples by setting various kinds of quotas, they asked one or two questions where today we would spend five or ten minutes on that one topic, and the low budgets and competitive pressures would have left little resources for interviewer training and supervision. Secrecy of exact methods used was commonplace, and it may be presumed that many first who hastened to set up their own survey shops in those days often engaged in fairly sharp practices.

With World War II survey research came of age. Men like Paul Lazarsfeld, Samuel Stouffer, Hadley Cantril and Rensis Likert lent their services to the government in the organization and direction of survey research activities. Such advances as probability sampling and Guttman scaling derived from this government work. More important, a whole second generation of research men who had worked under the "giants" I have named went back after the war to campuses or to commercial activities, and continued to teach and practice the art of survey research. Within two years after the way there had come into existence an American Association for Public Opinion Research, one of whose stated purposes was to improve standards. Leaders in the field had by that time concluded that the old ways of secrecy and every man for himself which prevailed in an earlier day were inappropriate for what had become a mature field of practice considerably affecting the public interest. AAPOR, as it is commonly called, has always been open to any individual with an interest in public opinion research, and while the majority of members are from the world of business or market research, more than one-third are from universities, government agencies, foundations and other not-for-profit institutions.

AAPOR in its early days almost foundered over the issue of standards. Some academic and government researchers considered some market researchers as rather shady types who ought to be made to shape up. Some market researchers considered the academics as ivory-tower types who were about to impose upom them a set of idealistic standards which could not possibly be fulfilled in the market place. Fortunately, cooler heads prevailed and AAPOR has grown increasingly strong and influential. But it took the Association several years to get around to discussing standards again and when it did it emerged, after another period of years, with a Code of Professional Ethics and Practices, with most of the emphasis on ethics. The American Marketing Association went through a similar experience, and it too now has a Code of Ethics.

These codes define the researcher's obligations to his respondents, to his client or sponsor, and to the public. To these latter two groups, the essential obligation can be summarized as full disclosure. No relevant information, either of substance or of method, may be withheld. In other words, the codes does not dictate the number of cases the researcher should collect on a given study, but it says he must disclose in his report the size of his sample; he is not told what is a satisfactory completion rate in a mail survey, but he is obligated to reveal the completion rate on which his data are based. This of course places the onus for getting good research, and perhaps properly, on the user of the survey, who is presumed to be sufficiently sophisticated, or to have access to sufficiently sophisticated advisers, to recognize loaded questions, faulty sample designs, and so on.

Caveat emptor.

While I can only speak for AAPOR, and not AMA, I can testify that the code has had some measure of success. Every AAPOR member attests that he will abide by the code when he joins the Association, and there is reason to believe that violators are rare. Non-members of AAPOR, of course, are not so bound, but even here there have been successes. AAPOR has an active Standards Committee and any member with a complaint can present the facts to that committee. The committee has no real enforcement power, but it has not yet been pushed to its ultimare sanction: publication of names of offenders in AAPOR's journal, the Public Opinion Quarterly. In one important case, a non-member was making large numbers of telephone calls in cities all over the country, ostensibly conducting a survey to find out the extent of the market for new automobiles. Actually, this was a cheap and easy way of compiling a list of prospects, and respondents who had indicated any intent to purchase, found themselves confronted a few days later by an aggressive salesman. This was an obvious violation to respondent anonymity in a purported survey and, if continued, could seriously affect public trust in all surveys. When a few members of AAPOR's Standards Committee arranged a visit with the offending researcher and discussed the matter with him, the practice was stopped. There have been many similar instances.

But the caveat emptor philosophy has certain limitations when there really is no emptor, or buyer, but when the survey is done for selfish or partisan purposes and the results fed directly to the public. The man in the street is hardly capable of recognizing poor sample design, bad question wording, or out-of-context interpretation. We have probably all been exposed to examples of private polls which purport to show that a particular political candidate is way ahead in a particular state, or mail surveys which indicate that large majorities wish to take some particular kind of political action. The professional political pollsters such as Gallup, Harris, Crossley, Mervin Field in California, and others, have been seriously concerned by the appearance of numerous "private surveys" in any major election campaign, and they have given much thought to how to impose "standards" in this area. Their response was much the same as that of AAPOR and of the American Marketing Association.

They set up a Council on Published Polls, which includes all of the major polltakers who regularly cover election contests, and they drafted some standards of disclosure which they have circulated to Congressmen, state and local officials, newspaper editors and other media representatives. The standards suggest questions which readers of the purported survey data should ask, and all members of the Council have pledged that they themselves will frankly answer all such questions: The questions themselves cover the usual areas: size and design of the sample, completion rate, question wording, means of data collection, dates of interviewing, and so on. Certainly such simple educational measures cannot stop the appearance of fradulent or misleading polls, but their existence cannot help but have a long-run effect as news editors become better informed about polls. Meanwhile, the very existence of the Council and its standards makes it much easier to discredit an inadequate poll. The guilty polltaker, instead of seeing his data widely published and accepted, is forced on the defensive when he cannot or will not answer legitimate questions about his method, and in many cases his releases may be thrown away without ever seeing publication at all.

In these various ways, then, professional associations have tried to grapple with the problem of standards. But what is the present situation? Surveys and polls are more popular than ever. Federal and state governments want "evaluations" of particular health, educational and welfare programs which can only be conducted by means of survey research. Business firms more and more rely on survey research data for marketing decisions. Often the sponsors of the surveys have little or no understanding of methodological problems and they are always highly resistant to large expenditures of cost and time. The pressures for cutting corner here and there inevitably lead to a lowering of standards.

The popularity and apparent simplicity of survey research even attracts what we might call the "innocent ignorant". A friend of mine, an eminent psychologist who has been working with survey data most of his professional career, was recently appalled to learn that a chemistry professor on his campus had sent his students into various local supermarkets to question housewives about their attitudes and knowledge concerning the presence of phosphates in detergents. The good professor had no knowledge of how to obtain a truly representative sample of housewives, or even whether housewives were really his intended universe; he had no experience with questionnaire construction; his students had not a glimmer of interviewing problems and techniques. His survey, besides giving the students a faulty understanding of survey research, probably antagonized some of his respondents who had already been approached by other amateur surveyors and who may have been put off by the inept questioning and approach of the student interviewers, and in the end this survey producted data which may have been quite seriously misleading. Similar surveys are doubtless conducted every day by high schoolers, local do-good societies, small business, and of course newspapers.

It is clear that the problem of advancing standards in the field of survey research is not a simple one. We can probably not insist that that chemistry professor be required to take courses in survey research and pann an examination, any more than we can demand the same for the high school students, the members of the local Women's Club, or the small businessman. What AAPOR has tried to do, in at least a preliminary attempt to solve the problem, is to make expert advice available to such groups. The AAPOR membership, while concentrated heavily in New York, Washington and other major research centers, is scattered widely over many cities and college campuses. Members are urged to be alert for opportunities to be of local professional service; inquiries direct to AAPOR are referred to a local member, or the names of several members are provided to the inquirer; mailings have gone out to college campuses advising them of AAPOR's desire to be of help in consultation on survey research matters. Individual members sometimes establish formal consultant relationships with local groups, or more often simply donate their advisory services as a professional obligation.

You will have noted that all of the efforts I have referred to, to establish or maintain standards, have dealt with ethics rather than techniques, and have been of an educational rather than punitive variety. Such efforts are necessarily slow and imperfect. One may ask, why cannot categorical rules be set down for survey research standards and those standards rigidly enforced? It should not be too difficult to set reasonable standards for sampling, completion rates, question wording, interviewer training, and so on. We can all recognize abuses when we see them. But, in fact, a moment's thought, or even less, reveals very clearly that operational standards for survey research are impossible to set.

We sometimes call our work a science, and indeed a good piece of survey research will meet the scientific requirement of being capable of replication by an independent researcher following the same methods. But it is obvious that certain areas of survey research are still very much of an art, and particularly so when we are dealing with measures of attitude, future intentions, or beliefs, rather than factual or behavioral information. There is no one correct way to write a survey question, nor even any way to know exactly which question or questions should be asked. In spite of a great deal of empirical and experimental research, there are no hard and fast rules which govern the selection of interviewers, nor is the same method of training and supervision appropriate for all of those hired. Even assuming such standards were possible and maintained, there is no way a researcher can prevent his data from being quoted out of context or manipulated in improper ways, so that all his methodological precautions come to naught.

Aside from the fact that we simply do not know enough to dictate standards in many areas or survey research, there is the fact that standards simply <u>have</u> to be relative to the time and cost resources available. The cost of quote - a survey - can range from a few hundred dollars to several hundred <u>thousant</u> dollars. Obviously, if one has unlimited time and money, one can design and perfect a survey with extremely high standards. While our resources are harly unlimited, we at NORC have had recent experience with an ambitious evaluation study of federal manpower training programs, sponsored by OEO and the Department of Labor. Enrollees in five programs in ten cities have been interviewed four times over a two-year period; this sample has been matched with a group of controls, selected from house-to-house screening, who have been similarly followed up. The demand for a final completion rate of 80% in this mobile and hard-to-find sample has required the setting up of separate NORC offices in each of the ten cities; rigorous training and supervision of an appropriate interviewing staff; and the expenditure of an enormous amount of time, effort and money to locate and interview reluctant or elusive individuals.

But the point is that such expenditures are hardly required on most surveys. The marketing man who wants to know which of two advertisements is more attractive to the public, or the community agency who wants to know how exercised the public is about environmental pollution, can make do with much less. Furthermore, the researcher rarely if ever has unlimited time and cost resources. He has a deadline and he must stay within a particular budget, whether it be generous or miniscule. In effect, then, he has to design his survey within those constraints. He can't afford a full probability sample, so he uses quotas or some means of weighting for persons not at home. He can afford only a half-hour interview, so he throws away all the batteries of questions he would like to ask about personality characteristics. He can't afford a full probability sample, so he uses quotas or some means of weighting for persons not at home. He can afford only a half-hour interview, so he throws away all the batteries of questions he would like to ask about personality characteristics. He can't afford to supervise the interviewers as well as he would like, so he pretty much accepts what they give him. His report is due in three weeks, so he can't pursue all the lines of analysis he had intended. All this is not necessarily bad. A survey can be over-designed and too perfect for the job intended; <u>some</u> reasonably accurate information is usually better than none at all. But is sure makes the job of setting standards difficult.

Finally, beyond the fact that we do not know enough about some areas and that the same standards are not appropriate to all surveys, any set of standards must always be based on past experient and present knowledge. The title of this paper refers to an "advancing field". Survey research has advanced tremendously in my own time, and it continues to advance. When Gallup first sent a national staff of interviewers out to select a representative sample by the use of sex-age-economic level quotas, this represented a tremendous advance over the Literary Digest's use of mail ballots. The growing demand for probability sampling represented a similar advance over quota sampling. The contributions of numerous individuals and agencies have improved our knowledge of all phases of survey research, and the introduction of the computer has of course revolutionized our means of processing survey data. We are perhaps on the verge of even more exciting discoveries.

The techniques of survey research are being employed in ever more sophisticated ways. Merely within the past six months, we at NORC have been faced with the following problems of research design and execution.

> Personal interviews with all practicing internists in a defined suburban area. since many of the physicians knew one another, the problem was to avoid contamination effects.

Identification of the population of working journalists in the United States and personal interviews with a representative sample of these.

Establishing a panel of patients suffering from Parkinson's disease and following these up over a period. Half of the group are receiving a new method of treatment, half are controls.

Sampling patients served by the emergency room of a large metropolitan hospital and following up several weeks later with interviews in their homes.

Validation of elementary school data submitted to the Office of Education, through interviewer inspection of records at a sample of schools and brief interviews with school officials and parents.

The list could go on and on. The point is that each of these assignments involved a great deal of fresh thought on such problems as defining the universe, inventing an appropriate method of sampling, gaining access to the intended respondents, and obtaining valid data from them. And naturally there were the usual constraints of time and cost in each case. Under these circumstances, no conceivable set of general operational standards for survey research could offer much guidance.

In sum, then, I hold that there has been and will continue to be advancing standards for survey research. In almost every aspect of research, our workmanship today is far superior to what it was twenty years ago, or even ten years ago. But these advancing standards have come about not through any codification of standards approved by some official group, but through the inventiveness and increasing sophistication of the sponsors, practitioners and users of survey research. There will always be abuses of the survey method, as of many other things in this imperfect world, but it is becoming more and more difficult for such surveys to gain any measure of acceptance. More than this, we can probably not reasonably expect.

CONFIDENTIALITY AND CO-OPERATION BETWEEN RESEARCHERS

John C. Scott, University of Michigan

Privacy is everybody's business in survey research. Our strong organizational commitments to protecting the confidentiality of survey data now serve to inhibit cooperation between research institutions. Yet, our shared concern should make confidentiality an area where uniform standards can be achieved and joint actions taken which would have considerable benefit to all survey organizations.

Briefly, I would like to review some of the ways in which the concern for confidentiality inhibits cooperation between survey facilities, and then indicate how steps can be taken toward standardizing procedures which should facilitate increased cooperation.

To begin with, let's talk a minute about the nature of privacy and how it is related to the idea of fully informed consent. Privacy, as I shall use it, is the control that a person has over what others know about him as an individual. Privacy has been called one of man's great privileges - with the stipulation that "it should only be surpassed by its invasion by the perfect person at the most auspicious moment." Obviously, the survey interviewer is not always the perfect person, nor, as our call records testify, does he always ring the bell at the most auspicious moment. However, our experience has been that more often than many realize the survey interviewer does approach these ideals. Most people have few opportunities in their lives when they can voluntarily express their feelings and reveal themselves without fear of what may result from this relaxation of their control over information regarding themselves. Responses to the interviewer's questions, while lacking the legal status of "privileged communications," are communications made by individuals who feel they have been released from many privacy considerations by the promise of anonymity and by the non-evaluative style of the interviewer. In essence, the respondent trusts the researchers to perform part of the task of guarding his privacy. For many this release makes the interviewer the perfect person; however, it is no simple matter for a researcher to promise confidentiality. One difficulty is that what an individual wishes to keep private varies tremendously from person to person, therefore researchers must assume that everything regarding a respondent that has become known to them in the course of gaining the interview must be guarded. This includes not just answers to questions, but also such facts as who granted an interview; the information that no one was home when a call to interview was made, or even how the house was decorated and furnished.

Another difficult concept to operationalize is that of fully informed consent. Speaking only from the standpoint of privacy, informed consent means:

1. that the respondent understands he is not obliged to answer,
2. that the researcher is obliged to guard all information obtained from the respondent. While the information may be published or shared with colleagues, this will be done in such a way that the linking of response to him is impossible,
3. that the respondent understands what risk is involved in giving information even under these conditions. This includes the risk that the respondents' answers, taken together, may reveal more than he intended, and that despite every reasonable effort by the researcher some compromise of confidentiality may occur.

Obviously, there are other aspects to informed consent beyond the focus of this discussion which I will not go into now.

These are heavy responsibilities which no survey organization can afford to take lightly and, as we know a violation of confidentiality can have serious consequences not only for the organization directly involved, but the repercussions from a widely publicized incident would be detrimental to the profession as a whole.

In order to maintain confidentiality a survey organization must:

1. be able to have a high degree of trust in all persons who could possibly link information with individual respondents,
2. be prepared to support employees against pressures brought to bear on them to divulge respondent information.
3. be in a position to discipline persons if they divulge information,
4. have an active program to instruct and remind all those with access to respondent information of the importance of confidentiality,
5. destroy linking information as soon as practical, and maintain barriers to identification as long as the link exists.

Trustworthiness is essential because close supervision of survey workers, especially interviewers, is impossible. As the Pentagon papers incident has demonstrated, no security system is perfect if a person with access to sensitive information wishes to breach confidentiality.

Agreements to cover the legal and other expenses of employees who refuse to hand over respondent information if subpoenaed are one type of support an organization can give employees. The existence of statements of professional ethics is another, and although at first pass it may not seem supportive, a written policy of dismissal in case respondent information is disclosed also strengthens an employee's bargaining position in dealing with attempts to force disclosure.

It is the necessity of knowing that the

promise you give respondents can be backed up that inhibits cooperation between survey organizations at the present time. While there is no question in my mind of the commitment of all of us to respecting the rights of respondents and maintaining confidentiality, I do not have enough familiarity with the field interviewers of other organizations to have the same assurance of trustworthiness that I have with my own organization's interviewers. I imagine each of you shares some of these feelings of disquietude when you use unfamiliar interviewers, but trust is only part of the picture. I am not in the same position of offering support and sanctioning persons who do not work directly for my organization. This is most difficult when data is gathered for researchers outside the organization, particularly when their study design requires linking the data we gather with other information they have. Another instance in which control is lessened is that of the researcher who leaves the organization, taking his data with him. Unless particular attention is paid to the problems of confidentiality which may arise in such circumstances, an organization may suddenly discover that the pledge of confidentiality given respondents has been violated, and there is very little the organization can do about it. The likelihood that confidentiality will be violated increases with the number of people who have access to the information and as control over them becomes weak.

There are several ways in which campus based survey groups may cooperate at the data gathering stage. Some are relatively free from problems of confidentiality; others are much more susceptible. Those that are free are:

1. Replication: Two or more organizations coordinate the design, conduct and analysis of a study so that each follows essentially the same course, but since each organization performs its own analysis there is no transfer or sharing of data files.
2. Independent Complementation: Two or more organizations study different populations following the same design and procedures. Analysis is conducted independently within each organization so that data files are not merged, each organization reports on its own sample, for example, independent state election studies, or Organization A studies blacks and Organization B studies whites, or Organization A does a national study and Organization B ties in intensive local studies.
3. Development: One organization handles developmental phases, doing pre-testing, pilot studies or methodological investigations -- the results of which are fed to others for final data gathering.

The types of cooperation which require greater safeguards are:

1. Archiving: When an organization makes individual level data available to others for analysis, after destroying identification and links.
2. Division of Services: The phases of a single study are split up among several organizations, for example, one does sampling and interviewing and another does coding and data processing.
3. Staff Sharing: An interviewer's pool is formed for common sample areas.
4. Supplementation: More than one organization gathers data which are fed to a single organization for analysis. For example, several organizations do regional or local studies using the same design and instruments, turning over the raw data to one organization for analysis; or two organizations do large parallel national surveys, merging the data from both samples to obtain a number of cases larger than either could easily produce independently.

Although we have at one time or another worked with other organizations in most of these ways, this cooperation has typically not been something well thought out in advance and sought; but more nearly something accepted as the best means of accomplishing the job.

I believe that there are benefits to be gained from greater cooperation, which should be planned for and sought. I realize that there are many other barriers to undertaking joint research efforts than those associated with maintaining confidentiality, but this is one obstacle which could easily be eliminated. In fact, if a formal set of standards and procedures regarding confidentiality were agreed to by academic survey organizations, there would be benefits in addition to the greater possibilities for joint research efforts.

Compared to what any of us can do singly, the profession as a whole can do a better job of educating the public in how survey information is protected:

1. We can work more effectively toward securing the status of privileged communication for survey interview data;
2. We can discourage attempts by sponsoring agencies to have data turned over to them in ways which may violate confidentiality;
3. A statement by the profession supports those who may face citations for contempt if data are subpoenaed. I realize the codes of professional ethics of many of our disciplines cover the treatment of subjects, but survey research is interdisciplinary and has interests extending beyond a single society.

In conclusion I would like to mention that over the last several months I have been working with a committee of the Institute for Social Research toward the development of such a statement for our own organization. I expect that this statement will be ready for distribution within the next month or so.

SAMPLING ERROR ESTIMATES FOR PANEL STUDIES: IMPLICATIONS FOR STANDARDIZATION

Harold N. Organic, Brown University

Introduction

The Population Research Laboratory at Brown University has for a number of years conducted a series of annual sample surveys of the population of Rhode Island. The annual samples are related to one another within the framework of an overall panel design that permits continuing surveillance of the population and provides both cross-sectional and longitudinal data on community social structure, health needs and health care delivery.

While panel designs have long been known and often imaginatively used, they have generally been given short shrift by social researchers who have shown a marked preference for the cross-sectional model. The reasons for this are not far to seek when one recognizes that the cross-sectional model is free from certain difficulties inherent in the panel design. Chief among these, for our present purposes, is the attrition phenomenon that results in the loss of sample cases over time as respondents die, refuse continued cooperation at some point after the initial contact, or move their residence leaving no forwarding address for future contact. The bias introduced by this attrition has been difficult to measure, and consequently corrective steps hard to apply.

Nevertheless, the unique advantages of the panel design for providing time series data are manifest, and are assuming increasing importance in the view of many social researchers. This is especially true of, but by no means limited to those who deal with such applied problems as the provision of human services --- health care delivery being a specific case in point. In practical terms elected officials, administrators and interested professionals seek time series data for the planning, execution and evaluation of public programs and services. A research strategy directed toward continuing surveillance of a population and/or system and the recording of data thereon, implies some effective variant of the panel design.

The Sample Design

An experimental design for this purpose has been developed at the Population Research Laboratory and has been in operation since 1967. The purpose of this paper is to suggest a method for dealing with the attrition problem mentioned above. Three successive annual samples in Rhode Island were drawn independently from a frame constructed on the familiar clustered, stratified, area probability, multi-stage model [1], [2]. In each year approximately 1,100 household interviews were conducted, and follow-up interviews (usually by telephone) were taken at annual intervals thereafter. In all, 3345 respondents were enlisted in the three rounds of initial contacts in the Fall of 1967, '68 and '69. By the Fall of 1970 the first panel (Sample I) was three years old, and the others (Samples II and III) were two years and one year old, respectively. As would be expected, the oldest sample suffered the greatest loss (17.66 per cent), the second sample was intermediate (8.88 per cent), and the third sample sustained the smallest loss (6.01 per cent). These dropout rates were computed as a proportion of initial panelists who failed for any reason to give a follow-up interview in the Fall of 1970. The number of cases (N) in the initial samples and the number of dropouts by the Fall of 1970 are shown in the bottom row of columns 1, 3, 4, 6, 7, and 9 of Table 1.

The Problem

The fact of attrition brings to mind a number of questions:

1. how comparable were the initial panels;
2. has there been differential dropout by a given characteristic;
3. if yes, was the differential patterned;
4. was the difference between loyal and dropout components so great as to impair the "representativeness" of the residual loyal sub-sample;
5. if yes, at what point in the life of the panel did the departure from representativeness occur --- in the first, second...<u>n</u>th year;
6. is there available an efficient method to replenish losses, thus extending the representative utility of the panel; and finally
7. if yes, how can we test whether the replenishment cases are unbiased replacements for the losses previously sustained?

Appropriate statistics for estimating the sampling error of cluster samples have been available for some time. The calculations required for such estimates are tedious and time consuming, however, and it remained for the appearance of both large-scale electronic computers and the appropriate software to permit such calculations with reasonable economy of time and effort. Now that machines and programs for this purpose are at hand we should expect that sampling error estimates will appear together with the substantive findings published by researchers who employ the cluster sample design in their work. When this is fully realized we shall have come far toward achieving the standardization in procedure and reporting so important in this field.

The Data and Discussion

In attempting to answer questions 1 to 5 above we employed a recently developed program [3] to estimate the standard error of ratio differences for cluster samples. Ratio (percentage) differences between samples are expressed in standard error equivalents (SEE's) and shown in columns 10, 11, and 12 of Table 1. Thus, in comparing the initial panel Samples I and II for the age group under 35 (columns 1 and 4) we note a difference of 29.64 - 27.18 = 2.46 percent. One standard error for this comparison, as computed by the estimator program, was 1.85 percent, and the comparison is presented in column 10 as 1.33 SEE (2.46 / 1.85 = 1.33). All other between-sample comparisons for the characteristics selected were similarly evaluated and appear in columns 10, 11, and 12. A value of 2.00 SEE or greater is taken to indicate a significant difference between the sub-groups involved.

The first question asked above was --- how comparable are the initial panels? We note that for age, sex, and religion none of the 30 SEE's equals or exceeds 2.00. For these characteristics, then, we can conclude that the three panels are well matched. For the remaining characteristics, however, the matter is not so clear-cut as marital status, education, and total family income exhibit three, three, and five SEE's, respectively, that exceed 2.00. It is significant to note that 10 of the 11 SEE's exceeding 2.00 occur in columns 10 and 11, and that each of these columns involves Sample I.

We must assume that any sample of SEE's will itself be subject to sampling error, and that some SEE's will exceed the value of 2.00 merely by chance. Nevertheless, the concentration of significant SEE's in columns 10 and 11 should lead us to suspect that while Samples II and III appear to be comparable, there may be something unusual about Sample I. A large number of comparisons involving many additional characteristics (variables) would throw light on this point. If further investigation along these lines indicates that one sample is poorly matched to the others, the researcher may well review the sampling and field operations employed for that sample. It would also be possible that poor sample matches on a number of variables could be attributed to the errant variables themselves. Thus, the scheme of categories he chose may be the source of the difference, or a previously unsuspected departure from comparability in the wording or administration of the item. Not to be excluded from consideration is the possibility that the observed differences reflect a real change in the population under study --- income differences, for example, could result from a change in economic activity and salary/wage levels.

Here, as in the discussion that follows, the purpose is not to explain in detail the preliminary empirical findings, but rather to illustrate some of the ways that analysis of the data using this method can throw light on the researcher's problem of dealing with his material. In any case, the application of this scheme will not of itself releave the researcher of the task of close examination of the materials, but it may prove a useful tool.

The second and third questions asked --- has there been differential dropout by a given characteristics and if yes, was the differential patterned? --- are similar to the first, but involve examining within-sample differences. SEE's for loyal-dropout comparisons were calculated and are shown in columns 13, 14, and 15. Examining this section of Table 1 we note that neither sex nor religious preference exhibits differential attrition. The remaining characteristics, however, show evidence of attrition differentials of two sorts.

On the one hand age, and to a lesser extent marital status and total family income, exhibit patterned attrition; the case of educational attainment is less clear. To take up the characteristic of age, it appears that respondents aged 65 and over are strikingly more likely to be dropouts than any other age group. Given the mortality of man this is not a surprising finding, but it is interesting that the within-sample SEE's for this age group are 3.23, 2.66, and 1.68, respectively, for the three samples. In this instance both temporal and directional patterns can be discerned. The temporal pattern is evident from the fact that the SEE's vary monotonically with initial sample year. It will be recalled that at the time of the Fall 1970 follow-up interviews Samples I, II, and III were three, two, and one year(s) old, respectively. The direction of the differences is negative in all three cases, indicating a larger proportion of loyal than dropout respondents in each sample. The opposite is true for the age group 35 to 49. Suggestions of patterning (through SEE's are small) can be detected in the remaining two age groups. Here, as with other characteristics amenable to such manipulation, smaller interval sizes (10-year age groups, for example) could yield more detailed information about the fine-grain behavior of the variable.

The widowed and the aged are, to a considerable extent, overlapping categories. It is not surprising, therefore, that the widowed display a pattern of within-sample SEE's congruent to that seen among persons 65 years of age and over. The presently married, for similar reasons, appear to display attrition patterns similar to the younger age group. Examination of the table in this fashion also reveals total family income to behave in the same general way --- low income respondents having a greater tendency to drop out and high income respondents to remain loyal. As pointed out above, to the extent that these characteristics are related, to that extent the present analysis may be partially obscured. A more refined analysis will be required to evaluate the relative contribution to attrition made by each variable independent of the others.

If the overall study design employed the panel model solely for studying cohorts and their

TABLE 1

SELECTED SAMPLE CHARACTERISTICS: FREQUENCY DISTRIBUTIONS FOR EACH OF THREE SUCCESSIVE ANNUAL PANEL SAMPLES AND THEIR LOYAL-DROPOUT COMPONENTS IN PERCENT; AND BETWEEN AND WITHIN SAMPLE DIFFERENCES IN STANDARD ERRORS.

SELECTED CHARACTERISTIC		FREQUENCY DISTRIBUTIONS IN PERCENT FOR ANNUAL AND COMPONENT SAMPLES									SAMPLE DIFFERENCES IN STANDARD ERRORS					
		SAMPLE I			SAMPLE II			SAMPLE III			BETWEEN			WITHIN		
		INITIAL	LOYAL	DROPOUT	INITIAL	LOYAL	DROPOUT	INITIAL	LOYAL	DROPOUT	I-II (1-4)	I-III (1-7)	II-III (4-7)	I (2-3)	II (5-6)	III (8-9)
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
AGE:	Under 35	29.64	30.70	24.63	27.18	27.51	23.76	28.31	28.15	30.77	1.33	0.69	0.58	1.60	1.04	0.42
	35-49	30.17	31.47	24.12	29.90	31.08	17.82	27.01	28.25	7.69	0.14	1.62	1.27	2.20	3.27	4.95
	50-64	22.63	22.20	24.62	25.42	25.00	29.70	25.07	24.51	33.85	1.69	1.39	0.17	0.73	1.09	1.56
	65, +	17.56	15.63	26.63	17.50	16.41	28.72	19.61	19.09	27.69	0.04	1.39	1.47	3.23	2.66	1.68
SEX:	Male	41.61	41.27	43.22	43.18	43.05	44.55	42.09	41.83	46.15	0.89	0.66	0.27	0.48	0.29	0.63
	Female	58.39	58.73	56.78	56.82	56.95	55.45	57.91	58.17	53.85	0.89	0.66	0.27	0.48	0.29	0.63
M.S.:	Married	68.59	70.91	57.79	72.56	74.32	54.46	69.66	69.98	64.62	2.29	0.57	1.45	2.99	4.47	0.89
	Widowed	13.22	11.42	21.61	13.02	12.16	21.78	13.41	13.09	18.46	1.49	0.15	0.33	3.69	2.25	1.14
	Separated	3.11	2.58	5.53	3.25	2.90	6.93	2.50	2.17	7.69	0.24	0.78	1.21	1.67	1.68	1.60
	Divorced	4.17	3.99	5.02	2.64	2.51	3.96	4.72	4.72	4.62	2.04	0.62	2.38	0.64	0.83	0.04
	Never Married	10.91	11.10	10.05	8.53	8.11	12.87	9.71	10.04	4.61	1.90	0.93	0.89	0.43	1.48	2.17
REL.:	Roman Catholic	64.33	65.30	59.80	66.40	66.51	65.35	63.55	63.68	61.54	1.04	0.51	1.56	1.61	0.23	0.37
	Protestant	29.81	29.53	31.16	28.49	28.57	27.73	29.05	29.13	27.69	0.69	0.48	0.30	0.48	0.19	0.24
	Jewish	2.22	2.26	2.01	1.93	2.03	0.99	3.33	3.05	7.69	0.48	0.19	0.16	0.23	0.97	1.50
	None	2.66	0.54	1.51	2.64	0.48	0.99	3.52	0.49	0.00	0.04	1.30	1.28	0.79	0.50	0.19
	Not Ascertained	0.98	2.37	5.52	0.54	2.41	4.94	0.55	3.65	3.08	....[a]	....[a]	....[a]	....[a]	....[a]	....[a]
ED.:	Under 8 Years	14.11	13.47	17.08	14.87	13.51	28.71	14.62	14.07	23.08	0.52	0.38	0.16	1.42	3.62	1.52
	8-11 Years	41.36	40.30	46.24	36.41	35.91	41.59	34.87	35.24	29.23	2.47	3.24	0.72	1.98	1.16	1.00
	High School Grad.	30.60	31.47	26.63	32.98	34.27	19.80	33.95	34.25	29.23	1.24	1.51	0.51	1.49	3.78	0.84
	Some College	4.53	4.63	4.02	6.95	7.05	5.94	6.11	5.91	9.23	2.78	1.89	0.86	0.45	0.41	0.91
	College Grad., +	8.61	9.38	5.03	8.44	8.88	3.96	9.25	9.35	7.69	0.15	0.58	0.74	2.63	2.80	0.49
	Not Ascertained	0.79	0.75	1.00	0.35	0.38	0.00	1.20	1.18	1.58	....[a]	....[a]	....[a]	....[a]	....[a]	....[a]
INC.:	Under $4500	27.15	24.57	39.20	22.51	20.56	42.58	22.66	21.86	35.38	2.36	2.35	0.08	4.04	3.67	2.47
	$4500 to 7499	29.72	30.50	26.13	27.53	27.22	30.69	25.62	25.59	26.15	1.07	2.13	1.34	1.31	0.74	0.11
	$7500 to 12,499	27.78	29.20	21.10	30.96	32.72	12.87	30.25	30.71	23.08	1.53	1.24	0.38	2.61	4.98	1.37
	$12,500, +	13.40	14.33	9.05	18.65	19.31	11.88	21.10	21.65	12.31	3.59	4.91	1.50	2.35	2.22	2.41
	Not Ascertained	1.95	1.40	4.52	0.35	0.19	1.98	0.37	0.19	3.08	....[a]	....[a]	....[a]	....[a]	....[a]	....[a]
	TOTALS	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00						
	(N)	(1127)	(928)	(199)	(1137)	(1036)	(101)	(1081)	(1016)	(65)						

[a] Standard errors not calculated for 'not ascertained' category.

experience over time, our last four questions would not be of crucial relevance. However, where an important purpose of the study is to monitor a representative sample of a defined population, the question of representativeness and how to cope with it can assume great significance. Using the case of age to illustrate the point, it is clear that as the total sample size grows through the addition of new annual samples there will be an upward shift in the age distribution. The cumulated sample of loyalists will grow older, and the annual addition of new respondents will not be sufficient to redress the balance. The problem becomes even more difficult as attrition introduces other biases and contributes to further departures of loyal respondents from representativeness of the universe under study.

Within-sample differences between initial and loyal respondents have not been calculated in this analysis, although this could be done without difficulty. Unhappily, however, there are no clear-cut guidelines for determining how great a difference can be tolerated, so that such within-sample comparisons would not presently be of much help. The existence of a significant difference between loyal and dropout sub-groups indicates that there is some bias due to differential dropout. Whether or not this bias has an effect on the follow-up results depends upon the proportion of dropouts. In order for the follow-up results to be biased substantially there must be both a high dropout rate and a substantial difference between loyalists and dropouts. In most cases where there is a significant difference between these two sub-groups the frequency distributions for the loyalists are still very close to those of the initial sample.

An independent test of the representativeness of the 1970 follow-up is currently being made by comparing the distributions of characteristics with those in the distributed 1970 Census tapes. This will provide a test for the combined effects of sampling error, original non-response, and attrition. By comparing the 1970 follow-up data for the three samples with the 1970 Census we can answer the question of when the dropout rate begins to affect substantially the representativeness of the loyal respondents.

If substantial departures from representativeness are indicated there are two general strategies that might be employed to deal with the matter. One of these strategies would be to drop the affected panel(s) --- probably the oldest. The decision to drop panels would be guided by age of the panels and/or some selected level of dropout.

The second strategy (that might be employed in conjunction with the first) would be to replenish the eroded samples with new respondents. It would be important to insure the suitability of the new respondents for this purpose and to test whether new respondents were, indeed, well matched with the dropouts they will represent. At the present time (August 1971) an experimental study of this replenishment strategy is under way.

A sample of replenishment respondents for Sample III dropouts is being drawn, and they will be interviewed in the current Fall follow-up interview round. Sampling procedures employed in the original Sample III selection are to be repeated, with some modifications discussed below. Criteria for a replenishment interview are as follows:

a. residence at an address that appeared on the original sample selection list, if the current occupants were not living in the dwelling unit at the time of initial interviewing. An exception is made in the case of a household still remaining if that address did not yield an initial interview due to refusal. The purpose of this rule is to allow original refusals, and the occupants of previously vacant addresses, and <u>new</u> occupants of a vacated interview address to fall into the sample;

b. residence at a selected dwelling unit that was constructed since the original sample was drawn. Such new construction will be placed at the bottom of the original list and the original selection interval applied; and

c. residence in an initial sample household (family) still remaining at the original address, <u>if</u> the initial respondent has for any reason left the household permanently.

In all cases the original criteria for respondent eligibility and selection will be applied. The procedures outlined here will be applied to a randomly selected half of the original sample segments for purposes of economy. An examination of the characteristics of initial respondents from the selected and unselected segments has revealed no differences that could be detected, and it is believed that this 'split-halving' will not introduce bias into the replenishment sample. After weighting to account for the half-sample employed the replenishment sample will be compared with the dropout component (the 65 dropouts identified at the end of the Fall 1970 round plus those newly identified at the end of the current round), and the results reported.

<u>Summary</u>

The increasing need for the surveillance of sample populations in studies of social change and for the planning, execution and evaluation of public programs is presented. Whether for theoretical or for applied purposes, however, this need implies some form of panel design, and the problem of panel attrition and its consequences is discussed. The current panel study at the Population Research Laboratory at Brown University is described, and a method for measuring and coping with attrition is suggested. Preliminary findings and plans for further analysis are presented in the hope that a standardized method for measuring, dealing with, and reporting attrition can be adopted by social researchers who employ the panel survey design in their work.

Acknowledgments

The study reported here is supported by United States Public Health Service Grant HS-00246. The author wishes to thank his colleagues Professors Speare and Roa for their helpful comments and suggestions on an earlier version of this paper presented at the meeting of the American Statistical Association at Fort Collins, Colorado, August 1971.

References

[1] Organic, Harold N. and Goldstein, Sidney, "The Brown University, Rhode Island Population Research Laboratory: Its Purposes and Initial Progress," in The Community as an Epidemiological Laboratory: A Casebook of Community Studies, Irving I. Kessler and Morton L. Levin, eds., (Baltimore: The Johns Hopkins Press, 1970), pp. 212-232.

[2] Kish, Leslie, Survey Sampling, (New York: John Wiley & Sons, 1965).

[3] User's Manual for PSALMS Computer Program, mimeographed, no author indicated (Ann Arbor: Survey Research Center, The University of Michigan, 1971).

INFERENCE FROM CLUSTERED SAMPLE

Martin R. Frankel, University of Michigan and CUNY

Social scientists often analyze a probability sample of a population rather than a complete census. They then face the task of inferring from sample results to the results they would have obtained if the analysis were performed over the entire population.

One of the keys to making these needed inferences is the availability of methods for estimating the sampling errors of the sample derived population estimates, coupled with generalizations that can be made about the distribution of these population estimates and their estimated sampling errors.

In this paper, we lay out three general methods that may be used to estimate sampling errors of population estimates from complex (clustered and stratified) probability samples. For each of these methods, we describe their implementability and discuss their reliability and validity. This discussion of reliability and validity is based on recently completed empirical research.

For the sake of expositional simplicity, let us assume a sample design that calls for the selection of two primary sampling units (psu's) from each of H strata. It is assumed that there are A primary units in each stratum, and that the selection of two of these A units is made by simple random sampling without replacement. Thus, we have a clustered and stratified sampling design where each population element has equal probability (f) of appearing in the sample (self-weighted sample). It should be noted that any of the three variance estimation methods can be generalized to accomodate unequal allocation between strata, PPS (or any non-epsem) selection of psu's within strata, as well as subsampling of psu's.

To avoid confusion, I will be using the term first-order to describe sample estimates g(S), where g is a function and S is a sample, and corresponding population parameters g(P), where P is the population, that are of primary interest to the substantive analyst. Some examples of these first-order estimates and parameters are ratio means, differences of ratio means, totals, ratios of ratios, simple correlations, partial correlations, multiple correlations, and regression coefficients (simple, multiple, path, MCA, dummy variable, etc.). The term second-order is used to describe estimates, also made from the sample, of the sampling variability (error) of the first-order estimates.

The three second-order estimation techniques described in this paper are labelled the Taylor expansion method (TAYLOR), the method of balanced repeated replication (BRR), and the method of jack-knife repeated replication (JRR). It should be noted, however, that this scheme of appellation is not unique.

[1] *This research was carried out under a Joint Research Project with the U.S. Bureau of the Census. This article draws from results to be found in* (3).

Taylor Expansion Method

The use of the Taylor expansion for obtaining an estimate of the variance of the first-order estimate of a ratio mean has been known for some time. All sampling textbooks describe its use in this context. Deming [2] and Kish [7] describe its use in the propagation of variance for other functions of the basic sample sums. The method is also known as the "delta" or δ-method, or simply as the linearization method. However, to my knowledge, the first detailed published extension, specific to survey sampling, of this method to more complex first-order estimates is due to Tepping [17].

When this method is used, we produce an approximate estimate of the sampling variance of a sample function that is the linear or first term of the Taylor series expansion of the first-order sample estimate of interest.

There are actually two, and sometimes three, approximation assumptions that are made when this method is used. Following Tepping's paper, the method can be described as follows:

> Let $y = (y_1,\ldots,y_k)$ be a vector of sample statistics that are linear combinations of the primary sampling unit values y_{iha}, where h is the index over strata and a is the psu within strata index. That is,
>
> $$y_i = \sum_{h=1}^{H} \sum_{a=1}^{2} y_{iha} = \sum_{h=1}^{H} y_{ih} . \quad (1)$$
>
> Similarly, let $E(y) = Y = (Y_1,\ldots,Y_k)$ be the corresponding vector of population values. Also, let g(Y) be the first-order parameter we wish to estimate by the first-order sample estimation function g(y).

The first assumption to be made is that the sampling variance of g(y) is approximately equal to the sampling variance of the first degree terms of the Taylor approximation of g(y) near Y. That is,

$$\mathrm{VAR}(g(y)) \doteq \mathrm{VAR}(g(Y) + \sum_{i=1}^{k} (y_i - Y_i) \frac{\partial g(Y)}{dY_i}), \quad (2)$$

where the partial derivatives are evaluated at y = Y. Since the terms g(Y) and $Y_i(\partial g(Y)/dY_i)$ are constant over all samples, this reduces to

$$\mathrm{VAR}(g(y)) \doteq \mathrm{VAR}(\sum_{i=1}^{k} (\frac{\partial g(Y)}{dY_i}) y_i) . \quad (3)$$

The terms y_i and Y_i are linear combinations of corresponding values of the psu's, thus of the stratum values $y_{ih.}$ and $Y_{ih.}$. Because selection is assumed independent between strata, we may rewrite (3) as

$$VAR(g(y)) \doteq \sum_{h=1}^{H} W_h^2 \, VAR\left(\sum_{i=1}^{k} \frac{\partial g(Y)}{dY_i} y_{ih.}\right) \quad , (4)$$

where W_h is the constant "weight" of the h^{th} stratum.

If there are two primary units selected without replacement and with equal probabilities f_h, from each stratum, then we may estimate the variance of the $y_{ih.}$'s by

$$var\left(\sum_{i=1}^{k} \frac{\partial g(Y)}{dY_i} y_{ih.}\right) =$$

$$(1-f_h)\left\{\sum_{i=1}^{k} \frac{\partial g(Y)}{dY_i} y_{ih1} - \sum_{i=1}^{k} \frac{\partial g(Y)}{dY_i} y_{ih2}\right\}^2 , (5)$$

where y_{ih1} and y_{ih2} are the sample totals from the two psu's of the h^{th} stratum. Keyfitz [6] called early attention to this simple form for two primary sampling units. If $W_h = 1$ and $f_h = f$ for all h=1,...,H, our estimate of VAR(g(y)) is

$$var(g(y)) =$$

$$(1-f) \sum_{h=1}^{H} \left\{\sum_{i=1}^{k} \frac{\partial g(Y)}{dY_i} y_{ih1} - \sum_{i=1}^{k} \frac{\partial g(Y)}{dY_i} y_{ih2}\right\}^2 . (6)$$

In order to use this estimate, we should ideally have values for the constants $\partial g(Y)/dY_i$. Of course, if these were known, we would probably know g(Y) and would not need to make the estimate g(y). These constants must be estimated from the sample at hand. Thus, we have the second approximation assumption associated with this method.

It is commonly assumed that this substitution of sample values for population values does not greatly increase the error in this estimate of variance. However, this is only a conjecture.

Balanced Repeated Replication (BRR) Methods

More complete descriptions and discussions of balanced repeated replication methods for computing estimates of sampling errors have already appeared in a number of developmental papers [4, 5, 8, 9, 10, 11, 12, 14]. However, as is the case with the other two variance estimation methods, with the exception of an unpublished study by Tepping [16] concerning the behavior of the Taylor expansion estimates of the variance of simple ratios and the research herein described [3], no empirical data have been collected that deal with the validity and precision of these methods.

The variance estimates produced by BRR can be described as follows: Assume that we have a stratified sample design with two primary sampling units selected from each stratum with equal probability f (srs). Let S denote the entire sample; let H_i denote the i^{th} half-sample formed by including one of the two primary units in each of the strata; and let C_i denote the i^{th} complement half-sample, formed by the primary units in S not in H_i.

If we form k half-samples $H_1,\ldots,H_k$ and corresponding complement half-samples $C_1,\ldots,C_k$, then we may produce four BRR type estimates of variance of the first-order estimate g(S).

Half Minus Total - $var_{BRR-H}(g(S)) =$

$$\frac{(1-f)}{k} \sum_{i=1}^{k} (g(H_i)-g(S))^2 \qquad (7)$$

Complement Minus Total - $var_{BRR-C}(g(S)) =$

$$\frac{(1-f)}{k} \sum_{i=1}^{k} (g(C_i)-g(S))^2 \qquad (8)$$

Sum of BRR-H and BRR-C - $var_{BRR-S}(g(S)) =$

$$\frac{var_{BRR-H}(g(S)) + var_{BRR-C}(g(S))}{2} \qquad (9)$$

Half Minus Complement - $var_{BRR-D}(g(S)) =$

$$\frac{(1-f)}{4k} \sum_{i=1}^{k} (g(H_i) - g(C_i))^2 \qquad (10)$$

There are several methods for choosing the pattern of primary units that form the repeated half and complement half samples H_i and C_i. The method used in my empirical research is known as "full-orthogonal balance." For a more complete description of the method, see [10, 11].

As previously noted, each of the second-order estimates actually estimates the variance of a linearized form of the first-order estimate. In the case of the four BRR estimates, this linearization of g(S) is

$$\frac{1}{2k} \sum_{i=1}^{k} (g(H_i) + g(C_i)) \quad . \qquad (11)$$

Because of the interchangeability of H_i with C_i, the forms BRR-H and BRR-C possess the same expectation. As a result, their mean, BBR-S, shares this equality. The BRR-H and BRR-C forms should be viewed as estimates of BRR-S, which are less costly to compute. For the moment, we will eliminate the -H and -C forms of BRR from our discussion. If the function (11) has bias which is linearly decreasing in the number of primary sampling units, the BRR-S form (as well as the -H and -C forms) gives an unbiased estimate of the mean square error of (11).

Under any circumstances, the form BBR-D is an unbiased estimate of variance for (11) [8, 9].

Jack-Knife Repeated Replication (JRR) Methods

The term jack-knife repeated replication describes a set of second-order estimation methods that were motivated by the Tukey jack-knife estimation procedure [1, 18] and by BRR.

With BRR methods, each of the k replications estimates the variance of the entire sample. With the JRR methods, each replication gives us a measure of the variance contributed by a single stratum. The technique used to measure this stratum variance contribution was suggested by the Tukey jack-knife method for variance estimates formed by leaving out replicates of the sample. The specific procedures described below appear in the literature for the first time here.

JRR estimates of the variance are computed as follows: Assume that we have an epsem, stratified sample design with two primary sampling units selected with equal probability f, from each of H strata. Let S denote the entire sample; let J_i, (i=1,...,H), denote the replicate formed by removing from S one of psu's in the i^{th} stratum, and including twice the other psu in the i^{th} stratum. Let CJ_i, (i=1,...,H) denote the complement replicate formed from S by interchanging the psu's in the i^{th} stratum that are eliminated and duplicated. The four JRR estimates of the variance of the first-order estimate g(S) are:

Estimate I (JRR-H)

$$var_{JRR-H}(g(S)) = (1-f)\sum_{i=1}^{H}(g(J_i)-g(S))^2 \quad (12)$$

Estimate II (JRR-C)

$$var_{JRR-C}(g(S)) = (1-f)\sum_{i=1}^{H}(g(CJ_i)-g(S))^2 \quad (13)$$

Estimate III (JRR-S)

$$var_{JRR-S}(g(S)) = \frac{var_{JRR-H}(g(S))+var_{JRR-C}(g(S))}{2} \quad (14)$$

Estimate IV (JRR-D)

$$var_{JRR-D}(g(S)) = \frac{(1-f)}{4}\sum_{i=1}^{H}(g(J_i)-g(CJ_i))^2 \quad (15)$$

From (15), the linearization associated, in a loose fashion, with the JRR estimates is of the form

$$\frac{1}{2H}\sum_{i}^{H}(g(J_i)+g(CJ_i)) \quad . \quad (16)$$

As is the case with BRR, the JRR forms suffixed with -H and -C share the same expectation with each other and with JRR-S. These two former forms should be considered as cheaper to compute but less precise forms of JRR-S. As we did with BRR-H and BRR-C, the forms JRR-H and JRR-C will be, for the moment, eliminated from our discussion.

Implementability

For all three variance estimation methods (TAYLOR-1 form only; BRR-2 forms, BRR-S and BRR-D; JRR-2 forms, JRR-S and JRR-D), as applied to first-order estimates which are functions of total sample first and second moments, much of the cost of computation is directly related to the number of strata; not to the total sample size. All three methods require only one pass (by the computer) over the entire set of individual cases. In this single pass sums, sums of squares, and sums of cross products, are computed for each of the 2 x H psu's that constitute the entire sample. A simultaneous accumulation over psu's yields intermediate statistics for the total sample. All subsequent computations are performed on these psu and total sample "intermediate statistics."

If the TAYLOR method of variance estimation is used, the intermediate statistics for the total sample, the y_i's in (1), are used to produce the sample estimates of the required partial derivatives $\partial g(Y)/dY_i$. Given these partials, we then use the psu intermediate statistics to form the terms $\sum_{i=1}^{k}(\frac{\partial g(Y)}{dY_i})y_{iha}$ for each psu. The paired squared differences of these terms (6) yield the estimate of sampling error.

When the BRR method of variance estimation is used, one pass over the set of 2 x H psu intermediate statistics is required to form a half-sample and its complement. The half-sample intermediate statistics are formed by the accumulation of one of the two sets of psu intermediate statistics from each of the H strata. The complement half-sample intermediate statistics are formed by subtracting the half-sample intermediate statistics from the intermediate statistics for the total sample. The required first-order estimates g(S), $g(H_i)$, and $g(C_i)$ are produced from the intermediate statistics from the total sample, the i^{th} half-sample, and the i^{th} complement half-sample. These terms are manipulated as in (9) and (10) to form the BRR-S or BRR-D estimate of sampling error.

The computation required for the JRR estimates of sampling error are essentially the same as those required for BRR, with the exception of the formation of the replicates and complement replicates. To form the i^{th} replicate, we subtract from the total sample intermediate statistics the intermediate statistics from one of the psu's in the i^{th} stratum, and add to this the intermediate statistics from the other psu in the stratum. Reversing the labeling of the psu's within the i^{th} stratum, we repeat this procedure to form the

i^{th} complement replicate.

In terms of time requirements, the TAYLOR method of variance estimation is optimal for relatively simple first-order estimates. This includes simple ratio means, differences of ratio means, and simple ratios of ratios. The TAYLOR method begins to lose its time advantage when the computations required to make sample estimates of the partial derivatives become more time-consuming than the time required to form the half or replicate samples. Although the point at which this occurs is somewhat dependent on the number of strata, we have found that the computation of sampling errors for simple correlation coefficients and simple or multiple regression coefficients is equally time-consuming with all three methods. For even more complex first-order estimates, the expression of the partial derivatives in closed form may be beyond our mathematical ability and in this case we must use either BRR or JRR.

At the University of Michigan Survey Research Center, we have not as yet found these forms for partial and multiple correlation coefficients, although this certainly does not mean that they do not exist.

This final observation points out a strength of JRR and BRR methods for variance estimation. If we can specify the first-order estimate g(S) and if we can assume that g(S) is reasonably close to $(g(H_i) + g(C_i))/2$, then we can compute an estimate of the sampling error of g(S) with BRR or JRR.

Reliability and Validity

So far I have described three methods of estimating sampling errors and have commented on their implementability and relative costs. Now we must deal with the question of how well these estimates perform. It would have been preferable if we had general analytic and non-assymtotic comparisons of these three methods. However, to date, efforts in this area have not yielded useful results. Following a tradition among statisticians that goes back at least as far as 1907, when W.S. Gossett, writing under the name "Student," selected 750 simple random samples from a population of criminals' left middle finger measurements in order to evaluate his theoretical derivation of the distribution of the sample mean divided by its estimated standard error [15], I empirically compared and evaluated all three variance estimation methods, using three clustered and stratified sample designs which called for the paired selection of primary sampling units (approx. 14 elements) from 6 strata (approx. 170 elements), 12 strata (approx. 340 elements) and 30 strata (approx. 847 elements). For a more complete description of this study, which made use of data from the Current Population Survey of the U.S. Bureau of the Census, the reader is directed to Frankel [3]. The three methods (five variants: TAYLOR, BRR-S, BRR-D, JRR-S, JRR-D) were used to estimate the sampling error of simple ratio means, differences of ratios, simple correlations, and multiple regression coefficients. BRR and JRR methods were used to estimate sampling errors for partial and multiple correlation coefficients.

Several criteria were used in evaluating the relative merits of the variance estimation methods. First, we looked at their bias, their variance and their mean squared error. None of the three methods appeared to be singularly optimal under any of these criteria.

Somewhere along the line, we realized that none of these criteria actually told us what we wanted to know. We decided that the designation of a statistically best variance estimation technique should be based on a measure of how well the technique allowed the analyst to make valid inference statements about first-order estimates. Put another way, we dècided that our prime interest was not in variance estimation, per se, but in variance estimation as an imput to inference statements.

For this reason, we chose as our ultimate evaluative criteria the degree to which these three variance estimation techniques would yield estimates, var (g(S)), that made the approximation

$$\frac{g(S) - E(g(S))}{\sqrt{var(g(S))}} \sim t_{(H)} \qquad (17)$$

most valid. For each of the five estimation forms, TAYLOR, BRR-S, BRR-D, JRR-S, and JRR-D, we computed the proportion of times this ratio, $(g(S) - E(g(S))/\sqrt{var(g(S))}$, computed for each sample selected under a particular design, fell within the symmetric limits $\pm$ 2.576, $\pm$ 1.960, $\pm$ 1.645, $\pm$ 1.280, and $\pm$ 1.000. Table 1 shows these proportions when this ratio is distributed exactly as a Student's t random variable, and Tables 2-6 show these proportions for the five different variance estimation forms. Since the expected proportions vary with the degrees of freedom, in this case equal to the number of strata, these proportions are shown separately for each of the three sample designs studied. The proportions were averaged over first-order estimates of the same type. There were 6 means, 12 differences of means (D.MEANS) and simple correlations (CORR.S), 8 multiple regression coefficients (BETAS), 6 partial correlation coefficients (PARTIAL R.S) and 2 multiple correlation coefficients (MULTIPLE R.) involved in these averages.

For all types of first-order estimates studied, we find the average proportions (rounded to two places) produced by the BRR-S estimates (Table 3) agree at least as well, and in most cases better, with proportions predicted by Student's t, than proportions produced by any of the other variance estimation methods (TAYLOR, BRR-D, JRR-S, JRR-D).

Although there is some variability between first-order estimate types and between the various sample designs (sizes), the proportions produced with BRR-S estimates, within symmetric intervals, agree excellently with those predicted by Student's t for all first-order estimates except the multiple correlation coefficients (See Tables 1 and 3).

Although the BRR-S method does produce estimates of variance that are optimal under the criteria we have chosen, we find that the other methods are often very close seconds. A measure of this closeness is given in Table 7 which is derived from Tables 2-6. Happily, this table indicates that when we are dealing with first-order estimates that are ratio means and differences of means, all methods perform about equally well. Thus, given the research at hand, we can tentatively recommend the following optimal (both in terms of computing costs and our chosen statistical criteria) strategy be followed for producing sampling errors of first-order population estimates.

1. Use the Taylor method for ratio means, differences of ratios and other similar forms.
2. Use BRR-S for more complex regression-related statistics; correlations and regression coefficients.
3. Given 1 and 2, one can feel fairly safe in using the approximation

$$\frac{g(S) - E(g(S))}{\sqrt{var(g(S))}} \sim t_{(H)}$$

in order to generate either classical or Bayesian inference statements.

TABLE 1

PROPORTION OF STUDENT'S T AREA WITHIN SELECTED INTERVALS

Degrees Of Freedom	Intervals				
	±2.576	±1.960	±1.645	±1.282	±1.000
6	.9580	.9023	.8489	.7529	.6441
12	.9757	.9264	.8741	.7760	.6630
30	.9848	.9407	.8896	.7903	.6747
∞	.9900	.9500	.9000	.8000	.6827

TABLE 2

SAMPLE ESTIMATE - EXPECTED VALUE, DIVIDED BY TAYLOR ESTIMATE OF STANDARD ERROR PROPORTION OF TIMES WITHIN STATED LIMITS

Statistic(s)	±2.576	±1.960	±1.645	±1.282	±1.000
6 STRATA DESIGN					
Means	0.9483	0.8879	0.8329	0.7379	0.6279
D. Means	0.9450	0.8842	0.8372	0.7381	0.6306
Corr.S	0.9158	0.8367	0.7744	0.6708	0.5631
Betas	0.9421	0.8733	0.8146	0.7167	0.6029
Partial R.S.					
Multiple R.					
12 STRATA DESIGN					
Means	0.9712	0.9192	0.8646	0.7625	0.6542
D. Means	0.9653	0.9078	0.8525	0.7539	0.6358
Corr.S	0.9333	0.8589	0.8028	0.7050	0.5992
Betas	0.9662	0.9121	0.8496	0.7437	0.6217
Partial R.S.					
Multiple R.					
30 STRATA DESIGN					
Means	0.9819	0.9431	0.8881	0.7844	0.6537
D. Means	0.9821	0.9433	0.8842	0.7742	0.6429
Corr.S	0.9650	0.8983	0.8362	0.7225	0.6025
Betas	0.9787	0.9319	0.8837	0.7781	0.6612
Partial R.S.					
Multiple R.					

TABLE 3

SAMPLE ESTIMATE - EXPECTED VALUE, DIVIDED BY BRR-S ESTIMATE OF STANDARD ERROR
PROPORTION OF TIMES WITHIN STATED LIMITS

6 STRATA DESIGN

Statistic(s)	±2.576	±1.960	±1.645	±1.282	±1.000
Means	0.9558	0.9042	0.8450	0.7562	0.6450
D. Means	0.9500	0.8997	0.8497	0.7578	0.6483
Corr.S	0.9475	0.8864	0.8358	0.7386	0.6250
Betas	0.9662	0.9150	0.8600	0.7683	0.6642
Partial R.S.	0.9567	0.9083	0.8550	0.7661	0.6511
Multiple R.	0.9350	0.8950	0.8233	0.7383	0.6133
		12 STRATA DESIGN			
Means	0.9721	0.9221	0.8700	0.7692	0.6612
D. Means	0.9658	0.9117	0.8617	0.7614	0.6458
Corr.S	0.9553	0.8967	0.8439	0.7578	0.6397
Betas	0.9733	0.9337	0.8746	0.7733	0.6529
Partial R.S.	0.9661	0.9117	0.8694	0.7544	0.6250
Multiple R.	0.9200	0.8500	0.7900	0.6767	0.5500
		30 STRATA DESIGN			
Means	0.9825	0.9444	0.8906	0.7894	0.6569
D. Means	0.9829	0.9462	0.8875	0.7783	0.6475
Corr.S	0.9725	0.9108	0.8617	0.7533	0.6325
Betas	0.9825	0.9381	0.8900	0.7887	0.6706
Partial R.S.	0.9550	0.8967	0.8442	0.7533	0.6450
Multiple R.	0.9125	0.8250	0.7350	0.6375	0.5275

TABLE 4

SAMPLE ESTIMATE - EXPECTED VALUE, DIVIDED BY BRR-D ESTIMATE OF STANDARD ERROR
PROPORTION OF TIMES WITHIN STATED LIMITS

6 STRATA DESIGN

Statistic(s)	±2.576	±1.960	±1.645	±1.282	±1.000
Means	0.9533	0.8996	0.8404	0.7487	0.6379
D. Means	0.9481	0.8950	0.8450	0.7503	0.6436
Corr.S	0.9411	0.8761	0.8189	0.7131	0.6069
Betas	0.9587	0.8996	0.8433	0.7446	0.6446
Partial R.S.	0.9467	0.8900	0.8283	0.7272	0.6111
Multiple R.	0.9033	0.8217	0.7583	0.6417	0.5467
		12 STRATA DESIGN			
Means	0.9721	0.9208	0.8687	0.7667	0.6579
D. Means	0.9656	0.9097	0.8594	0.7586	0.6422
Corr.S	0.9492	0.8883	0.8344	0.7397	0.6264
Betas	0.9700	0.9250	0.8654	0.7646	0.6412
Partial R.S.	0.9583	0.9006	0.8456	0.7267	0.6011
Multiple R.	0.9067	0.8150	0.7400	0.6067	0.5067
		30 STRATA DESIGN			
Means	0.9819	0.9437	0.8894	0.7881	0.6569
D. Means	0.9825	0.9454	0.8867	0.7779	0.6462
Corr.S	0.9696	0.9083	0.8550	0.7467	0.6212
Betas	0.9812	0.9369	0.8881	0.7831	0.6687
Partial R.S.	0.9533	0.8925	0.8350	0.7433	0.6300
Multiple R.	0.8975	0.8100	0.7175	0.6125	0.4975

TABLE 5

SAMPLE ESTIMATE - EXPECTED VALUE, DIVIDED BY JRR-S ESTIMATE OF STANDARD ERROR
PROPORTION OF TIMES WITHIN STATED LIMITS

6 STRATA DESIGN

Statistic(s)	±2.576	±1.960	±1.645	±1.282	±1.000
Means	0.9508	0.8942	0.8362	0.7421	0.6329
D. Means	0.9464	0.8939	0.8397	0.7428	0.6367
Corr.S	0.9311	0.8633	0.8047	0.6992	0.5906
Betas	0.9521	0.8833	0.8304	0.7312	0.6200
Partial R.S.	0.9367	0.8683	0.8100	0.7050	0.5950
Multiple R.	0.9117	0.8400	0.7800	0.6600	0.5600
12 STRATA DESIGN					
Means	0.9712	0.9200	0.8662	0.7650	0.6554
D. Means	0.9653	0.9083	0.8558	0.7561	0.6375
Corr.S	0.9439	0.8750	0.8261	0.7308	0.6167
Betas	0.9675	0.9162	0.8542	0.7496	0.6283
Partial R.S.	0.9494	0.8883	0.8256	0.7106	0.5822
Multiple R.	0.8950	0.8133	0.7383	0.6333	0.5167
30 STRATA DESIGN					
Means	0.9819	0.9431	0.8887	0.7856	0.6537
D. Means	0.9821	0.9433	0.8842	0.7742	0.6433
Corr.S	0.9658	0.9021	0.8471	0.7346	0.6137
Betas	0.9800	0.9325	0.8844	0.7787	0.6631
Partial R.S.	0.9458	0.8792	0.8192	0.7250	0.6183
Multiple R.	0.8950	0.7925	0.7025	0.5950	0.4950

TABLE 6

SAMPLE ESTIMATE - EXPECTED VALUE, DIVIDED BY JRR-D ESTIMATE OF STANDARD ERROR
PROPORTION OF TIMES WITHIN STATED LIMITS

6 STRATA DESIGN

Statistic(s)	±2.576	±1.960	±1.645	±1.282	±1.000
Means	0.9500	0.8912	0.8337	0.7396	0.6329
D. Means	0.9458	0.8889	0.8389	0.7400	0.6353
Corr.S	0.9292	0.8553	0.7944	0.6892	0.5814
Betas	0.9454	0.8796	0.8258	0.7262	0.6142
Partial R.S.	0.9300	0.8561	0.7961	0.6906	0.5772
Multiple R.	0.8850	0.8033	0.7350	0.6133	0.5133
12 STRATA DESIGN					
Means	0.9712	0.9196	0.8658	0.7642	0.6550
D. Means	0.9653	0.9083	0.8544	0.7558	0.6369
Corr.S	0.9428	0.8719	0.8211	0.7217	0.6108
Betas	0.9671	0.9142	0.8508	0.7471	0.6250
Partial R.S	0.9450	0.8800	0.8133	0.6994	0.5694
Multiple R.	0.8850	0.7933	0.7067	0.5933	0.4950
30 STRATA DESIGN					
Means	0.9819	0.9431	0.8881	0.7850	0.6537
D. Means	0.9821	0.9433	0.8842	0.7742	0.6433
Corr.S	0.9658	0.9008	0.8442	0.7333	0.6112
Betas	0.9794	0.9319	0.8844	0.7787	0.6619
Partial R.S	0.9450	0.8775	0.8158	0.7225	0.6108
Multiple R.	0.8875	0.7925	0.6975	0.5825	0.4700

TABLE 7

AVERAGE DEVIATION OF PROPORTIONS FROM THOSE PRODUCED BY BRR-S ESTIMATES
(IN UNITS OF 0.01)

First Order Estimates	Second Order Estimate															
	Taylor				BRR-D				JRR-S				JRR-D			
	Number Of Strata				Number Of Strata				Number Of Strata				Number Of Strata			
	6	12	30	Total	6	12	30	Total	6	12	30	Total	6	12	30	Total
Means	1.2	0.6	0.4	0.7	0.2	0.0	0.0	0.1	1.0	0.0	0.2	0.4	1.2	0.2	0.2	0.5
Differences of Means	0.8	0.8	0.6	0.7	0.0	0.2	0.0	0.1	0.6	0.2	0.6	0.5	1.2	0.2	0.6	0.5
Simple Correlations	5.8	4.0	1.8	3.9	1.8	1.2	0.2	1.1	3.4	2.0	1.4	2.3	4.0	2.8	1.4	2.7
Regression Coefficients	3.2	2.0	0.8	2.0	0.4	0.4	0.4	0.4	1.8	1.4	0.8	1.3	2.0	1.4	0.8	1.4
Partial Correlation Coefficients	-	-	-	-	1.6	1.8	1.0	1.5	3.2	3.4	2.0	2.9	3.4	4.4	2.4	3.4
Multiple Correlation Coefficients	-	-	-	-	7.0	4.0	2.2	4.4	5.4	3.8	3.2	4.1	9.4	6.1	4.4	6.6

NOTE: Total is the average for all three sample designs.

In the few cases where BRR-S proportions were greater than Student's t values, the deviation was measured from the hypothesized t value.

REFERENCES

1. Brillinger, D.R., "The Application of the Jack-Knife to the Analysis of Sample Surveys." Commentary, 8 (1966), 74-80.
2. Deming, W.E., Sample Design in Business Research. New York: John Wiley and Sons, 1960.
3. Frankel, M.R., Inference from Survey Samples: An Empirical Investigation. Ann Arbor: Institute for Social Research, University of Michigan, 1971.
4. Gurney, M., "The Variance of the Replication Method for Estimating Variances from the CPS Design." Dittoed memorandum, U.S. Bureau of the Census, 1962.
5. Gurney, M., "McCarthy's Orthogonal Replications for Estimating Variances, with Grouped Strata." Dittoed memorandum, U.S. Bureau of the Census, 1964.
6. Keyfitz, N., "Estimates of Sampling Variance When Two Units are Selected from Each Stratum." Journal of the American Statistical Association, 52 (1957), 503-510.
7. Kish, L., Survey Sampling. New York: John Wiley and Sons, 1965.
8. Kish, L., and Frankel, M., "Balanced Repeated Replication for Analytical Statistics." Proceedings of the Social Statistics Section of ASA, 1968.
9. Kish, L., and Frankel, M., "Balanced Repeated Replications for Standard Errors." Journal of the American Statistical Association, 65 (1970), 1071-1094.
10. McCarthy, P.J., Replication: An Approach to the Analysis of Data from Complex Surveys. Washington: National Center for Health Statistics, Series 2, No. 14, 1966.
11. McCarthy, P.J., Pseudo-Replication: Further Evaluation and Application of the Balanced Half-Sample Technique. Washington: National Center for Health Statistics, Series 2, No. 31, 1966.
12. McCarthy, P.J., "Pseudo-Replication: Half Samples," Review of the International Statistical Institute, 37 (1969), 239-264.
13. Plackett, R.L. and Burman, P.J., "The Design of Optimum Multifactorial Experiments." Biometrika, 33 (1946), 305-325.
14. Simmons, W.R. and Baird, J., "Use of Pseudo-Replication in the NCHS Health Examination Survey." Proceedings of the Social Statistics Section of ASA, 1968.
15. "Student," "The Probable Error of a Mean." Biometrika, 6 (1908), 1-25.
16. Tepping, B.J., internal memorandum, U.S. Bureau of the Census, 1967.
17. Tepping, B.J., "The Estimation of Variance in Complex Surveys." Proceedings of the Social Statistics Section of the American Statistical Association, 1968.
18. Tukey, J.W., "Bias and Confidence in Not-Quite Large Samples: Abstract." Annals of Mathematical Statistics, 29 (1958), 614.

SAMPLING ERRORS IN PERIODIC SURVEYS

I.P. Fellegi and G.B. Gray, Statistics Canada

Introduction

The Canadian Labour Force Survey is the largest continuing household survey conducted by Statistics Canada. Each month about 13,000 households are selected in cities and towns 15000 and over in a two-stage sample of city blocks and households within and about 13,000 are selected in 139 geographic strata in the rural areas in four stages beginning with 2 primary sampling units in each stratum. More details are found in [1].

Estimation and Variance Estimation

Totals of characteristics are estimated by formula (1) of the appendix for each of ten provinces in Canada. This is a multiple ratio estimation formula using projected population estimates by age-sex within each province.

The variance of each characteristic is defined by formula (2) while the calculation is accomplished by formula (3) which may be abbreviated by formula (3a).

Covariances between estimates one or more months apart (commonly twelve months' intervals are considered) are calculated by formula (3b) by letting X and Y refer to estimates of a given characteristic of two different months.

Variances of the average of successive months (say 3 or 12 months) may be readily found by formula (3c) by putting a_i = 1/3 or 1/12 and x_i the estimate for any one of the successive months being averaged.

Each month variance estimates are obtained for over 40 characteristics and the results are produced in TABLES A and B as illustrated in the appendix. Similar tables are also produced for each province and regions comprising groups of provinces.

Also variance estimates of differences, quarterly and annual averages are produced for the same characteristics and an illustration of the variance of the annual average is given in TABLE G in the appendix.

Uses of the Monthly Variance Calculations

(1) Consultation by subject matter analysts: TABLES A at the Canada level and sometimes at provincial level are distributed to different departments who consult these regularly along with the monthly bulletin. Departments include Economic Council of Canada, Bank of Canada, Department of Finance, Department of Labour, and the Department of Manpower and Immigration.

(2) Attachment of alphabetic code to each characteristic in the monthly bulletin to warn users of the extent of sampling error the estimates may be subject to. These codes range from "a" for the smallest coefficient of variation (such as Total Employed at the Canada level to "g" for the largest coefficient of variation (such as Temporary Layoffs in a small region) and the definitions are in the appendix. These are not revised each month but each year by averaging the coefficients of variation for the preceding calendar year.

(3) Ongoing control of the survey. Two replicated population estimates within each stratum should be almost the same if the measures of size used to derive inverse probabilities for blowing up the sample counts in each replicate were up-to-date. Varying growth rates and operational problems such as errors in the field result in large differences between the population estimates attributable to other than sampling variability of sample take. A monthly control and feedback system is set up for the control and correction of such operational problems as pinpointed by the monthly variance programs. For 300 monthly replicated population estimates, investigations are undertaken whenever the differences exceed fifty percent and in 1970 this phenomenon occurred in more than ten percent of the replications. Operational errors occurred in about 1/5 of these or about 6 to 8 errors each month, indicating over time at any rate a significant effect upon the operation as a result of the regular control and feedback.

(4) Time Series Information on the Behaviour of the Survey Design. Each variance estimate is divided by the variance estimate for an unstricted random sample to determine what Kish calls in [3] the design effect and what we have been calling the binomial factors. These are given in TABLE B (See appendix). These binomial factors provide useful information on the behaviour of survey design, its possible deterioration over time, the effectiveness of various measures introduced during the life time of the survey aimed at diminishing the variance and even possibly to a better understanding of the behaviour of the labour force characteristics themselves. While variances tend to increase, the coefficients of variation decrease as the sample size or the size of estimates increases. The binomial factors thus tend to remain stationary with respect to a given sample selection and estimation procedure so that the factors measure the sample design performance with the effect of sample size and size of estimate as it varies over time removed.

Further Remarks on the Binomial Factors

Since 1966, a monthly time series of binomial factors has been built up for over 40 characteristics at the Canada and provincial levels, broken down by self and non-self representing areas. The binomial factors measure the combined effects of stratification, sampling with probabilities proportional to size, and multistage sampling upon the variances of the estimates compared with the values when simple random sampling is effected. Stratification and sampling with probabilities proportional to size tend to reduce the variance if these are effective while multi-stage sampling tends to increase it, the extent of this increase being dependent on the degree of clustering.

Analysis of Binomial Factors

In order to derive any meaningful analysis of the binomial factors, it was necessary to separate self and non-self representing areas. The reason for this lies in the two domains being different and in the sample design pertaining to the two domains being distinct. The inflation of the variance in city blocks of cities and towns over the random sample variance of a similar estimate tends to be lower than the inflation of the variance in the PSUs of the rural and small urban areas so that the factors in the cities and towns tend to be lower. With the two types of areas separated, useful and revealing information on the design performance may be obtained from a series of factors with or without seasonal adjustments. A few illustrations will be given, having the impact of some adjustments in the design as well as that of some substantive developments.

Illustrations of the Use of Binomial Factors in Cities and Towns (SRU Areas)

In the cities and towns three major events occurred in the last few years, all three of which would be expected to reduce the binomial factors: in August 1968 the sample was reduced by a factor of 1/3 but this was carried out by reducing the within-block sample only, thus reducing clustering; in August 1969 to December 1969 a revision of the size measures was carried out and this was repeated in the period July to December 1970.

Table I below illustrates the binomial factors for Canada SRU Employed from January 1966 to December 1970 without seasonal adjustment. The data was seasonally adjusted (not shown) and despite the lack of evidence of significant seasonality, some interesting information was revealed. The seasonally adjusted factors averaged 1.06 in the interval Jan-July 1968 but dropped to 0.99 in the interval Aug-Dec 1968 presumably as a result of reduction in clustering. The size measure revision appeared to have had little effect in Aug.1969 but a slight one after July 1970. This does not mean that the program for size revision is not beneficial. Total employed in cities might be too aggregated a measure to affect significantly the binomial factors. If we look at Table 2 below, where the binomial factors are averaged for each calendar year between 1966 and 1970 for various disaggregations of employed, it is found that the tendency is for the factors to increase from 1966 to 1968 and then decline in 1969 and 1970. This phenomenon is particularly noticeable in "Construction" and "Community, Business, and Public Services", industries whose employees one would expect to find in the suburbs. It might thus be hypothesized that revision of size measures has a particularly beneficial effect on the variance of those characteristics one intuitively associates with the suburbs.

For "Unemployed" and "Unemployed Men", the size measure revisions appeared to have benefited these characteristics in 1969 but not in 1970. However, the average figure in 1970 is misleading since the general level of an estimate is not supposed to affect the binomial factor but a special effect has occurred which has been revealed in the X-11 seasonally adjusted data. Unemployed appears to have become more clustered, at least in the cities, resulting in higher binomial factors as indicated by higher than average seasonal factors in December and January, especially in the winters of 1969-70 and 1970-71. Thus the recent high levels of unemployment is seen to be accompanied by increased clustering of the unemployed. This is an example of how an analysis of binomial factors might shed some light on some substantive developments related to shifts in the geographical concentration of people with certain characteristics.

It appears that subject matter analysts of the labour force scene are unaware of this analytical tool. Further examples of this pertaining to the cities are (1) High binomial factors for "Construction" consistently in April, indicating an uneven start in construction, (2) High binomial factors for "Teenage Employed" in May and August, indicating uneven commencement and termination of summer employment, (3) High binomial factors for Agriculture Employed from December to March indicating a higher clustering effect in the winter months. We can only offer passing judgement on the value of the above information but the analysis might be explored further.

Use of the Binomial Factors in Rural Areas

In the rural areas, there has been a noticeable but irregular increase in the binomial factors between 1966 and 1970, for Total Employed and Employed by various industries. This increase is presumably attributable to a gradual deterioration of stratification and in the measures of size used for selection purposes at various stages of sampling. Some seasonalities also occur in the rural areas such as: (1) "Manufacturing" with a high binomial factor in the winter months, attributable perhaps to rural population commuting to large cities in the winter months, (2) "Unemployed", as in the cities has higher binomial factors in the summer. Another interesting phenomenon with respect to "Unemployed" has been a reduction in the binomial factors over the past two years. So it appears that, without any changes in the sample design in the rural areas since redesign, the level of high unemployment in the past couple years has, in the rural areas, become more widespread and hence less clustered than in earlier years, resulting in lower binomial factors. A third case is "Teenage Unemployed" with high binomial factors in June rather than in May as in the cities and towns, thus indicating a tendency for rural teenagers to look for work closer to the end of the school year.

Possible Extension of Binomial Factor Calculations

The binomial factors show the combined effect of stratification, clustering, ratio estimation, etc. and it is difficult to separate these out. Alternative binomial factors are given in (i) formula (8a) in the appendix in which the variance of a stratified random sample estimate is shown, and by dividing the actual variance by (8a), one arrives at a new binomial factor (8b) which is free of the effect of stratification but combines the impact of all other design factors, (ii) formula (9) which is a "pure" stratification index, (iii) formula (10) which assumes unrestricted random sampling in the whole province but ratio estimation as in formula (1) and (iv) formula (11), which provides a "pure" ratio estimation index.

Cost of Variance Estimation Programme

The annual cost of running the programmes for which both variances and covariances are produced monthly is about $22,000 annually, compared with $1.5 million for the annual budget of the Field Operation and processing connected with the Labour Force Survey as a whole.

Table 1: Canada SRU Employed Binomial Factors

Month	1966	1967	1968	1969	1970
January	1.000	0.810	1.110	0.850	0.960
February	0.970	0.820	0.900	0.720	0.920
March	0.850	0.920	1.060	0.900	0.910
April	0.940	0.870	1.080	0.900	0.890
May	0.840	0.910	1.080	0.910	0.930
June	0.990	1.000	0.920	0.860	1.110
July	1.030	0.860	1.070	0.840	0.900
August	1.120	1.020	0.940	0.970	0.810
September	0.900	0.890	1.020	0.850	0.830
October	1.140	0.970	1.150	0.890	1.060
November	0.910	1.000	0.950	0.910	0.980
December	0.720	1.170	1.050	0.890	1.010
Average	0.951	0.937	1.028	0.874	0.943

Table 2: Average Annual Binomial Factors by Year and Type of Area for Specified Characteristic

Characteristic	NSRU Areas					SRU Areas				
	1966	1967	1968	1969	1970	1966	1967	1968	1969	1970
Employed	1.582	1.502	1.480	1.537	1.738	0.951	0.937	1.028	0.874	0.943
Employed Agric.	3.545	4.440	4.684	4.115	4.501	3.163	3.281	4.383	4.718	3.368
Employed Non-Agric.	2.379	2.584	2.429	2.313	1.917	1.013	1.017	1.093	0.932	0.949
Other Prim.Ind.	3.485*	4.234	4.953	4.449	5.032	3.708*	1.698	1.827	1.822	1.883
Manufacturing	2.713*	3.291	3.343	2.956	3.292	1.336*	1.301	1.614	1.404	1.347
Construction	2.002*	2.302	1.846	2.023	1.582	1.382*	1.243	1.351	1.278	1.260
Comm,Bus.Pers.Serv.	2.541*	2.621	2.495	2.755	2.883	1.929*	1.651	2.069	1.527	1.457
Public Admin. and Defence	2.255*	2.428	2.224	2.571	2.374	1.240*	1.691	1.548	1.915	1.742
Married Women in L.F.	1.509	1.763	1.410	1.373	1.357	1.072	0.994	1.162	0.927	0.942
Unemployed	2.269	2.366	2.233	2.162	2.005	1.358	1.347	1.466	1.398	1.514
Unemployed Women	1.269	1.535	1.507	1.403	1.682	1.214	1.167	1.243	1.165	1.261
Unemployed Men	2.327	2.195	2.270	2.022	1.756	1.329	1.320	1.342	1.365	1.423
Unemployed 14-19	1.381	1.530	1.519	1.646	1.563	1.255	1.216	1.293	1.163	1.274
Seeking <1 month	1.715	1.644	1.744	1.503	1.638	1.279	1.262	1.253	1.249	1.253
Seeking >6 months	1.690	1.928	1.650	2.131	1.886	1.163	1.265	1.346	1.341	1.249

* 10 months only (January and February excluded)

Appendix
Formulas Used in Estimation and Variance Estimation

Consider a characteristic (such as Unemployed, Employed, or any minor characteristic) whose sample total in a balancing unit b (sub-unit or sections of large cities or metropolitan areas, urban or rural portions of selected primary sampling units) with type of area j pertaining to age-sex category a = x_{jba}.

Let p_{jba} = age-sex category "a" sample total in (p,j),

P_a = projected Census population in age-sex category a at provincial level,

W_j = weight pertaining to type of area j (large cities, urban or rural), and

B_{jb} = balancing unit b factor to compensate for non-response.

Then estimated total for a province is given by:

(1) $\hat{X} = \sum_a P_a \, (\hat{X}_a/\hat{P}_a)$, where

(1a) $\hat{X}_a = \sum_j W_j \sum_b B_{jb} \, x_{jba}$ and similarly for $\hat{P}_a$.

For characteristics pertaining to only some of the age-sex categories, only those age-sex categories referred to in the characteristic are summed over in (1).

(2) $V(\hat{X}) = \sum_a P_a^2 R_a^2$ Rel.Var $(\hat{X}_a/\hat{P}_a) + \sum_{a=a'} P_a P_{a'} R_a R_{a'}$ Rel.Cov. $(\hat{X}_a\hat{P}_a)\cdot(\hat{X}_{a'}/\hat{P}_{a'})$ where $R_a = X_a/P_a$ and is estimated by $\hat{R}_a = \hat{X}_a/\hat{P}_a$.

(3) $\hat{V}(\hat{X}) = \sum_h [\Delta\hat{X}_h - \sum_a \hat{R}_a \, \Delta\hat{P}_{ha}]^2$, h being a stratum or group of sub-units, where $\hat{X}_{hk} = \sum_a P_a \, (\hat{X}_{hak}/\hat{P}_a)$ and $\hat{P}_{hak} = \sum_{b\varepsilon h} \sum_j W_j B_{jb} \, x_{jba(k)}$ and similarly for $\hat{P}_{hak}$; $\Delta\hat{X}_h = \hat{X}_{h1} - \hat{X}_{h2}$ and similarly for $\Delta\hat{P}_{ha}$. k=1 or 2 denotes either the two selected primary sampling units of a stratum or two groups of selected segments in several sub-units. "h" here denotes a group of sub-units or stratum instead of an individual sub-unit as in the definition for the estimate. Hence, the term "paired area" for h in the variance estimation formula.

By defining $D_{hx} = \Delta\hat{X}_h - \sum_a \hat{R}_a \, \Delta\hat{P}_{ha}$, we may

(3a) write $\hat{V}(\hat{X}) = \sum_h D_{hx}^2$ and

(3b) $\hat{CV}(\hat{X},\hat{Y}) = \sum_h D_{hx} D_{hy}$

Let $X^* = \sum_i a_i X_i$, a linear combination of estimates which may or may not be correlated.

Then $V(\hat{X}^*) = \sum_i a_i^2 V(\hat{X}_i) + 2 \sum_i \sum_{j>i} a_i a_j \, CV(\hat{X}_i,\hat{X}_j)$

and $\hat{V}(\hat{X}^*) = \sum_i a_i^2 \sum_h D_{hx_i}^2 + 2 \sum_i a_i a_j \sum_h D_{hx_i} D_{hx_j}$

by substituting 3a and 3b.

Finally, the estimated variance of $\hat{X}^*$ may be simplified to:

(3c) $\hat{V}(\hat{X}^*) = \sum_h [\sum_i a_i D_{hx_i}]^2$ so that to find the variance of any linear combination of estimates with constant factors a_i such as $\hat{X}^* = \sum_i a_i \hat{X}_i$, we may define a corresponding D-value by:

(3d) $D_{hx}^* = \sum_i a_i D_{hx_i}$

Design Effect

A random sample variance within type of area j (urban, rural self-representing areas) derived by assuming a random sample of $1/W_j$ of each individual without replacement in type of area j is given by:

(4) $\hat{B}_j = (W_j - 1) \hat{X}_j (1 - \hat{X}_j/\hat{P}_j)$, neglecting the variance in $\hat{P}_j$ and the factor $\hat{P}_j/(\hat{P}_j - 1)$ employed in the variance corresponding to a hypergeometric distribution.

Here, $\hat{X}_j = \sum_a (P_a/\hat{P}_a) \, W_j \sum_b B_{bj} \, x_{jba}$ and similarly for $\hat{P}_j$.

At the self and non-self representing area level T say

(5) $\hat{B}_T = \sum_{j\varepsilon T} \hat{B}_j$ and by defining the variance estimate for type of area T,

(6) $\hat{V}_T = \sum_{j\varepsilon T} D_{hx}^2$, we can derive a binomial factor given by

(7a) $F_T = \hat{V}_T/\hat{B}_T$ for a specific type of area level and at the provincial level p.

(7b) $F_p = \sum_{T\varepsilon p} \hat{V}_T / \sum_{T\varepsilon p} \hat{B}_T$ and similarly for regions (groups of provinces) and the Canada level by adding the variances and binomial variances.

Other binomial variances and hence binomial factors may be derived. These have not yet been tried out but are presented here as a matter of interest.

Assuming a simple random sample of $1/W_h$ of persons in each stratum, we may derive the corresponding binomial variance by:

(8a) $\hat{B}_{ST} = \sum_{h \in T} (W_h - 1) \hat{X}_h (1 - \hat{X}_h/\hat{P}_h)$. and

the binomial factor by:

(8b) $F_{ST} = \hat{V}_T/\hat{B}_{ST}$.

Also, the ratio of two binomial variances may be calculated to permit a "pure" stratification index pertaining to type of area T as given by:

(9) $I_{ST} = \hat{B}_T/\hat{B}_{ST}$.

Finally, to permit measurement of the effect of ratio estimation, one can derive a variance assuming unrestricted random sampling and undertaking a ratio estimate procedure and the variance is given by:

(10) $\hat{B}_R = (W_{p-1}) \sum_a \hat{X}_{pa} (1-\hat{X}_{pa}/P_{pa})$,

appropriate only at the provincial level since no ratio estimation and the effectiveness of the ratio estimation may be obtained by deriving a so-called ratio estimate index given by:

(11) $I_R = \hat{B}_p^*/\hat{B}_R$ where $\hat{B}_p^* = (W_{p-1}) \hat{X}_p$ $(1 - \hat{X}_p/P_p)$ based on the assumption of unrestricted simple random sampling in the whole province instead of summing binomial variances over types of areas as now undertaken.

Below are illustrations of tables that are produced monthly or annually.

EXAMPLE 1: TYPICAL PAGE IN THE MONTHLY LABOUR FORCE BULLETIN SHOWING LETTERED SYMBOLS FOR STANDARD DEVIATIONS

Table 1 Summary	S.D. (1)	1971 July 24	1971 June 19	1970 July 18	1970 June 20	1969 July 19	1969 June 21
Total							
Population 14 years of age and over(2)		15,408	15,372	15,030	15,000	14,651	14,619
Labour force	a	9,068	8,859	8,819	8,677	8,550	8,403
Employed	a	8,554	8,308	8,301	8,148	8,201	8,020
Agriculture	d	612	544	619	569	644	580
Non-agriculture	a	7,942	7,764	7,682	7,579	7,557	7,440
Unemployed	d	514	551	518	529	349	383
Not in the labour force	a	6,340	6,513	6,211	6,323	6,101	6,216
Participation rate (3)	a	58.9	57.6	58.7	57.8	58.4	57.5
Unemployment rate (4)							
Actual	d	5.7	6.2	5.9	6.1	4.1	4.6
Seasonally adjusted		6.3	6.4	6.6r	6.3	4.6r	4.8
Men							
Population 14 years of age and over(2)		7,632	7,614	7,448	7,433	7,262	7,246
Labour force	a	6,137	5,989	6,020	5,885	5,867	5,743
Employed	a	5,769	5,595	5,649	5,501	5,613	5,469
Agriculture	d	519	469	526	487	543	493
Non-agriculture	a	5,251	5,126	5,123	5,013	5,070	4,976
Unemployed	d	368	394	371	384	254	274
Not in the labour force	b	1,495	1,625	1,428	1,548	1,395	1,503
Participation rate (3)	a	80.4	78.7	80.8	79.2	80.8	79.3
Unemployment rate (4)	d	6.0	6.6	6.2	6.5	4.3	4.8

(1) "S.D." = Standard deviation. For explanation, see "Reliability of Estimates", page 8.
(2) Excludes inmates of institutions, members of the armed services, Indians living on reserves and residents of the Yukon and Northwest Territories.
(3) The labour force as a percentage of the population 14 years of age and over.
(4) The unemployed as a percentage of the labour force.
r Revised.
Note: With the exception of Table 2, all statistics refer to a specific week, the last day of which is indicated. The sums of individual items may not always equal the total because of rounding.

EXAMPLE 2: EXPLANATION OF SAMPLING ERRORS AND NON-SAMPLING ERRORS GIVEN ON LAST PAGE OF MONTHLY BULLETIN

Reliability of Estimates

(a) Sampling Error

The estimates in this report are based on a sample of households. Somewhat different figures might have been obtained if a complete census had been taken using the same questionnaires, interviewers, supervisors, processing, etc. as those actually used in the Labour Force Survey. This difference is called the sampling error of the estimates. In the design and processing of the Labour Force Survey extensive efforts have been made to minimize the sampling error. The sampling error (expressed as a per cent of the estimate it refers to) is not the same for all estimates; of two estimates the larger one will likely have a smaller per cent sampling error and of two estimates of the same size the one referring to a characteristic more evenly distributed across the country will tend to have a smaller per cent sampling variability. Also, estimates relating to age and sex are usually more reliable than other estimates of comparable size.

(b) Non-sampling Errors

Errors, which are not related to sampling, may occur at almost every phase of a survey operation. Interviewers may misunderstand instructions, respondents may make errors in answering questions, the answers may be incorrectly entered on the questionnaires and errors may be introduced in the processing and tabulations of the data. All these errors are called non-sampling errors. Some of the non-sampling errors will usually balance out over a large number of observations but systematically occuring errors will contribute to biases. Non-sampling errors can be reduced by a careful design of questionnaires, intensive training and supervision of interviewers and a thorough control of the processing operation. In general, the more personal and more subjective inquiries are subject to larger errors. Also, data referring to persons with less stable labour force status will have relatively large non-sampling errors.

(c) Alphabetic Indicators of Standard Deviation

The sampling error, as described under (a) is not known. A quantity, called the standard deviation, can however be estimated from sample data itself. The standard deviation of an estimate is a statistical measure of its sampling error. It also partially measures the effect on non-sampling errors, but does not reflect any systematic biases in the data. The chances are about 68 out of 100 that the difference between a sample estimate and the corresponding census figure would be less than the standard deviation. The chances are about 95 out of 100 that the difference would be less than twice the standard deviation and about 99 out of 100 that it would be less than 2 1/2 times as large.

The standard deviations of the estimates, expressed as a per cent of the estimates, are indicated by letters. The letter "a" indicates that the standard deviation is smaller than 0.5% of the estimate, the letter "b" indicates that the standard deviation is between 0.6% and 1.0% of the estimate and so on as shown in the table below.

Alphabetic designation of per cent standard deviations

Alphabetic indicator	Per cent standard deviation
a	0.0% - 0.5%
b	0.6% - 1.0%
c	1.1% - 2.5%
d	2.6% - 5.0%
e	5.1% - 10.0%
f	10.1% - 15.0%
g	15.1% -

The actual standard deviation of an estimate is not the same each month. Since the standard deviations of the current estimates are not available at the time when this report is published, the alphabetic indicators are based on the average standard deviations during the last year. They should, therefore, be interpreted only as indications of the order of magnitude of the standard deviations.

(d) Standard Deviation of Month-to-Month Changes

A rough upper limit for the standard deviation of the difference (change) between two estimates referring to two months up to a year apart may also be indicated using the table above. For most characteristics published in this report the standard deviation of the difference between two estimates is likely to be somewhat smaller than the standard deviation of the smaller of the two estimates or in the immediately preceding range.

For example, suppose that a hypothetical estimate in May and June was 513,000 and 625,000 respectively and the per cent standard deviation of both estimates was indicated by the letter "c", i.e. it was between 1.1% and 2.5%. The difference between the May and June estimates (112,000) would, therefore, have a standard deviation which would likely be smaller than 2.5% of 513,000, i.e. it would likely be smaller than 12,800.

(e) Current Estimates of Standard Deviations

Standard deviations are computed monthly for several estimates and month-to-month changes. These are available usually in a few weeks after the publication of this report and can be obtained on request. Beginning with 1966, an annual report on the standard deviations during the last year will be released.

EXAMPLE 3: ANALYSIS OF WEIGHTED AVERAGE FROM SURVEY 235 TO SURVEY 246 OR FROM JAN. 1970 TO DEC. 1970. WEIGHTS AT BOTTOM OF PAGE. TABLE G

CANADA

CHARACTERISTIC	WTD. AVER.	S.D. AVER.	PER CENT S.D.	VARIANCE OF AVER. NSRU	VARIANCE OF AVER. SRU
1. TOTAL	15016.1	0.0	0.0	0.0	0.0
2. EMPLOYED	7879.5	19.45	0.2	158.6	219.8
3. UNEMPLOYED	494.5	7.40	1.5	16.0	38.8
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
44. UNEMPLOYED, MEN	373.6	6.21	1.7	11.2	27.3
45. UNEMPLOYED, WOMEN	120.9	2.81	2.3	2.3	5.6
46. TOTAL MEN	7440.9	0.0	0.0	0.0	0.0

WTD. AVER. AND S.D. AVER. IN THOUSANDS. VARIANCE IN MILLIONS
DENOMINATOR OF WEIGHTS = 12000
NUM. WTS. = 1 1 1 1 1 1 1 1 1 1 1 1 0

EXAMPLE 4: MONTHLY VARIANCE SUMMARY TABLE B FOR SURVEY 247 JAN. 1971

CANADA

CHARACTERISTIC	ESTIMATE		FACTORS VAR.EST/BIN.VAR.		
	NSRU	SRU	NSRU	SRU	COMB.
1. TOTAL	4842.2	10379.8	0.0	0.0	0.0
2. EMPLOYED	2151.4	5516.3	1.94	0.95	1.16
3. UNEMPLOYED	240.6	427.6	2.18	1.23	1.47
4. NOT IN LABOUR FORCE	2450.2	4435.8	1.37	0.83	0.95
5. EMPLOYED, MEN	1567.0	3535.5	1.08	0.43	0.57
6. EMPLOYED, WOMEN	584.4	1980.8	1.44	0.81	0.92
7. EMPLOYED, AGRICULTURE	405.7	32.6	5.02	3.47	4.80
8. EMPLOYED, NON-AGRICULTURE	1745.7	5483.7	2.21	0.96	1.21
9. USUALLY WORK 35-99 HRS.	1527.9	4779.0	1.99	0.87	1.09
10. AT WORK 35 HRS. OR MORE	1357.6	4353.2	1.97	0.90	1.10
11. AT WORK LESS THAN 35 HRS.	170.2	425.8	2.59	1.20	1.47
12. DUE TO ECONOMIC REASONS	25.8	58.1	1.43	1.04	1.12
13. DUE TO OTHER REASONS	144.4	367.7	2.95	1.22	1.55
14. USUALLY WORK 0-34 HRS.	217.8	704.7	1.72	1.12	1.22
15. WITHOUT WORK & SEEKING WRK	214.7	395.3	2.44	1.19	1.50
16. SEEKING FULL TIME WORK	210.4	374.4	2.45	1.16	1.49
17. SEEKING PART TIME WORK	4.3	20.9	1.58	1.17	1.22
18. ON TEMPORARY LAYOFF	25.9	32.3	3.37	1.99	2.43
19. SEEKING UNDER 1 MONTH	42.2	88.0	1.80	1.29	1.40
20. SEEKING 1-3 MONTHS	106.3	148.9	2.56	1.21	1.61
21. SEEKING 4-6 MONTHS	35.1	81.6	1.84	1.42	1.51
22. SEEKING MORE THAN 6 MOS	31.2	76.9	1.80	1.19	1.32
23. PERSONS 14-19, EMPLOYED	205.7	439.4	1.39	0.82	0.94
24. PERSONS 14-19, UNEMPLOYED	52.4	76.0	1.43	0.95	1.09
25. PERSONS 14-19, NOT IN L.F.	656.2	1092.2	0.45	0.37	0.39
26. UNEMPLOYED,CONSTRUCTION	51.0	90.9	1.72	0.89	1.09
27. UNEMPLOYED,LABOURERS	40.1	61.4	1.30	0.94	1.04
28. OTHER PRIMARY INDUSTRIES	115.2	89.3	4.25	1.72	2.85
29. MANUFACTURING	394.2	1320.3	2.44	1.38	1.56
30. CONSTRUCTION	120.0	268.9	1.75	1.26	1.36
31. TRANSP.& OTHER UTILITIES	176.6	500.8	2.11	1.58	1.68
32. TRADE	318.4	1009.6	1.77	1.31	1.39
33. FINANCE INSUR.& REAL ESTATE	51.4	331.8	1.87	1.33	1.38
34. COMMUNITY,BUS.& PERS. SER	474.1	1576.3	2.31	1.14	1.34
35. PUBLIC ADMIN. & DEFENCE	95.8	386.8	2.48	1.75	1.85
36. IN LABOUR FORCE,MEN	1768.8	3851.8	0.58	0.34	0.39
37. IN LABOUR FORCE,WOMEN	623.2	2092.1	1.33	0.74	0.84
38. IN LABOUR FORCE,MALES 55&	300.4	557.1	0.68	0.28	0.38
39. IN L.F.,MARRIED WOMEN	397.3	1167.8	1.10	0.90	0.93
40. PAID NON-AGRIC. WORKERS	1520.6	5116.1	1.93	1.22	1.36
41 PAID WORKERS	1575.9	5134.6	1.86	1.20	1.33
42. PAID WORKERS,MEN	1090.7	3246.9	1.27	0.60	0.72
43. PAID WORKERS,WOMEN	485.2	1887.8	1.44	0.89	0.97
44. UNEMPLOYED,MEN	201.7	316.3	1.95	1.23	1.43
45. UNEMPLOYED,WOMEN	38.8	11.3	1.39	1.12	1.17
46. TOTAL MEN	2481.6	5059.8	0.0	0.0	0.0

ESTIMATES IN THOUSANDS

EXAMPLE 5: MONTHLY VARIANCE SUMMARY TABLE A FOR SURVEY 247 JAN. 1971

CANADA

	CHARACTERISTIC	EST.	S.D.	PER CENT S.D.	VARIANCE ESTIMATES NSRU	SRU
1.	TOTAL	15222.0	0.0	0.0	0.0	0.0
2.	EMPLOYED	7667.7	31.64	0.41	361.5	639.9
3.	UNEMPLOYED	668.2	14.50	2.17	78.7	131.4
4.	NOT IN LABOUR FORCE	6886.0	28.51	0.41	259.5	553.2
	.	.	.	.	.	.
	.	.	.	.	.	.
	.	.	.	.	.	.
	.	.	.	.	.	.
44.	UNEMPLOYED,MEN	518.1	12.53	2.42	58.9	98.2
45.	UNEMPLOYED,WOMEN	150.1	6.46	4.31	9.0	32.8
46.	TOTAL MEN	7541.4	0.0	0.0	0.0	0.0

EST AND S.D. EQUALS ESTIMATE AND STANDARD DEVIATION IN THOUSANDS
VARIANCE ESTIMATE IN MILLIONS

References

[1] Fellegi, I.P., Gray, G.B. and Platek, R.: "The New Design of the Canadian Labour Force Survey", Journal of the American Statistical Association, 62(1967), pp 421-453.

[2] Keyfitz, N.: "Estimates of Sampling Variance where Two Units are Selected from Each Stratum", Journal of the American Statistical Association, 52(1957), pp 503-510.

[3] Kish, L.: "Survey Sampling", J. Wiley and Sons Ltd., p.88 etc.

JACKKNIFING FOR VARIANCE COMPONENTS IN COMPLEX SAMPLE SURVEY DESIGNS

Ralph E. Folsom, David L. Bayless, and Babu V. Shah
Research Triangle Institute, Research Triangle Park, N. C. 27709

1. INTRODUCTION

The variance components methodology presented in sections two and three of this paper extends results of Seeger (1970) which were developed for sampling designs with equal probability selection from effectively infinite populations at each stage of sampling. We have shown that Seeger's simple analysis of unweighted means also works for linear statistics from a class of highly stratified three stage designs allowing PPS selection. The trick will be to use properly expanded-up last stage responses as the basic variables of analysis.

In order to accommodate the non-linear statistics which are commonly used with such complex designs, we have developed in section four a multi-stage extension of the Quenouille (1956)-Tukey (1958) Jackknife.

2. NOTATION AND MODEL

The class of sampling designs that we have considered are stratified three stage designs with PPS selection at the first two stages and equal probability sampling at the last stage. To simplify our presentation, we assume that first stage units are sampled "with" replacement and are subsampled independently each time they are selected. Second and third stage units are selected "without replacement."

To establish the link with Seeger's variance components methodology, we will work with the expanded up last stage responses in equation (1).

$$y_{ijk} = M_{ij}Y_{ijk}/p_i p_{j/i} \tag{1}$$

where the small p's are relative size measures for the first and second stage units and M_{ij} is the number of third stage units in the (ij)-th secondary unit. The cap-Y represents some characteristic of population unit (ijk). Notice that if only one unit was selected at each stage of sampling then y_{ijk} would be the Horvitz-Thompson (1952) estimator for the population total

$$Y_{+++} = \sum_{i=1}^{N} \sum_{j=1}^{S_i} \sum_{k=1}^{M_{ij}} Y_{ijk} \tag{2}$$

In general, the Horvitz-Thompson estimator for Y_{+++} can be written as the average of our small-y variables; that is,

$$\hat{Y}_{+++} = y_{...} = \sum_{i=1}^{n} \sum_{j=1}^{s_i} \sum_{k=1}^{m_{ij}} y_{ijk}/ns_i m_{ij} \tag{3}$$

We have defined five variance components associated with the various stages of sampling in our design. Four of these components can be defined simply in terms of the "effects" presented in the "model identity" of equation four.

$$y_{ijk} = \mu + \rho_i + \eta_{j/i} + \varepsilon_{k/ij} \tag{4}$$

where

$$\mu = Y_{+++}$$

$$\rho_i = (Y_{i++}/p_i - Y_{+++})$$

$$\eta_{j/i} = (Y_{ij+}/p_i p_{j/i} - Y_{i++}/p_i)$$

$$\varepsilon_{k/ij} = (y_{ijk} - Y_{ij+}/p_i p_{j/i}).$$

For a balanced sample selected "with replacement" at each stage, one can show that

$$\mathrm{Var}(\hat{Y}_{+++} = y_{...}) = \sigma_P^2/n + \sigma_{S/P}^2/ns + \sigma_{K/S}^2/nsm \tag{5}$$

where the components are defined

$$\sigma_P^2 = \sum_{i=1}^{N} p_i \rho_i^2$$

$$\sigma_{S/P}^2 = \sum_{i=1}^{N} \sum_{j=1}^{S_i} p_i p_{j/i}\, \eta_{j/i}^2 = \sum_{i=1}^{N} p_i \sigma_{S/P}^2(i)$$

$$\sigma_{K/S}^2 = \sum_{i=1}^{N} \sum_{j=1}^{S_i} \sum_{k=1}^{M_{ij}} p_i p_{j/i}\, \varepsilon_{k/ij}^2/M_{ij}$$

$$= \sum_{i=1}^{N} \sum_{j=1}^{S_i} p_i p_{j/i}\, \sigma_{K/S}^2(ij).$$

For our "without" replacement sampling at stages two and three, we need two more components. The second stage component involves normalized joint inclusion probabilities $\theta_{jj'/i} = \{\pi_{jj'/i}/s_i(s_i-1)\}$ and squared differences

$\varepsilon_{jj'/i}^2 = (Y_{ij+}/p_i p_{j/i} - Y_{ij'+}/p_i p_{j'/i})^2$ as follows

$$V_{S/P}^2 = \sum_{i=1}^{N} p_i \sum_{j=1}^{S_i} \sum_{j'<j} \theta_{jj'/i}\, \varepsilon_{jj'/i}^2$$

$$= \sum_{i=1}^{N} p_i V_{S/P}^2(i).$$

The third stage "without replacement" component is

$$V^2_{K/S} = \sum_{i=1}^{N} \sum_{j=1}^{S_i} P_i P_{j/i} \sum_{k=1}^{M_{ij}} \varepsilon^2_{k/ij}/(M_{ij}-1)$$

$$= \sum_{i=1}^{N} \sum_{j=1}^{S_i} P_i P_{j/i} V^2_{K/S}(ij). \tag{7}$$

The variance of the Horvitz-Thompson estimator for a balanced version of our design can now be written in the simple form of equation (8) where the cap-sigmas are linear composites of our five separate components

$$\mathrm{Var}(\hat{Y}_{+++} = y_{...}) = \Sigma^2_P/n + \Sigma^2_{S/P}/ns + \Sigma^2_{K/S}/nsm \tag{8}$$

with

$$\Sigma^2_P = \{\sigma^2_P - (V^2_{S/P} - \sigma^2_{S/P})\}$$

$$\Sigma^2_{S/P} = \{V^2_{S/P} - (V^2_{K/S} - \sigma^2_{K/S})\}$$

$$\Sigma^2_{K/S} = V^2_{K/S} .$$

For explicit derivations of these results, see Folsom, Bayless, and Shah (1971).

3. UNBIASED ESTIMATION

With the sampling structure and components definitions outlined in section 2, we can show that the following simple unbiased estimators are available for our cap-sigmas

$$\hat{\Sigma}^2_P = (MS_P - M'_{S/P})$$

$$\hat{\Sigma}^2_{S/P} = (MS_{S/P} - MS'_{K/S}) \tag{10}$$

$$\hat{\Sigma}^2_{K/S} = MS_{K/S}$$

where the MS's denote the following "analysis of unweighted means" type mean squares

$$MS_P = \sum_{i=1}^{n} (y_{i..} - y_{...})^2/(n-1)$$

$$MS_{S/P} = \sum_{i=1}^{n} \sum_{j=1}^{s_i} (y_{ij.} - y_{i..})^2/n(s_i-1)$$

$$MS'_{S/P} = \sum_{i=1}^{n} \sum_{j=1}^{s_i} (y_{ij.} - y_{i..})^2/ns_i(s_i-1) \tag{11}$$

$$MS_{K/S} = \sum_{i=1}^{n} \sum_{j=1}^{s_i} \sum_{k=1}^{m_{ij}} (y_{ijk}-y_{ij.})^2/ns_i(m_{ij}-1)$$

$$MS'_{K/S} = \sum_{i=1}^{n} \sum_{j=1}^{s_i} \sum_{k=1}^{m_{ij}} (y_{ijk}-y_{ij.})^2/ns_i m_{ij}(m_{ij}-1)$$

The derivation of expected mean squares which leads to the estimators in equation (10) is detailed in Folsom, Bayless, and Shah (1971). Unbiased estimates for the five separate components are also presented in the report cited above.

4. MULTIPLE-STAGE JACKKNIFING

Our contribution to the Jackknife procedure involves partitioning the variance of a non-linear statistic such as $\hat{\theta}$ in equation (12)

$$\hat{\theta} = f[y_{+...}(1), \cdots, y_{+...}(g)] \tag{12}$$

into components like our cap-sigmas. The (plus) on the little y's (sample totals) in (11) indicate summation over h = 1(1)H strata. Estimates for θ are first formed from pseudo-replicates obtained by successively deleting the data from sampling units at a particular level of the design. These estimates as they occur in equation (13) are subscripted by a minus sign followed by labels for the deleted sampling unit.

$$J\theta_{hijk} = n_h s_{hi} m_{hij}\hat{\theta} - (n_h-1)\hat{\theta}_{-hi} - n_h(s_{hi}-1)\hat{\theta}_{-hij}$$

$$- n_h s_{hi}(m_{hij}-1)\hat{\theta}_{-hijk} \tag{13}$$

Equations (14) and (15) demonstrate the form of the replicate estimator when a first stage unit is deleted

$$\hat{\theta}_{-hi} = f\{Y_{-hi}(1), \ldots, Y_{-hi}(y)\} \tag{14}$$

with

$$Y_{-hi}(r) = y_{+...}(r) - [y_{hi..}(r) - y_{h...}(r)]/(n_h-1) \tag{15}$$

If results from classical theory hold up in this finite population context, then we would expect the average of our pseudo-values shown in equation (16) to have less bias than $\hat{\theta}$ in (12).

$$\hat{\theta}_{JK} = \sum_{h=1}^{H} \sum_{i=1}^{n_h} \sum_{j=1}^{s_{hi}} \sum_{k=1}^{m_{hij}} J\theta_{hijk}/Hn_h s_{hi} m_{hij} \tag{16}$$

$$= J\theta_{....}$$

For a linear statistic, the Jackknife estimate in (15) reduces to (12). To estimate variance components for the jackknifed statistic, we substitute unweighted means of the pseudo-values into the mean squares in equations (11). This is spelled out for the first-stage component in equation (16)

$$J\Sigma_P^2 = (JMS_P - JMS'_{S/P}) \qquad (16)$$

where

$$JMS_P = \sum_{h=1}^{H} \sum_{i=1}^{n_h} (J\theta_{hi..} - J\theta_{h...})^2/(n_h - 1)$$

$$JMS'_{S/P} = \sum_{h=1}^{H} \sum_{i=1}^{n_h} \sum_{j=1}^{s_{hi}} (J\theta_{hij.} - J\theta_{hi..})^2/n_h s_{hi}(s_{hi} - 1)$$

5. EMPIRICAL RESULTS

The P-Values in Table I represent ratio estimates computed from a stratified three-stage sample of High School Seniors conducted by the Research Triangle Institute for the National Center for Educational Statistics. Although the small sample sizes involved in this pretest make it impossible to draw any general empirical conclusions, it is interesting to note that

1. The Jackknife and Standard P-Values are numerically equivalent, indicating little or no bias in the combined ratio estimate.

2. The Jackknife Components for the last two stages are numerically equivalent to corresponding "Taylor Series" estimates with only a slight difference at the PSU stage.

The "Taylor Series" linearization alluded to in point 2 above is a direct extension of Tepping's (1968) results to our variance components setting.

6. DISCUSSION

Although our variance components methodology was developed for a particular sample, it applies to a fairly wide class of stratified three-stage designs. The "with replacement" at the first-stage simplifies the mean squares, but it is not crucial to the application of our Multi-stage Jackknife. This Jackknife shares with Taylor series linearization the property of producing a pseudo-value which is associated with a particular last stage unit. By borrowing the form of variance and variance components estimators already available for linear statistics, the Jackknife and Taylor series linearizations provide direct extensions of these results to non-linear statistics. Our limited empirical results show that these two methods produce very similar results for ratios. In summary, we feel that the Jackknife replication technique with our extension will prove to be a very useful method of variance and variance components estimation for complex sample statistics.

REFERENCES

1. Folsom, Ralph E., Bayless, David L., and Shah, Babu V. Jackknifing for Variance Components in Complex Sample Survey Designs. Presented at the American Statistical Association Meeting at Fort Collins, Colorado, August 23, 1971.

2. Quenouille, M. H. (1956). Notes on bias in estimation. Biometrika 43, 353-360.

3. Seeger, P. (1970). A method of estimating variance components in unbalanced designs. Technometrics 12, 207-218.

4. Tepping, B. (1968). The estimation of variance in complex surveys. Proceedings of the Social Statistics Section of the American Statistical Association. 11-18.

5. Horvitz, D. G., and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. Journal of the American Statistical Association. 47, 663-685.

6. Tukey, J. W. (1958). Bias and confidence in not-quite large samples: Abstract. Ann. Math. Statist. 29, 614.

TABLE I

TAYLOR SERIES (TS) AND JACKKNIFE (JK)
P-VALUES AND VARIANCE COMPONENTS

Item Code	Description of Item	P-Value[1/] x 10^2		Variance Components x 10^4							
		P		PSU		Pair		Student		Total[2/]	
				$V_1=\hat{\Sigma}^2_P(+)$		$V_2=\hat{\Sigma}^2_{S/P}(+)$		$V_3=\hat{\Sigma}^2_{K/S}$		$V=Var(\hat{P})$	
		$\hat{P}$	JP....	TS	JK	TS	JK	TS	JK	TS	JK
A	Highest Education of Parents is Less Than High School	15	15	35.61	35.46	8.72	8.73	91.18	98.18	14.84	14.92
B	Definite or Likely Goer to College	43	43	3.66	3.93	21.49	21.49	246.4	246.4	8.81	9.01
C	Plan to Attend College	59	59	13.20	13.47	17.67	17.67	242.9	242.9	11.39	11.50
D	Don't Belong to a Minority Group	88	88	4.43	4.44	9.55	9.55	88.77	88.77	4.55	4.56
+	Average over Four Items	51	51	14.22	14.32	14.36	14.36	169.1	169.1	9.90	10.00

1/ $P = \frac{\text{Sum of Student Weights with Attribute}}{\text{Sum of All Student Weights}}$

2/ $10^4V = MS_P(+)/3$. See Section 5 for the formulas of $MS_P(+)$.

VARIANCES OF THE CURRENT POPULATION SURVEY, INCLUDING WITHIN- AND BETWEEN-PSU COMPONENTS AND THE EFFECT OF THE DIFFERENT STAGES OF ESTIMATION

Martha J. Banks and Gary M. Shapiro, U.S. Bureau of the Census

I. Introduction

The important results in this paper are given in the tables in Section IV. The tables contain various results from the Census Bureau variance program used for the Current Population Survey (CPS) that are felt to be useful in planning for and estimating variances for other household surveys.

Section II gives a summary of the sample design and estimation procedures of CPS which are useful in an understanding of the results of this paper. These features, as well as other aspects of the sample not covered here, are described more completely elsewhere/8/. This section can be skipped by persons already fully familiar with the survey. Section III describes the variance program now in use.

II. Description of the Current Population Survey (CPS)

A. Nature of the CPS. CPS is a sample survey conducted monthly for the Bureau of Labor Statistics by the Bureau of the Census to obtain national estimates of employment, unemployment, and other characteristics of the labor force. Because of the sample design, the survey can be and is used to produce estimates for a wide variety of other demographic characteristics for the population as a whole, as well as for various subgroups of the population.

B. Sample Design. Since January 1967, the monthly CPS sample has consisted of about 48,000 eligible households in 449 first-stage sampling units (primary sampling units, or PSU's) comprising 863 counties and independent cities. Beginning in July 1969, the overall sampling rate for the survey has been 1 in 1240. The 449 PSU's were selected out of 357 strata. Of the 357 strata, 112 consist of only 1 PSU. Such PSU's are necessarily in sample and are called self-representing (SR). The sampling rate within each of these PSU's is 1 in 1240. The other 245 strata contain more than one PSU each, and the sample PSU's from these strata are called non-self-representing (NSR) since a sample PSU from one of these strata also represents the other PSU's in the same stratum.

The 245 strata were grouped into 122 pairs of strata with one stratum left over. From each pair of strata, one stratum was picked at random, each with equal probability. From each selected stratum one PSU was chosen with probability proportionate to the 1960 populations of the PSU's. The sample size to be taken from the chosen PSU was determined such that the effective sampling rate within the stratum was 1 in 1860 (i.e., 3/2 x 1240).

From each of the remaining 122 strata not selected, two PSU's per stratum were independently chosen with probability proportionate to size. Since the choices were independent, it was possible for the sample PSU's to be either the same or different. In the thirty strata where the two choices were the same, this procedure simply results in twice as large a sample within each twice-chosen PSU as would have been the case if the PSU were chosen only once. The one unpaired stratum left over was handled similarly.

C. Selection of the Sample Within the Sample PSU's. The object of subsampling within each of the sample PSU's was to obtain a self-weighting probability sample of housing units and units in special places. The housing units were selected in segments containing, on the average, about six housing units.

The selection of a sample of segments within each PSU proceeded in several stages as follows:

1. Selection of a sample of the enumeration districts used in the 1960 Census of Population and Housing. The enumeration districts used in the Census were geographic areas, usually with well-defined boundaries and containing, on the average, about 250 dwelling units.
2. Subdivision of each enumeration district into segments.
3. Selection of a sample of segments in each of the selected enumeration districts.

There were two types of sampling used within enumeration districts. List sampling of units enumerated in 1960 was used primarily in urban areas and area sampling was used primarily in rural areas.

A subsample of building permits from a sample of areas where such permits were required and available was used for most of the newly constructed units. Where such permits were either not required or not available, the newly constructed units were picked up in area-sample enumeration districts. In list-sample enumeration districts, however, the newly constructed units were picked up by a successor check, which is described in /8/.

D. Rotation of the Sample. The rotation system used in the CPS may be described as follows:

1. The entire sample is divided into eight equal, separate, systematic subsamples, referred to as rotation groups. One new rotation group is introducted into the survey each month, and one old one is replaced.
2. Each new rotation group is included in the survey for four months, then is excluded for eight months, then is returned for an additional four months. The chart presented in exhibit A below, shows in a simplified form, how the rotation system operates. Examination of the chart will make clear the important characteristic that in any month, six of the eight rotation groups in sample will have been in the survey for the previous month, i.e., there will always be a 75 percent month-to-month overlap. Also, half of the rotation groups in any month will have been in the survey exactly a year before.

EXHIBIT A CPS ROTATION CHART

```
1971 March . . . 1 2 3 4 . . . 5 6 7 8
     April. . . . 2 3 4 5 . . . 6 7 8 1
      May . . . . . 3 4 5 6 . . . 7 8 1 2
      :      . . . . .         . . .
      :       . . . . .         . . .
        1972 March . . . . . 5 6 7 8 . . . 1 2 3 4
```

E. Estimation Procedure. To arrive at a final estimate, adjustment for nonresponses is made, two stages of ratio estimation are applied, and the preceding month's data is utilized to form a composite estimate.

1. Adjustment for Nonresponse.-For all units except large special places, strata are combined into 76 groups containing from one to nine strata each. The ratio of the designated sample to the interviewed sample within each of six race-residence categories is used as the adjustment for noninterview.

The basic weight of 1240 (which is the inverse of the overall sampling fraction) for an individual record is then multiplied by the appropriate factor. The adjustment for nonresponse is a little different in the large special places, but the effect is about the same.

2. First Stage Ratio Estimate Adjustment.-The purpose of the first-stage ratio estimate is to reduce the contribution to the variance arising from the sampling of PSU's, i.e., the variance that would still be associated with the estimates if all households in every sample PSU were included in the survey each month.

The first-stage ratios are based on 1960 Census data and are applied only to sample data for the NSR PSU's. For the NSR PSU's in each of the four Census regions, a ratio is computed for each of six race-residence groups as follows:

$$\frac{\text{1960 Census Pop. in Race-Residence Group for NSR Strata in Region}}{\text{Estimates of this Pop. Based on 1960 Census Pop. for Sample PSU's}}$$

The basic weight for all records from NSR strata is then multiplied by the appropriate ratio of this type, on top of the previously applied noninterview factor.

3. Second-Stage Ratio Estimate Adjustment.-The second-stage ratio estimate factor adjusts sample estimates of the U.S. population in a number of age-sex-race groups to independently derived current estimates of the population in each of these groups. These independent estimates are prepared each month by carrying forward the most recent Census data to take account of subsequent aging of the population, mortality, and migration between the U.S. and other countries. The CPS sample returns, after application of the noninterview adjustment and first-stage ratios are, in effect, used only to determine the percentage distribution of the population within each age-sex-race group by employment status and various other characteristics.

4. Composite Estimate.-The last stage in the preparation of estimates is the derivation of a composite estimate. The composite estimate is a weighted average of two estimates for the current month for any particular item. The first of these two estimates is the result of the two stages of ratio estimates described above. The second estimate consists of the composite estimate for the preceding month to which has been added an estimate of the change from the preceding month to the present month, where the estimate of change is based on the six rotation groups common to the two months (about 75 percent of the households in sample in the current month).

For such a composite estimate to be unbiased, the weights for the two components must add to 1. In CPS, the weights used are each ½.

III. Variance Estimation in the CPS

A. Background. In the past decade, the Census Bureau has used several methods of variance computation for the CPS data. The two most reliable of these have been a replication method and a paired difference method based on work by Keyfitz. [2]

The CPS replication variance program used 20 replications. Although some attempt at balancing to improve the reliability of the variance estimates was made, the more effective balancing procedures developed by McCarthy [3] were not used. [8] contains a description of the replication method as used by the Census Bureau.

The Keyfitz method estimates variances analytically. Basically, it ". . . amounts to calculating a linear combination of sample totals for each primary sampling unit, and then estimating the variance of the sum of those linear combinations." [5] Since Keyfitz's original article, Tepping [5] has given a more eloquent and general formulation of the variance method. [4], [6] and [7] contain detailed descriptions of the Keyfitz method as used by the Census Bureau.

B. Basic Theory. In the Keyfitz formulation, the basic idea is that

$E(x_1 - x_2)^2 = VAR(x_1 + x_2)$ provided that $E(x_1) = E(x_2)$.

Key theorems in [4] give simple expressions for the relvariance of a ratio and of a sum of ratios, e.g.,

$$V^2_{\frac{x}{y}} = \sum_s \left(\frac{x_{s1} - x_{s2}}{E(x)} - \frac{y_{s1} - y_{s2}}{E(y)} \right)^2$$

where x_{si} is an estimate for the i^{th} half of the s^{th} "Keyfitz cluster." (The meaning of "Keyfitz cluster" is clarified below.)

In the Tepping formulation the Taylor series approximation, to terms of the first degree, is written out for the estimate of interest. The variance of the Taylor approximation is then computed directly.

C. Treatment of SR PSU's. Variances are computed differently for SR and NSR PSU's. The SR PSU's are collapsed into 18 clusters. Subcluster 1 of a Keyfitz cluster (the x_{1s} in the above formula is an estimate for sub-cluster 1 of the s^{th} Keyfitz cluster) consists of four of the eight rotation groups, and subcluster 2 consists of the other four. The Keyfitz method is applied four different times for different combinations of the rotation groups. The four resulting variance estimates are then averaged, giving a more reliable result than if only one combination had been used.

D. Treatment of NSR PSU's. For NSR PSU's two techniques are applied. One technique is the one described above, with each subcluster consisting of four rotation groups. This technique, applied to SR PSU's, gives an estimate of total SR variance, but applied to NSR PSU's, it gives an estimate of the NSR within-PSU variance only.

For estimating the total NSR variance, there are 123 Keyfitz clusters, each consisting of one of the pairs of strata as described in II.B. For each pair of strata, there are three sample PSU's, two from one stratum (denoted by A1 and A2) and one from the second stratum (denoted by B). For variance purposes, A1 and B can be thought of as representing only the PSU's in their respective strata; while A2 can be thought of as representing all PSU's in both strata. Thus, for all PSU's, there is a between-PSU-within-stratum variance component, but for only 1/3 of the PSU's there is also a between-stratum component. In order to reflect this in an unbiased fashion, a weighted average of two variance estimates is formed /6/. The first is of the form $(A_1 - A_2)^2$, and includes an unbiased estimate of the between-PSU variance but no between-stratum variance.

The second is of the form $\left(\frac{A_1 + A_2}{2} - B\right)^2$, and includes a between-stratum variance as well. Weights of 7/12 and 1/9, respectively, were derived for these two terms to produce the desired unbiased variance estimate.

E. Census Computer Program. The computer program is written to estimate variances for 45 simple totals (such as total unemployed persons). (The figure of 45 is an upper limit determined by computer storage space considerations.) However, the program can also compute a limited number (10) of covariances, so that with a little arithmetic, variances of simple ratios (such as the unemployment rate) can also be estimated. Estimates of within-PSU variance, between-PSU variance, and between-stratum variance are computed as well as total variance. Separate variance estimates are also produced for each of the several estimators for the unbiased estimate (includes noninterview adjustment), the first-stage ratio estimate, the second-stage ratio estimate, the first- and second-stage combined ratio estimate, and the composite estimate. Further, variances of both monthly level and month-to-month change are produced. (The latter is produced only for the first-and second-stage combined ratio estimate and for the composite estimate.) Also, for the unbiased estimate and the first-stage ratio estimate, the variances for estimates for SR PSU's only and for NSR PSU's only are given.

IV. Data from Keyfitz Variance Program

A. Introduction. Now we get to the heart of the paper: Presentation of tables. All of the tables contain actual data. Nothing completely new and unexpected is presented. Most tables substantiate theoretical work for which there previously has been little or no empirical verification.

It should be noted that the number of digits shown in the tables are not an indication of the reliability of the estimates. In general, the last digit or two are of doubtful significance, but the figures were left unrounded so that the reader can manipulate them as desired before rounding.

B. Tables 1 and 2. Both Tables 1 and 2 record the design effects (Deff's)1/ for each of the items in the variance program for unbiased estimates and the first- and second-stage combined ratio estimate (i.e., noncomposite estimate). The figures are ratios of the actual monthly CPS variances (using an annual average of the monthly data) divided by the variances appropriate for a simple random sample of persons and an unbiased estimate ($\frac{pq}{n}$). Here, p represents the proportion computed from the sums of the twelve monthly CPS estimates of totals. In Table 2, for a characteristic like "Unemployed males of Negro and other races," the denominator of p is males of Negro and other races, 16 and over (14 and over for 1965 and 1966). For the same characteristic (and all other characteristics) in Table 1, the denominator of p is Total Persons, 16 and over. Only those characteristics which are subsets of age-sex-race groups for which independent control totals are used in the ratio estimation are included in Table 2.

In making comparisons between Deff's in Tables 1 and 2, there are small increases in Table 2 for relatively rare characteristics like "Males who are agriculture employed", but rather large increases in Table 2 for the more frequent characteristics like "Females in civilian labor force". These more frequent characteristics are the ones most helped by the second-stage ratio estimate factors, as indicated by the unusually small Deff's for the first- and second-stage combined ratio estimate in Table 1. In contrast, the Deff's in Table 2 for these more frequent characteristics are more in line with the Deff's for other types of characteristics in Table 1.

Table 2 Deff's are primarily applicable under two circumstances. One circumstance is if one is drawing a sample from a universe consisting of only a restricted age-sex-race group (e.g., males 16-19). The second is if you are actually interested in the percentage of persons in a restricted age-sex-race group. (In this case, however, only the Deff's for the first and second-stage combined estimate are applicable.)

In comparing the unbiased estimate and ratio estimate, note that without a single exception, the ratio estimate reduces the variance. In general, as would be expected, the characteristics possessed by a relatively large percentage of the population are helped most by the ratio estimate. The effect of the ratio estimation is rather dramatic for these "large" characteristics. For the unbiased estimate, these "large" characteristics have among the largest design effects, while for the ratio estimate they have the lowest Deff's. The one characteristic possessed by a small percent of the population that is very much helped is "Employed persons of Negro and other races", but this characteristic constitutes a large percentage of total persons of Negro and other races in certain age-sex categories for which independent controls are used.

For the unbiased estimate, only rural population and agriculture employed characteristics have Deff's as high as the "large" characteristics. The Deff's for these rural items are significantly reduced by the ratio estimate, but they still remain relatively high. This obviously reflects the highly clustered nature of these populations.

1/ Within the Census Bureau, the term "Factor over Random" is used instead of design effects, since it is the factor that expresses the amount of variance over and above simple random sampling variance.

Since 1967 is the year that the present sample design was instituted, one other thing to look for in Table 1 is the difference between the Deff's for 1967 through 1969, and those for 1965 and 1966. The only really significant differences occur for agriculture employed and rural farm items and for the items "At school" and "Self-employed". For all of these items, the Deff's are smaller for the recent years for the unbiased estimate, but interestingly enough, not for the ratio estimate. (In fact, for "At school", the Deff's are actually larger for the ratio estimate for recent years.)

An important change which was made in the method of estimating variances for 1967 is almost certainly responsible for the differences for the agriculture employed and rural farm items. Beginning in 1967, the program provided an unbiased estimate of variance, whereas previously the program used a collapsed-stratum estimate which is upward-biased because it contained a between-stratum component. A change in the sample design in 1967 permitted this improvement. Table 3 can be used to estimate the magnitude of this bias (see Section IV.C). Its elimination is probably the reason for the large reductions among agriculture employed and rural farm characteristics, since these characteristics have very large between-PSU components of the total variance.

This change is also undoubtedly partially responsible for the difference for the "Self-employed" item, since about 25 percent of the self-employed are in agriculture.

A change in the population base used for labor force data is probably responsible for the difference for the "At school" item. The tabulation change is that all characteristics after 1966 are tabulated for the civilian noninstitutional population 16 and over; whereas in the previous years, 14 and 15 year olds were also included in the tabulations. Obviously, this cange can be expected to have important effects for this item, while at the same time affecting other items only slightly.

C. Table 3. Table 3 gives the ratio of the total between-variance (sum of the between-PSU and "between-stratum variance") to the total variance, and the ratio of the between-stratum variance to the total variance. See Section III.D. for the meaning of "between-stratum variance."

Since both of these between-variance estimates are derived by the subtraction of one variance estimate from another, they are not very reliable. For this reason the ratios of between to total variance vary so from year to year that even the 3-year average shown in the table is only a crude measurement device, as the negative ratios indicate. It can be clearly seen, however, that the overwhelming component of variance is within-PSU variance rather than between-PSU or between-stratum variance.

Items involving agriculture employed and those employed as wage and salary workers have relatively high total between variance; while for unemployment items, it is relatively low. The average ratio of total between variance to total variance is .114 for agriculture employed, .121 for wage and salary workers, and .011 for the unemployed.

It is rather surprising to see the results for the rural population. One would expect these items to have high positive ratios rather than the largest of all the negative ratios. Although the ratios for all three years are negative, only the ratio for one year is a large negative.

When a collapsed-stratum variance estimation procedure is used for a sample design where there is only one sample PSU in each stratum, the variance estimate is an overestimate due to the inclusion of a "between-stratum variance." The data in Table 3 can be used to estimate the relative magnitude of this bias in the variance estimate for the CPS design that was in effect prior to 1967. In that design the strata were the same as the present strata, but there was only one sample PSU per stratum.

The magnitudes of the between-PSU-within-stratum variance and the "between-stratum variance" (as calculated by a collapsed-stratum procedure) for the previous design are approximated by multiplying the corresponding estimates from Table 3 by 1.5 and 9.0 respectively.

Thus, in the former design, for the "Average of all items", the ratio of between-PSU variance to total variance can be estimated as .159 [.159 = (1.5) (.051 - .011) + 9(.011)], with the (.099) term representing the bias of the between-PSU variance estimate. Note that in this case, the bias represents about 60 percent of the between-PSU variance estimate.

D. Table 4. The composite estimate is the weighted average of two estimates for the same characteristic, as explained in Section II.E.4. Table 4 presents the ratio of the variance of the composite estimate to the variance of the first- and second-stage combined ratio estimate (i.e., the noncomposite estimate) for both estimates of monthly level and of monthly change.

In general, the composite estimate reduces the variances somewhat, but this is not always the case. For unemployment items, "Part-time usually full time" items, and for the item "With a job, not at work", the use of the composite estimate increases the variance.

A composite estimate that weighted the two component estimates differently would be better for these items. In 1963, Gurney [1] determined the optimal weights for various items. She estimated, for example, that for unemployment data, weighting the noncomposite estimate by .7 and the other estimate by .3 would result in a variance lower than either the present composite estimate or the noncomposite estimate. In considering the estimation procedure for CPS, it was decided to use a single pair of weights that would be reasonably good for all items, rather than different weights for different items, in order to minimize complications.

For each characteristic, the use of the composite estimate helps the estimate of monthly change more (or hurts it less) than it does the estimate of monthly level. This is as expected, since the composite estimate makes use of the previous month's estimate. On the average, the variance on the estimate of monthly level is reduced by 4 percent, while the variance on monthly change is reduced by 19 percent.

E. Table 5. Ofentimes, in order to increase reliability, several months' worth of data are accumulated and averaged. The first four columns of Table 5 present the reduction in variance obtained by using 3, 6, 9 or 12 consecutive months' worth of data rather than a single month's data. It is also frequently necessary to estimate the variance of the estimated difference between two months' data. The ratios presented in the last two columns are of the form:

$\frac{VAR(x - y)}{\frac{1}{2}(VAR(x) + VAR(y))}$. For the first of these columns, x and y represent estimates for adjacent months; while for the last column, they represent estimates for two months a year apart. The data in this table results from a special computer program, as well as the regular Keyfitz program.

It should be remembered that for CPS, each rotation group is in sample four months, excluded eight months, and then returned for four (see Exhibit A in Section II.D). This results in a 75 percent monthly overlap of rotation groups and a 50 percent overlap of rotation groups for year-apart data. The figures in the table would, of course, be different with a different rotation scheme.

Looking at the "average" part of the table (the first four columns) there are wide differences among items. Items for which the correlation over time is low, such as unemployment items, are helped considerably by multiple-month data; while items for which the correlation over time is high, such as agriculture employed and rural items, are helped only a little. The CPS sample design is such that when a segment drops out of sample, it is replaced by another segment which is in close geographic proximity. Thus, there is a correlation not only between identical rotation groups but between non-identical rotation groups. For example, if all the people in a segment are rural farm, then all the people in the replacement segment are also likely to be rural farm. Because of the rotation scheme, a hypothetical item with perfect correlation between identical rotation groups ($\rho_s = 1.00$) and no correlation between different rotation groups ($\rho_s = 0.00$) would cause the 3-month variance to be reduced to .78 of the monthly variance. (See appendix for the calculation of .78) Since, for example, the agriculture-employed figure is .83, this means that the correlation between non-identical rotation groups must be at least .23, as can be calculated from formula 5 in the appendix.

The fifth column of the table gives a ratio of the variance for a difference of two adjacent months to the variance of monthly level. Note that for a hypothetical item with perfect correlation between identical rotation groups and no correlation between non-identicals, the ratio would be .50. If there were no correlation at all, the ratio would be 2.00. The last column gives a similar ratio for two months a year apart, instead of two adjacent months. For this ratio, perfect correlation between identicals and no correlation between non-identicals would yield a ratio of 1.00, and no correlation at all would yield a ratio of 2.00.

For the "Average of all items", the ratio for adjacent months is about 45 percent larger than the ratio for data a year apart. This is as expected, since not only are there fewer identical rotation groups for the latter, but the passage of time usually reduces the correlation both between identicals and between non-identicals. There seems to be one major exception to this in the table - a person's likelihood of being on vacation. Thus, "With a job, not a work" is the only item that does not have a higher ratio for data a year apart than for adjacent months.

As an example of how to use these columns, a good approximation to the variance between adjacent months' unemployment levels can be obtained from a single month's variance by multiplying by 1.47. Use a factor of 1.56 to estimate the variance of yearly differences in teenage employment.

G. Table 6. The Keyfitz and replication methods of estimating variances were referred to in Section III. The sixth table uses 1964 data to compare the relvariances (variances divided by the squares of the estimates) calculated by the replication method and by the Keyfitz method. This is shown for each of the items the two computer programs have in common. Variances for both the first- and second-stage combined ratio estimate and the composite estimate are compared.

The two methods appear to give consistent results, as can be seen from columns 3 and 8. However, the replication relvariances vary more from month to month than do the relvariances computed by the Keyfitz method. Columns 4, 5, 9 and 10 give one-sixth of the range of the monthly figures. This approximation to the standard error of the relvariance estimates shows, as expected, that the Keyfitz method provides much more reliable estimates than does the replication method (20 replications). Had the McCarthy method [3] of choosing the half-sample been used, the replication method would compare more favorably.

Prior to 1968, there was a minor error in the computation of the composite estimate. This has been taken into account by a slight adjustment in columns 6 and 9.

V. Acknowledgements

The authors wish to thank Robert H. Hanson and Joseph Waksberg for their many valuable comments on the preparation of this paper.

TABLE 1 Design Effects for the Unbiased Estimate and Ratio Estimate for 1965 through 1969

Characteristic	Percent of Population possessing characteristic					Design Effects 1/									
	Population 14+		Population 16+			Unbiased Estimate					1st & 2nd Stage Combined Est.				
	1965	1966	1967	1968	1969	1965	1966	1967	1968	1969	1965	1966	1967	1968	1969
<u>Civilian Labor Force</u>															
Total	57	57	60	60	60	7.65	9.29	9.49	9.87	10.09	1.16	1.06	.97	1.03	1.15
Females	20	21	22	22	23	2.54	2.60	2.63	2.72	2.79	1.12	1.07	.94	.93	1.01
Under 20 yr.	5.4	5.9	5.1	5.1	5.2	1.71	1.90	1.66	1.75	1.97	.82	.79	.62	.67	.69
<u>Employed</u>															
Negro and Other Races	5.8	5.9	6.2	5.8	6.2	6.03	6.30	5.25	5.90	6.23	.49	.54	.48	.53	.59
Males	35	35	36	36	36	3.50	4.48	4.19	4.23	4.23	.32	.32	.28	.28	.31
Working 1-14 hours	3.4	3.4	3.0	3.1	3.0	1.73	1.65	1.41	1.62	1.58	1.45	1.31	1.21	1.39	1.29
Part-Time (1-34 hr.), usually full-time	3.4	3.0	3.6	4.5	4.1	1.97	1.75	1.64	1.85	2.10	1.69	1.51	1.40	1.53	1.78
Part-Time (1-34 hr.), usually full-time, part-time for economic reasons	.85	.72	.92	.78	.77	1.81	1.45	1.42	1.90	1.49	1.63	1.32	1.33	1.77	1.39
Self-employed	6.4	6.1	5.6	5.4	5.3	2.51	2.38	1.85	1.83	1.86	1.46	1.38	1.38	1.45	1.41
<u>Nonag. Employed</u>															
Total	51	52	54	55	55	6.63	8.25	8.45	8.81	9.15	1.27	1.18	1.16	1.17	1.23
Male	32	32	34	34	34	3.44	4.39	4.26	4.27	4.33	.48	.48	.45	.43	.44
Female	18	19	21	21	21	2.35	2.48	2.50	2.62	2.69	1.11	1.10	.95	.97	1.01
Working 35 hr.+	39	39	41	40	40	4.54	5.65	6.00	5.66	5.62	1.07	1.10	1.13	1.12	1.17
With a job, not at work	2.5	2.6	2.9	3.1	3.2	1.44	1.60	1.58	1.66	1.75	1.25	1.32	1.31	1.36	1.44
<u>Ag. Employed</u>															
Total	3.5	3.1	3.0	2.9	2.6	6.81	6.46	4.30	4.43	4.13	3.21	3.14	3.12	3.30	3.04
Male	2.9	2.6	2.4	2.4	2.2	5.18	5.06	3.30	3.47	3.23	2.42	2.46	2.48	2.68	2.45
Female	.64	.57	.52	.51	.47	3.21	3.17	2.51	2.55	2.35	2.28	2.34	2.09	2.21	2.04
Working 35 hr.+	2.3	2.1	2.0	1.9	1.7	5.43	5.22	3.58	3.45	3.37	2.72	2.69	2.77	2.65	2.60
Unpaid family workers and self-employed	2.3	2.1	2.0	1.9	1.8	6.63	6.48	3.95	3.91	3.73	2.90	3.02	2.69	2.69	2.53
<u>Employed as Wage and Salary Worker</u>															
In nonag.	45	47	50	50	50	5.80	7.14	7.45	8.09	8.42	1.32	1.27	1.16	1.24	1.39
In ag.	1.2	1.0	1.0	1.0	.87	4.08	3.45	3.16	3.61	2.99	3.16	2.82	2.95	3.38	2.86
In durables or nondurables (mfg.)	14	15	16	16	15	2.78	3.57	3.13	3.41	3.42	1.71	2.10	1.73	1.88	2.05
In durables	8.1	8.6	9.3	9.2	9.2	2.49	3.34	2.68	2.72	2.67	1.92	2.46	1.89	1.84	1.82
In construction	2.9	2.9	3.0	3.0	3.0	1.56	1.66	1.54	1.61	1.84	1.28	1.30	1.27	1.31	1.46
In retail trade	6.8	6.9	7.5	7.5	7.5	1.65	1.77	1.89	1.78	2.08	1.22	1.21	1.33	1.32	1.52
In service industry, including private household workers	13	14				2.63	2.63	2/	2/	2/	1.80	1.75	2/	2/	2/
As private household workers			1.6	1.5	1.4	2/	2/	1.54	1.53	1.52	2/	2/	1.30	1.28	1.22
<u>Unemployed</u>															
Total	2.7	2.3	2.5	2.3	2.2	1.48	1.43	1.50	1.55	1.56	1.25	1.29	1.27	1.35	1.37
Wage and salary workers in durables or nondurables (mfg.)	.60	.60	.65	.56	.56	1.23	1.26	1.23	1.31	1.34	1.16	1.17	1.18	1.22	1.27
<u>White Unemployed</u>															
Males			1.0	.91	.88	2/	2/	1.35	1.27	1.23	2/	2/	1.25	1.16	1.15
Females			1.0	.91	.91	2/	2/	1.24	1.26	1.29	2/	2/	1.13	1.14	1.17
Males, 16-19			.28	.28	.27	2/	2/	1.17	1.23	1.15	2/	2/	1.08	1.13	1.08
Females, 16-19			.26	.26	.26	2/	2/	1.16	1.19	1.17	2/	2/	1.05	1.07	1.05

See footnotes at end of table.

TABLE 1 (Cont'd) Design Effects for the Unbiased Estimate and Ratio Estimate for 1965 through 1969

Characteristic	Percent of Population possessing characteristic: Population 14+ 1965	Population 14+ 1966	Population 16+ 1967	Population 16+ 1968	Population 16+ 1969	Design Effects1/ Unbiased Estimate 1965	Unbiased Estimate 1966	Unbiased Estimate 1967	Unbiased Estimate 1968	Unbiased Estimate 1969	1st & 2nd Stage Combined Est. 1965	1st & 2nd Stage Combined Est. 1966	1st & 2nd Stage Combined Est. 1967	1st & 2nd Stage Combined Est. 1968	1st & 2nd Stage Combined Est. 1969
Negro and Other Races Unemployed															
Males			.24	.22	.20	2/	2/	1.40	1.53	1.72	2/	2/	1.07	1.26	1.32
Females			.28	.26	.25	2/	2/	1.49	1.53	1.52	2/	2/	1.19	1.21	1.28
Males, 16-19			.08	.08	.08	2/	2/	1.23	1.29	1.39	2/	2/	1.00	1.12	1.14
Females, 16-19			.08	.07	.08	2/	2/	1.23	1.24	1.39	2/	2/	1.10	1.14	1.13
Rural															
Total nonfarm	29	24	24	25	25	15.29	17.36	17.88	18.66	18.34	7.95	11.12	10.01	10.69	10.62
Total farm	6.4	6.1	5.6	5.4	5.1	12.35	12.66	7.77	7.66	7.48	5.37	5.19	5.65	5.63	5.48
Male farm	3.4	3.1	2.9	2.8	2.6	6.67	6.61	4.07	4.03	4.00	2.94	2.77	2.94	3.00	2.93
Miscellaneous															
Total Males	47	47	46	47	47	4.67	5.74	5.23	5.20	5.20	0	0	0	0	0
Total Females	53	53	54	53	53	5.00	6.14	5.92	6.02	5.78	0	0	0	0	0
Total Persons of Negro & Other Races	11	11	11	11	11	10.89	12.01	8.95	9.97	10.24	0	0	0	0	0
Household Heads	44	44	46	46	46	3.15	4.12	3.89	3.78	3.75	.61	.65	.58	.52	.54
At school	8.2	8.1	5.0	5.1	5.0	2.17	2.30	1.77	1.97	1.83	.71	.68	.90	1.04	1.05

TABLE 2 Design Effects for Characteristics which are Subsets of Age-Sex-Race Groups for the Unbiased Estimate and Ratio Estimate for 1965 through 1969

Characteristic	Population 14+ 1965	Population 14+ 1966	Population 16+ 1967	Population 16+ 1968	Population 16+ 1969	Unbiased Estimate 1965	Unbiased Estimate 1966	Unbiased Estimate 1967	Unbiased Estimate 1968	Unbiased Estimate 1969	1st & 2nd Stage Combined Est. 1965	1st & 2nd Stage Combined Est. 1966	1st & 2nd Stage Combined Est. 1967	1st & 2nd Stage Combined Est. 1968	1st & 2nd Stage Combined Est. 1969
Civilian Labor Force															
Female	38	39	42	42	43	3.28	3.39	3.56	3.64	3.78	1.44	1.39	1.26	1.25	1.38
Under 20 Yr.	36	38	49	49	50	2.53	2.88	3.11	3.28	3.77	1.21	1.20	1.17	1.24	1.33
Employed															
Negro and Other Races	50	54	57	54	58	12.25	12.88	12.13	13.76	15.59	.92	1.10	1.04	1.09	1.33
Males	73	73	77	78	78	8.87	11.44	12.32	12.16	12.50	.78	.78	.77	.82	.90
Nonag. Employed															
Males	67	68	71	73	73	7.35	9.57	10.44	10.33	10.79	.98	1.00	1.04	1.05	1.08
Females	34	36	38	39	40	2.94	3.14	3.27	3.40	3.53	1.38	1.39	1.26	1.26	1.34
Ag. Employed															
Males	5.9	5.3	5.1	5.1	4.7	5.35	5.21	3.40	3.57	3.32	2.50	2.53	2.55	2.76	2.51
Females	1.2	1.1	1.0	1.0	.9	3.22	3.18	2.46	2.56	2.36	2.29	2.36	2.10	2.22	2.05
White Unemployed															
Males			2.3	2.2	2.1	2/	2/	1.37	1.28	1.25	2/	2/	1.26	1.17	1.16
Females			2.0	1.9	1.9	2/	2/	1.25	1.27	1.30	2/	2/	1.14	1.16	1.18
Males, 16-19			6.6	6.2	6.2	2/	2/	1.25	1.30	1.22	2/	2/	1.15	1.20	1.14
Females, 16-19			5.6	5.7	5.8	2/	2/	1.22	1.26	1.23	2/	2/	1.11	1.13	1.11
Negro & Other Races Unemployed															
Males			4.7	4.6	4.3	2/	2/	1.48	1.61	1.76	2/-	2/	1.12	1.32	1.43
Females			4.9	4.4	4.2	2/	2/	1.57	1.60	1.58	2/	2/	1.25	1.26	1.33
Males, 16-19			12	12	11	2/	2/	1.42	1.47	1.57	2/	2/	1.14	1.27	1.28
Females, 16-19			10	10	10	2/	2/	1.40	1.39	1.55	2/	2/	1.23	1.27	1.26
Rural															
Male farm	6.9	6.5	6.0	6.0	5.7	6.95	6.86	4.21	4.18	4.13	3.06	2.87	3.03	3.10	3.03

1/ Twelve months of Keyfitz data were averaged for the numerators of the factors.
2/ No Keyfitz variance estimate available.

TABLE 3 Total Between[1/] and Between Stratum Variance as Proportions of Total Variance, 1967-1969 Averages 1st and 2nd Stage Combined Ratio Estimate

TABLE 4 Ratio of the Variance of the Composite Estimate to the Variance of the Noncomposite Estimate, 1969 Data

Characteristic	Total Between	Between Stratum	Ratio of the Annual Averages of Monthly Level	Ratio of the Annual Averages of Monthly Change
Civilian Labor Force				
Total	.099	.019	.8560	.6895
Females	.082	.013	.8384	.6619
Ages, 16-19	.037	.006	.9788	.8537
Employed (Average)	.076	.010		
Negro and Other Races	.062	.002	.8746	.7799
Males	.065	.019	.8976	.8043
Working 1-14 hrs.	.069	.002	1.0620	.9583
Part-time (1-34 hrs) usually full-time	.117	.016	1.2226	1.1478
Part-time for economic reasons, usually full-time	.090	.010	1.1834	1.0866
Self-employed	.053	.013	.8663	.6992
Nonagriculture Employed (Average)	.048	.012		
Total	.036	.014	.8628	.6957
Males	.027	.013	.9061	.8045
Females	.071	.011	.8385	.6398
Working 35+ hours	.050	.014	.9725	.9035
With a job, not at work	.054	.007	1.1485	1.0633
Agriculture Employed (Average)	.114	.017		
Total	.123	.016	.9033	.7162
Males	.111	.021	.8899	.7003
Females	.255	.003	.9274	.8223
Working 35+ hours	.105	.026	.8861	.7655
Unpaid family and self-employed	-.026	.018	.9095	.6889
Wage & Salary Workers (Average)	.121	.020		
In nonagriculture	.043	.016	.8560	.6881
In agriculture	.288	.016	.9353	.8791
In durables or nondurables (mfg.)	.224	.059	.8373	.5940
In durables	.225	.039	.8280	.6245
In construction	.035	.002	.8319	.7390
In retail trade	.011	.008	.8305	.6893
As private household workers	.024	.000	.8796	.8271
Unemployed				
Total	.023	.008	1.1096	.9939
Wage and salary workers in durables or nondurables (mfg.)	.008	.010	1.0514	.9714
White Unemployed				
Males	.006	.008	1.0663	.9972
Females	.034	.006	1.0824	1.0044
Males, 16-19	-.003	.008	1.1371	1.0453
Females, 16-19	.039	.002	1.0926	1.0391
Negro & Other Races Unemployed				
Males	.005	.003	.9706	.9584
Females	.012	-.002	1.1172	.9973
Males, 16-19	.014	-.003	1.0223	1.0004
Females, 16-19	-.026	-.009	1.1594	.9959
AVERAGE OF ALL UNEMPLOYMENT ITEMS	.011	.003	1.0809	1.0003
Rural				
Total nonfarm	-.136	-.007	.9081	.4382
Total farm	-.130	.019	.8785	.5345
Male farm	-.133	.017	.8736	.5359
Miscellaneous				
At school	.024	.008	.9245	.7996
Household heads	.008	.004	.7693	.5358
AVERAGE OF ALL ITEMS	.051	.011	.9558	.8139

1/ Total between variance is the sum of between PSU variance and "between stratum" variance.

TABLE 5 Ratio of the Variances of Monthly Averages for Three, Six, Nine and Twelve Months' to One Month's Estimate and Ratio of the Variance of Differences Between Two Months to the Variance of Monthly Level 1st and 2nd Stage Combined Ratio Estimate, 1969 and 1970 Data[1/]

Characteristic	Averages of: Three Months	Six Months	Nine Months	Twelve Months	Differences of: Two Adjacent Months	Two Months A Year Apart
Civilian Labor Force						
Total	.75	.58	.49	.46	.66	1.00
Females	.76	.58	.48	.43	.63	1.11
Ages, 16-19	.60	.41	.33	.29	1.12	1.56
Employed						
Negro and Other Races	.72	.53	.44	.40	.70	1.31
Males	.68	.49	.41	.37	.82	1.26
Working 1-14 hrs.	.53	.34	.25	.21	1.31	1.65
Part-time (1-34 Hrs) usually full-time	.43	.26	.20	.18	1.67	1.82
Part-time for economic reasons, usually full-time	.43	.25	.18	.14	1.67	1.88
Self-employed	.77	.60	.50	.47	.61	1.16
Nonagriculture Employed						
Total	.76	.60	.51	.49	.61	1.03
Males	.74	.58	.50	.47	.66	1.12
Females	.76	.58	.49	.44	.62	1.21
Working 35+ hours	.65	.48	.41	.38	.97	1.38
With a job, not at work	.43	.25	.16	.13	1.66	1.62
Agriculture Employed						
Total	.83	.69	.61	.58	.42	.73
Males	.84	.70	.63	.60	.39	.71
Females	.71	.51	.41	.34	.74	1.12
Working 35+ hours	.77	.62	.55	.51	.56	.81
Unpaid family and self-employed	.83	.68	.59	.55	.42	.79
Wage and Salary Workers						
In nonagriculture	.77	.61	.53	.51	.58	1.03
In agriculture	.77	.61	.53	.48	.57	1.06
In durables or nondurables (mfg)	.80	.65	.58	.54	.45	.79
In durables	.78	.63	.55	.50	.49	.79
In construction	.72	.50	.38	.31	.68	1.39
In retail trade	.75	.55	.44	.37	.64	1.44
As private household workers	.65	.45	.35	.30	.94	1.43
Unemployed						
Total	.51	.31	.23	.20	1.33	1.71
Wage and salary workers in durables or nondurables (mfg.)	.49	.30	.22	.19	1.38	1.75
White Unemployed						
Males	.48	.30	.23	.19	1.42	1.66
Females	.49	.30	.21	.16	1.40	1.88
Males, 16-19	.45	.26	.19	.16	1.59	2.00
Females, 16-19	.45	.25	.17	.12	1.60	1.78
Negro & Other Races Unemployed						
Males	.49	.29	.21	.18	1.45	1.84
Females	.47	.28	.20	.17	1.52	1.80
Males, 16-19	.46	.29	.22	.20	1.55	1.86
Females, 16-19	.47	.27	.19	.16	1.44	1.86
AVERAGE OF ALL UNEMPLOYMENT ITEMS	.48	.29	.21	.17	1.47	1.81
Rural						
Total nonfarm	.90	.79	.76	.73	.22	.64
Total farm	.87	.73	.66	.63	.31	.78
Male farm	.86	.71	.64	.62	.34	.78
Miscellaneous						
At school	.60	.42	.34	.30	1.06	1.45
Household heads	.74	.53	.43	.40	.57	1.10
AVERAGE OF ALL ITEMS	.66	.48	.40	.36	.92	1.33

1/ January 1969 through June 1970 data were used.

TABLE 6

Comparison of Relvariances Computed by the Keyfitz and Replication[1] Methods, 1964 Data

Item[3]	1st & 2nd Stage Combined Ratio Estimate					Composite Estimate				
	Keyfitz	Replication	Column (2) / Column (1)	Keyfitz	Replication	Keyfitz[2]	Replication	Column (7) / Column (6)	Keyfitz[2]	Replication
	Average Relvariance ($x10^{-8}$)			1/6 of range of monthly figures ($x10^{-8}$)		Average Relvariance ($x10^{-8}$)			1/6 of range of monthly figures ($x10^{-8}$)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Total labor force	1,049	781	.74	60	163	1,068	654	.61	60	119
Females in labor force	5,604	4,355	.78	300	365	5,301	4,359	.82	360	681
Employed males	855	987	1.15	70	86	647	915	1.41	60	142
Total nonwhite employed	12,575	15,032	1.20	1,300	3,276	11,787	14,191	1.20	1,300	3,869
Total ag. employed	126,176	137,657	1.09	7,000	24,923	113,742	127,708	1.12	6,700	29,242
Male ag. employed	109,862	125,492	1.14	10,000	19,855	96,723	113,007	1.17	7,200	22,501
Female ag. employed	559,273	545,447	.98	54,000	110,163	538,806	522,559	.97	60,000	93,949
Total nonag. employed	1,669	1,275	.76	70	169	1,411	1,154	.82	60	94
Employed males in nonag.	1,375	1,378	1.00	55	271	1,130	1,344	1.19	75	267
Employed females in nonag.	6,702	5,563	.83	550	797	5,743	5,279	.92	320	1,026
Full-time nonag. employed	2,741	2,180	.80	650	483	2,339	2,526	1.08	900	604
Employed in nonag. - With job, not at work	63,577	81,612	1.28	16,700	26,032	73,953	95,287	1.29	19,000	22,645
Unemployed	53,913	55,909	1.04	6,150	8,842	46,612	63,127	1.35	5,300	8,747
Average			.98					1.07		

1/ Twenty replications were used. 2/ These have been adjusted by applying correction factors. See text.
3/ All items used were for population 14 and over.

APPENDIX A.-Percent Reduction in Variance Due to Use of Three Months' Data Rather than One

For the estimate of a total (x) based on a single rotation group,

(1) $VAR(x) = N^2 \frac{\sigma^2}{n}$, where n is the number of sample cases in 1 rotation group and N is total population for the country.

For the estimate of a total based on 1 full month's data,

(2) $VAR(x) = \sum_{i=1}^{8} \frac{N^2}{(8)^2} \frac{\sigma^2}{n} = \frac{N^2\sigma^2}{8n}$

For the estimate of a total based on 3 consecutive months' data,

(3) $VAR(x) = \sum_{i=1}^{24} \frac{N^2\sigma^2}{(24)^2 n} + 2\sum_{i=1}^{16} \frac{N^2\rho_s\sigma^2}{(24)^2 n} + 2\sum_{i=1}^{8} \frac{N^2\rho_d\sigma^2}{(24)^2 n}$

where ρ_s is the correlation between identical rotation groups, and ρ_d is the correlation between non-identical, but matching, rotation groups.

In other words,

(4) $VAR(x) = \frac{N^2\sigma^2}{8n}\left(\frac{3}{9} + \frac{4}{9}\rho_s + \frac{2}{9}\rho_d\right)$

(5) $R_x = \frac{1}{3} + \frac{4\rho_s}{9} + \frac{2\rho_d}{9}$,

where R_x = the ratio of the variance of the monthly average for three months to the variance for one month's estimate.

If we assume ρ_s =1 and ρ_d = 0, we get $R_x = \frac{7}{9}$,

i.e., there is a $\frac{2}{9}$, or 22% reduction in the variance compared to 1 month's data.

REFERENCES

1. Gurney, Margaret, "Sampling for Time Series", Internal Census Bureau memo. to Messrs. Wm. Hurwitz & Max Bershad, April 26, 1963.
2. Keyfitz, Nathan, "Estimates of Sampling Variance When Two Units are Selected From Each Stratum," Journal of the ASA 52 (Dec. 1957) pp. 503-510.
3. McCarthy, P. J., "Pseudo-Replication: Half-Samples," Review of the Internat'l. Statistical Inst. 37, No. 3 (1969), pp. 239-264.
4. Shapiro, Gary M., "Keyfitz Method of Estimating Variance & Its Application to the Current Pop. Survey" Internal Census Bureau Document, Sept.1, 1966.
5. Tepping, Benj. J., "Variance Estimation in Complex Surveys" Proceedings of the Social Statistics Section 1968 ASA, pp. 11-18.
6. Woltman, Henry & Turner, Anthony G., "Two Additional Methods for Estimating Total NSR Variance by the Keyfitz Method for the CPS-MLS Merged Surveys", Internal Census Bureau memo. to Robert Hanson, Jan. 13, 1967.
7. Woltman, Henry, "Variance Estimation Formulas to be Used in Computing Total NSR PSU Variances for the 449-Area Design", Internal Census Bureau document, March 12, 1968.
8.__________U.S. Bureau of the Census, the Current Population Survey-A Report on Methodology, Technical Paper No. 7, 1963.

DISCUSSION

Walt R. Simmons, National Center for Health Statistics

1. Comment on Fellegi and Gray Paper.

1.1 Let me say first that this paper was a pleasure to read for at least three reasons:

A. It presents a *simple* and *operational* procedure for calculating the approximate sampling variance of a statistic from a complex survey, without going into a welter of the complicating factors which are always present, and which worry many of us who deal with these matters.

B. I'm especially pleased to see the procedure designed and used not alone for securing the standard error of a statistic to be published, but as an organized system of operational control of the survey. This reviewer believes that one of the most serious weaknesses in survey work today is the failure of survey execution to mirror faithfully the design. The control system described by Fellegi and Gray, based on the variance analysis patterns is good medicine for this ailment.

C. I give a cheer for the recognition that there is substantive information in the time series of variances compared with the simple random model variances that goes beyond the usual measure of precision of the sample statistics. I agree fully that this is a real contribution to the information provided to the analyst and economist.

1.2 Let me turn, then, to a few criticisms and questions.

A. I found the notation and presentation in the Appendix less clear or clean than it might have been. This began with a couple of typos in my draft, and with a very skimpy definition of the "balancing unit factor" B_{jb} at the beginning. It wasn't helped by the choice of B for both "balancing factor" and "variance," nor the oddly worded introduction of the symbol h.

B. A more substantial issue is closely allied to one of the strengths of the recommended procedure. The estimates of variance themselves have a sampling error, of course. Thus observed differences between two variances may represent real differences, or they may reflect only random variation. Analysts need be warned of this circumstance, lest they draw unwarranted conclusions; just as they might from looking at primary statistics, and disregarding their standard errors. Perhaps there should be advice to analysts to look at *patterns* of variances rather than a simple variance or pair of variances in drawing conclusions.

C. This variance of estimated variances raises another point which is perhaps a matter of taste. If one plots estimated relative standard errors against a series of statistics from a multi-purpose survey, he finds generally that the former decreases in a fairly smooth pattern with increases in the latter. There are exceptions or deviations from a smooth curve. Some are real, some are themselves random variation. It seems to me that when the statistical agency publishes a very large number of estimated variances, not all of which will have an internal consistency, the consumer is puzzled. I lean in favor of publishing a more limited number of average or "typical" variances or relative standard errors. This point has importance, too, when one is focusing attention on the ratio of estimated variance to the variance of the simple random model. If this point of view is not accepted, I would argue that seven levels of coded published variances are too many. It is too many for the consumer to keep in mind, and too many because it overstates the precision with which the variances are estimated. Might not this be better:

a for $CV < 5\%$

b for $5\% < CV < 15\%$

c for $CV \geqslant 15\%$

1.3 The Canadian linearized variance procedure is efficient for estimating variances of means, difference of means, and simple ratios. This paper does not discuss the problem of estimating variance of more elaborate statistics such as regression coefficients, or of position statistics. There another technique—perhaps balanced pseudo-replication—is needed. Pseudo-replication is useful, too, as a device for discovering the relative impact on overall variance of different design features and different estimator features.

1.4 May I say again, I liked this paper.

2. Comment on Frankel Paper.

2.1 The Frankel paper is in my judgment one of the most significant and satisfying pieces of research to come out in recent years in the realm of applied survey sampling. On the analytic side, in survey work, our ultimate objective is usually to estimate from sample data a first order statistic, say θ', of the parameter θ and then a standard error S_θ of θ', form the ratio $\frac{\theta'}{S_{\theta'}}$, *assume* that the ratio is distributed as t- or normal, about a mean of $\frac{\theta}{S_{\theta'}}$ with unit variance. When the sampling is simple-random, and θ is, say a mean, both theory and empirical evidence have justified this approach. When the survey design is more complex, involving ratios, clustering, stages, phases, post-stratification—and θ is a more elaborate parameter such as a median, ratio, or correlation coefficient, we have not been sure how to calculate $S_{\theta'}$ and have been quite unsure of the real distribution $\theta/S_{\theta'}$. Frankel's paper, and his dissertation behind the paper, have taken a long and welcome step toward resolution of both problems.

2.2 All survey samplers should be grateful for these results. I'm especially pleased, perhaps, because the results confirm the practice that is being followed today in the National Center for Health Statistics. We use both the linear scheme and BRR—which we call Pseudo-replication, and which we prefer for its analytic capabilities. The BRR approach was evolved by Philip McCarthy, building on the work of others while he was searching for "methods for analyzing data from complex surveys," working under an NCHS contract.

2.3 Frankel did find BRR "best" under his (I think, appropriate) criterion in every one of 90 plus estimates tested. But he found, also, that both the Taylor expansion technique and Jackknife gave acceptable results. For simple statistics such as means, or ratios, or their differences, he endorses the linear approach because it is cheaper. For more elaborate statistics, he chooses BRR. That seems to be a sound course.

3. Comment on Banks and Shapiro Paper.

3.1 To me, the most significant feature of the paper is the way in which it extends a Census Bureau tradition, and exhibits not merely how a sampling variance of a statistic can be estimated, but how that estimate or estimating process can be further analyzed to tell more about the survey design. Recall that their paper explores the Design Effect; the contribution of between-PSU and within-PSU components; the "bias" of between-PSU estimates of variance in the usual collapsed stratum operation; the impact of the rotation patterns of CPS; the impact of the composite estimator; comparative estimators: linear and replication; and variance of the variance estimators. All of these matters—and others—are important, and deserve attention.

3.2 Banks and Shapiro say that the linear scheme and replication give essentially the same result—i.e. they have empirically the same expected value. I would not quarrel with their conclusion; indeed, my own experience tends to support the declaration. But their evidence in the paper for the statement is not overly strong. For example, for what should be the best estimate (total labor force), the two estimated rel-variances differ by 34% of the smaller one (about 1.5 times the estimated standard error of that difference.) This isn't entirely comforting.

3.3 The authors conclude that the linear scheme provides much more reliable estimates than the replication because typically, the former has a standard error of the order of ½ that of the latter. The evidence on this point is the consequence of a good many factors, some of which I'm in no position to assess. But I would note two:

A. The replication variability would be reduced if more replicates were used—for example, if 40 rather than 20 were used, variance of the variance would be greatly reduced.

B. What effect on estimating variance of the variance in the linear case does the dropping of 2nd order terms have?

4. Comment on Folsom, Bayless and Shah Paper.

4.1 First, I apologize for not having had time to study this paper as carefully as the other three. In particular, I have not attempted to verify the rather extensive mathematics in the paper.

4.2 If, indeed, they have developed a technique for producing unbiased estimates of the contribution to sampling variance from each of three stages of sampling in a complex design, that is a very definite contribution to the design of any similar subsequent survey.

4.3 I'm unable to express a judgment on the relative validity or impact on precision of several factors in the development of this paper, but I might call attention to three which may be significant.

A. Although the authors declare that the method is applicable to any number of sampling units at any stage, a part of the development depends on sampling with replacement in the 1st stage, and an assumption that at least 2 PSU's appear in each stratum, with no PSU appearing more than once in the sample.

B. This study, like some others, appears to secure results from Jack-knifing and from Taylor expansion that are clearly similar. Indeed, the degree of similarity (practical identity) is surprising, considering the quite complex algebraic formulations of both methods. This is particulary notable, since the estimated relative standard errors of the estimated variances are quite substantial, running for the components mostly from 50 percent upward. Is there some fictitious element or redundancy in computation which makes results from the two methods more alike in the numerical example given than might usually be the case?

C. The complexity of the approach suggests that the required computer programming likely is also complex—expecially when non-response is taken into account—and so one must be careful to be watchful for risk of error in this direction.

4.4 It's no detraction from this paper to note that effective design use of variance components depends not only upon knowledge of these components but, also, upon good unit cost data. And the latter are often not known with precision.

5. Summary Comments.

5.1 In summarizing my thoughts on these four papers, I can say as I did in beginning: this is a strong set of papers which I enjoyed and which make real contributions. Collectively they solve, or take significant steps toward solution of a considerable variety of important survey problems. I'm certainly not going to try at this point to catalog those problems or solutions, but may I remind the audience of several important features under discussion today.

5.2 Of first importance is the putting on a sounder basis the drawing of inferences from complex surveys--and as the current saying goes, "that's what it is all about."

5.3 The papers shed a good bit of light on what are today the three leading methods of calculating sampling variance for statistics from complex surveys, and on the comparative advantages of each.

5.4 The authors offer a number of examples and a wealth of leads in showing how analysis of variance estimates can produce added information about sampled universes beyond that commonly obtained in the first order statistic.

5.5 Lest we be too complacent about our successes in these areas, may I note that none of the papers today dealt explicitly with that other fundamental problem in survey work: measurement error—though the proposed estimators include a part of measurement variance.

5.6 Finally, I should like to take advantage of having the floor to point out one of the special advantages of the replication schemes. This is a feature which I and associates at the National Center for Health Statistics have described in earlier papers, and which we find useful and convenient. In using half-sample pseudo-replication, we print out, say, a 2-way table from not only the parent sample but also the same table as estimated from each of the perhaps 20 half samples. Thus for *any* statistic, such as a median or difference between domain means, we can compute a variance on a desk calculator, with a simple computation, using the 20 replicate estimates. This course avoids the necessity of foreseeing every need beforehand, or of a possibly substantial reprogramming of the variance run. It is, if you like, a "Jack-knife of a different color."

DID INTERCENSAL ESTIMATES GO WRONG IN THE 1960's? A VIEW FROM THE NATIONAL LEVEL

Meyer Zitter and David L. Word, U.S. Bureau of the Census

This paper reviews the performance of estimates of population of States and large SMSA's and their constituent counties published by the Census Bureau during the past decade in light of the results of the 1970 Census of Population. Differences between the estimates and the census counts are summarized, examined and reasons for such differences are explored in specific cases. Some attention is paid to the factors possibly contributing to the errors in the 1960's compared with the 1950's.

Background of Test Program

After both the 1950 and the 1960 censuses, the Census Bureau conducted comprehensive tests on the accuracy of methods of preparing population estimates. These tests not only reviewed the accuracy of census-prepared estimates but also considered the accuracy and reliability of alternative estimating procedures. The methods adopted for preparing State and metropolitan area estimates in the 1960's were chosen on the basis of the results of the 1960 test program. 1/ A similar test of even larger scale is now underway covering State and county estimates and is being carried out jointly and cooperatively with many States as part of the Federal-State Cooperative Program for Local Population Estimates. 2/ The present report is only a small part of that study and focuses on estimates that have been published and on the methods now in use by the Census Bureau. The accuracy of alternative procedures is not included here.

The present report covers both estimates that were prepared prior to the 1970 census as well as sets that represent some updating of the numbers on the basis of data available after the census, but none of the estimates incorporate any of the census results. By way of background, in January 1970, the Census Bureau published "provisional" estimates of State population as of July 1, 1969. These estimates, based mainly on the average of the results of what is commonly referred to as "Component Method II" and the "Regression", incorporated reported data series reflecting on migration and on population change up through the period ending July 1, 1968 and on extrapolations to July 1, 1969. 3/ For present purposes, these estimates (labelled "Set I" in the tables) were further extrapolated nine months to April 1, 1970 for comparison with 1970 Census. In spite of their very provisional nature, they are included in the review since they were the main figures available to the public at the time of the census and were providing the first impressions on the adequacy of intercensal estimates as the census counts were being announced.

By the time the final 1970 census results were becoming available in mid- and late 1970, it was possible to update the 1969 provisional estimates to incorporate reported data reflecting on migration up through the period ending July 1, 1969. This was done and the new figures in turn were extrapolated to the April 1, 1970 date (Set II). A final set (Set III) represents uniformly the average of the two methods used without adjustment for any States with special data. In effect it is essentially the same as the second set for all States except for seven States where supplemental local data or special censuses were used. 4/

Thus, at the State level we have for comparison three sets of estimates of State population, one set reflecting extrapolation for a 21-month period and the other two for a 9-month period. None of these estimates, in theory, represent the <u>best</u> estimates since it is now possible to develop yet another set of estimates which would incorporate all the reported available indicators of population change up through 1970.

The use of multiple estimates here is not intended to confuse (although it probably will) but rather to point up the situation that prevails. Estimates can be and are revised regularly to take advantage of the latest available data and their "illusive" nature should be kept in mind when assessing their accuracy.

In addition to the States, estimates are also reviewed and evaluated for the 100 largest SMSA's and their approximately 290 constituent counties. Only one set of estimates is evaluated here--a set consistent with published estimates for these areas for 1968. They have been updated from 1968 to 1969 by incorporating symptomatic data series for the period ending July 1, 1969 and extrapolated nine months to April 1970. The estimates incorporate the results of three estimating procedures--Component Method II, Composite, and Housing Unit--for all or part of the estimating period. 5/

Accuracy of State Estimates

Did the intercensal State estimates go wrong during the 1960's? How does the performance of the estimates in this decade compare with that of the preceding period? In general the State estimates were reasonably accurate, mainly within the margin of error expected of such estimates, and compare favorably with past experience. However, there were a sufficient number of exceptions to the generally favorable performance with some evidence of selective deterioration and regional bias to give cause for concern.

On an overall basis the average error of the provisional State estimates published before the census (table A) ran 2.3 percent--a very respectable showing in historical perspective. The corresponding error in 1960 was 2.4 percent and in 1950, 3.9 percent. The 1970 level of accuracy improves considerably as we move from the provisional series to the updated regular set of estimates (based uniformly on the average of the results of the two standard estimating procedures). The average error drops to 1.9 percent on this basis (table B). This appears to be an increase over the error of 1.6 percent by these methods when tested against 1960, but the latter has the advantage of one more year's worth of current data so we need to await the final updating of the 1970 estimates before determining the extent of difference in the two sets. Incidentally, it is noteworthy that the reduction in the 1970 error from 2.3 percent for the provisional estimates to 1.9 percent based on the standard techniques cannot be attributed entirely to updating of the estimates. Part of the difference is due to the "special treatment" given to a number of States for which, in our wisdom, we saw fit to take advantage of additional local data designed to improve the estimates. In four States we were misled, viz. in Massachusetts and Kansas, estimates tied to the State censuses led to larger deviations from the census counts than the standard techniques; in Hawaii and D.C. the additional data series proved to be unreliable. In three other States, however, Delaware, Rhode Island, and New York, Federal special censuses in the second half of the 1960's provided solid bases for later estimates. Such Federal censuses are always to be preferred over "estimates".

The overall average, however, provides only a partial measure on the accuracy of the estimates. Looking at the individual State differences we find a serious worsening of the estimates between 1960 and 1970 in a selected number of cases. A simple stratification of the States on a regional basis shows that on the average estimates for States in the South performed far more poorly than those for States outside the South. There appears to be significant deterioration in the former and improvement in the latter between 1960 and 1970. In 1960 these methods for the States in the South yielded an average error of 1.9 percent, slightly higher than the average error of 1.5 percent for States outside the South (table C). By 1970, however, the average error for the States in the South was 3.1 percent compared with 1.3 percent for the balance of the country--a significant widening of the gap. Within the South the poorest estimates were: Mississippi, with a deviation from the census of +6.6 percent, Arkansas, +5.2, South Carolina, +5.1, West Virginia, +3.6, and Florida with a -4.2 percent (table 2). Maryland and Louisiana also had errors in excess of 3 percent. In addition to the relatively large differences for these Southern States, there also appears to be a "high bias". The estimates for thirteen of the seventeen States in the South exceeded the census counts. For the 34 States outside the South, of course, a low offsetting bias resulted with only 10 positive errors.

The average error of 1.3 percent for the 34 States outside of the South represents a very commendable showing even though four of these States still had errors in excess of 3 percent (but none over 5 percent).

Explanation of Poor State Estimates

Although there is no precise answer as to why the 1970 estimates for these Southern States were so poor and much worse than in 1960, a number of factors emerge which are suggestive of the elements contributing to the errors in specific instances.

(1) Probable overcorrection of births for South Carolina and Arkansas. These States had relatively low completeness of birth registration at the time of the 1950 Birth Registration Test and the estimation system for updating may have failed to reflect real improvements in birth registration. A recent study of completeness of birth registration indicates some improvements above the allowance used in our estimates and consequently we may have overcorrected births in the 1960's in the South on an overall basis.[6]

(2) An underlying assumption of Component Method II is that the procedure provides a very accurate estimate of the school-age population and of net migration for that group. The major uncertainty of the procedure lies in extending the school-age migration rate to the migration rate for the total population. Yet, contrary to expectations, in the cases of Mississippi and South Carolina, relatively large errors occurred in the basic estimates of the school-age population. The extension to the total population merely aggravated the situation. In the case of South Carolina, enrollment rates at the elementary school level were among the lowest of the States in 1960; thus, any improvement in these rates between 1960 and 1970 should result in an overestimate for both the school-age population and for the total population.

General deterioration of the school enrollment time series underlying Component Method II may have contributed to increased errors. There were many changes during the 1960's in the type of series available for estimates and lack of consistency and comparability over the period strains the methodology to distinguish between spurious and real changes. Furthermore, school data series of poor quality could seriously affect the estimate since enrollment data carries significant weight in both methods. This particular failing was not limited to the States in the South, however.

(3) In the case of Florida which unlike the other States had a low estimate, the underestimate of the population by Method II could be expected in view of the heavy net inmigration to the State at the older ages--a migration pattern that hardly could be reflected by this method. (This problem should disappear in the 1970's as MEDICARE statistics are used to measure the older population separately).

(4) The behavior of the regression model warrants particular attention. It is quite likely that a good part of the increase in the error in the regression-based estimates for the South is due to a change in the relationship of the input variables with a resulting upward bias in the estimate for selected States. The regression

technique for population estimation, as we know, is an imprecise instrument, and depends on the assumption that the general relationships of the variables that existed in the base period will continue. The model used to generate estimates during the 1960's was built upon data series covering the 1950-1960 period. Even here the States in the South deviated slightly more from the regression line, on the average, than States outside the South. The average deviation was 2.3 percent (root-mean-square error of 3.2 percent) compared with an average of 2.0 percent for the other States (root-mean-square error of 2.3 percent). So even at best we might expect slightly larger errors for States in the South in 1970 than the other States. When "projected" to 1970, however, the average deviation for the States in the South increased while the average for the States outside of the South improved, 2.9 percent for the former and 1.6 percent for the latter. But, in fact, the data for 1960-1970 (substituting 1970 census counts for estimates) indicate an improvement in the regression model with a significant decrease in the deviation about the regression line. The improvement in the average deviation was particularly noticeable for the States in the South (table D). The basic weights also changed with the economic variables dropping in importance. At the same time, the economic variables for the States in the South were increasing well above national averages.

1960-70 data indicate significant covergence toward national averages of the economic variables underlying the regression. In South Carolina, for example, the number of automobiles per capita in 1960 was 88 percent of the national average. By 1970, this had increased to 97 percent. Nonagricultural employment per capita increased from 81 to 93 percent of the national average, and the number of income tax returns rose from 74 to 87 percent. Similar types of convergence exist for the other States mentioned. Such faster growth rates of these variables in the South for the 1960's yielded "high" population estimates when substituted into the 1950-60 regression model used to generate 1970 estimates.

The upward bias and large errors of the regression were averaged in with the larger-than-usual errors by Method II, also on the high side, so that the ultimate result was substantially poorer estimates than in the earlier decade. On the brighter side, since it appears that the 1960-70 regression model is a much better expression of the relationship of the variables and population than the model reflecting the 1950-60 period, we could expect improved performance from the regression in the 1970's. 7/

(5) The errors by each of the methods for the States indicated were generally in the same direction, thus losing an important advantage of the averaging technique. In 1970 this was true for 15 of the 17 States in the South; in 1960 this occurred in only 7 cases. This failure to receive the "breaks of statistical averaging" (due to some common bias of the methods, no doubt) also contributed to the poorer performance of the estimates.

Accuracy of SMSA Estimates

Shifting our attention to estimates for SMSA's and their constitutent counties we find much to be optimistic about with their accuracy, to some extent better than expected. On an overall basis the average error for the one hundred largest SMSA's was 2.0 percent, with no apparent regional bias, as with the States (table E). The 31 SMSA's in the South included in the test had the same average error, 2.0 percent, as the other areas. We find it an interesting anomaly that the SMSA estimates in the South were more accurate than their corresponding States--2.0 percent versus 3.1 percent. In the past we generally expected more accurate estimates for States since size of area usually has an important effect on level of accuracy. Implicitly, then, it is apparent that the nonmetropolitan or rural parts of the South were overestimated.

For the central counties, the average error was a relatively low 2.3 percent, again with no regional bias; "suburban" or ring counties had a much larger error, 3.7 percent. Here the South, non-South difference was significant--4.6 percent for the South versus 3.2 percent elsewhere, but the difference, we believe, is due more to size differentials than to any regional bias. The ring counties in the South in the test areas are generally much smaller in population than the ring counties outside the South.

In spite of the relatively small error on an overall basis, there still remained too many large errors for us to be complacent about estimates for such areas. The estimates for about 10 percent of the SMSA's differed from the census counts by more than 5 percent, about one-fourth of the ring counties had errors of this magnitude. Yet, this pattern and level of errors was about what could be expected based on observations of earlier tests of accuracy of estimates for metropolitan areas and counties. 8/

There doesn't seem to be any particular pattern or common element for the 10 SMSA's tested here with errors in excess of 5 percent. The estimates were generally low. In two instances the estimating technique cannot be faulted since the figures were tied-in with local State censuses (Massachusetts and Kansas) which eventually proved to be poor in relation to decennial census counts.

In other words, the current state of the art is such that one should expect some relatively large errors in at least a number of instances when making estimates for a large number of areas. In some instances the errors result from the inappropriations of the methodology to specific areas, such as the use of Component Method II to fast-growing resort and retirement-type areas (St. Petersburg, for example); in other cases poor or inconsistent data input could be the cause. Furthermore, most estimating procedures seem to fall short for areas of very rapid growth (table F) regardless of reason for growth.

A word about the methodology of the estimates for the SMSA's and their counties is also in order since the particular application may also affect accuracy.

In general, as stated earlier, the estimates were prepared by averaging together the results of three estimating procedures using largely independent and separate input--Component Method II, Composite, and a Housing Unit Method. 9/ However, the estimates were prepared within the framework of our State estimates program so that the procedure involved working with the SMSA's (of the 100 largest) in each State and a "balance of State" category treated as separate units. The resulting estimates were adjusted to independently derived State totals. Aside from a practical need for providing consistent sets of SMSA and State estimates, our experience has been that imposing summary control totals of larger areas over smaller areas tends to reduce the overall average error of estimates. 10/ The test program now underway with the States should provide additional evidence on this point.

In light of the evidence indicated above for States and SMSA's, why the increased doubts about the adequacy of intercensal estimates? For one thing, of course, the number of geographic areas included in the above review is only a very small percentage of the thousands of separate areas such as counties and cities for which local estimates are available. Also, these findings are not representative of the accuracy of local estimates. Furthermore, it is clear to me that the topic is sparked, in large part, by the many controversies (complaints?) that arose when preliminary census field counts were announced and so many local officials were surprised and disappointed at the results for their areas. Invariably, local opinion was that the census counts were much too low--sometimes said as a matter of faith--others were being guided (or misguided) by their own city or county estimates.

The Census Bureau has not published any extensive estimates for cities, but if we had, I suspect that the results would also not be encouraging. Our experience, based on selected test studies, is that present methodology used for city estimates tend to greatly overstate the population. We haven't made an extensive review of local city estimates against the 1970 census, but it's obvious the question that needs to be answered is why estimates prepared locally tended to overstate the population (as indicated by the census). Perhaps there are different reasons for each specific area but let me generalize for this occasion based on knowledge on how many of these estimates are usually prepared and the expected accuracy of such methods.

The Census Bureau periodically conducts surveys on types of methods and kinds of estimates prepared by local agencies. 11/ We find that by and large city estimators tend to rely on a single method and single data source for making its population estimate, most often a "housing unit method" using building permit data. Reliance on a single method is in itself a serious weakness. Basing the method entirely on building permit data (and/or utility data) compounds the inadequacies of the estimates.

Some years ago we conducted a special study on the use of the housing unit method for making population estimates for cities. Even though we concluded that the housing unit method was a useful approach to population estimation, the results indicated a "positive" bias and high average error of the method. 12/

A major problem is that building permits, which are most often used as input, give us only one side of the picture and leads to some gross estimates of the housing inventory. Converting housing units to households and to population is no simple, straightforward task since the necessary components, i.e., changes in vacancy rates and in size of household since the last benchmark are not available on a current basis. Even if a fair estimate of the number of housing units is obtained, it has not always been possible to arrive at accurate estimates of the number of households.

One interesting fact that the tests show is that the error in the number of households was also very high and, in effect, contributes as much to the overall error in the estimate of population as the error introduced by the estimate of the size of household. The error introduced by lack of adequate data on current size of household is particularly significant in 1970 because of the rather sharp decline in the average size of household in the 1960's brought about by the steep decline in fertility and the large increase in number of one-person households.

In summary, then, it appears that intercensal estimates are still viable, particularly for large areas, but considerable improvement is needed if the margin of error is to be maintained at reasonable levels. Unfortunately, accuracy of performance in one decade does not guarantee similar performance in later decades. Estimators need to be continually on the look-out for, or to arrange to develop, improved or new data series reflecting on population and new techniques of data manipulation if adequate estimates are to become available on an extensive and regular basis.

Footnotes

* The research underlying this report was carried out in the State and Local Population Estimates and Projections Branch, (Population Division), under the direction of Donald E. Starsinic, Branch Chief.

1/ For discussions, descriptions and citations of earlier studies, see "Accuracy of Methods of Preparing Postcensal Population Estimates for States and Local Areas", Meyer Zitter and Henry Shryock, Jr., Demography, Vol. I, No. 1, 1964.

2/ See, Meyer Zitter, "Federal-State Cooperative Program for Local Population Estimates, Status Report, January 1971", The Registrar and Statistician, Vol. 36, No. 4, April 1971.

3/ Although most State estimates were based uniformly on the results of the average of the two methods cited, this was not the case for 7 States where special kinds of data were available (e.g., special censuses). See, Current Population Reports, Series P-25, No. 436.

4/ Op. cit., Footnote 3/.

5/ See, Current Population Reports, Series P-25, No. 432.

6/ See, Bureau of the Census, Current Population Reports, Series P-25, No. 460, p. 5.

7/ The 1960-70 model is not final at this point since 1969 was the last year for which the variables were available. Consequently, the results are labelled "preliminary".

8/ Accuracy of Methods of Preparing Postcensal Estimates for Counties: A Summary Compilation, by Meyer Zitter, Donald E. Starsinic, and David L. Word, paper presented at annual meeting of Population Association of America, Boston, Massachusetts, August 18-20, 1968.

9/ Op. cit., Series P-25, No. 432.

10/ See article by Zitter and Shryock, "Accuracy of Methods....", op. cit., Footnote 1.

11/ The most recent inventory is published in Current Population Reports, Series P-25, No. 454.

12/ "Accuracy of the Housing Unit Method in Preparing Population Estimates for Cities", Donald E. Starsinic and Meyer Zitter, Demography, Vol. V, No. 1, 1968.

Table A.--SUMMARY OF DEVIATION (PERCENT) OF PROVISIONAL STATE ESTIMATES FROM CENSUS COUNTS: 1970, 1960, AND 1950

(Alaska and Hawaii included in 1970 only)

Summary Measure	1970	1960	1950
All States			
Average deviation	2.30	2.44	3.93
Root means square deviation	2.86	3.39	5.53
Deviation in excess of 3%	18	14	19
Deviation in excess of 5%	3	4	13
Positive deviations	25	28	21
South			
Average deviation	3.17	2.49	4.04
Root means square deviation	3.68	3.45	4.87
Deviation in excess of 3%	9	5	10
Deviation in excess of 5%	2	2	6
Positive deviations	14	9	3
Non-South			
Average deviation	1.86	2.41	3.87
Root means square deviation	2.34	3.35	6.51
Deviation in excess of 3%	9	9	8
Deviation in excess of 5%	1	2	7
Positive deviations	11	19	18

Table B.--SUMMARY OF DEVIATIONS (PERCENT) OF ALTERNATIVE SETS OF 1970 STATE ESTIMATES FROM 1970 CENSUS

Summary Measure	Set I Provisional 1969 Estimates Published in P-25, No. 436 Extrapolated to 1970	Set II Revised 1969 Estimates Extrapolated to 1970	Set III Average of Methods for 1969 and Extrapolated to 1970
All States			
Average deviation	2.30	2.06	1.86
Root means square deviation	2.86	2.56	2.37
Deviation in excess of 3%	18	15	11
Deviation in excess of 5%	3	3	3
Positive deviations	25	26	23
South			
Average deviation	3.17	3.22	3.05
Root means square deviation	3.68	3.57	3.40
Deviation in excess of 3%	9	6	7
Deviation in excess of 5%	2	0	3
Positive deviations	14	12	13
Non-South			
Average deviation	1.86	1.48	1.26
Root means square deviation	2.34	1.88	1.65
Deviation in excess of 3%	9	9	4
Deviation in excess of 5%	1	3	0
Positive deviations	11	14	10

Table C.--SUMMARY OF DEVIATION (PERCENT) OF STATE ESTIMATES BY COMPONENT AND REGRESSION METHODS: 1970 AND 1960

(Includes Alaska and Hawaii in 1970 only)

Summary Measure	Method II 1960	Method II 1970	Regression 1960	Regression 1970	Average of Methods	
					1960	1970
All States						
Average deviation	2.31	2.23	2.72	2.05	1.64	1.86
Root means square deviation	3.52	2.82	3.66	2.62	2.41	2.37
Deviation in excess of 3%	10	13	17	12	6	11
Deviation in excess of 5%	4	5	8	2	2	3
Positive deviations	28	30	20	20	25	23
South						
Average deviation	3.16	3.55	2.79	2.90	1.88	3.05
Root means square deviation	5.03	4.09	3.98	3.39	2.84	3.40
Deviation in excess of 3%	5	9	5	8	2	7
Deviation in excess of 5%	3	5	3	1	1	3
Positive deviations	10	13	10	13	12	13
Non-South						
Average deviation	1.87	1.58	2.68	1.62	1.51	1.26
Root means square deviation	2.34	1.95	3.47	2.13	2.14	1.65
Deviation in excess of 3%	5	4	12	4	4	4
Deviation in excess of 5%	1	0	5	1	1	0
Positive deviations	18	17	10	7	13	10

Table D.--SUMMARY OF DISPERSION OF CENSUS COUNTS ABOUT LEAST SQUARE REGRESSION LINE: 1950-60 AND 1960-70

(Figures are expressed as percent deviations of estimates derived from regression (Y_c) from Census counts)

Summary Measure	1960 Census Deviation: 1950-60 Projected from 1940-50	1960 Census Deviation: 1950-60 Actual a/	1970 Census Deviation: 1960-70 Projected from 1950-60	1970 Census Deviation: 1960-70 Actual b/
All States				
Average deviation..............	2.72	2.07	2.05	1.52
Root means square deviation....	3.66	2.59	2.62	1.99
Deviation in excess of 3%......	17	7	12	5
Deviation in excess of 5%......	8	3	2	2
Positive deviations............	20	22	20	23
South				
Average deviation..............	2.79	2.29	2.90	1.44
Root means square deviation....	3.98	3.15	3.39	1.78
Deviation in excess of 3%......	5	2	8	1
Deviation in excess of 5%......	3	3	1	2
Positive deviations............	10	8	13	12
Non-South				
Average deviation..............	2.68	1.95	1.62	1.67
Root means square deviation....	3.47	2.32	2.13	2.34
Deviation in excess of 3%......	12	5	4	4
Deviation in excess of 5%......	5	0	1	0
Positive deviations............	10	14	7	11

a/ Regression Equation: $Y_c = .06+.30X_1+.14X_2+.22X_3+.08X_4+.07X_5+.12X_6$

X_1 = Births; X_4 = Income Tax Returns (Federal)

X_2 = Deaths; X_5 = Passenger Auto Registration

X_3 = Ele. Enrollment; X_6 = Nonagricultural Employment

b/ Provisional Regression Equation: $Y_c = -.08+.23X_1+.25X_2+.46X_3+.09X_4+.03X_5+.05X_6$

Table E.--DIFFERENCES (PERCENT) BETWEEN CENSUS COUNTS AND ESTIMATES FOR 100 LARGEST SMSA'S: 1970

	Average Percent Difference: SMSA's	Average Percent Difference: Central Counties	Average Percent Difference: Suburban Counties	No. of errors in excess of 5%: SMSA's	No. of errors in excess of 5%: Central Counties	No. of errors in excess of 5%: Suburban Counties
Total, 100 largest SMSA's	2.0	(120) 2.3	(169) 3.7	10	12	45
South (N=31)	2.0	(37) 2.3	(57) 4.6	3	4	21
Non-South (N=69)	2.0	(83) 2.3	(112) 3.2	7	8	24
100 SMSA's by size group						
25 largest	2.1					
Next 25 largest	1.8					
3rd group	1.8					
4th group	2.2					

Distribution of Errors for 100 Largest SMSA's

	All SMSA's	SMSA's Outside South	South
N =	100	69	31
Less than 1 percent	33	24	9
1.0 to 1.9 percent	27	18	9
2.0 to 2.9 percent	18	11	7
3.0 to 4.9 percent	12	9	3
5.0 percent and over	10	7	3

Table F.--PERCENT DISTRIBUTION OF COUNTY ERRORS, BY RATE OF GROWTH

Average error \ Growth rate	Gains of: Less than 10%	Gains of: 10 to 25%	Gains of: 25 to 50%	Gains of: 50% and over	Population loss
N =	(69)	(86)	(72)	(41)	(22)
Under 3 percent.....	64	73	54	46	64
3 to 5 percent......	23	17	15	19	18
5 percent and over..	13	10	31	34	18

Table 1.--PERCENT DEVIATION OF ALTERNATIVE SETS OF STATE ESTIMATES FROM CENSUS COUNTS: 1970

Region, Division, and State	Population April 1, 1970 1/ (Census) (In thousands)	PERCENT DIFFERENCE		
		Set I Provisional 1969 Estimates Published in P-25, No. 436 Extrapolated to 1970	Set II Revised 1969 Estimates Extrapolated to 1970	Set III Average of Methods for 1969 and Extrapolated to 1970
UNITED STATES, TOTAL.......	203,185	0.3	0.3	0.2
Regions				
Northeastern States......	49,001	0.2	-0.2	0.4
North Central States.....	56,577	-0.3	-0.3	-0.5
The South................	62,797	1.5	1.8	1.6
The West.................	34,810	-1.0	-0.9	-1.2
Northeast				
New England..............	11,848	-2.2	-2.7	-1.8
Middle Atlantic..........	37,153	1.0	0.6	1.2
North Central				
East North Central.......	40,253	-0.1	-0.6	-0.6
West North Central.......	16,324	-0.5	0.4	-0.1
South				
South Atlantic...........	30,671	0.6	0.5	0.1
East South Central.......	12,804	3.0	3.1	2.9
West South Central.......	19,322	2.0	2.9	2.9
West				
Mountain.................	8,282	-2.0	-1.0	-1.2
Pacific..................	26,528	-0.7	-0.8	-1.2
New England				
Maine....................	994	-1.5	-2.0	-2.2
New Hampshire............	738	-1.6	-1.5	-1.2
Vermont..................	445	-0.5	-1.1	-0.5
Massachusetts............	5,689	-3.4	-3.9	-2.7
Rhode Island.............	950	-3.7	-3.7	0.1
Connecticut..............	3,032	0.2	-0.7	-1.0
Middle Atlantic				
New York.................	18,191	1.4	0.6	2.2
New Jersey...............	7,168	0.9	1.1	0.7
Pennsylvania.............	11,794	0.4	-0.2	-0.2
East North Central				
Ohio.....................	10,652	1.6	0.9	0.9
Indiana..................	5,194	-0.8	-1.0	-1.0
Illinois.................	11,114	0.1	-0.6	-0.9
Michigan.................	8,875	-0.4	-0.4	-0.5
Wisconsin................	4,418	-3.7	-3.5	-3.6
West North Central				
Minnesota................	3,805	-2.2	-1.7	-1.5
Iowa.....................	2,825	-1.5	1.1	1.4
Missouri.................	4,677	(Z)	0.7	0.7
North Dakota.............	618	-0.7	1.0	0.2
South Dakota.............	666	-1.4	-0.3	-0.6
Nebraska.................	1,484	-2.2	-1.7	-1.6
Kansas...................	2,249	3.7	4.1	-0.4
South Atlantic				
Delaware.................	548	-	-1.4	-2.8
Maryland.................	3,922	-2.6	-3.0	-3.4
District of Columbia.....	757	5.8	4.2	-1.3
Virginia.................	4,648	1.7	1.4	0.8
West Virginia............	1,744	4.1	3.4	3.6
North Carolina...........	5,082	3.5	2.8	2.9
South Carolina...........	2,591	4.9	5.1	5.1
Georgia..................	4,590	2.4	2.7	2.2
Florida..................	6,789	-4.7	-4.2	-4.2
East South Central				
Kentucky.................	3,219	0.9	1.5	1.6
Tennessee................	3,924	2.4	2.4	2.2
Alabama..................	3,444	3.1	3.1	2.6
Mississippi..............	2,217	7.1	6.9	6.6
West South Central				
Arkansas.................	1,923	4.6	5.4	5.2
Louisiana................	3,643	3.9	3.5	3.2
Oklahoma.................	2,559	1.1	1.2	1.3
Texas....................	11,197	1.1	2.7	2.8
Mountain				
Montana..................	694	0.3	-0.5	-0.9
Idaho....................	713	1.3	2.2	2.4
Wyoming..................	332	-3.9	-1.3	-0.7
Colorado.................	2,207	-3.6	-3.2	-3.0
New Mexico...............	1,016	-1.9	0.7	0.4
Arizona..................	1,772	-2.7	-1.1	-1.3
Utah.....................	1,059	-0.1	0.1	-0.1
Nevada...................	489	-3.7	-2.3	-3.8
Pacific				
Washington...............	3,409	1.1	0.1	0.7
Oregon...................	2,091	-1.8	-0.4	-0.3
California...............	19,956	-1.1	-1.1	-1.7
Alaska...................	302	-5.0	-3.5	-3.4
Hawaii...................	770	4.8	1.9	-0.2

1/ Figures from 1970 Census of Population PC(V2)-1, United States, Advance Report, Table 1. See, PC(1)-A, U.S. Summary, for final corrections.

TABLE 2.--PERCENT DEVIATION OF STATE ESTIMATES FOR CENSUS COUNTS BY COMPONENT AND REGRESSION METHODS: 1970 AND 1960

Division and State	Method II 1960	Method II 1970	Regression 1960	Regression 1970	Average of methods 1960	Average of methods 1970
UNITED STATES, TOTAL..	-0.01	+0.24	-0.01	+0.24	-0.01	+0.24
New England:						
Maine	+0.12	-2.64	-1.18	-1.73	-0.53	-2.18
New Hampshire	+0.33	-0.91	+0.56	-1.48	+0.44	-1.19
Vermont	-2.91	+0.22	-6.18	-1.26	-4.54	-0.52
Massachusetts	-0.18	-2.82	+2.01	-2.62	+0.92	-2.72
Rhode Island	+0.75	+0.37	+0.06	-0.12	+0.40	+0.12
Connecticut	-2.55	-0.98	+2.98	-1.09	+0.22	-1.04
Middle Atlantic:						
New York	-0.11	+2.96	+0.40	+1.52	+0.14	+2.24
New Jersey	-1.54	+1.74	+1.43	-0.30	-0.06	+0.72
Pennsylvania	+1.22	+0.07	+0.21	-0.55	+0.72	-0.24
East North Central:						
Ohio	+2.08	+0.88	-0.40	+0.84	+0.84	+0.86
Indiana	-1.00	-1.50	-4.53	-0.41	-2.76	-0.95
Illinois	+2.03	-1.77	+1.21	+0.02	+1.62	-0.87
Michigan	+2.47	-0.88	-4.15	-0.02	-0.84	-0.45
Wisconsin	+2.86	-2.42	-1.38	-4.84	+0.74	-3.63
West North Central:						
Minnesota	+0.34	-0.46	-2.20	-2.45	-0.93	-1.45
Iowa	+2.59	+2.55	-4.67	+0.26	-1.04	+1.41
Missouri	-0.29	+0.45	-1.35	+1.01	-0.82	+0.73
North Dakota	+4.38	+2.66	-6.43	-2.23	-1.02	+0.22
South Dakota	+4.15	+1.16	-2.20	-2.29	+0.98	-0.57
Nebraska	+4.21	-0.94	-1.45	-2.28	+1.38	-1.61
Kansas	-1.01	-0.75	-4.03	-0.11	-2.52	-0.43
South Atlantic:						
Delaware	+3.87	-3.78	-4.20	-1.80	-0.16	-2.79
Maryland	-1.57	-2.86	+2.08	-3.88	+0.26	-3.37
District of Columbia	-17.10	-5.54	-1.24	+2.77	-9.17	-1.34
Virginia	+0.46	+1.73	-0.09	-0.09	+0.18	+0.82
West Virginia	+5.97	+4.55	-2.59	+2.63	+1.69	+3.59
North Carolina	+1.60	+2.16	+2.35	+3.53	+1.98	+2.85
South Carolina	+2.50	+7.11	+1.75	+3.11	+2.12	+5.11
Georgia	-1.81	+1.20	+3.05	+3.28	+0.62	+2.24
Florida	-6.06	-7.42	+10.15	-1.01	+2.04	-4.21
East South Central:						
Kentucky	+4.86	+0.86	-0.26	+2.35	+2.30	+1.60
Tennessee	-0.58	+1.12	-0.39	+3.24	-0.48	+2.18
Alabama	+0.26	+3.38	+8.71	+1.87	+4.48	+2.62
Mississippi	+2.23	+5.05	+0.78	+8.19	+1.50	+6.62
West South Central:						
Arkansas	-0.94	+5.42	+5.90	+4.90	+2.48	+5.16
Louisiana	-2.39	+4.45	+0.77	+1.85	-0.81	+3.15
Oklahoma	+0.55	+1.50	-0.77	+1.15	-0.11	+1.33
Texas	+0.90	+2.06	+2.41	+3.53	+1.66	+2.79
Mountain:						
Montana	+1.89	+2.75	-3.04	-4.54	-0.58	-0.89
Idaho	+1.62	+4.33	-6.53	+0.48	-2.46	+2.41
Wyoming	-1.41	+1.12	-2.16	-2.46	-1.78	-0.67
Colorado	-1.43	-1.95	-0.95	-4.07	-1.19	-3.01
New Mexico	-5.77	+3.06	-7.60	-2.21	-6.68	+0.42
Arixona	-0.92	-0.20	-0.18	-2.31	-0.55	-1.25
Utah	+1.08	+0.10	-3.31	-0.33	-1.12	-0.12
Nevada	+2.58	-2.54	+7.25	-5.14	+4.92	-3.84
Pacific:						
Washington	-0.91	+0.75	-2.09	+0.57	-1.50	+0.66
Oregon	+0.84	+0.63	+0.05	-1.26	+0.44	-0.32
California	-4.11	-1.90	-3.45	-1.42	-3.78	-1.66
Alaska		-4.18		-2.53		-3.35
Hawaii		-0.74		+0.41		-0.17

Table 3.--DISPERSION OF CENSUS COUNTS ABOUT LEAST SQUARE REGRESSION LINE: 1950-60 AND 1960-70

(Figures are expressed as percent deviations of estimates derived from regression (Y_c) from Census counts)

Region, Division and State	1950-60	1960-70 1/ Preliminary
New England		
Maine	+1.0	-2.4
New Hampshire	+2.6	-1.4
Vermont	-0.4	-3.2
Massachusetts	+3.7	-1.3
Rhode Island	+0.7	+0.4
Connecticut	+2.8	+0.5
Middle Atlantic		
New York	+1.9	+3.3
New Jersey	+1.6	+1.9
Pennsylvania	-0.1	-0.2
East North Central		
Ohio	-0.7	+2.0
Indiana	-2.4	-0.4
Illinois	+2.7	+1.7
Michigan	-4.4	+1.1
Wisconsin	+1.2	-3.2
West North Central		
Minnesota	+0.6	-1.7
Iowa	-0.7	-0.3
Missouri	+1.1	+2.3
North Dakota	-1.4	-1.0
South Dakota	+1.6	-1.1
Nebraska	+2.7	-1.2
Kansas	-0.2	+0.5
South Atlantic		
Delaware	-2.3	+0.7
Maryland	-0.5	-2.7
District of Columbia	+5.2	+5.5
Virginia	-1.0	-0.3
West Virginia	-5.2	-1.1
North Carolina	+0.6	(Z)
South Carolina	-3.0	+1.0
Georgia	+1.8	+1.0
Florida	-0.1	+0.8
East South Central		
Kentucky	+0.2	-0.4
Tennessee	-0.4	+0.9
Alabama	+2.0	+0.7
Mississippi	+4.0	+5.2
West South Central		
Arkansas	+8.3	+2.1
Louisiana	-1.9	+2.7
Oklahoma	+1.5	+0.1
Texas	-1.0	+3.3
Mountain		
Montana	-2.1	-2.0
Idaho	-3.2	+0.3
Wyoming	-0.4	-2.0
Colorado	-1.7	-4.0
New Mexico	-4.7	-0.4
Arizona	-1.8	-0.6
Utah	-2.6	-0.1
Nevada	+4.8	-0.9
Pacific		
Washington	-2.5	-1.0
Oregon	-1.5	-2.7
California	-2.7	-0.1
Alaska		+1.4
Hawaii		-2.6

1/ Preliminary; Independent variables are for 1960 and 1969.

Table 4.--PERCENT DEVIATION OF ESTIMATES FOR 100 LARGEST SMSA'S FROM CENSUS: 1970

SMSA	April 1, 1970 Census (In thousands)	Percent Difference
New York, N.Y.	11,570	0.1
Los Angeles-Long Beach, Calif.	7,032	-0.6
Chicago, Ill.	6,979	-1.9
Philadelphia, Pa.	4,818	1.1
Detroit, Mich.	4,200	-0.1
San Francisco-Oakland, Calif.	3,110	-2.1
Washington, D.C.-Md.-Va.	2,861	-0.5
Boston, Mass. 1/	3,375	-4.4
Pittsburgh, Pa.	2,401	-1.0
St. Louis, Mo.-Ill.	2,363	0.7
Baltimore, Md.	2,071	-3.4
Cleveland, Ohio	2,064	0.6
Houston, Texas	1,985	0.5
Newark, N. J.	1,857	2.7
Minneapolis-St. Paul, Minn.	1,814	-5.0
Dallas, Texas	1,556	1.0
Seattle-Everett, Wash.	1,422	-2.0
Anaheim-Santa Ana-Garden Grove, Calif.	1,420	-5.1
Milwaukee, Wis.	1,404	-4.6
Atlanta, Ga.	1,390	0.9
Cincinnati, Ohio	1,385	-1.7
Paterson-Clifton-Passaic, N. J.	1,359	1.0
San Diego, Calif.	1,358	-5.0
Buffalo, N. Y.	1,349	-1.5
Miami, Fla.	1,268	-5.1
Kansas City, Mo.-Kan.	1,257	0.7
Denver, Colo.	1,228	-5.2
San Bernadine-Riverside-Ontario, Calif.	1,143	-0.7
Indianapolis, Ind.	1,110	-2.7
San Jose, Calif.	1,065	-3.2
New Orleans, La.	1,046	2.1
Tampa -St. Petersburg, Fla.	1,013	-3.9
Portland, Ore.-Wash.	1,009	-0.9
Phoenix, Ariz.	968	-4.2
Columbus, Ohio	916	-2.3
Providence-Pawtucket-Warwick, R.I.-Mass. 2/	770	-2.0
Rochester, N. Y.	883	-1.0
San Antonio, Texas	864	1.6
Dayton, Ohio	850	0.8
Louisville, Ky.-Ind.	827	-0.7
Sacramento, Calif.	801	-2.5
Memphis, Tenn.-Ark.	770	2.2
Fort Worth, Texas	762	-3.2
Birmingham, Ala.	739	0.1
Albany-Schenectady-Troy, N. Y.	722	0.4
Toledo, Ohio-Mich.	692	-1.0
Norfolk-Portsmouth, Va.	681	1.0
Akron, Ohio	679	1.6
Hartford, Conn. 3/	817	-0.5
Oklahoma City, Okla.	641	-1.9
Syracuse, N. Y.	637	-0.3
Gary-Hammond-East Chicago, Ind.	633	-1.9
Honolulu, Hawaii	631	1.3
Ft. Lauderdale-Hollywood, Fla.	620	-5.5
Jersey City, N. J.	609	0.8
Greensboro-Winston-Salem-High Point, N. C.	604	0.7
Salt Lake City, Utah	558	-0.9
Allentown-Bethlehem-Easton, Pa.-N. J.	544	-1.7
Omaha, Nebr.-Iowa	541	-1.9
Nashville, Tenn.	541	1.3
Grand Rapids, Mich.	539	-3.3
Youngstown-Warren, Ohio	536	-0.4
Springfield-Chicopee-Holyoke, Mass. 4/	583	-5.5
Jacksonville, Fla.	529	-0.6
Richmond, Va.	518	1.9
Wilmington, Del.-N. J.	499	-0.4
Flint, Mich.	497	-0.2
Tulsa, Okla.	476	0.8
Orlando, Fla.	428	-2.8
Fresno, Calif.	413	1.8
Tacoma, Wash.	411	0.7
Harrisburg, Pa.	411	-2.4
Charlotte, N. C.	409	2.4
Knoxville, Tenn.	400	1.8
Wichita, Kansas	389	5.9
Bridgeport, Conn. 5/	793	0.5
Lansing, Mich.	378	-2.7
Mobile, Ala.	377	1.3
Canton, Ohio	372	-0.8
Davenport-Rock Island-Moline, Iowa-Ill.	363	2.2
El Paso, Texas	359	-1.1
New Haven, Conn. 6/	745	-1.8
Worcester, Mass. 7/	638	-4.2
Wilkes-Barre-Hazelton, Pa.	342	-0.3
Peoria, Ill.	342	-0.9
Utica-Rome, N. Y.	341	6.2
York, Pa.	330	-3.0
Bakersfield, Calif.	329	2.4
Little Rock-North Little Rock, Ark.	323	2.2
Lancaster, Pa.	320	-3.1
Beaumont-Port Arthur-Orange, Texas	316	0.9
Chattanooga, Tenn.-Ga.	305	2.0
Binghamton, N. Y.	303	1.3
Reading, Pa.	296	-0.7
Shreveport, La.	294	2.4
Spokane, Wash.	287	-4.5
South Bend, Ind.	280	-1.4
Duluth-Superior, Minn.-Wis.	265	1.1
Johnstown, Pa.	263	(Z)
Corpus Christi, Texas	285	5.6

1/ Mass. SEA C
2/ Rhode Island SEA A
3/ Conn. SEA C
4/ Mass. SEA A
5/ Conn. SEA A
6/ Conn. SEA B
7/ Mass. SEA B

(Z) Less than 500 or 0.05 percent.

DID INTERCENSAL ESTIMATES GO WRONG? A VIEW FROM ARKANSAS

Forrest H. Pollard, University of Arkansas

The purpose of this presentation is to examine the degree of reliability of postcensal population estimates constructed for the 1960-1970 intercensal period for the State of Arkansas and its respective counties.

To measure the reliability or accuracy of the estimative techniques, a comparison was made between estimates prepared for April 1, 1970, and United States Bureau of the Census counts as of this date. Three estimative techniques, standardized methods recommended by the Bureau of the Census for participants in its Federal-State Cooperative Program for Local Population Estimates, were used in the analyses.

The techniques were Component Method II, Composite Method and Ratio-Correlation Method [1, p. 66].

The remainder of the report contains: 1) a brief description of each of the methods used and the procedure followed in making the estimates; 2) background information concerning Arkansas' population; 3) the comparison of estimate results with 1970 census counts; and 4) presentation of findings and recommendations.

1. ESTIMATIVE PROCEDURE

Only a summary description of each of the three methods used in preparing the estimates is provided, since detailed published explanations of these techniques are readily available.

In applying Component Method II, the latest Bureau of the Census civilian population count for the area is used as the estimate base. Adjustments are made to this count to account for changes, resulting from natural increase and net migration, occurring to the area's population over the estimate time interval. The specific population components accounting for change are births, deaths, net civilian migration, net movement of civilians into the Armed Forces, and military personnel stationed in the area as of the estimate date. A symptomatic data series, school enrollment, is utilized as the base for measuring net civilian migration.[1]

By contrast the Composite Method consists of estimates prepared by age, sex, and color, utilizing various symptomatic data to estimate the population of the several age groups. A frequently used age classification is: under 5; 5-17; 18-44; 45-64; and 65 years of age and over. Birth registrations, census counts and school enrollment data are used to estimate the size of the populations under five years of age and 5-17 years of age; births for females ages 18-44, and estimated sex ratios applied to the expected female population to determine the number of males estimated to be in this age group; and death registrations for those groups 45 years of age and over. Estimates for each of these age groups are then summed to provide a total expected population. The Census Bureau's Composite estimate technique is a variation of the Bogue-Duncan Composite Method.[2]

The third estimating technique utilized was the Ratio-Correlation Method. A multiple regression equation based on data for the 1950-1960 intercensal period was derived for use in preparing post-1960 census annual county population estimates. The equation states the relationship between five independent variables, which are expressed as ratios reflecting the change over the intercensal period, in a county's share of the State total for each of the symptomatic data series and the dependent variable, which is a ratio reflecting the change in the county's share of the State population over the intercensal period.[3] The symptomatic data series included in the equation were births, deaths, school enrollment, employment, and motor vehicle registrations.[4]

County estimates, as of July 1, 1969, were prepared utilizing each of the three methods. The resulting three estimates for each county were then averaged to derive a single county number. These county estimates were adjusted to agree with the Census Bureau's estimate of the State's total population. As a final step, the county estimates were then extrapolated to April 1, 1970, on the basis of the annual rate of population change estimated to have occurred within each county between April 1, 1960, and July 1, 1969.

2. BACKGROUND INFORMATION ABOUT ARKANSAS' POPULATION

Arkansas' 1970 census count of 1,923,295 ranks thirty-second in size among the 50 states, reflecting a population growth of 137,023 (7.7 percent) during the 1960's. Although this rate of increase (7.7 percent) is not as great as that of the Nation (13.3 percent), it does represent a significant reversal of a pattern experienced during the previous twenty years in which Arkansas' population fell from a 1940 count of 1,949,387 to 1,786,272 in 1960, reflecting a population loss of 163,115 (8.4 percent). Prior to this period the State's population had increased in each decennial census from an 1890 count of 1,128,211 to 1,949,387 in 1940.

---

[1] For a detailed discussion of this method see [2].

[2] The Bogue-Duncan Method is described in [3] and the Census Bureau's Composite Method in [4].

[3] For example, the value of an independent variable X, births, would be expressed as follows:

$$\frac{\text{Percent of total state births in county i, 1960}}{\text{Percent of total state births in county i, 1950}}$$

[4] A description of this method is contained in [5, pp. 279-281] and [6, pp. 36-39].

Only 12 of Arkansas' 75 counties had larger populations in 1960 than in 1940. However, 46 (61.3 percent) of the State's counties experienced growth between 1960 and 1970.

Of Arkansas' 75 counties the 1970 census showed: one having a population larger than 100,000; 21 with populations ranging from 25,000 to 100,000; 20 with populations ranging from 15,000 to 25,000; and 33 counties with populations of less than 15,000.

There are four SMSA's whose central cities are located entirely (three) or partially (one) within the State. The SMSA's are Little Rock-North Little Rock; Pine Bluff; Fort Smith, Arkansas-Oklahoma; and Texarkana, Texas-Arkansas.

3. COMPARISON OF ESTIMATES TO CENSUS

As previously stated, this presentation involves a comparison of county population estimates prepared as of April 1, 1970, with Bureau of the Census counts as of the same date. The basic summary measure used in examining the relative reliability of the estimates is the average percent error. This measure is the average of the total of percentage deviations of each county estimate from its census count, signs disregarded.

Two sets of county estimates, each utilizing the three estimate techniques were prepared. One set of estimates was adjusted to agree with the State estimate total prepared by the Bureau of the Census and the other set was not adjusted.[5]

Tables 1 and 1-A show the average percent error and other summary measures of percentage error pertaining to the 1970 Arkansas county population estimates.

The average error for the adjusted county estimates (Table 1), considering each technique separately, was 7.0 percent for Ratio-Correlation Method, 7.8 percent for the Composite Method, and 9.1 percent for Component Method II. However, averaging the results of either the Composite and Ratio-Correlation or the results derived from applying all three techniques provides an even smaller average error, 6.8 percent.

Significantly smaller average errors were obtained when the estimates were not adjusted (Table 1-A). The average error for the Ratio-Correlation Method was 5.5 percent; Component Method II, 6.5 percent; and the Composite Method, 7.1 percent. An even smaller average error, 5.0 percent, is realized by averaging the results of either the application of all three techniques or the average of Component Method II and the Ratio-Correlation Method.

Distributions of the State's 75 counties by size of error, utilizing an average of the three techniques (Table 1-A), indicated that five of the counties had errors in excess of ten percent, 34 in excess of five percent, and 48 in excess of three percent.

Tables 2 and 2-A seem to indicate that the average percentage error is less for metropolitan (SMSA) county estimates than for other urban and for rural counties, although population size may be a factor in this observation. For the adjusted estimates the smallest average error observed was 4.1 percent for the metropolitan (SMSA) counties; 6.5 percent for those counties whose population is 50 percent or more urban; and 6.8 percent for the rural counties. Smallest average errors for the nonadjusted estimates were 1.9 percent, 4.2 percent, and 5.2 percent respectively.

Average percentage errors by county population size are also shown in Tables 2 and 2-A. The adjusted estimates indicate that the smallest average error was 0.1 percent for the State's only county with a population of over 100,000; 6.1 percent for the 21 counties with populations between 25,000 and 100,000; 7.6 percent for the 20 counties with populations between 15,000 and 25,000; and 6.6 percent for the 33 counties with populations of 15,000 or less. For the unadjusted estimates the percents were 0.3; 4.1; 5.2; and 5.4 respectively. These average errors seem to indicate that no improvement is secured in estimates of counties with populations of 15,000 to 25,000 when compared to estimates for counties with populations of 15,000 or less.

Counties with increasing populations experienced smaller average percentage errors than counties that were losing population. Tables 3 and 3-A show that for Arkansas' 29 counties that lost population, between 1960 and 1970, the smallest average percentage error was 8.2 for the adjusted estimates and 5.3 percent for the unadjusted estimates. These percents compare to average percentage errors of 5.7 and 3.5 respectively for the 13 counties growing at an intercensal rate of less than 7.7 percent and to average percentage errors of 5.5 and 4.7 respectively for the 33 counties growing at a rate greater than 7.7 percent.

4. FINDINGS AND RECOMMENDATIONS

1. The preceding analysis shows that the adjustment of the county estimates to agree with the Bureau of the Census' State control total resulted in estimate errors (the smallest of which was an average deviation of 6.8 percent for the counties and a percentage deviation of 5.4 for the State) that were greater than desired.

2. However, the estimate errors computed on the basis of the unadjusted estimates (the smallest of which was an average error of 5.0 percent for the counties and a percentage deviation of 2.1 for the State) would have been acceptable.

3. It is recommended that the desirability of the school enrollment adjustment procedure be reevaluated prior to its continued use. This adjustment contributed substantially to the upward bias in the Arkansas population estimate.

---

[5]The adjusted 1970 State estimate, based on the Component Method II and Ratio-Correlation Techniques, differed from the Bureau of the Census Count by 5.4 percent whereas, the unadjusted State estimate, based on the Component Method II and Composite Techniques differed by only 2.1 percent.

An upward adjustment in reported school enrollment to derive the population 7.5 through 14.5 years of age and the resulting effect on estimated net migration accounted for a substantial part of the difference between the adjusted and unadjusted estimates.

4. Births occurring in Arkansas were corrected for under-registration on the basis of a 1950 Birth Completeness Test. This correction procedure should be modified by the Bureau of the Census to reflect improvements occurring in birth registrations, since it also contributed to the upward bias in the State estimate.

5. To summarize, results of the evaluation appear to indicate that the method utilized in constructing the estimates was appropriate (an average of Component Method II, Composite and Ratio-Correlation techniques); that the input data were sufficiently reliable; and that had the assumptions concerning the adjustments for school enrollment and births been more realistic, then the Arkansas estimates would have been within an acceptable range of error.

REFERENCES

[1] Zitter, Meyer, "Federal-State Cooperative Program for Local Population Estimates: Status Report, January 1971," *The Registrar and Statistician,* Vol. 34, No. 4 (April 1971).

[2] U. S. Department of Commerce, Bureau of the Census, *Current Population Reports,* "Methods of Population Estimation: Part I–Illustrative Procedure of the Census Bureau's Component Method II," Series P-25, No. 339 (Washington: Government Printing Office, June 1966).

[3] Bogue, Donald J., and Duncan, Beverly, "A Composite Method for Estimating Postcensal Populations of Small Areas by Age, Sex, and Color," *Vital Statistics–Special Reports,* Vol. XLVII, No. 6, U. S. Department of Health, Education and Welfare, National Office of Vital Statistics (Washington: Government Printing Office, August 1959).

[4] U. S. Department of Commerce, Bureau of the Census, *Current Population Reports,* "Estimates of the Population of Counties and Metropolitan Areas, July 1, 1966, A Summary Report," Series P-25, No. 427 (Washington: Government Printing Office, August 14, 1967).

[5] Schmitt, Robert C., and Crosetti, Albert H., "Accuracy of the Ratio-Correlation Method for Estimating Postcensal Population," *Land Economics,* Vol. XXX, No. 3 (August 1954).

[6] Goldberg, David, Feldt, Allen, and Smit, J. William, "Michigan Population Studies, No. 1," *Estimates of Population Change in Michigan, 1950-1960* (Ann Arbor, Michigan: The University of Michigan, 1960).

TABLE 1

SUMMARY MEASURES OF PERCENTAGE ERRORS FROM 1970 CENSUS
OF 1970 COUNTY POPULATION ESTIMATES, ARKANSAS

(Adjusted)

Summary Measures	Component Method II (X_1)	Composite Method (X_2)	Ratio Correlation (X_3)	Averages (X_1,X_2)	(X_2,X_3)	(X_1,X_3)	(X_1,X_2,X_3)
Average error (percent)	9.1	7.8	7.0	7.4	6.8	7.3	6.8
Root mean square error	11.1	9.9	8.4	9.1	8.3	8.7	8.2
Number of errors in excess of 3.0 percent	61	55	56	58	56	57	55
Number of errors in excess of 5.0 percent	50	42	44	43	41	48	46
Number of errors in excess of 10.0 percent	31	25	18	22	16	23	21
Number of positive errors	63	56	60	61	58	66	62

TABLE 1-A

(Not Adjusted[1])

Average error (percent)	6.5	7.1	5.5	5.5	5.6	5.0	5.0
Root mean square error	8.0	9.0	6.9	6.9	6.9	6.0	6.1
Number of errors in excess of 3.0 percent	53	54	47	52	50	46	48
Number of errors in excess of 5.0 percent	41	38	39	35	37	33	34
Number of errors in excess of 10.0 percent	19	23	9	12	10	8	5
Number of positive errors	49	46	43	49	49	48	52

[1] The State totals resulting from these county estimates were not adjusted to agree with State population estimate totals prepared by the U.S. Bureau of the Census.

Source: Industrial Research and Extension Center, University of Arkansas.

TABLE 2

AVERAGE PERCENTAGE ERRORS FROM 1970 CENSUS OF 1970 COUNTY POPULATION ESTIMATES, ARKANSAS, BY COUNTY SIZE CLASS

(Adjusted[1])

County Size Class	Number of Counties	Component Method II (X_1)	Composite Method (X_2)	Ratio Correlation (X_3)	Averages (X_1,X_2)	Averages (X_2,X_3)	Averages (X_1,X_3)	Averages (X_1,X_2,X_3)
Group I								
Metropolitan counties, SMSA's	7	6.4	5.5	4.1	5.2	4.1	5.2	4.7
Other urban counties (50 percent or more of population urban)	10	6.7	10.2	6.7	7.8	8.2	6.5	7.4
Rural counties	58	9.8	7.7	7.4	7.5	6.8	7.7	7.0
All counties	75	9.1	7.8	7.0	7.4	6.8	7.3	6.8
Group II								
Population 100,000 or more	1	0.1	5.0	3.7	2.5	4.3	1.9	2.9
Population between 25,000 and 100,000	21	7.1	6.2	6.8	6.2	6.1	7.0	6.4
Population between 15,000 and 25,000	20	9.4	8.5	7.7	8.0	7.8	7.6	7.6
Population less than 15,000	33	10.4	8.6	6.7	7.9	6.6	7.6	6.8
All counties	75	9.1	7.8	7.0	7.4	6.8	7.3	6.8

[1] The State totals resulting from these county estimates were adjusted to agree with State population estimate totals prepared by the U.S. Bureau of the Census.

Source: Industrial Research and Extension Center, University of Arkansas.

TABLE 2-A

AVERAGE PERCENTAGE ERRORS FROM 1970 CENSUS
OF 1970 COUNTY POPULATION ESTIMATES, ARKANSAS, BY COUNTY SIZE CLASS

(Not Adjusted[1])

County Size Class	Number of Counties	Component Method II (X_1)	Composite Method (X_2)	Ratio Correlation (X_3)	Averages (X_1,X_2)	(X_2,X_3)	(X_1,X_3)	(X_1,X_2,X_3)
Group I								
Metropolitan counties, SMSA's	7	5.5	4.1	1.9	3.8	2.7	3.4	3.1
Other urban counties (50 percent or more of population urban)	10	4.3	8.8	5.2	5.7	6.6	4.2	5.3
Rural counties	58	7.0	7.1	6.0	5.7	5.8	5.3	5.2
All counties	75	6.5	7.1	5.5	5.5	5.6	5.0	5.0
Group II								
Population 100,000 or more	1	4.4	3.1	0.4	0.7	1.7	2.0	0.3
Population between 25,000 and 100,000	21	4.4	5.0	4.7	4.1	4.4	4.2	4.1
Population between 15,000 and 25,000	20	6.3	7.5	5.5	5.9	6.4	5.2	5.6
Population less than 15,000	33	8.0	8.2	6.2	6.3	6.0	5.4	5.4
All counties	75	6.5	7.1	5.5	5.5	5.6	5.0	5.0

[1] The State totals resulting from these county estimates were not adjusted to agree with State population estimate totals prepared by the U.S. Bureau of the Census.

Source: Industrial Research and Extension Center, University of Arkansas.

TABLE 3

AVERAGE PERCENTAGE ERRORS FROM 1970 CENSUS OF
OF 1970 COUNTY POPULATION ESTIMATES, ARKANSAS, BY RATE OF GROWTH

(Adjusted)

County Growth Rate Class, 1960 to 1970	Number of Counties	Component Method II (X_1)	Composite Method (X_2)	Ratio Correlation (X_3)	Averages (X_1,X_2)	(X_2,X_3)	(X_1,X_3)	(X_1,X_2,X_3)
Fast growing counties (more than 7.7 percent)	33	9.1	6.9	6.0	6.4	5.5	7.0	5.8
Slow growing counties (less than 7.7 percent)	13	8.3	7.7	5.7	7.2	5.8	6.3	6.2
Counties losing population	29	9.4	9.0	8.6	8.5	8.7	8.2	8.3
All counties	75	9.1	7.8	7.0	7.4	6.8	7.3	6.8

TABLE 3-A

(Not Adjusted[1])

Fast growing counties (more than 7.7 percent)	33	7.5	6.4	5.1	5.3	4.7	5.3	4.7
Slow growing counties (less than 7.7 percent)	13	5.2	6.7	4.1	4.9	4.7	3.5	4.2
Counties losing population	29	5.9	8.0	6.6	6.1	7.0	5.3	5.8
All counties	75	6.5	7.1	5.5	5.5	5.6	5.0	5.0

[1] The State totals resulting from these estimates were not adjusted to agree with State population estimate totals prepared by the U.S. Bureau of the Census.

Source: Industrial Research and Extension Center, University of Arkansas.

POPULATION ESTIMATING AND THE CENSUS OF 1970, THE CALIFORNIA EXPERIENCE

Walter P. Hollmann, State of California, Department of Finance

I. CITIES

After the years of working in darkness the California population estimators were at last able to confront their work systematically with the census. Perhaps it was not total darkness, since The State of California has a limited census program of its own. A series of test estimates were prepared for practically all of the cities participating in a program of local estimates designed for the redistribution of tax monies. The test estimates were compared with the results of the census in terms of a number of common elements:- total population, household population, housing units of several types, households and persons per household.

For the purposes of this paper, and by way of illustration, estimates for the cities within Orange and Santa Clara counties and the City of San Diego were scrutinized. One of the selected counties lies in the north, about forty miles south of San Francisco, the other in the south, adjacent to Los Angeles. Both have experienced unusual growth and they are served by different major power companies. One or more cities in each of the counties operates a municipal electric utility, while electric service is provided to the City of San Diego by a third major stockholder-owned utility.

The estimates were all prepared by a housing unit method, for two rather critical reasons, which, on reflection, are the same...no data options. Since the purpose of the program is the distribution of tax money, a minimum time must elapse between the reference date and the date at which the new estimate can be used for entitlement calculations. Utility records and construction statistics are available on a nearly current basis. Other types of estimates, which I will discuss later, require data that are far too slow in coming and some are all but impossible to obtain for incorporated cities. Secondly, the data usually available for cities do not permit estimates by other methods.

The housing unit method, for those unfamiliar with this type of effort, is very simple. Total housing units are estimated by adding to those recorded in the latest census new construction and annexations and subtracting demolitions and the rare disincorporations. Households or occupied housing units are estimated in a similar fashion by using residential electric customer increase since the census year. Group quarters population, handled separately, is based upon local data while average household size is artfully increased or decreased from the benchmark on the basis of observed trend, or type of construction or occasionally upon school enrollment statistics, if the latter are collected for the incorporated area. House trailer or mobilhome population is handled separately from other household population because of the popularity of electric master metering in mobilhome parks. For those who prefer formulae:

$$HU_t = HU_o + HUC_{o-t} - HUD_{o-t} + HUG_{o-t} - HUL_{o-t},$$

where HU_t is the estimate of total Housing Units (less mobilhomes) at the time of the estimate, and

HU_o = Units at the benchmark date
HUC_{o-t} = Housing Units constructed between o and t
HUD_{o-t} = Housing Units demolished between o and t
HUG_{o-t} = Housing Units annexed or gained between o and t
HUL_{o-t} = Housing Units de-annexed or lost between o and t

Construction includes units moved in; demolition includes those moved out.

$$HH_t = HH_o + \Delta REC_{o-t} + \Delta MMHH_{o-t}$$

where HH_t and HH_o represent households (less those in mobilhomes) at the time of the estimate and the benchmark respectively

ΔREC_{o-t} = change in Residential Electric Customers, suitably corrected for lags in recording annexations
$\Delta MMHH_{o-t}$ = change in master metered housing units (other than mobilhomes) to which an estimated vacancy has been applied.

Finally, $P_t = HH_t \times PP_t + TrP_t + GQP_t$

or, the total estimated population equals the product of the estimated households and the average household size plus the populations in mobilhomes and in group quarters.

A comparison of the total estimated populations with the census leads to the inescapable conclusion that the method contains a sharp upward bias. Table 1, following shows the percentage difference from the corresponding statistic in the Census of 1970 of each of five elements related to housing unit estimates from their estimated values. In Orange County the average error of the population estimates was 2.33 percent high; eleven cities were estimated to have 17,900 more inhabitants than the census reported while another 11 were estimated to have 8,200 persons fewer than were reported. Among sixteen cities the estimates fell short by 6100 households while among the remaining six, 1900 too many were estimated. The method underestimated households but its failure (or rather the estimators' failure) to perceive the full effect of the drop in household size resulted in high estimates of population in spite of low numbers of households. Only five of the 22 were estimated low, the remaining 17 were high. In Santa Clara County the total population was high by an aggregate of 33,500 in eight cities, low by 6900 in seven cities. Households were estimated high by 4200 in five cities, and low by 4400 in ten; characteristically the average household size was underestimated in only two of the cities.

It is clear that at least within our hands, the method carries a serious upward bias and much more study is required not only of the quality of the statistics used but also of the determinants of changing household size. Were the Bureau to provide persons per household by units in structure one might be able to estimate household size with more precision by use of the change in housing composition. Which gives rise to another concern with the Census. The column headed "Singles" in Table 1 presents the percent error in our estimates of single housing units in the housing inventories of the cities in Orange and Santa Clara counties. The results are dismal; 20 of the 25 cities on the list were underestimated, and seriously. These were cities within the mail-out area and a study of the wording of the question on units-in-structure suggests overreporting of singles. If this is so, substantial doubt is cast upon the units-in-structure data.

In a recent set of provisional estimates for California counties, the housing unit method was used but with a more vigorous attempt to establish an average household size on the basis of past trends in the variable itself and the partial indicator, school-enrolled children per household. Although we can not know if we were more accurate, the fact that the results of the population calculation by this method for the 58 counties of California seemed to agree more closely with other methods than in the past was encouraging.

II. COUNTIES

Each summer, estimates of county populations in California are prepared for the current July 1 on a provisional basis and are revised for the preceding year. The methods used recently were based upon experimental work performed during the decade 1960-70; four were used for provisional estimates and six for revisions.

Two of the methods, designated Department of Finance Regression I and II are component techniques. The equations which appear in the handout are

$$Z = 1.103 + .734X + .374Y$$

$$Z = .172 + .215X + .921Y$$

and were derived from 1950-60 experience. The independent variables are, X equal to the percent change in occupied housing units and percent change in residential electric customers, respectively, in the two equations and Y equal to the percent change in school enrollment in grades 3 through 8 over the enrollment in grades 2 through 7 in the prior year in both equations, for each of the ten years, multiplied geometrically. The dependent variable Z represents net civilian migration as a percent of civilian population at the beginning of the year. It is apparent that the second equation weights school enrollment change more heavily than the first. In the postcensal estimating period, both equations perforce use the change in residential electric customers since estimates of household change are based upon that available statistic.

The housing unit method, described earlier, is the third method, while Census Bureau Component Method II is the fourth.[1] Two methods which present problems in timely data acquisition are the Composite Method of Bogue and Duncan[2] and the Ratio Correlation Method described by Crosetti and Schmitt.[3] The former method generates broad age groups of the population to be estimated from symptomatic indicators for which rates can be assumed. The latter is used to divide the total population of the State among the counties on the basis of the relationship of shares of seven independent variables to population shares--births, deaths, elementary school enrollment, fee-paid auto registrations, income tax returns, covered employment and taxable sales.

For the tests, the first four methods were used to calculate the populations of the state and its 58 counties as of April 1, 1970, and the averages of the four as well as the averages of the four adjusted for the assumed effects of the two additional methods were calculated. The standards of comparison were the state and county final total census populations less estimated military, a civilian figure subject to later change when sample data are ready for use. Two methods for which current data could not be obtained--Ratio Correlation and Composite--were calculated for July 1, 1969 and compared with the latest intercensal estimates for that date, estimates which were based upon as much of the detail of the Census of 1970 as was available. Thus, there are eight comparisons possible--each of four methods, their arithmetic mean and an adjusted mean, six with the 1970 census, and two additional methods with a 1969 standard based upon the 1970 census.

Statewide, the Housing Unit Method estimated 4.8 percent high while Census Bureau Method II estimated 3.0 percent low. The average of the four methods was .16 percent high when adjusted for the presumed effects of the two missing methods and only .09 when left unadjusted. Apparently the four methods for which reasonably current data can be used, when averaged, yield the best estimate of the State's population, but this is not necessarily the most satisfactory technique if minimum error in county populations is the criterion.

California has 58 counties of which 24 are metropolitan, i.e. they are Standard Metropolitan Statistical Areas or parts thereof, and 34 are nonmetropolitan. Of the 24, those of 500,000 inhabitants or more are 10 in number; of the 34 those with 20,000 or more inhabitants number 18. On this basis, four strata were identified, large and small metropolitan counties and large and small nonmetropolitan.

The fourth stratum contained one county (Alpine) of 484 and one of 2365 inhabitants (Sierra); mean errors were calculated with and without these two. Table 2 displays the mean errors from the different methods and combinations of methods for the strata. Except in the state totals, absolute values of errors were averaged, hence no effect of upward or downward bias is presented. No Ratio Correlation value is shown for the State since use of the method is restricted to allocation by county of a predetermined whole. The best performance in estimating the large metropolitan counties shown in the table is that of Ratio Correlation, with a mean error of one percent. The best method for all counties was the Adjusted Average with a mean error of five percent and a very creditable two percent for the largest counties; it estimated Los Angeles County within .06 percent, but this is not shown.

Table 2 compares the success of the methods in another way in comparing for each method, separately for the metropolitan and nonmetropolitan, the number of counties estimated high with the number estimated low. In the same table a comparison of the methods with respect to the magnitude of errors of estimating is shown. Four ranges were selected which might be subjectively characterized as "excellent", "good", "fair to poor" and "unacceptable", or in percent error ranges, 0 to 1.99, 2.00 to 4.99, 5.00 to 9.99 and 10 percent or greater.

It is apparent that for California counties the composite method seems to estimate low as often as high for metropolitan and nonmetropolitan counties, taken together, although it tends to be slightly high in nonmetropolitan counties. The results show that it was "excellent" in half the metropolitan counties but less than one-sixth of the nonmetropolitan; in fact this otherwise successful method was 10 or more percent off the mark in nearly a third of the nonmetropolitan counties. Ratio Correlation, with a slight tendency to estimate high seems to have yielded the best results but the fact that it is controlled to an accepted total renders this less than miraculous. Unfortunately, neither of these methods are satisfactorily timely, i.e., they can not be used for developing current-year estimates.

Turning to the four methods that can be used for current-year estimates, the tendency of the Housing Unit Method to estimate high is obvious in both metropolitan and nonmetropolitan counties. Its accuracy, when measured against the census is only fair for metropolitan counties, (14 of 24 metropolitan counties within five percent) and poor for nonmetropolitan (seven of 34). Census Bureau Method II, one of the poorer methods for estimating the State, was about as accurate (13 of 24 metropolitan counties within five percent, nine of 34 nonmetropolitan).

Although the first of the two Department of Finance regression methods performed rather well with 19 of 24 metropolitan counties and 15 of 34 nonmetropolitan counties (within five percent), the adjusted average was slightly better. The case might be made, indeed has been made, that averages of methods are preferable to single methods in this type of estimating.

What remains to be done in California is a standardized test of Method II, the Composite and Ratio Correlation to the specifications of the Bureau to enable those in the Cooperative Program[5] to evaluate methods for counties across the nation.

Inquiries were made of several other western states in order to compare our experience with that of others. Although it was determined that no two states were engaged in the same estimating activities with respect to their counties, the efforts of the several states were not without common elements and concerns. Utah, for example, reported that its estimates of major counties were satisfactory but trouble was encountered with the smaller ones. Its errors were between five thousandths of a percent to 20 percent with a median of 3.7 percent and their methods a modified Method II and another component method as well as membership statistics from the L. D. S. Church.

The State of Washington is in the process of preparing tests of methods. Census Bureau Method II underestimated the State by eight percent and yielded the poorest distribution by county, especially in the larger ones. Ratio Correlation gives the best basis for distributing the States' population. Of the 39 counties of the State, it came closest in 20, the Composite Method came closest in 14 and Method II in 10.

Hawaii, with four counties, used a still different approach. The State total was estimated by the Bureau, using a weighted average; 25 percent Ratio Correlation, 25 percent Method II and 50 percent a special Hawaii component method using arrival and departure data for net migration. As the decade progressed, the two traditional methods diverged increasingly from the special component method and in 1970 were substantially closer to the Census. It was concluded that departures were less scrupulously recorded than arrivals by the steamship companies and airlines; the result was an overestimate of the population by the Hawaii method. Table 4 shows the extent of the error of estimation and the percent distribution among Hawaii's four counties.

Hawaii analysts feel that a part of the error is the result of a census undercount and they are encouraged in this belief by surveys of selected areas and covered worker statistics.

In closing my remarks to this panel on "Why Did Intercensal Estimates Go Wrong in the 1960s?" I should allude to some of the problems we have had in our local estimates work with the Census benchmark. In the material from the first count which was hungrily consumed by local planners, a substantial number of misallocations were detected, errors attributable to the address coding guide or to other sources of inaccurate coding. Such errors, which may be negligible when dealing with a unit as large as a county, loom very large when attempting to evaluate an estimating method for a city. We are indeed fortunate in having the summary tape information for without it we would have been unable to understand, and sometimes to detect census errors. With it we can occasionally reconstruct what probably happened. Is it not fair to ask that in local estimates, at least, on the basis of an active program of data collection and estimates of population and housing, "Where did the Census go wrong?". It is also fair to ask whether small area data from the second and subsequent counts can be very useful to local people unless a substantial effort toward their improvement is undertaken.

1
Bureau of the Census, Current Population Reports, Series P-25 No. 339, 1966.

2
Bogue, Donald J. and Duncan, Beverly "A Composite Method for Estimating Postcensal Population of Small Areas by Age, Sex and Color". U. S. Department of Health, Education and Welfare, Selected Studies, Vol. 47 No. 6, 1969.

3
Crosetti, Albert H. and Schmitt, Robert C., "A Method of Estimating the Intercensal Population of Counties", Journal of the American Statistical Association, December 1956, pp 587-590.

4
Table 4 furnished by Robert C. Schmitt, Department of Planning and Economic Development, State of Hawaii.

5
Meyer Zitter, "Federal-State Cooperative Program for Local Population Estimates," The Federal Registrar and Statistician, U. S. Department of Health, Education and Welfare, January 1968.

Table 1. COMPARISON OF ELEMENTS OF HOUSING UNIT METHOD, CENSUS AND TEST ESTIMATES, SELECTED CALIFORNIA CITIES, PERCENT DEVIATION

ORANGE COUNTY	HOUSEHOLDS	HOUSEHOLD POP.	HOUSING UNITS	SINGLES	PERSONS PER.
Anaheim	2.04	4.79	- .36	- 4.56	22.33
Brea	-1.49	-6.35	5.60	3.42	- 5.07
Buena Park	- .36	2.82	.01	-12.47	3.29
Costa Mesa	-1.20	3.39	1.38	--	4.66
Cypress	-2.05	1.79	- 4.54	--	4.02
Fountain Valley	-3.20	- .91	- 8.00	--	2.37
Fullerton	-3.97	-2.39	- 1.34	- 3.29	.16
Garden Grove	-1.28	.78	- 1.23	- 2.99	2.09
Laguna Beach	-6.46	-4.62	- 1.32	1.17	2.00
La Habra	-6.00	3.01	- 3.54	--	9.57
Los Alamitos	-8.31	-6.77	- 6.52	--	1.67
Newport Beach	-2.77	-2.34	4.76	- 3.30	.47
Orange	-2.67	-2.03	- 2.45	- 6.90	.64
Placentia	1.67	.86	4.39	- .44	- .80
San Clemente	-1.35	.44	- 7.29	--	1.90
San Juan Capistrano	-5.11	-6.37	4.66	--	- 1.36
Santa Ana	-1.93	- .23	2.01	- 3.70	1.73
Seal Beach	.27	-1.64	.09	--	- 1.90
Stanton	6.19	5.72	5.75	- 7.53	- .45
Tustin	3.39	8.82	4.17	3.61	6.09
Westminster	- .51	.06	- .80	- 3.32	.56
Yorba Linda	2.28	2.51	.26	--	.27

SANTA CLARA COUNTY	HOUSEHOLDS	HOUSEHOLD POP.	HOUSING UNITS	SINGLES	PERSONS PER.
Campbell	3.04	- 1.27	4.9	-16.8	- 4.13
Cupertino	1.73	8.41	- 2.8	-10.1	6.32
Gilroy	- 9.44	- 8.44	- 4.8	- 5.6	1.16
Los Altos	- .25	4.16	1.4	2.6	4.36
Los Altos Hills	- 3.92	.46	- 3.89	--	4.65
Los Gatos	- 9.64	- 8.31	- 9.3	-15.8	1.69
Milpitas	- .91	3.18	- 0.3	- .5	4.24
Monte Sereno	-15.13	-20.05	12.2	--	5.85
Morgan Hill	-14.34	-11.69	-16.60	--	2.87
Mountain View	3.85	9.07	10.60	- 8.40	5.16
Palo Alto	- 5.73	1.36	- 4.79	- 3.43	7.87
San Jose	1.88	4.46	3.57	- 1.17	2.39
Santa Clara	- 4.77	- 2.27	- 5.56	- 8.48	2.52
Saratoga	- 3.96	- 2.07	.18	- 1.56	1.86
Sunnyvale	2.33	6.08	7.83	1.56	3.50
City of San Diego	2.26	2.91	2.54	--	.89

Table 2. COMPARISON OF ERRORS[1] OF SIX ESTIMATING METHODS AND AVERAGES OF METHODS, COUNTY GROUPINGS, CALIFORNIA

METHOD	1	2	3	4	5	6	7	8
	State as a whole	58 counties	10 Metro counties 500,000 and Over	14 Metro counties of less than 500,000	18 Non-Metro counties 20,000 and over	16 Non-Metro counties under 20,000	15 small counties less Alpine	14 small counties less Alpine, Sierra
Four-Method Average, 1970, Adjusted to Six	+0.161	4.977	2.021	3.707	5.277	7.599	6.439	5.974
Four Method Average, 1970, Unadjusted	-0.089	5.761	2.804	3.915	6.258	8.667	7.867	8.042
HHR D of F Regression I, 1970[2]	-0.217	5.043	2.689	3.203	5.114	8.046	6.489	6.877
REC D of F Regression II, 1970[3]	-2.031	5.667	3.973	5.229	5.404	7.406	7.651	8.150
Housing Unit, 1970	+4.844	8.215	4.702	4.649	8.347	13.382	11.188	11.184
Census Bureau Method II, 1970	-2.951	7.079	5.882	6.090	8.183	7.450	7.395	7.154
Composite, 1969	1.18	6.96	2.60	3.55	6.28	13.42	10.85	7.38
Ratio Correlation, 1969		4.87	.99	3.54	4.59	8.78	7.27	7.72

1. Excepting for State as a whole, errors are averaged without regard for sign.
2. Z = -1.103 + .734X + .374Y; X = Household ratio, HHR; Y = School enrollment ratio Grades 3-8/Gr.2-7
3. Z = .172 + .215X + .921Y; X = Residential electric customer ratio, REC; Y = School enrollment ratio

Population Research Unit
1623 10th Street
Sacramento, California 95814

Table 3. DIRECTIONS AND MAGNITUDES OF ERRORS OF SIX ESTIMATING METHODS AND AVERAGES OF METHODS, METROPOLITAN AND NONMETROPOLITAN COUNTIES, CALIFORNIA

Frequency distributions of 24 Metropolitan and 34 Nonmetropolitan counties.

	DIRECTIONS					MAGNITUDE							
	METRO		NONMETRO			METROPOLITAN				NONMETROPOLITAN			
	High	Low	High	Low		0-1.9	2-4.9	5-9.9	10%& over	0-1.9	2-4.9	5-9.9	10%& over
Four-Method Average, 1970, Adjusted to Six	14	10	30	4		11	9	3	1	6	9	13	6
Four-Method Average, 1970, Unadjusted	14	10	31	3		8	10	5	1	4	6	16	8
HHR D of F Regression I, 1970	15	9	26	8		11	8	4	1	7	8	13	6
REC D of F Regression II, 1970	9	15	28	6		7	9	5	3	6	14	5	9
Housing Unit, 1970	23	1	33	1		7	7	9	1	1	6	9	18
Census Bureau Method II, 1970	10	14	26	8		3	10	8	3	8	1	16	9
Composite, 1969	11	13	19	15		12	8	3	1	5	8	11	10
Ratio Correlation, 1969	15	9	17	17		15	7	1	1	8	7	13	6

TABLE 4

STATE OF HAWAII

County	1969 estimate		1970 census	
	Number	Percent	Number	Percent
State total.......	793,747	100.0	769,913	100.0
Hawaii.................	67,229	8.5	63,468	8.2
Honolulu...............	645,319	81.3	630,528	81.9
Kauai..................	31,666	4.0	29,761	3.9
Maui...................	49,533	6.2	46,156	6.0

APPENDIX

HOUSING UNIT METHOD CALCULATIONS

$$HU_t = HU_o + HUC_{o-t} - HUD_{o-t} + HUG_{o-t} - HUL_{o-t}$$

$$HH_t = HH_o + \Delta REC_{o-t} + \Delta MMHH_{o-t}$$

$$P_t = HH_t \times PP_t + TrP_t + GQP_t$$

DEPARTMENT OF FINANCE REGRESSION EQUATIONS

I. $Z = 1.103 + .734X + .374Y$

II. $Z = .172 + .215X + .921Y$

DISCUSSION

Peter A. Morrison, The Rand Corporation

Population estimates usually are prepared with practical rather than experimental objectives foremost in mind. The contexts of many comparative studies so far have arisen fortuitously, leaving little room for experimental design in advance. Even when such forethought has been possible, adherence to a rigid scientific design usually proves impractical. The following points illustrate complications that typically arise:

- The quality of the data used for estimation generally varies from one population to another, with uneven effect on individual methods. The accuracy of Component Method II, for example, is vulnerable to poor school-enrollment data since these form the basis for its migration estimate. The quality of data and the precision of a method, therefore, cannot be separately distinguished.

- In small-area applications, modifications often are necessary to adapt an estimating method to the local data environment. Instances where one method has been applied to every population in a computationally consistent manner are the exception rather than the rule. These variations reduce strict comparability to an unknown degree.

- Comparisons among methods that have been applied to separate universes are especially hazardous, since the accuracy of an estimate varies systematically with a population's absolute size and its relative rate of growth. One method may appear more precise than another simply because the former was applied to a disproportionate number of heavily populated or slowly growing areas, both of which lend themselves to more precise estimation.

- Statistical measures used to gauge the relative precision of estimating techniques are inadequate. Effective comparisons are difficult, and statistical appraisals of differences are rarely conducted. The conventional measure adopted in most studies is the mean of percentage deviations, neglecting signs, between estimated and enumerated populations (symbolized hereafter as $\bar{D}$).[1] This measure indicates relative error independent of an area's absolute population size, thereby weighting large and small study populations equally. As a result, a few numerically small populations--for which relative error can be large--may swamp the measure, overstating the actual degree of imprecision.

- The short estimating periods used in some comparative studies occasionally favor simple extrapolative procedures over more analytical techniques. These results must be viewed with caution. A minimum imprecision is inherent in several of the latter methods; and although simple extrapolation will sometimes outperform them in the short run, it should not be concluded that this advantage will hold for longer intervals.

All of these complications discount the value of most comparative evidence now available. Exhaustive performance tests based on 1970 census data are planned by the Bureau of the Census. For now, the current evidence lends itself to only a few general conclusions.

CONCLUSIONS SUPPORTED BY COMPARATIVE EVIDENCE

1. *No single method of estimating local population shows consistently greater accuracy, although the Regression Method continues to look most promising.*
2. *Evidence consistently shows that lower average error can be attained by averaging together estimates made by different methods.*
3. *Average error tends to be lower for counties whose populations are large or metropolitan.*
4. *Average error varies with rates of population growth.* $\bar{D}$ is lowest among slowly growing counties, followed by rapidly growing counties, followed by counties losing population.

---

1 $$\bar{D} = \frac{\sum_{i=1}^{N} \left| \frac{\text{Estimated} - \text{Enumerated}}{\text{Enumerated}} \times 100 \right|}{N},$$

where N is the number of populations for which estimates are prepared.

Reference:

Morrison, Peter A., *Demographic Information for Cities: A Manual for Estimating and Projecting Local Population Characteristics*, R-618-HUD, The Rand Corporation (June 1971).

FERTILITY AND MORTALITY ESTIMATION FROM SINGLE-ROUND SURVEY DATA

Hilary J. Page, Princeton University

A majority of the world's population lives in areas in which universal birth and death registration systems are either absent or else so incomplete as to yield records that are practically useless for fertility and mortality estimation. Modernizing countries will doubtless need in the long run to develop efficient civil registration systems; but for most developing countries the attainment of this ideal is many years, if not decades, away. In the meantime, attempts are being made to develop specifically demographic data-collection and analysis procedures that will produce more reliable estimates of vital rates than are permitted by existing civil registration systems. Strategies to improve direct recording of vital events include intensive registration schemes in sample areas; multi-round surveys, or repeated recording of the individual elements in sample populations; or some combination of these as in dual-records systems of the PGE type, in which events recorded by one observation system are matched against those recorded by an independent system and an estimate of the total events (including those missed by both systems) is obtained from the results of the matching process. This type of strategy is the focus of Seltzer's paper presented later in this session, evaluating PGE studies [15].

An alternative tactic in the absence of adequate registration records is to avail oneself of the indirect evidence on fertility and mortality that can be gleaned from single-round censuses or surveys. The principle forms of evidence are the reported age-distributions and retrospective reports of reproductive and bereavement experiences. Such data are typically fragmentary and defective, besides being indirect, and they certainly cannot be taken at face value. However, in the last two decades a large body of estimation techniques has been developed to extract the most reliable and informative components of these data and to derive estimates of vital rates from them, by exploiting the relationships that must hold between various demographic variables.

This paper reviews several of the indirect estimation techniques that are most widely applicable in the developing world at the present time - that is, techniques appropriate when the only data available are those from one or very few single-round censuses or surveys and are more or less defective in quality. The papers by Cho [3] and Zachariah [21] that follow discuss special procedures that are applicable under less inauspicious circumstances: situations in which the data either exhibit exceptionally high quality with respect to age-reporting (in this case, Korea) or form a series extending over a comparatively long time-period (several Arab countries).

It is obviously impossible to present a fully comprehensive review of the state of the art in a short paper. The literature on indirect estimation of vital rates is substantial; for a detailed description of the basic principles and procedures the reader can be referred to a U.N. manual devoted to the subject [18]. The present paper is restricted to little more than a sketch of some of the possibilities that are based on analyses of aggregate age-distribution data, and concentrates on use of the more detailed information gathered from retrospective fertility and mortality reports. My practical assessment of these techniques derives largely from experience with their application to materials from Africa south of the Sahara: major analyses of these data have been carried out at the Institut National des Statistiques et des Etudes Economiques in Paris [11] and at Princeton [2]. Many of the problems that arise are not peculiar to these particular data sets but are commonly found, with greater or less intensity, in other developing areas.

Estimation from age-distributions

Given age-distributions from two censuses or surveys (and the assumption that the population is closed to migration) one can readily estimate apparent intercensal survival rates for each age-group and hence build up a life table for the intercensal period. provided there are no inequalities of coverage between the two censuses to bias the estimates. systematic patterns of age-misreporting often give rise to wild fluctuations between the estimates for adjacent age-groups (or even reported survival rates exceeding unity) making the results unusable as they stand. Mortality estimation under these conditions is greatly facilitated by our knowledge of the strong tendency for death rates at different ages in any population to be intercorrelated. Thus if we can determine the mortality level for one age-group it is possible to estimate upper and lower limits for mortality rates at other ages. Exact determination of mortality rates at other ages is not feasible because the inter-correlations are not perfect; in fact, several distinct age-structures of mortality have been detected, and from these various sets of model life tables have been prepared incorporating each age-structure at different levels of mortality. Three sets of model tables have been developed to date. The first set was prepared by the United Nations in 1955 [17], and has been widely used; this set has now been largely replaced by subsequent sets because it does not embody variations in the age-structure of mortality at the same overall level as do the later tables. The later models consist of the four 'families' (age-structures of mortality) prepared by Coale and Demeny at Princeton [7], and the various combinations of mortality patterns incorporated in the tables prepared by Lederman at the Institut National des Etudes Démographiques in

Paris. The most striking difference between the various observed age-structures of mortality lies in the relationship between infant and early childhood mortality (under age 5) on the one hand, and adult mortality on the other. These differences imply that it is useful to have separate estimates of adult and of early childhood mortality.

Adult mortality rates can be estimated fairly reliably by comparison of the reported intercensal survival rates and those embodied in model life tables. In practice it is necessary to use cumulated sections of the age-distribution rather than individual cohorts, in order to minimise the effects of age-misreporting. The technique of forward projection can be used to project successive cumulated segments of the age-distribution recorded in the first census (for example, the population aged over 10 years, over 15 years, etc.) according to the survival rates of the most likely model life tables; the resulting projected populations can be compared with the population actually recorded in the later census. The model life table producing the best fit can then be taken to represent mortality rates over all age-groups - except the very youngest age-group, for which no survival ratio is reported. If we are prepared to assume that childhood and adult mortality bear the same relation to each other in the population as they do in the model life table selected, then the overall average intercensal death rate can be estimated by applying the model's age-specific mortality rates to the mid-period population. Since the intercensal growth rate is also known, the overall birth rate can be obtained by subtraction. However, the assumption about childhood mortality is questionable: it would be preferable to make separate estimates of this using other forms of data and other estimation techniques.

Recent fertility levels can also be estimated from one age-distribution by reverse projection of the population under age 5, assuming that this number is accurately reported (which is unlikely) and that childhood mortality conditions are known from some other source. Alternatively, since the age-distribution of a closed population depends only on its recent history of fertility and mortality, if a stable situation can be assumed then one has the range of stable population theory at one's disposal for estimation [18,19]. If either the rate of population increase is known or an index of mortality is available, then fertility levels can be inferred from an age-distribution. Tabulation of stable age-distributions by Coale and Demeny [7] has greatly simplified this sort of estimation; also tabulation of approximate adjustment factors for quasi-stable situations in which mortality has been declining [18]. However, the assumption of stability (or some conjecture as to the nature and extent of departure from stability) limits successful application of the method. Furthermore, choice of a particular age-structure of mortality may introduce errors. Fertility estimates from model stable populations are not severely affected by this. Mortality estimates are very sensitive to it, and consequently cannot be accorded much credibility unless there is independent evidence as to the appropriate age-structure to use. Finally, there is the problem of age-misreporting: birth rate estimates, for example, vary widely depending upon which index of the age-distribution is taken as the most reliable; detailed analysis of the age-distribution data is really required in each individual case to select this index.

Retrospective fertility data

Given highly accurate information from women regarding the birth dates of both themselves and each of their children, one can determine the age-specific fertility rates that have been experienced by each cohort. If we then assumes that the fertility experience of those who have died was not exceptional, then we can reconstruct in detail the population's fertility experience over roughly half a century. For most developing countries we must rule out the possibility of obtaining large numbers of individual reproductive histories of sufficiently high quality. In societies where respondents usually know neither their own nor their children's chronological ages, the intensive interviewing and highly trained enumerators required to produce full and accurate records are beyond the scope of most large-scale demographic enquiries. Bogue [1] has developed techniques of recording and analysing full pregnancy histories, but these would generally appear to be limited in application to small and very intensive studies or to the younger and better educated segments of the population.

The data that can be, and have been, more widely collected are much less ambitious. They consist of mean parity (average number of children ever born per woman up to the date of the survey) and current fertility (average number of children born per owoman during the year preceding the survey) - both sets of data being tabulated typically for 5-year age-groups. Both types of data are subject to severe weaknesses and cannot be used directly.

Current fertility data would permit straightforward calculation of the prevailing age-specific fertility schedule if they were accurate; but accuracy is unlikely on several grounds. Patterns of age-misreporting systematically related to physical maturity and preconceptions about age at puberty, to marital status, or to apparent fertility are all likely, since age is frequently estimated on the basis of these more readily observable characteristics. As one official report expressed it (Central African Republic), a married woman's age is too often estimated by rule of thumb as being '14 years plus twice the number of children she has borne' [6, p. 13]. Such biases may account for the improbably high mean age of the fertility schedule reported in the 1960 Ghana Census ([10, pp. 7-9] and also noted in the 1962 Kenya Census [2, p. 173], for example. Even estimates of the overall population birth rate from the current fertility data are highly dubious because of misreporting of the occurence or non-occurence of a birth in the preceding twelve months. There seems to be a tendency in many populations for either over- or under-reporting of these events - the reason, it is usually suggested, being a misperception of the specified reference period. In addition, there is the problem of straightforward omission of some births: infants that died very shortly after birth are probably particularly likely to be omitted, especially if the birth occurred a long time before the survey. Analysis of the births reported for a one year reference period in Upper Volta provide an illustration of typical data defects. When tabulated by month of reported occurence, the monthly total fell off quite regularly and rapidly with increasing remoteness from the survey date [19, p.246]. This feature might indicate either a propensity to misperceive time periods, or a tendency to omit births with extension of the recall period. Whatever the cause, the observation enhances one's levels of skepticism with respect to the raw data.

Accurate age-specific parity data could be used directly to provide a summary measure of the average fertility to date of each cohort - a measure of completed fertility for those cohorts that have passed the childbearing age. Strictly, it would be a measure for the surviving portion of the cohort, but the effects of mortality differentials between relatively fertile and relatively infertile women are presumed to be negligible. The age-reporting problem is at least as severe here as with current fertility data, if not more so. The parity data are likewise subject to omission; this usually appears to be of considerable magnitude in older women's reports, which consequently have little credibility.

The simple fact that the average parity of any cohort of women corresponds to the cumulation of the age-specific fertility rates they have experienced can be exploited in a number of ways; though in practice, difficulties often arise from the curvilinearity of typical age-specific fertility schedules combined with the fact that data are frequently available only for 5-year age-groups, which necessitates some form of curve fitting. Straightforward differencing of the parity data to obtain estimates of age-specific fertility rates is one possibility if data are available by single years of age. This requires, though, that reporting of both age and parity be very accurate; moreover, unless fertility has been extremely stable in the past, the results confound the effect of the rise in parity for each cohort as it passes through its reproductive span with the effect of differences in fertility experience between successive cohorts. The conditions required for this easy procedure to work are rarely encountered.

Where moderately complete registration records exist and where age is reported quite accurately, the more reliable portions of the parity data (those relating to younger women, who are less likely to omit a significant proportion of births) can be used to provide some check and ajustment of the registration data. Comparison of the younger women's reported parity with the average parity levels implied by the cumulated age-specific fertility rates recorded for the same cohorts in the registers provides an adjustment factor for these registration data. If we can assume that the level of registration completeness is uniform at all ages and for all cohorts, then this adjustment factor can be applied to all the registered births, permitting estimates of cohort fertility rates and hence of both current levels and of past trends (if the registration series covers any length of time). Registration data are rarely of sufficiently high quality to be amenable to this form of adjustment, however.

With the assumption that fertility rates have been constant in the recent past, the cumulation of current fertility data from a cross-sectional survey corresponds to the past experience of the various cohorts of women, summarised in their average parity. Under these conditions we can estimate prevailing age-specific fertility rates from the current fertility data in a manner analogous to that in which both prevailing and past rates can be obtained from adjusted registration data. In this case the current fertility data are assumed correct in age-structure though not necessarily in level; that is, it is assumed that any tendency to over- or under-report current births is the same for all age-groups. Brass has developed an ingenious procedure on the basis of this, by which the current fertility rates by 5-year age-groups can be converted into the average parity levels they imply for the same 5-year age-groups, taking into account the curvilinearity of fertility schedules. [18, chapters 2 and 7; 2, chapter 3]. The procedure utilises a model age-structure of fertility for populations not practising fertility control; it is allowed to vary only in its starting age. The location of the model that is appropriate for any particular population is determined from the ratio of the current fertility rates for the two youngest reproductive age-groups. A model that was allowed to vary in form as well as starting point might better describe the variety of fertility schedules. Coale [8] has recently shown that the age-pattern of entry into marriage follows a standard pattern that varies in both starting point and form; since the rising portion of the fertility schedule in populations not practising fertility control is determined largely by the age-pattern of entry into marriage, both parameters should probably be included in model fertility schedules. However, in practice this point is very minor: the improvement any such refinement might introduce is probably paltry in relation to the errors that arise from defects in the data.

The Brass method is very appealing and has been used extensively. But it is sensitive to fertility trends and to data defects - especially it is very sensitive to massive and systematic age-misstatements like those common to data from tropical Africa [2, chapter 3, appendix B], and its practical value is therefore limited.

Despite the availability of retrospective fertility data in many areas, we are often thrown back to inferring fertility from age-distribution data.

Retrospective mortality data

Current mortality data from single-round surveys - that is, deaths by age during the preceding twelve months - are of notoriously poor quality. Errors and omissions are rife and the data are rarely usuable.

Mortality at early ages can, however, be estimated from data on the number of children ever born and the number of these surviving, tabulated by age of mother. The basic method, again developed by Brass [18, chapters 2 and 7 ; 2, chapter 3], involves conversion of the proportions surviving into conventional life table functions. The essential idea is that, if the time distribution of births that have occurred to a group of women is known, then the proportion surviving for these women at any point in time can be viewed as a function of the mortality schedule to which the children have been subject. Roughly speaking, the proportion dead for women 20-24 years of age corresponds to the probability of dying before age 2; that for women aged 25-29 years to the probability of dying before age 3, and so forth. The correspondence is only approximate because of the effects of different age-structures of mortality and of different time distributions of births. The time distribution of births that have occurred to a cohort of women is given quite simply by the age-specific fertility schedule they have experienced; if we knew this then we could determine the age-distribution the children would have in the absence of mortality. If we also knew the probability of dying to which the children have been subject we could calculate the proportion of them that would have died before the survey; conversely, given the proportion dead, we can calculate the probability of dying before a certain age, for any particular age-structure of mortality. The method involves the input of two models, therefore; one for the assumed age-structure of fertility and one for that of early mortality.

Brass used the same model age-structure of fertility to determine the time-distribution of births as is used in his fertility estimation method; here the ratio of parity levels reported for the two youngest reproductive age-groups is used as an index of the appropriate starting age. Unfortunately, this index is rather sensitive to the forms of misreporting that are probably common. Sullivan [16] has developed a modification of the Brass basic procedure using an alternative way of determining the time distribution of births. This is based on empirical single-year age-specific schedules rather than on a standard schedule, and hence to some extent circumvents the problem that the standard does not allow for variations in form as well as starting age. Moreover, the modification uses the ratio of parity levels for the second and third reproductive age-groups, which is probably sless sensitive to data errors than the ratio for the first and second groups. The modification is based on regression analysis (for various levels and age-patterns of mortality) of the relationships between the proportion dead of children ever born for women in each age-group and this index of the fertility schedule. Sullivan found very close relationship between this index on the one hand and the conversion factor needed to transform the proportion dead into the corresponding life table function. He found this over the entire range of mortality levels within each of the Coale-Demeny families of model life tables, but slightly different relationships between these families. Regression coefficients for the North and West families are given in: -

North

$${}_2q_0/D_2 = 1.30 - .63(P_2/P_3)$$
$${}_3q_0/D_3 = 1.17 - .50(P_2/P_3)$$
$${}_5q_0/D_4 = 1.15 - .42(P_2/P_3)$$

West

$${}_2q_0/D_2 = 1.30 - .54(P_2/P_3)$$
$${}_3q_0/D_3 = 1.17 - .40(P_2/P_3)$$
$${}_5q_0/D_4 = 1.13 - .33(P_2/P_3)$$

where ${}_xq_0$ is the proportion dying before exact age x in the life table;
P_i is reported parity for women in the i'th 5 year age-group (i=1 for age-group 15-19);
D_i is the proportion of children dead as reported by women in the i'th age-group

Comparison of results of the two procedures for African data [13] shows that the estimated probabilities of dying before ages 2, 3, and 5 for any given population are similar, but that they are more consistent with the model life tables on which their derivation is based when calculated from Sullivan's modification than from the original procedure. Consistency is not necessarily a virtue: if mortality had been declining then the estimated probabilities of dying before older ages should be too high relative to the estimated probabilities of dying before the youngest ages, since the former refer to mortality conditions of a less recent period. Such a tendency is found very faintly in the Sullivan estimates for those populations where the estimated probabilities of dying before age 2 are particularly low: one would scarcely expect to find such a tendency where early mortality is still extremely high.

Several problems remain with this procedure. Firstly there are the problems of misreporting; both omissions and age-misstatement. Secondly, the assumption that the cross-sectional parity data truly reflect past experience of the younger cohorts of women - that is, that fertility has not changed during the last ten to fifteen years - may be unwarranted. Thirdly, the estimates refer to past mortality conditions rather than prevailing ones, and so are incapable of providing up-to-date information. The estimated probability of dying before age 2, for example, is based on average conditions over the preceding 4-5 years. Finally, the choice of a particular model age-structure of mortality may introduce uncertainty, for the choice must be made on the basis of limited clues and alternative models yield rather different results. Determination of an appropriate age-structure is a problem throughout most of the developing world; it seems to be of particular severity in Africa where there are indications that early child mortality between ages one and five may be exceptionally high relative to infant mortality. One can question whether any of the Coale-Demeny or U.N. models represent early mortality patterns adequately. Recent multi-round surveys by Cantrelle and his associates in Senegal [4, 5] for example, suggest that the pattern may be far more extreme than any of the Coale-Demeny models. Examination of the estimates obtained from the Sullivan procedure using the two most likely Coale-Demeny families (North and West) shows that the estimates of proportions surviving to age five are almost the same whichever family is used, but that the implied levels of infant mortality are divergent [14]. Clearly, whilst retrospective data of this kind can indicate roughly the levels of early mortality, we can have very little confidence in our estimates until we know more about the age-structure of early mortality. For this, multi-round surveys and intensive registration schemes are probably the only answer, though national resources in most countries are not adequate to develop these on a large-scale.

With respect to adult mortality patterns and levels we know even less. The possibility exists of estimating adult mortality from single-round surveys by exploiting data on the proportion of people in each age-group whose mothers and fathers have died. Again we owe the development of a technique to Brass, but in this case the procedure is considerably more complex (especially for males) and has not been fully worked out and tested. The essential idea is analogous to that used in estimation from proportions of children surviving, and again it involves the input of model age-structures of fertility and mortality; the former is rarely available for males. If we can assume that the structure of fertility is known for some time in the past - say five years ago - then we can calculate the approximate age-distribution of the women who gave birth during that year. If we knew also the age-structure and level of adult female mortality, we could then calculate what proportion of those women would have died during the five years before the survey. Conversely, given the proportion of children whose mothers have died, and making some assumption as to the age-structure of mortality, we could estimate the level of that mortality. The procedure suffers similar problems as that for estimating childhood mortality: in addition, the assumption that survival of the children is independent of survival of the mother (analogous to the assumption that survival of the mother is independent of her fertility in the childhood mortality case) may introduce a bias. To date this procedure has not been widely applied, even for females: in Africa, data are available for some five or so countries, and are currently being worked on by Brass.

Conclusion

Indirect estimation techniques have been applied with varying degress of success, their usefulness being restricted most frequently by one or more of three sets of limitations.

Firstly, the sensitivity of most procedures to age-misstatement. This problem bedevils any detailed demographic analysis in much of the developing world, and will continue to plague demographers for a long time. It even prevents reliable indirect estimation of such simple summary measures as the overall population birth and death rates inasmuch as estimation of these must depend at present on techniques sensitive to age mis-statement: overal current fertility and mortality reports cannot be used because of over- or under-reporting.

Secondly, the use of population models to compensate for incompleteness of data. Model age-structures of fertility and mortality, and model stable age-distributions permit estimation in cases where otherwise it would be impossible; but they must all be selected on the basis of fragmentary information, and the appropriateness of any particular model is often open to question. Only intensive attempts to gather reliable direct data on fertility and mortality age-structures can ease this problem.

Thirdly, the assumption in some procedures that the population has experienced constant fertility and mortality in the recent past. A broad theoretical insight has been gained on the relations between age-distribution and trends in vital rates, notably by Coale [9], but at present we are able to make only limited practical application of this: we can estimate current vital rates using age-distribution data provided that the tempo and magnitude of the trends are known and have been comparatively regular. Procedures for detecting and measuring trends are clearly going to become of increasing importance in the future as mortality conditions improve and as fertility levels may start to decline.

It is clear that any estimates obtained from single-round surveys should be viewed with some degree of skepticism; also that we should use and compare all possible pieces of information and methods of estimation, rather than relying on a single item or technique. For many countries, however, they are the only sort of estimates presently available. For those countries where direct recording of vital events is more nearly adequate or where intensive efforts are being made to obtain accurate direct observation on a sample of the population, single-round survey estimates will provide a useful check on, and complement to, the direct data for some time to come.

REFERENCES

[1] Bogue, D.J. and Bogue, E., A Model Interview for Fertility Research and Family Planning Evaluation, Family Planning Manual No. 3; and Techniques of Pregnancy History Analysis, Family Planning Manual No. 4, Community and Family Study Center, University of Chicago, 1970.

[2] Brass, W. et al., The Demography of Tropical Africa, Princeton University Press, 1968.

[3] Cho, L.J., "On Estimating Annual Birth Rates from Census Data on Children," paper presented at the Annual Meeting of the American Statistical Association, Fort Collins, August, 1971.

[4] Cantrelle, P. Etude démographique dans la région du Siné-Saloum (Sénégal), 1963-1965. O.R.S.T.O.M. Travaux et Documents, No. 1, Paris, 1969.

[5] Cantrelle, P. et al., Mortalité de l'enfant dans la région de Khombol-Thiénaba (Sénégal), 1964-1968, O.R.S.T.O,M. Cahiers, Série Sciences Humaines, Vol. VI, Paris, 1969.

[6] Central African Republic, Mission Socio-Economique Centre-Oubangui; France, Secrétariat d'Etat aux Relations avec les Etats de la Communauté, Enquête démographique Centre-Oubangui, 1959. Méthodologie, résultats provisoires, Paris, 1969

[7] Coale, A.J. and Demeny, P., Regional Model Life Tables and Stable Populations, Princeton University Press, 1966.

[8] Coale, A.J., "Age Patterns of Marriage," Population Studies, Vol. XXV (July, 1971), pp. 193-214.

[9] Coale, A.J., Fertility, Mortality and Age Composition, Princeton University Press, forthcoming.

[10] Gaisie, S.K., Dynamics of Population Growth in Ghana, University of Ghana Demographic Unit, Ghana Population Studies No. 1, Legon, undated.

[11] I:N.S.E.E., Afrique noire, Madagascar, Comores: démographie comparée, Paris, 1967.

[12] Lederman, S., Nouvelles tables-types de mortalité, I.N.E.D. Travaux et Documents No. 53, Presses Universitaires de France, Paris, 1969.

[13] Page, H.J. and Coale, A.J., "Estimates of Fertility and Child Mortality in Africa south of the Sahara," in S.H.Ominde (ed.), Studies in the Population of Africa, 1971.

[14] Page, H.J., "Infant and Child Mortality," paper prepared for the African Regional Population Conference, Accra, December,1971.

[15] Seltzer, W., "PGE Studies: Costs and Effectiveness," paper presented at the Annual Meeting of the American Statistical Association, Fort Collins, August, 1971.

[16] Sullivan, J.M., Estimates of Childhood Mortality Conditions from Childhood Survival Statistics, unpublished Ph.D. dissertation, Princeton University, 1970.

[17] United Nations, Age and Sex Patterns of Mortality, ST/SOA/Series A/22, New York,

[18] United Nations, Methods of Estimating Basic Demographic Measures from Incomplete Data, prepared by A.J.Coale and P.Demeny, ST/SOA/Series A/42, New York,1967.

[19] United Nations, The Concept of a Stable Population: Application to the Study of Countries with Incomplete Demographic Statistics, prepared by J. Bourgeois- Pichat, ST/SOA/Series A/39, New York, 1968.

[20] Upper Volta, Service de Statistique et de la Mécanographie; France, Service de Coopération, I.N.S.E.E., Enquête démographique par Sondage en République de Haut-Volta, 1960-1961, Tome II, Paris,1970.

[21] Zachariah, K.C., "Application of Stable Population Techniques for Estimating Demographic Measures for Arab Countries," paper presented at the Annual Meeting of the American Statistical Association, Fort Collins, August, 1971.

PROBLEMS OF APPLYING STABLE POPULATION TECHNIQUES IN ESTIMATING DEMOGRAPHIC MEASURES FOR ARAB COUNTRIES

K.C. Zachariah, International Bank for Reconstruction and Development

Introduction

One of the common procedures for estimating birth rate and other fertility measures of a population which lacks reliable birth registration statistics is the stable population method. In this method, the observed age distribution is assumed to be stable and is used along with the rate of population growth or information on the level of mortality to estimate the crude birth rate or gross reproduction rate. The conditions necessary for the application of the method are more or less fulfilled in several Arab countries of North Africa and South West Asia. Therefore, in a recent study of the demographic measures of Arab countries the stable population procedure was widely used.[1] The application is greatly facilitated by the recent publication of a United Nations manual which not only describes the methodology of the stable population method but also gives numerical examples and discusses many practical problems in applying the method to the data of developing countries.[2] Yet, there are some problems which the manual has not dealt with adequately and some of these were encountered in the study of the Arab countries. This paper describes the approach taken in that study in meeting one of these problems, namely, the effect of a recent decline in mortality on the estimates of birth rates.

As mentioned above, the conditions necessary for the application of the stable population approach are more or less fulfilled in most of the Arab countries. All of them except Saudi Arabia, Lebanon and Yemen have taken at least one census in recent years, and single-year or five-year age distributions are available. It is also possible to estimate the rate of population growth for some recent period. Fertility has remained fairly constant and external migration has been negligible. However, mortality has fallen at a comparatively rapid rate during the last 20 to 30 years. The assumption of constant mortality level implied in the stable population method of estimating birth rates is, therefore, not valid for these countries at the present time. The birth rates obtained by assuming a stable age distribution require corrections for the effects of the recent decline in mortality.

A procedure for adjusting the estimate of the birth rate for the effects of mortality decline is given in U.N. Manual IV (pp. 25-28), but its application to a particular situation requires that the period of mortality decline t, and the degree of mortality decline, k, be known and the growth rate, or the mortality level used in the calculation, measure the average situation during a five-year or a ten-year period prior to the census. As a result, several problems arise in applying this procedure to some countries. First, the intercensal period may not be 5 or 10 years. It is 12 years for Algeria, 13 years for the UAR, 8 years for Iraq, etc. Second, it is rather difficult to determine the year of onset of the decline of mortality. Undoubtedly, most of the improvements in mortality took place in very recent years, but mortality has been falling, though slowly, even in earlier years, and the period of its decline may be taken as a relatively short period of 15 or 20 years when most of the improvements took place or a longer one including the incipient period. Third, the rate of mortality decline has been changing. Inasmuch as the period of mortality decline is indefinite, the amount of decline per year is also uncertain.

Empirical rule for mortality corrections when growth is used

For countries which have taken several censuses in recent years (e.g. the UAR, Algeria, India, Mexico, etc.), an alternate approach to the problem of adjustment of birth rate for mortality decline is given below. In this approach there is no need to make any assumption about the period of mortality decline or to consult any standard table of mortality corrections.

From the tables of mortality corrections given in U.N. Manual IV (p. 119) and from other considerations, it is evident that the mortality corrections are negative at younger ages and positive at older ages. In other words, there is an age at which no correction is required. Moreover, this age of no correction varies according to the period for which the growth rate is calculated; the longer the period, the higher the age at which the correction is zero. Thus, in Manual IV when the growth rate is calculated for a 5-year period, for all values of t above 15 years the correction is negative at ages 5 and 10 and positive at higher ages. In this case, the amount of correction is a minimum (almost zero) at age 10 for all values of k (amount of decline) and for all values of t (period of decline) above 10 years. When the growth rate is calculated for a 10-year period, for all values of t above 20 years the correction is negative at ages 5, 10 and 15 and positive at all other ages. In this case, the amount of correction is at a minimum (almost zero) at age 15, irrespective of the degree and the period of mortality decline. Thus, corresponding to each period (5 years, 10 years, 15 years, etc.) for which the growth rate is calculated, there is an age at which no mortality correction is needed. Alternately, corresponding to each age x (5, 10, 15, ...) there is a period n_x (n_5, n_{10}, n_{15} ....) such that the birth rate estimated from the cumulative age distribution C_x and the growth rate during the period n_x will not require any correction for the effects of mortality decline. The problem of the adjustment of the birth rate for mortality declines is thus reduced to one of finding n_x and estimating the growth rate corresponding to this period.

It is seen that, for values of t above 10 years n_{10} is nearly 5 years and n_{15} is nearly 10 years. Experimental calculations on model populations have shown that for long periods of mortality decline n_5 is not far from zero, n_{20} is nearly 15 years, n_{25} is nearly 20 years, etc. On this evidence, an approximate empirical rule to estimate the birth rate which does not require any significant correction for mortality decline may be stated as follows:

Estimate the birth rate using C_x and the growth rate calculated for the period n_x where

$$n_x = x - 5 \quad ; \quad x = 5, 10, ... 40 \qquad (1)$$

According to this rule, the birth rates are estimated by using different growth rates at different ages thus eliminating the need for mortality corrections, instead of using the same growth rate at all ages and then correcting the estimates for the effects of the decline. However, it should be noted that the period n_x given against ages 5, 10, 15, etc. in equation (1) are approximate values and more exact values can be obtained empirically.

The growth rates for the 5-year period, 10-year period, 15-year period ... 35-year period prior to the census, are generally not available, and these have to be estimated from intercensal growth rates. The following procedure is recommended:

(i) plot the available intercensal growth rates on a graph (e.g. 1950-1960 at 1955, 1940-1950 at 1945, etc.);

(ii) join the points by a smooth curve and extend the curve to the date of the latest census (1960 in the example);

(iii) read off the growth rates at the mid point of each 5-year interval starting from the latest census date (mid point of 1955-60, 1950-55, 1945-50, etc.);

(iv) calculate the average growth rate for longer periods taking the arithmetic mean of the estimated 5-year rates covering the period (e.g. the growth rate for 1940-60 may be taken as equal to the arithmetic mean of the rates of 1940-45, 1945-50, 1950-55 and 1955-60).

In the numerical examples given below the above procedure appears to be satisfactory for all rates except those at the time of the latest census and consequently, the birth rate estimated from C_5 by the empirical rule may not be wholly reliable. The growth rates at higher ages are averages of a number of rates; the errors of interpolation tend to cancel out in the averaging process.

To check whether the procedure proposed in equation (1) gives reasonably accurate estimates of birth rates at ages other than 10 and 15 years, some tests were made and the results are given in Tables 1 and 2. In Table 1, the difference between the actual birth rate and that estimated by the empirical rule in 27 model quasi-stable (constant fertility rates and declining mortality rates) populations is given. All the models were obtained by the component method of population projection starting with initial stable age distributions with mortality level 5.5 (South Family of Princeton Model[3]) and GRR 2.5, 3.0 and 3.5 and assuming that the GRR remains constant through the entire period of projection (20 years, 30 years and 40 years) and expectation of life at birth increases at the rate of 0.5, 0.625 and 0.75 year per calendar year. The three initial levels of fertility in combination with three rates of decrease of mortality and three periods of projection give the 27 model quasi-stable age distributions.

The actual birth rates were obtained by dividing the number of births during a period by the average population of that period. For example, the actual birth rate corresponding to age 25 was the ratio of the average annual births during the 25-year period prior to the end of the projection period to the average population of that period. The births and growth rates for periods before the base year of the projection were taken as those of the corresponding stable population. The rates according to the empirical rule were obtained by using the final age distribution for each model and different growth rates as stipulated in the empirical rule (equation 1 above).

It is seen from Table 1 that the birth rates derived according to the empirical rule are very close to the actual rates. The maximum difference between the actual and estimated birth rates is of the order of 0.8 units (less than 2 percent) but most of the differences are of much smaller magnitude. The average absolute deviation varies between 0.2 units and 0.5 units. On the whole, the error in the estimate of the birth rate obtained by the empirical rule is minimum when the period of mortality decline is about 30 years, when the rate of increase of expectation of life at birth has been about 0.5 year per year (Table 4). When the period of mortality decline is less than 30 years the empirical rule tends to underestimate the birth rate and when the period of mortality decline is more than 30 years, the method tends to overestimate the birth rate. The level of fertility does not appear to be a relevant factor. Thus, at least for the range of variation of GRR and the rates and periods of mortality decline implied in the projections, the emplirical rule seems to give fairly satisfactory estimates of the birth rate.

In Table 2, comparisons are made in two actual populations, namely, the female population of India, 1961 and the male population of Mexico, 1960. The rates estimated by the Manual IV procedure are taken from that publication. They are close to the values obtained by the empirical rule, the maximum difference being 0.9 units or about 2 percent. The range of variation of the rates is, in fact, slightly smaller for the set of values obtained by the empirical rule. For India, the estimates obtained by this new approach are as good as, if not better than, those obtained by the Manual IV procedure. For Mexico, the Manual IV procedure seems to give more consistent estimates. At the same time, the new approach has an advantage in that there is no need to consult any standard table of mortality corrections.

Empirical rule when mortality level is used

Independent data to estimate the level of mortality were not available for most of the Arab countries. Therefore, the stable population method using mortality level instead of intercensal growth rate was not attempted in that study, although its use in place of the growth rate would have improved the accuracy of the estimates. The use of mortality level, however, would not have eliminated the need for corrections for mortality decline; the estimates obtained by the use of a single mortality level at all ages require correction for mortality decline. The nature of the corrections is somewhat similar to the situation when growth rates are used. They are negative at younger ages and positive at older ages. Corresponding to each mortality level used, there is an age where no correction is required. The longer the period to which the mortality level refers (5 years before the census, 10 years before the census, etc.) the higher is the age at which the transition from negative correction to positive correction takes place.

A working rule similar to that proposed above when growth rate is used may be given when the mortality level is used. It consists in the use of different mortality levels at different ages; the mortality level at the time of the census to be used along with the cumulative age distribution at age 5 (C_5), the average mortality level during the 5-year period prior to the census to be used along with C_{10}, that during the 10-year period prior to the census along with C_{15}, etc.

The validity of this rule is checked with model quasi-stable populations (the same 27 models mentioned earlier) and the results are given in Table 3. The differences between the actual birth rates and those estimated by the empirical rule are on the whole small, none exceeding 0.9. Most of the differences are negative suggesting that the empirical rule, in general, underestimates the birth rate. The difference varies from one age to another and from one model population to another. The empirical rule appears to be most accurate when the period of mortality decline is about 40 years, the rate of increase of expectation of life at birth is about 0.5 years, and when GRR is 3.5 (Table 4). However, the influence of the rate of mortality change and the level of fertility is, on the whole, negligible compared with the influence of the period of mortality decline.

The tests given in Tables 1, 2 and 3 do not establish the general validity of the empirical rule. In fact, the procedure may not be valid for all situations, but it appears to be satisfactory as a working rule for populations for which the GRR ranges between 2.5 and 3.5 and the expectation of life at birth has increased at a rate of one-half to three-fourths of a year per year for about 20 to 40 years, a situation which is closely approximated in most of the developing countries.

[1]Cairo Demographic Centre, Demographic Measures and Population Growth in Arab Countries, (Cairo) 1970.

[2]United Nations: Manual IV, Methods of Estimating Basic Demographic Measures From Incomplete Data, (New York) 1967.

[3]Coale, A.J. and Demeny, P., Regional Model Life Tables and Stable Populations, Princeton University Press, (Princeton) 1966.

Table 1: DIFFERENCE BETWEEN ACTUAL BIRTH RATES AND THOSE ESTIMATED BY THE EMPIRICAL RULE IN MODEL POPULATIONS

(Using growth rates)

Age	Rate of increase of expectation of life at birth								
	0.50			0.625			0.75		
	Period of mortality decline			Period of mortality decline			Period of mortality decline		
	20 years	30 years	40 years	20 years	30 years	40 years	20 years	30 years	40 years
	GRR = 2.5								
5	-0.15	-0.04	+0.12	-0.25	-0.17	+0.17	-0.18	-0.08	+0.12
10	-0.05	+0.19	+0.38	-0.07	+0.21	+0.42	-0.05	+0.23	+0.41
15	-0.06	+0.28	+0.56	-0.07	+0.34	+0.53	-0.08	+0.40	+0.58
20	-0.28	+0.35	+0.65	-0.32	+0.36	+0.64	-0.38	+0.43	+0.62
25	-0.35	+0.27	+0.70	-0.45	+0.21	+0.73	-0.48	+0.26	+0.63
30	-0.33	-0.08	+0.56	-0.44	0.00	+0.69	-0.58	-0.12	+0.64
35	-0.39	-0.23	+0.47	-0.52	-0.27	+0.51	-0.64	-0.33	+0.49
40	-0.44	-0.49	+0.37	-0.62	-0.47	+0.08	-0.74	-0.56	+0.09
	GRR = 3.0								
5	-0.21	-0.13	-0.10	-0.29	-0.18	+0.10	-0.36	-0.16	-0.01
10	-0.08	+0.15	+0.29	-0.12	+0.13	+0.34	-0.07	+0.16	+0.29
15	-0.06	+0.27	+0.49	-0.15	+0.36	+0.53	-0.17	+0.36	+0.49
20	-0.25	+0.35	+0.61	-0.33	+0.43	+0.65	-0.44	+0.45	+0.64
25	-0.30	+0.25	+0.75	-0.41	+0.30	+0.65	-0.46	+0.39	+0.75
30	-0.35	+0.08	+0.75	-0.41	-0.04	+0.78	-0.46	+0.01	+0.73
35	-0.40	-0.08	+0.59	-0.53	-0.22	+0.68	-0.55	-0.34	+0.53
40	-0.44	-0.31	+0.29	-0.69	-0.28	+0.37	-0.72	-0.52	+0.23
	GRR = 3.5								
5	-0.40	-0.23	-0.05	-0.34	-0.26	-0.02	-0.36	-0.25	0.00
10	-0.19	+0.05	+0.20	-0.12	+0.06	+0.29	-0.09	+0.08	+0.30
15	-0.15	+0.26	+0.38	-0.10	+0.32	+0.49	-0.09	+0.30	+0.47
20	-0.31	+0.35	+0.54	-0.40	+0.44	+0.59	-0.50	+0.45	+0.62
25	-0.28	+0.30	+0.66	-0.40	+0.39	+0.74	-0.50	+0.41	+0.80
30	-0.30	+0.07	+0.79	-0.35	+0.08	+0.74	-0.39	+0.17	+0.83
35	-0.32	-0.03	+0.31	-0.42	-0.17	+0.74	-0.41	-0.19	+0.72
40	-0.45	-0.15	+0.47	-0.61	-0.21	+0.48	-0.55	-0.46	+0.31

Table 2: COMPARISON OF BIRTH RATES ESTIMATED BY THE EMPIRICAL RULE WITH THOSE OBTAINED BY THE MANUAL IV PROCEDURE; FEMALE POPULATION OF INDIA, 1961 AND MALE POPULATION OF MEXICO, 1960

(Using growth rates)

Age	India, 1961			Mexico, 1960		
	Empirical rule	Manual IV	Difference	Empirical rule	Manual IV	Difference
5	38.9	38.3	0.6	37.5	37.0	0.5
10	46.2	46.5	-0.3	42.0	41.1	0.9
15	44.6	44.6	0.0	43.7	43.4	0.3
20	40.5	40.0	0.5	42.9	43.3	-0.4
25	41.8	41.0	0.8	42.7	43.1	-0.4
30	44.0	43.7	0.0	43.0	43.3	-0.3
35	45.3	45.4	-0.1	42.7	43.1	-0.4
40	45.3	45.7	-0.4	44.1	44.4	-0.3

Note: The estimates by the Manual IV procedure are taken from Manual IV pp. 69 and 71, and those by the empirical rule are based on West Family of stable populations, C_x values and the growth rates as stipulated by equation (1). For India the intercensal growth rates of 1921-31, 1931-41, 1941-51 and 1951-61 were first calculated, and those of 1951-56, 1946-51 ..... 1926-31 were obtained from them by interpolation and of 1956-61 and 1961 by graphical extrapolation. The growth rates for other periods (e.g. 1926-1961) were obtained as simple average of the 5-year rates covering that period. For Mexico a similar procedure was followed using the growth rates of 1950-60, 1940-50, 1930-40 and 1921-30.

Source: Demographic Measures and Population Growth in Arab Countries, op. cit. Appendix I, Table 4, p. 334.

Table 3: DIFFERENCE BETWEEN ACTUAL BIRTH RATES AND THOSE ESTIMATED BY THE EMPIRICAL RULE IN MODEL POPULATIONS

(Using mortality levels)

Age	Rate of increase of expectation of life at birth								
	0.50			0.625			0.75		
	Period of mortality decline			Period of mortality decline			Period of mortality decline		
	20 years	30 years	40 years	20 years	30 years	40 years	20 years	30 years	40 years
	GRR = 2.5								
5	-0.35	-0.39	-0.21	-0.09	-0.42	-0.20	-0.42	-0.32	-0.21
10	-0.35	-0.20	-0.13	-0.12	-0.23	-0.12	-0.39	-0.24	-0.13
15	-0.30	-0.16	-0.05	-0.26	-0.19	-0.08	-0.44	-0.21	-0.09
20	-0.40	-0.19	-0.02	-0.39	-0.20	-0.06	-0.54	-0.21	-0.08
25	-0.41	-0.21	-0.02	-0.52	-0.26	-0.05	-0.64	-0.30	-0.08
30	-0.30	-0.33	-0.06	-0.60	-0.41	-0.08	-0.71	-0.47	-0.11
35	-0.05	-0.45	-0.11	-0.66	-0.55	-0.16	-0.78	-0.65	-0.22
40	+0.23	-0.49	-0.26	-0.72	-0.73	-0.36	-0.84	-0.86	-0.46
	GRR = 3.0								
5	-0.38	-0.33	-0.24	-0.47	-0.35	-0.25	-0.52	-0.37	-0.23
10	-0.36	-0.23	-0.14	-0.42	-0.26	-0.14	-0.45	-0.27	-0.15
15	-0.35	-0.18	-0.05	-0.41	-0.20	-0.08	-0.48	-0.21	-0.10
20	-0.41	-0.18	-0.03	-0.47	-0.20	-0.02	-0.57	-0.21	-0.07
25	-0.43	-0.20	0.00	-0.52	-0.23	-0.03	-0.59	-0.26	-0.04
30	-0.26	-0.28	-0.02	-0.58	-0.36	-0.02	-0.70	-0.42	-0.05
35	+0.10	-0.36	-0.05	-0.64	-0.46	-0.08	-0.75	-0.56	-0.14
40	+0.43	-0.38	-0.17	-0.70	-0.62	-0.24	-0.80	-0.75	-0.33
	GRR = 3.5								
5	-0.49	-0.34	-0.26	-0.52	-0.38	-0.23	-0.58	-0.41	-0.24
10	-0.43	-0.24	-0.15	-0.46	-0.28	-0.15	-0.48	-0.30	-0.16
15	-0.38	-0.18	-0.07	-0.44	-0.22	-0.11	-0.50	-0.25	-0.10
20	-0.44	-0.18	-0.03	-0.52	-0.19	-0.06	-0.60	-0.23	-0.07
25	-0.38	-0.18	+0.02	-0.52	-0.21	+0.04	-0.61	-0.26	-0.03
30	-0.24	-0.26	+0.02	-0.58	-0.35	-0.02	-0.69	-0.40	-0.03
35	+0.21	-0.28	+0.02	-0.62	-0.40	-0.03	-0.72	-0.52	-0.10
40	+0.63	-0.29	-0.10	-0.69	-0.52	-0.18	-0.77	-0.69	-0.27

Table 4: AVERAGE ABSOLUTE DEVIATIONS OF ACTUAL AND ESTIMATED BIRTH RATE BY LEVEL OF FERTILITY, RATE OF INCREASE OF EXPECTATION OF LIFE AT BIRTH, AND PERIOD OF MORTALITY DECLINE

GRR	Δe_o	Period of mortality decline		
		20	30	40
		Using growth rates		
2.5	0.500	0.26	0.24	0.47
2.5	0.625	0.34	0.25	0.47
2.5	0.750	0.39	0.31	0.45
3.0	0.750	0.26	0.20	0.48
3.0	0.625	0.37	0.24	0.51
3.0	0.750	0.40	0.30	0.46
3.5	0.500	0.30	0.18	0.42
3.5	0.625	0.34	0.24	0.51
3.5	0.750	0.36	0.29	0.51
		Using mortality levels		
2.5	0.500	0.30	0.30	0.11
2.5	0.625	0.42	0.37	0.14
2.5	0.750	0.60	0.41	0.17
3.0	0.500	0.34	0.27	0.09
3.0	0.625	0.53	0.34	0.11
3.0	0.750	0.61	0.38	0.14
3.5	0.500	0.40	0.24	0.08
3.5	0.625	0.54	0.32	0.10
3.5	0.750	0.62	0.38	0.12

Source: Calculated from Tables 1 and 3.

ON ESTIMATING ANNUAL BIRTH RATES FROM CENSUS DATA ON CHILDREN

Lee-Jay Cho, East-West Population Institute and University of Hawaii

I. INTRODUCTION

For the majority of the world's population, the registration of vital events is incomplete, hence measurement of fertility based on birth statistics is virtually impossible. Most nations, however, conduct population censuses and surveys. Data from such sources may be used to fill in a large part of the gap in our knowledge on fertility. This paper presents procedures for estimating recent current fertility from the census data on young children. Their applications do not require special questions to be added to normal census schedules, merely simple tabulations of young children by age of mother. The techniques described here are further elaborations of an earlier work on estimating current fertility from census or survey data on young children.[1] This paper demonstrates the method of estimating annual birth rates with the data from the 1966 census for Korea, which like the majority of the countries in the world does not have a complete and reliable vital registration system.

Fortunately, the age data from Korean censuses have been tested in various studies and found to be accurate. Koreans remember their ages in the lunar calendar system quite accurately, because the year, month, day and even the hour of birth are needed to meet important requirements in their cultural tradition.

1966 Census Data on Own Children

The 1966 Korean Census was conducted on 1 October 1966; by using the census data on own children under 10 years of age, estimates of fertility for each year of the ten-year period preceding the census date can be made.

All 1966 census respondents were asked basic items such as age, marital status, and level of education, whereas sample enumeration was carried out for such socioeconomic characteristics as fertility, occupation, and labor force participation. Of the total 37,530 enumeration districts (EDs)--excluding special enumeration districts, such as military and other institutions--10 per cent were enumerated for detailed characteristics.

The 10 percent sample census allows us to estimate age-specific birth rates for the ten years preceding the census date for administrative areas, down to Gun level. (A Gun is an intermediate administrative area with an average population of about 130,000.) For Myun (the smallest administrative area with an average population of about 12,000) we are able to derive the average age-specific birth rates for the two five-year periods preceding the census date. The computation of the age-specific fertility rates for Myun from the ratios of own children under 5 and 5-9 years of age requires the use of interpolation procedures as well as the development of regression equations from the data for larger areas. The methodological work on this aspect is presently being carried out by the author at the East-West Population Institute.

Accuracy of Age Data

The majority of Koreans, like the Chinese, believe in the 12-year cycle of 12 different animals. Therefore, an early assumption of Westerners that rural Koreans would not know their exact age is false. Even in the rural areas, age is accurately reckoned. Korean parents keep records of their children's dates of birth (and in the majority of cases the hour of birth as well) according to the lunar calendar for specific cultural and traditional requirements.[2]

As in the case of Japan, therefore, age heaping is probably less common in Korea than in the United States. (Age heaping is a kind of age misreporting in which certain ages, often those ending in a specific digit, such as 5, 10, 15, 20, 25, etc., are overreported, and other ages, such as 19, 29, 39, etc., are underreported.)

In order to obtain accurate age data in Korea, one must not ask only for the age of an individual but rather also for the date of birth, according to either the lunar or the Western calendar. In the case of the lunar calendar, the name of the animal of the birth year should be obtained. The following question was asked to obtain information on age in the 1966 Korean Census: "How old are you?" "Specify Lunar or Western calendar birth date."

A serious mistake was committed in the 1960 census in Korea by simply asking the question "How old are you?" without allowing for the different calendars. Consequently, the age data from the 1960 census require a great deal of adjustment and modification for use in demographic analysis.

According to the Koreans and the Chinese, a child is one year old at the time of birth (implying that the nine-month gestation period is counted as one year of life). Furthermore, age in Korea is not reckoned from the last birthday, but the New Year (1 January). Thus, if a child is born on 1 January 1970, he will be considered two years old on 1 January 1971, exactly one year after birth. Thus, if one simply asks how old a person is, the information obtained will be substantially different from the information obtained by asking the age in completed years at last birthday. For example, in one extreme case, a child born on 31 December 1969, will be two years old on 1 January 1970 (while actually, the child is one day old according to age in completed years).

II. CURRENT FERTILITY MEASURES FROM THE AGE DATA

Census or survey data on own children who live with their mothers provide valuable material for estimating fertility when birth statistics are inadequate. By counting the number of children who live in the same household by age of mother, one can easily relate own children under a certain age to mothers or women in childbearing ages.

The principal techniques employed here require the knowledge of age-specific ratios of own children to women. For example, using data on own children aged 0-4 and 5-9 years, the procedure yields average annual fertility rates for the two five-year periods prior to the survey or census date. The estimates of fertility will be as accurate as the census or survey data on which they are based. The most important requirement is the accuracy of age reporting for young children.

Adjustments of Data

Necessary adjustments of the raw data are allowances for mortality of children and women in associated ages, and for those children not living with mothers. For the age data on children that are subject to age-misstatement, correction factors must be developed to adjust for age-misstatements of own children. In the case of the census data, an additional adjustment for undercount of children and women must also be made.

1. Mortality of Children and Women

a. Life Table

Mortality has declined in most countries. In recent years it has been so low in Korea that plausible variations of the adjustment factors would have little effect on the estimated fertility rates.

The Korean Bureau of Statistics has published two sets of life tables.[3] The 1955-1960 life table was calculated on the basis of the census data for 1955 and 1960. By comparing the age distribution of the two censuses, it was possible to estimate the proportion surviving in each age group, and (with certain adjustments) mortality rates. The 1955-1960 life table would reflect mortality during the five-year intercensal period.

The second life table is based on the enumerated number of deaths in the 1966 Special Demographic Survey (SDS). Because of the usual underenumeration of deaths of young children by the survey, some adjustments were made in the mortality figures for young children on the basis of the mortality curve for older ages. The life table based on the SDS data on deaths indicates the mortality situation in 1966. By employing Keyfitz's new iteration method,[4] a new life table based on the same survey data was prepared by the present author. The new life table differed little from the 1966 table published by the Bureau of Statistics.

The life table for 1955-1960 and the two life tables based on the 1966 survey indicate a substantial decline in mortality, and this trend appears quite reasonable. If the two sets of life tables, 1955-1960 and 1966, are to be employed for adjusting mortality of children for each of the ten years covered (1957-1966), interpolation of mortality rates between the two life tables must be made.

b. Brass technique

A census or survey usually provides data on the survival of children ever born by age of mother, which can be used to generate the adjustment factors for mortality by employing the procedure developed by William Brass.[5]

In the Brass procedure for estimating childhood mortality from reports of the number of children ever born who had died previous to the census, it is assumed that age-specific fertility and mortality rates have remained constant for the required age range and time period. The Brass estimates of child mortality are affected by the age pattern of fertility but are not affected by the level of fertility. In the case of Korea, the level and age pattern of fertility have substantially been changing in the recent years, and what is needed here is the adjustment factors that allow for changing age pattern of fertility brought about by rise in age at marriage in Korea.

Even without the adjustment for changes in the age pattern of fertility, the Brass estimates of childhood mortality from the 1966 census data on survival of children ever born appear reasonable, and deviate little from those child mortality rates based on the life tables cited above. (See part A of Table 2.)

c. Own Children and Children ever born to Women under 25 years of Age.[6]

Estimates of childhood mortality can be made from the data on number of own children living by age, for each age of mother, and on number of children ever born for each age of mother.

The general strategy is to use model life tables, and to determine the level of mortality which would: (1) yield the number of own children living for women at a particular age if employed to determine (by reverse-surviving) the number of births, and (2) provide a number of births equal to the total number of children ever born reported by age of these women. The procedure would allow for declining mortality.

By using the model life table, it is possible to find the fractional level of model life-table that would give an estimate of lifetime births equal to reported number of children ever born to women at a particular age, and that would indicate a level of mortality accounting for the reported difference between living own children and children ever born, for each age of women. In general, the estimated mortality will be lower for younger women, because their children will not have been exposed to the higher mortality in the past.

Since the data on children ever born to women by age in single years are not available from the 1966 census, the procedure is presently being experimented on the data from the 1970 census.

When mortality is relatively low in a country as in the case of Korea, further refinement of mortality adjustments will have little effect in the final estimate of fertility. For example, if the estimates of fertility for 1965 are made by using life tables reflecting two substantially different levels of mortality--notably, (1) $e_o^o = 64$ for females and 60 for males (1966 and (2) $e_o^o = 54$ for females and 51 for males (1955-60) showing a difference in life expectancy of about 10 years--we find that the estimated total fertility rates (TFR) differ by less than 5 percent and that the differences in the age-specific fertility rates are also very small:

Estimated Fertility Rates for Korea 1965

Age-specific Fertility Rates per 1000 Women

	Life Table with	TFR	15-19	20-24	25-29	30-34	35-39	40-44	45-49
1955-60	(1) Female $e_o^o = 54$	4.687	17	200	293	215	140	63	11
1966	(2) Female $e_o^o = 64$ Male $e_o^o = 60$	4.478	16	190	279	206	134	60	10
	Difference:	.208							

2. Children not living with mother

Most young children live in the same household as their parents and are, therefore, enumerated with their parents. In Asian countries, the proportion of young children who do not live with their mothers appears to be very small. For example, according to the 1966 Korean census of population, 98.2 percent of all persons under five years old were living with their mothers. Similarly, 95.3 percent of children five to nine years old were enumerated with their mothers.[7] The non-own children by each age were proportionately distributed to each age of women.

3. Census undercount

The Post-Enumeration Survey (PES) estimates in 1966 for the population are about the same as the 1966 census count. The census count of the population 0-4 years old equalled the PES estimate, but the PES estimate of the population 5-9 years old slightly exceeded the census count. If the PES estimate for the age groups 5-9 is used to correct for underenumeration in the 1966 census, fertility estimates for the period 5-9 years prior to the census would be slightly inflated, resulting in fertility estimates indicating a sharper decline. For this reason, no corrections are made for underenumeration in the present report.

The single-year age distribution from the 1970 census, however, will enable us to determine the extent of undercount of children in each age class in 1966. Preliminary estimates of fertility from the 1970 PES shown in Figure 1 indicate that children under 1 and 1 year of age appear to be somewhat underenumerated.

4. Age-misstatement

One must not entirely ignore the possibility of age-misstatement. The extent of age-misstatement of Korea, fortunately, has not been a major one in the previous censuses. A preliminary examination of the 1966 census age data for children indicates that there may be a small extent of age-misstatements in certain ages, but these would, at most, be a magnitude of 2 or 3 percent for which correction factors can be calculated using the ages distributions from the 1966 and 1970 census. The fertility estimates from the 1966 census presented in this paper have not been corrected for negligibly small extent of age misstatement.

Estimating Procedure

The essence of the estimating procedure is the reconstruction of the fertility experience of women enumerated in a survey or census in the ten-year period preceding the enumeration. Retrospective fertility estimates are made for the single-year cohorts of women from 15 to 54 years old at the time of the enumeration, and then, by simple interpolation of these estimates and by translation of the age of women at the time of the census to age at the reference period, the conventional period measures of age-specific fertility are derived. The following discussion describes in detail the procedure by which the fertility rates for each of the ten years preceding the 1966 Korean census date were estimated.

(1) Own children tabulation.

Usually, a census or survey operation requires coding the age of children living in the household. Own children can easily be tabulated by age of child and that of mother. For Korea, the data on own children by age were cross-tabulated by age of mother using single year classes between ages 0 and 10 for children, and 14 and 54 for women.

(2) Estimate of births: the numerator.

The women enumerated in the census represent a set of single-year female birth cohorts. Fortunately, as the census was taken on October 1, the age at the time of the census is approximately indentifiable by single calendar years. For example, women 15 years old at the census date, in this case 1966, may be taken as the birth cohort of 1952, women 16 years old as the cohort of 1951, and so forth. For each cohort, we have the number of own children under ten years of age by single years of age, as shown in Table 1 for Korea. These children can easily be "reverse-survived" to estimate the number of births for each of the ten years preceding the census. Thus, the number of children born to each birth cohort of women in the first year prior to the census is estimated by reverse-surviving children under one year of age and by making an allowance for the proportion of children not living with their mothers. In the same manner, children one year old can be "reverse-survived" to estimate the annual number of births in the second year prior to the census date. In general,

$$B^a_{-i} = C^a_i \; S^c_i \; M_i \; U_i \qquad i = 0,1,2,\text{--}9$$

where B^a_{-i} is the number of births $\underline{i}$ years prior to the survey date to women age $\underline{a}$ at the census date;

C_i is the number of own children $\underline{i}$ years old living in the household at the time of the census;

S^c_i is the reverse-survival factor for children from age $\underline{i}$ to birth, calculated as L_0/L_i from an appropriate life table;

M_i is the inverse of the proportion of children aged $\underline{i}$ living with their mothers;

U_i is the adjustment factor for underenumeration for persons aged $\underline{i}$ years. (This allowance for underenumeration of children is usually required if the data originate from a census.)

(3) Estimate of single year female cohort: the denominator.

For each set of birth estimates for each of the ten calendar years preceding the census, cohort sizes for each of the cohorts are estimated at the midpoints of the same ten years. This is done by taking the number of women by single years of the census age and "reverse-surviving" them with appropriate adjustment for mortality. The resulting estimates are the denominators for birth estimates, yielding fertility rates for each of the ten calendar years. The estimated female population by the census age can easily be "reverse-survived" to estimate the necessary cohort sizes for each of the same ten years, by the following step:

$$W^a_{-i} = W_a \; S^w_i \; U_a \qquad i = 0,1,2,\text{--}9$$

where W^a_{-i} is the number of women of the census age $\underline{a}$, $\underline{i}$ years before the survey;

W_a is the number of women age $\underline{a}$, at the time of the census;

S^w_i is the reverse-survival factor for $\underline{i}$ years calculated as $L_a/L_{a-i+\frac{1}{2}}$ from an appropriate life table.

(Again, use of the adjustment factor U_a is usually required if data are from a census.)

Thus, for instance, the ratio of children nine years old to mothers 30 years old at the census date represents the fertility of these women 9½ years before the census, when allowances are made for mortality of both children and women, and for children not living with mothers. It represents the fertility rate for women 20 years old ten years prior to the survey date. The denominator in this case is estimated by "reverse-surviving" women 30 years old at the census date for 9½ years. (Similarly, when computing the fertility rate for women for the first year preceding the census, the census estimate of the female population must be "reverse-survived" for half a year; for the second year, women must be "reverse-survived" for 1½ years; and so forth.) Table 4 presents estimates of cohort sizes of the single-year female birth cohorts at the midpoint of each of the ten calendar years preceding the census date for Korea.

The assumption that the fertility of women living at the time of the census is representative of the fertility of all women, including those who died during the period under study, would obviate the need for making allowance for mortality of women in associated ages, and also for those children not living with their mothers, insofar as the mothers of children not living with their mothers are dead. This would mean that allowance is made only for mortality of own children.

This assumption technically facilities the estimation procedures. The problem, however, is what proportion of non-own children are of dead mothers. If the assumption that mothers of all non-own children are dead is made, this would slightly underestimate the level of fertility, particularly estimates derived from the data for older children, because the proportion of non-own children increase with age, and it is very likely that not all the mothers of non-own children are dead.

(4) Fertility rates by birth cohort of women.

The elements of the birth matrix in Table 3 are divided by the corresponding elements of each of the vectors in Table 4, to obtain the single year age-specific fertility rate for each of the years under study, namely,

$$f^a_{-i} = \frac{B^a_{-i}}{W^a_{-i}} \qquad i = 0,1,2,\text{--}9$$

where f^a_{-i} is the fertility rate for women of the census age $\underline{a}$, $\underline{i}$ years before the census date, namely the fertility rate $\underline{i}$ years preceding the census date.

The estimated fertility rates (expressed for 1,000 women) are presented in Table 5. These rates indicate the fertility experienced by each of the female cohorts in each of the ten calendar years preceding the census.

(5) Fertility rates by age of women.

The fertility rates for 1966 in Table 5 represent the fertility of women from 14½ to 53½ years old (by single years) at the time of actual childbearing, and the fertility rates for 1966 represent the fertility of women from 15½ to 52½ years old at the time of actual childbearing, and so forth. By simple linear interpolation, fertility rates were estimated for the conventional single year age of women. This was done by taking the moving average of the fertility rates by age in Table 5 for each calendar year, and then moving up the fertility rate column of 1966 by one cell (year of women's age), that of 1965 by two cells, and so forth. The results as shown in Table 6 are estimates of age-specific fertility rates (period measures) by conventional single year age for the ten years preceding the census.

It is, however, preferable to produce the five-year age-specific fertility rates simply out of convenience in handling and analysis. This is done either by taking the weighted average fertility rate[8] from the single year age-specific rates for each of the five-year age groups from 15-19 to 45-49 years of age, or by performing the necessary calculations separately for the numerator and the denominator of the conventional five-year age-specific fertility rates, i.e., from the sizes of female cohorts (Table 4) and the number of births (Table 3). Dividing the consolidated number of births by corresponding women in the five-year age groups and making an allowance for non-own children yields the five-year age-specific fertility rates which are shown in Table 7.

Comparison with the Fertility Rates from Other Sources

Estimates of fertility rates for a recent period should always be followed by an effort to check their accuracy. The fertility estimates from the 1966 census data will be validated when fertility estimates from the 1970 census data are made for the overlapping period from 1960-1966 using the same methodology. The 1970 census data on own children are presently being tabulated.

One kind of check can be made by using the Korean 1968 Fertility and Family Planning Survey, which covered a total of 8,500 households from 1 September to 31 October. The fertility estimates for the ten years preceding the survey date would enable us to provide a check on the census estimates for the period 1959-1966. Thus, utilizing the household roster and the pregnancy histories obtained by this survey, the number of children was tabulated by age of child and age of mother. The same mortality adjustments that were employed for the 1966 census data on own children were applied to the survey data. Since the field work for the survey was done in the month of September, the annual fertility estimates from the survey refer to the years preceding 15 September, which corresponds to roughly the time when the 1966 census was taken (1 October). Therefore, the estimated annual fertility rates from both the survey and the census refer to about the same time periods. For the five-year period from 1961-1966, the two sets of estimates agree quite well. (See Table 8.)

Following the population census conducted in October, 1970, the PES was conducted using a random sample of about 8,000 households; it was taken during the ten-day period 20-30 November to check the completeness and accuracy of the census count.

From the PES schedules, the own children under 10 years old were tabulated according to the age of their mothers. Preliminary estimates of the fertility rate for the period preceding the census date were derived from these data. The PES estimates of fertility rates also compare very well with the census estimates as shown in Figure 2; the enumeration was done 4 years apart.

III. CONCLUDING OBSERVATIONS

The own children method of estimating recent fertility can be a powerful technique in countries where the following principal requirements are met:

(a) The census or survey data on children's ages are reasonably accurate.

(b) Most of the young children live with their mothers, and their relationship to the head of the family or household is clear.

(c) Mortality in the years prior to the census is relatively low.

Considering age accuracy, for example, Malaysian data on age[9] indicate that the age records of children of such ethnic groups as Malays and Indians, for which the age-reporting had not previously been accurate, have in recent years improved to such an extent that reasonably accurate birth rates for these ethnic groups can be estimated. And in the case of the age data with minor extents of age-misstatement, plausible correction factors can be developed. With the general social development in Asia, and particularly with the introduction of a compulsory educational system in an increasing number of countries, there is greater need for reckoning children's ages. Consequently, the 1970-71 censuses of population will undoubtedly produce more accurate age data, particularly for younger children, than the previous ones.

In most Asian societies the relationship of young children to the head of family tends to be clearly defined. Therefore there appears to be little difficulty in relating own children to their mothers in the census or survey schedule if some effort is made to provide precoded relationships.

Mortality in most countries had declined to such an extent that plausible variations in the adjustment factors for mortality would cause only small errors in the estimated fertility rates.

The foregoing observations suggest that countries with poor vital statistics would do well to produce tabulations on own children from their forthcoming population censuses. Such tabulations do not require adding special questions to the census questionnaire and can be obtained at moderate expense if done on a sample basis. The own children method, if carefully applied, will not only generate good estimates of recent fertility trends but may also facilitate studies of differential fertility.

FOOTNOTES

[1]Wilson H. Grabill and Lee-Jay Cho, "Methodology for the Measurement of Current Fertility from Population Data on Young Children," Demography, II (1965).

[2]For example, "Sajoo" (literally translated as the four pillars of a person): meaning that the time, day, month, and year of birth determine the child's future, and provide a basis for favorable marriage match. In the past, some mothers would try to postpone a birth until a certain hour of the day that would be propitious to a good combination of "Sajoo", although this practice is rare today.

[3]Korean Bureau of Research and Statistics, A Comprehensive Study on 1966 Population Census, Seoul, Korea, 1970.

[4]Nathan Keyfitz, "Finding Probabilities from Observed Rates or How to Make a Life Table," The American Statistician, 24(1):28-31 (February, 1970).

[5]William Brass et al., The Demography of Tropical Africa, Princeton: Princeton University Press, 1968, pp. 104-132; and United Nations, Manual IV, Methods of Estimating Basic Demographic Measures from Incomplete Data, ST/SOA/A/42, New York: United Nations, 1967.

[6]Suggested by Professor Ansley J. Coale of the Princeton University.

[7]Lee-Jay Cho and Man Jun Hahm, "Recent Change in Fertility Rates of the Korean Population," Demography, V. 2, (1968).

[8]The weights being the number of women in each age.

[9]Lee-Jay Cho, Estimates of Population for West Malaysia, Department of Statistics, Government of Malaysia, Kuala Lumpur, 1969.

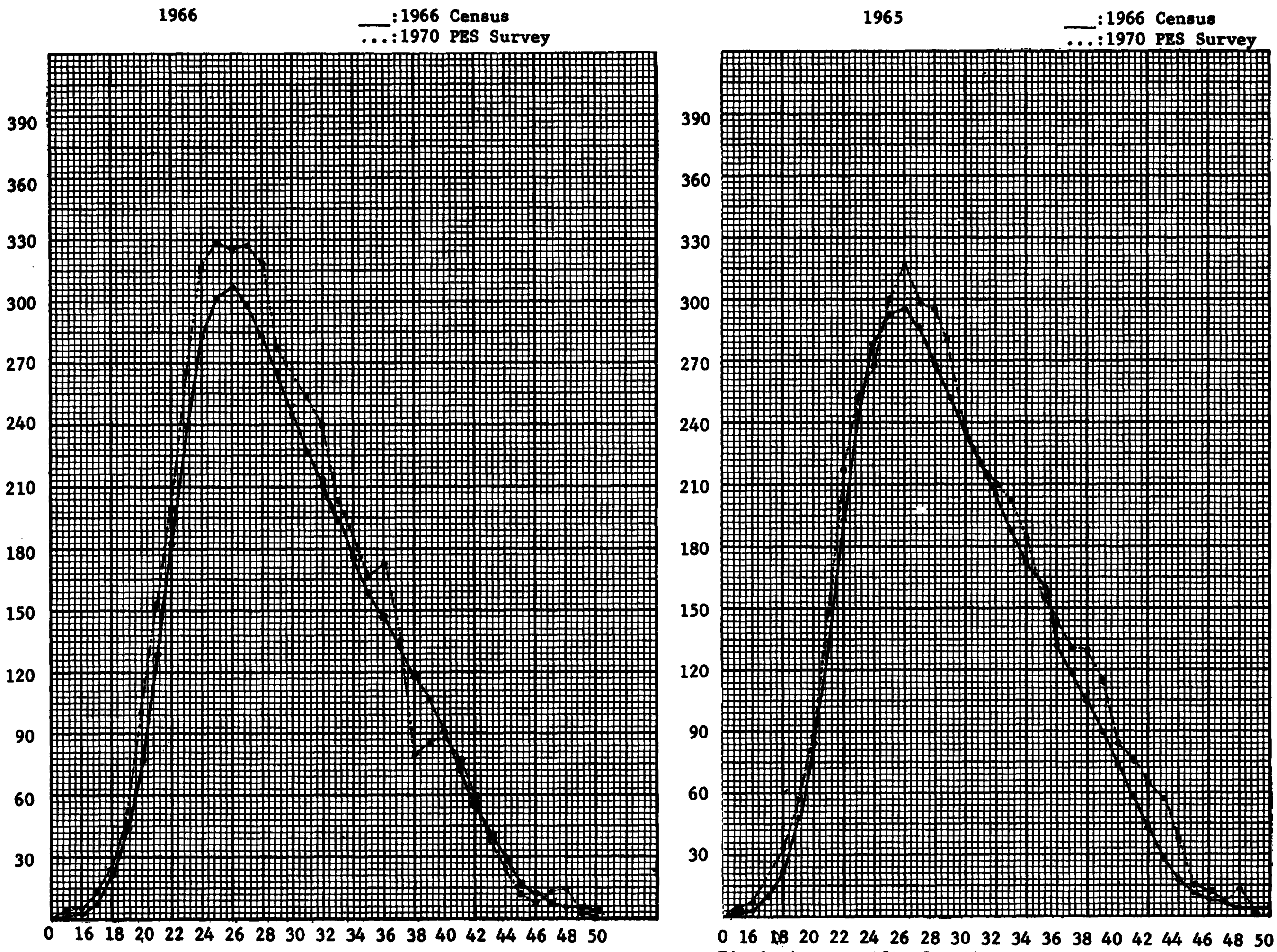
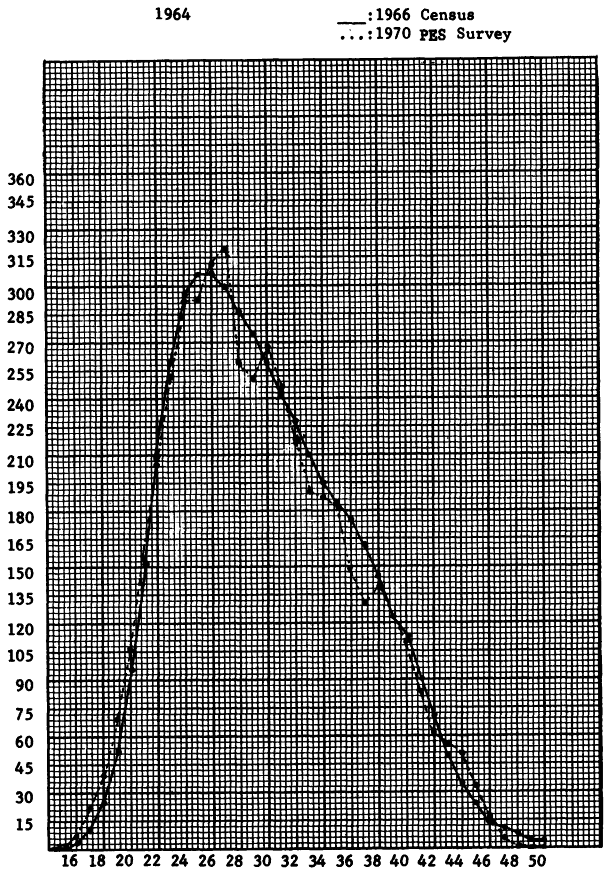


Fig.1 Age-specific fertility rates estimated from the 1966 Census and 1970 post-enimeration survey for Korea: 1961-1966



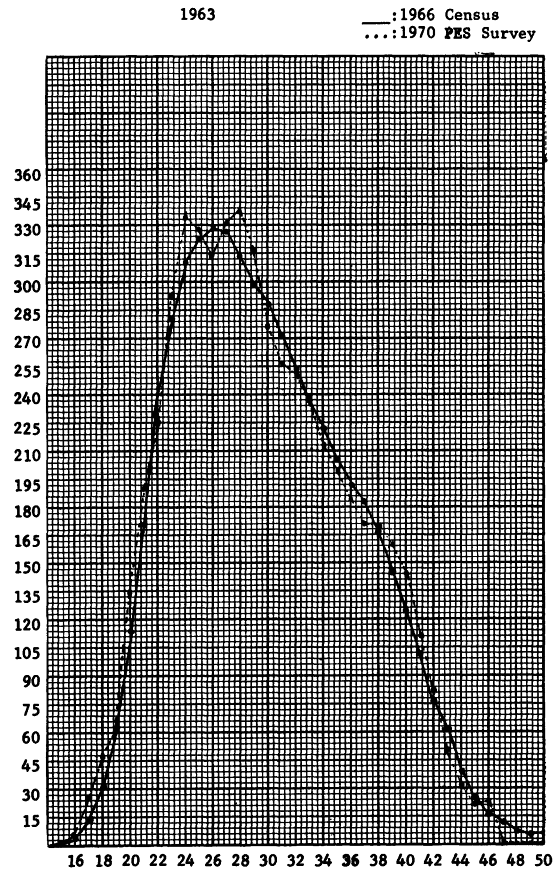
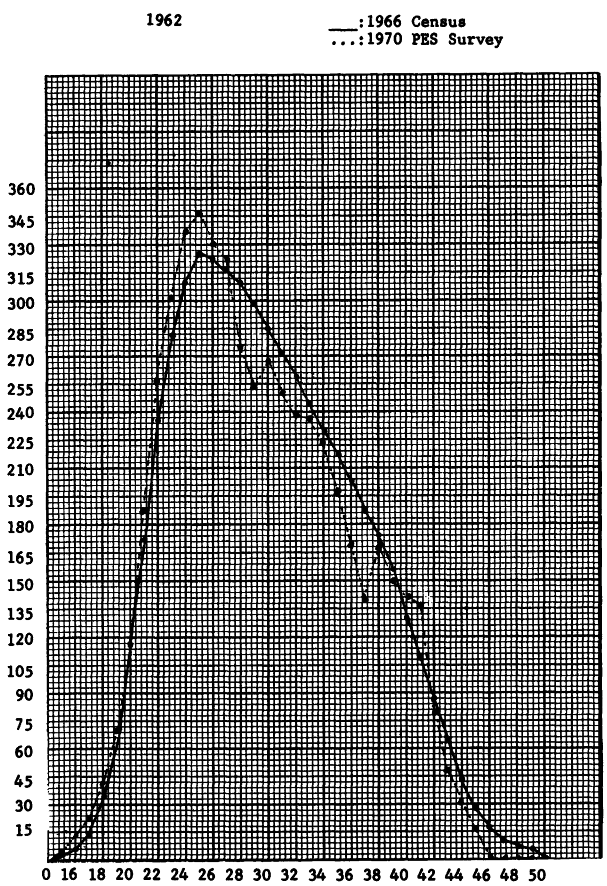


Fig.1 - Cont.



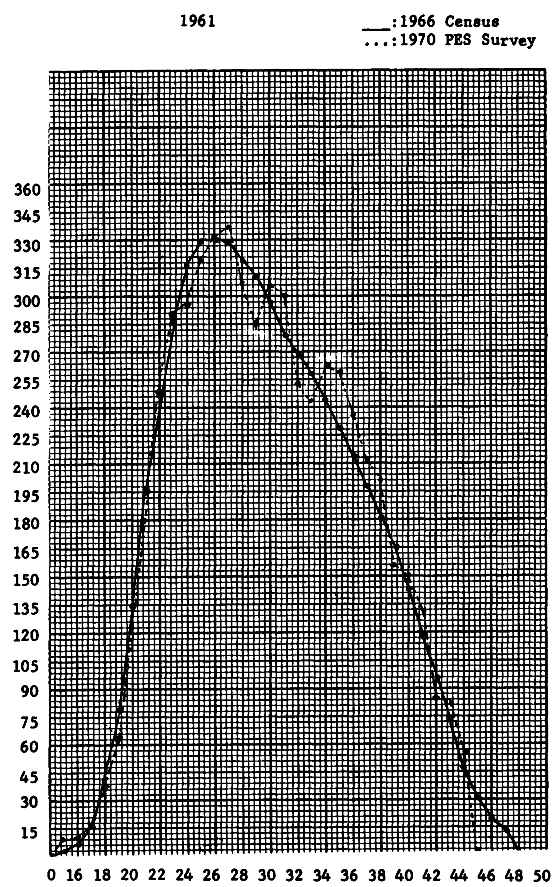


Fig. 1 Cont.

TABLE 1. Own children under 10 years old by exact age and approximate birth cohort of mother: Korea 1966 census

Exact age of women	Approximate birth cohort	Number of women	Own children up to 10 years old by age									
			0	1	2	3	4	5	6	7	8	9
15	1951	25539	7	0	0	0	0	0	0	0	0	0
16	1950	26029	21	7	0	0	0	0	0	0	0	0
17	1949	25219	89	33	9	0	0	0	0	0	0	0
18	1948	24966	285	123	45	11	0	0	0	0	0	0
19	1947	24568	698	327	118	48	26	0	0	0	0	1
20	1946	21170	1075	633	274	151	71	30	0	0	1	0
21	1945	20081	1867	1227	652	357	136	87	45	0	2	0
22	1944	20464	2997	1984	1329	808	366	172	107	58	2	1
23	1943	20979	4070	3158	2308	1658	917	440	240	106	90	0
24	1942	24058	6000	4959	4031	3066	2020	1254	617	314	157	95
25	1941	22945	6283	5669	4932	4209	2869	2033	1228	620	303	166
26	1940	21774	6247	5821	5648	5092	4008	3300	2293	1325	666	266
27	1939	22462	6358	6181	6295	6163	5357	4505	3683	2563	1438	704
28	1938	21263	5758	5793	5981	6145	5705	5204	4447	3505	2359	1306
29	1937	21887	5558	5636	6156	6503	6325	5892	5627	4787	3830	2421
30	1936	20047	4801	4804	5407	6017	5915	5956	5701	5214	4594	3305
31	1935	20409	4403	4633	5234	5989	5835	6003	6311	5858	5560	4161
32	1934	19303	3954	4063	4805	5334	5469	5756	5751	5794	5556	4564
33	1933	17531	3261	3481	3987	4698	4827	5131	5068	5186	5272	4363
34	1932	18761	3260	3419	4063	4790	4939	5308	5487	5451	5595	4846
35	1931	17028	2583	2824	3453	4064	4251	4718	4793	4876	4977	4388
36	1930	16173	2306	2482	2965	3604	3891	4193	4406	4462	4600	4114
37	1929	16975	2240	2427	2970	3553	3888	4158	4464	4535	4709	4179
38	1928	15134	1750	1928	2476	2952	3230	3660	3783	3872	4139	3671
39	1927	15031	1566	1779	2393	2686	3052	3408	3578	3743	3886	3527
40	1926	13897	1305	1436	1943	2386	2664	3006	3292	3323	3463	3268
41	1925	13665	1022	1258	1633	2234	2440	2741	3043	3137	3353	3049
42	1924	13386	808	1016	1445	1901	2176	2541	2844	2864	3178	2870
43	1923	13658	605	835	1249	1714	2081	2348	2659	2768	3068	2707
44	1922	12842	412	612	965	1305	1672	2057	2312	2438	2735	2518
45	1921	12141	239	411	659	1035	1316	1745	1994	2216	2378	2309
46	1920	11713	145	243	427	783	1077	1433	1728	1904	2162	2037
47	1919	9767	93	123	275	446	657	941	1247	1427	1584	1550
48	1918	10062	61	100	166	268	503	813	981	1175	1442	1407
49	1917	10642	53	80	121	187	308	572	811	1069	1277	1355
50	1916	10268	56	51	85	132	216	371	570	792	1023	1059
51	1915	9647	37	42	54	88	109	206	322	508	734	793
52	1914	8824	24	35	38	52	75	124	207	290	450	545
53	1913	9123	37	28	42	49	54	87	149	212	338	420
54	1912	9322	32	35	28	55	48	78	88	135	222	295
unknown	--	--	1911	1824	2176	2571	2787	3306	3613	3898	4595	4478

TABLE 2. Survival ratios for children aged 0-9 years and women aged 15-54 years based on the life tables, and for children only based on the Brass estimate of childhood mortality

i	0	1	2	3	4	5	6	7	8	9
	Part A. Survival ratios for children from birth to age i (L_i/l_o).									
Brass estimate	.9479	.9414	.9321	.9220	.9118	.9041	.8964	.8887	.8810	.8733
Life tables	.9505[1]	.9315[2]	.9244[2]	.9181[2]	.9126[2]	.8761[3]	.8708[3]	.8666[3]	.8635[3]	.8482[4]
Age a	Part B. Survival ratios for women from age a-i+½ to age i ($L_i/L_{a-i+½}$) based on the life tables.									
15...	.99935	.00000	.00000	.00000	.00000	.00000	.00000	.00000	.00000	.00000
16...	.99933	.99757	.00000	.00000	.00000	.00000	.00000	.00000	.00000	.00000
17...	.99917	.99725	.99563	.00000	.00000	.00000	.00000	.00000	.00000	.00000
18...	.99894	.99664	.99487	.99325	.00000	.00000	.00000	.00000	.00000	.00000
19...	.99875	.99605	.99390	.99213	.99052	.00000	.00000	.00000	.00000	.00000
20...	.99869	.99572	.99316	.99102	.98925	.98438	.00000	.00000	.00000	.00000
21...	.99872	.99566	.99285	.99029	.98816	.98315	.98089	.00000	.00000	.00000
22...	.99868	.99561	.99274	.98993	.98738	.98192	.97957	.97732	.00000	.00000
23...	.99868	.99556	.99266	.98979	.98699	.98102	.97832	.97599	.97374	.00000
24...	.99866	.99551	.99257	.98968	.98681	.98044	.97737	.97468	.97235	.96764
25...	.99862	.99537	.99240	.98946	.98658	.97992	.97660	.97354	.97086	.96602
26...	.99858	.99519	.99214	.98918	.98626	.97933	.97588	.97257	.96953	.96422
27...	.99857	.99508	.99191	.98887	.98592	.97869	.97517	.97173	.96843	.96262
28...	.99856	.99500	.99175	.98858	.98555	.97799	.97441	.97091	.96749	.96122
29...	.99855	.99492	.99161	.98837	.98522	.97717	.97356	.97000	.96651	.95991
30...	.99853	.99479	.99143	.98814	.98491	.97626	.97254	.96895	.96541	.95850
31...	.98848	.99457	.99114	.98779	.98451	.97527	.97135	.96765	.96408	.95684
32...	.99843	.99430	.99073	.98731	.98398	.97409	.97001	.96611	.96242	.95481
33...	.99836	.99399	.99025	.98669	.98328	.97262	.96843	.96437	.96049	.95235
34...	.99830	.99367	.98971	.98599	.98245	.97086	.96654	.96238	.95835	.94954
35...	.99826	.99340	.98924	.98531	.98160	.96897	.96449	.96020	.95606	.94659
36...	.99822	.99317	.98882	.98468	.98076	.96702	.96230	.95785	.95359	.94345
37...	.99815	.99288	.98838	.98406	.97994	.96503	.96000	.95531	.95089	.94006
38...	.99808	.99258	.98791	.98343	.97913	.96307	.95767	.95268	.94803	.93641
39...	.99803	.99229	.98742	.98277	.97832	.96115	.95535	.94999	.94504	.93254
40...	.99799	.99201	.98694	.98209	.97747	.95922	.95307	.94732	.94201	.92853
41...	.99795	.99177	.98651	.98147	.97665	.95723	.95079	.94470	.93901	.92444
42...	.99792	.99151	.98609	.98086	.97585	.95517	.94841	.94204	.93600	.92031
43...	.99788	.99124	.98566	.98027	.97507	.95306	.94596	.93927	.93295	.91615
44...	.99784	.99097	.98520	.96965	.97430	.95087	.94343	.93641	.92978	.91190
45...	.99765	.99045	.98451	.97878	.97327	.94855	.94076	.93340	.92646	.90748
46...	.99758	.99001	.98377	.97787	.97218	.94616	.93803	.93033	.92306	.90294
47...	.99787	.99026	.98369	.97750	.97163	.94397	.93547	.92744	.91983	.89841
48...	.99822	.99087	.98437	.97785	.97169	.94202	.93314	.92475	.91681	.89389
49...	.99844	.99131	.98519	.97873	.97224	.94022	.93095	.92217	.91388	.88932
50...	.99835	.99114	.98536	.97928	.97286	.93836	.92863	.91947	.91080	.88454
51...	.99792	.99017	.98437	.97863	.97259	.93617	.92601	.91641	.90737	.87949
52...	.99720	.98851	.98216	.97640	.97071	.93351	.92306	.91305	.90358	.87436
53...	.99632	.98640	.97904	.97275	.96704	.93013	.91954	.90925	.89939	.86909
54...	.99546	.98416	.97548	.96820	.96198	.92598	.91523	.90481	.89469	.86362

[1] 1966 Life Table; [2] 1964 Life Table; [3] 1959 Life Table; [4] 1955-60 Life Table.

TABLE 3. Estimated number of birth cohort of women (defined by age of woman at the time of the census) for 10 years preceding the census

Birth cohort	Age of women at census date	Calendar year 1966	1965	1964	1963	1962	1961	1960	1959	1958	1957
1951	15...	8	0	0	0	0	0	0	0	0	0
1950	16...	23	8	0	0	0	0	0	0	0	0
1949	17...	96	36	10	0	0	0	0	0	0	0
1948	18...	308	134	50	12	0	0	0	0	0	0
1947	19...	753	355	130	54	29	0	0	0	0	1
1946	20...	1160	687	301	168	80	34	0	0	1	0
1945	21...	2014	1333	717	398	154	100	52	0	2	0
1944	22...	3233	2155	1462	901	414	197	124	68	2	1
1943	23...	4391	3430	2538	1848	1036	504	278	124	107	0
1942	24...	6473	5386	4433	3417	2283	1436	715	389	187	115
1941	25...	6779	6157	5423	4691	3243	2328	1423	728	362	201
1940	26...	6740	6322	6211	5675	4530	3779	2657	1555	795	322
1939	27...	2860	6713	6923	6869	6055	5159	4267	3008	1716	853
1938	28...	6212	6291	6578	6849	6448	5959	5153	4114	2814	1583
1937	29...	5996	6121	6770	7248	7149	6747	6520	5619	4570	2934
1936	30...	5180	5217	5946	6706	6685	6821	6606	6120	5482	4005
1935	31...	4750	5032	5756	6675	6595	6874	7313	6876	6634	5043
1934	32...	4266	4413	5284	5945	6181	6592	6664	6801	6629	5531
1933	33...	3518	3780	4385	5236	5456	5876	5872	6087	6291	5288
1932	34...	3517	3713	4468	5339	5582	6078	6358	6398	6676	5873
1931	35...	2787	3067	3797	4230	4805	5403	5554	5723	5939	5318
1930	36...	2488	2696	3261	4017	4398	4802	5105	5237	5489	4986
1929	37...	2417	2636	3266	3960	4394	4762	5172	5323	5619	5065
1928	38...	1888	2094	2723	3290	3651	4191	4383	4545	4939	4449
1927	39...	1690	1932	2632	2994	3449	3903	4146	4393	4637	4274
1926	40...	1408	1560	2137	2659	3011	3442	3814	3900	4132	3961
1925	41...	1103	1366	1796	2490	2758	3139	3526	3682	4001	3695
1924	42...	872	1103	1589	2119	2459	2910	3295	3362	3792	3478
1923	43...	653	907	1374	1910	2352	2689	3081	3249	3661	3281
1922	44...	445	665	1061	1455	1890	2356	2679	2862	3263	3052
1921	45...	258	446	725	1154	1487	1998	2310	2601	2837	2798
1920	46...	156	264	470	873	1217	1641	2002	2235	2580	2469
1919	47...	100	134	302	497	743	1078	1445	1675	1890	1878
1918	48...	66	109	183	299	569	931	1137	1379	1721	1705
1917	49...	57	87	133	208	348	655	940	1255	1524	1642
1916	50...	60	55	94	147	244	425	661	930	1221	1283
1915	51...	40	46	59	98	123	236	373	596	876	961
1914	52...	26	38	42	58	85	142	240	340	537	661
1913	53...	40	30	46	55	61	100	173	249	403	509
1912	54...	35	38	31	61	54	89	102	159	265	358

TABLE 4. Reported female population in the census and estimated mid-year female population by birth cohorts (defined by age of women at the time of the census) for the 10 years preceding the census: Korea, 1966 Census

Birth cohort	Age of women at census	Reported in the 1966 census	Calendar year 1966	1965	1964	1963	1962	1961	1960	1969	1958	1957
1951	15	25539	25556	0	0	0	0	0	0	0	0	0
1950	16	26029	26046	0	0	0	0	0	0	0	0	0
1949	17	25219	25240	25330	0	0	0	0	0	0	0	0
1948	18	24993	25050	25095	25136	0	0	0	0	0	0	0
1947	19	24568	24599	24665	24719	24763	24803	0	0	0	0	0
1946	20	21170	21198	21261	21316	21316	21362	21400	21506	0	0	0
1945	21	20081	20107	20169	20226	20278	20322	20425	20472	0	0	0
1944	22	20464	20491	20554	20614	20672	20726	20841	20891	20939	0	0
1943	23	20979	21007	21073	21134	21195	21256	21385	21444	21495	21545	0
1942	24	24058	24090	24167	24238	24309	24380	24538	24615	24683	24742	24863
1941	25	22945	22977	23052	23121	23189	23257	23415	23495	23569	23634	23752
1940	26	21774	21805	21879	21947	22012	22077	22234	22312	22388	22458	22582
1939	27	22462	22494	22573	22645	22715	22783	22951	23034	23116	23194	23334
1938	28	21263	21294	21370	21440	21509	21575	21742	21821	21900	21978	22121
1937	29	21887	21219	21999	22072	22145	22215	22398	22481	22564	22645	22801
1936	30	20047	20077	20152	20220	20288	20354	20535	20613	20689	20765	20915
1935	31	20409	20440	20520	20591	20661	20730	20927	21011	21091	21169	21130
1934	32	19303	19333	19414	19484	19551	19617	19816	19900	19980	20057	20217
1933	33	17531	17560	17637	17704	17768	17829	18025	18103	18179	18252	18408
1932	34	18761	18793	18881	18956	19028	19096	19324	19411	19494	19576	19758
1931	35	17028	17058	17141	17213	17282	17347	17573	17655	17734	17811	17989
1930	36	16173	16202	16384	16356	16425	16490	16725	16807	16885	16960	17142
1929	37	16975	17007	17097	17175	17250	17323	17590	17682	17769	17852	18057
1928	38	15134	15163	15247	15319	15389	15457	15714	15803	15886	15964	16162
1927	39	15031	15061	15148	15223	15295	15364	15639	15734	15822	15905	16118
1926	40	13897	13925	14009	14081	14150	14217	14488	14581	14670	14753	14967
1925	41	13665	13693	13778	13852	13923	13992	14276	14372	14465	14553	14782
1924	42	13386	13414	13501	13575	13647	13717	14014	14114	14210	14301	14545
1923	43	13658	13687	13779	13857	13933	14007	14331	14438	14541	14640	14908
1922	44	12842	12870	12959	13035	13109	13181	13506	13612	13714	13812	14083
1921	45	12141	12170	12258	12332	12404	12474	12800	12906	13007	13105	13379
1920	46	11713	11741	11831	11906	11978	12048	12380	12487	12590	12689	12972
1929	47	9767	9788	9863	9929	9992	10052	10347	10441	10531	10618	10871
1918	48	10062	10080	10155	10222	10290	10355	10681	10783	10881	10975	11256
1917	49	10642	10659	10735	10802	10873	10946	11319	11431	11540	11645	11966
1916	50	10268	10285	10630	10421	10485	10554	10943	11057	11167	11273	11608
1915	51	9647	9667	9743	9800	9858	9919	10305	10418	10527	10632	10969
1914	52	8824	8849	8927	8984	9037	9090	9453	9560	9664	9766	10092
1913	53	9123	9157	9249	9318	9379	9434	9808	9921	10034	10144	10497
1912	54	9322	9365	9472	9556	9628	9690	10067	10185	10303	10419	10794

TABLE 5. Estimated age specific fertility rates per 1000 women for single year female birth cohorts (defined by age of women at the time of the census) for the 10 calendar years preceding the census

Birth cohort	Age of women at census date	Calendar year 1966	1965	1964	1963	1962	1961	1960	1959	1958	1957
1951	15...	0	0	0	0	0	0	0	0	0	0
1950	16...	1	0	0	0	0	0	0	0	0	0
1949	17...	4	1	0	0	0	0	0	0	0	0
1948	18...	12	5	2	1	0	0	0	0	0	0
1947	19...	31	14	5	2	1	0	0	0	0	0
1946	20...	55	32	14	8	4	2	0	0	0	0
1945	21...	100	66	36	20	8	5	3	0	0	0
1944	22...	158	105	71	44	20	10	6	3	0	0
1943	23...	209	163	120	87	49	24	13	6	5	0
1942	24...	269	223	183	141	94	59	29	15	8	5
1941	25...	295	267	235	202	139	99	61	31	15	9
1940	26...	309	289	283	258	205	170	119	70	35	14
1939	27...	305	297	306	302	267	225	185	130	74	37
1938	28...	292	294	307	318	299	274	236	188	128	72
1937	29...	274	278	307	327	322	301	290	249	202	129
1936	30...	258	259	294	331	328	332	321	296	264	192
1935	31...	232	245	280	323	318	329	348	326	313	236
1934	32...	221	227	271	304	315	333	335	340	331	274
1933	33...	200	214	248	295	306	326	324	335	345	287
1932	34...	187	197	236	281	292	315	328	328	341	297
1931	35...	163	179	221	262	277	307	315	323	333	296
1930	36...	154	166	199	245	267	287	304	310	324	291
1929	37...	142	154	190	230	254	271	293	300	315	281
1928	38...	125	137	178	214	236	267	277	286	309	275
1927	39...	112	128	173	196	225	250	264	278	292	265
1926	40...	101	111	152	188	212	238	262	266	280	265
1925	41...	81	99	130	179	197	220	245	255	275	250
1924	42...	65	82	117	155	179	208	234	237	265	239
1923	43...	48	66	99	137	168	188	213	223	250	220
1922	44...	35	51	81	111	143	174	197	209	236	217
1921	45...	21	36	59	93	119	156	179	200	217	209
1920	46...	13	22	39	73	101	133	160	178	203	190
1919	47...	10	14	31	50	74	104	138	159	178	173
1918	48...	7	11	18	29	55	87	105	127	157	152
1917	49...	5	8	12	19	32	58	82	109	131	137
1916	50...	6	5	9	14	23	39	60	83	108	111
1915	51...	4	5	6	10	12	23	36	57	82	88
1914	52...	3	4	5	6	9	15	25	35	55	65
1913	53...	4	3	5	6	7	10	17	25	40	49
1912	54...	4	4	3	6	6	9	10	15	25	33

TABLE 6: Estimated single-year age-specific fertility rate for Korea: 1957-66

Age of women*	Calendar year 1966	1965	1964	1963	1962	1961	1960	1959	1958	1957
15	1	1	1	1	3	3	4	5	6	7
16	2	3	4	5	6	7	10	10	11	11
17	8	10	10	14	14	17	21	23	25	25
18	22	23	25	32	34	41	45	50	55	54
19	43	49	53	65	71	79	90	100	101	100
20	77	85	96	114	117	135	152	159	165	160
21	129	134	152	171	172	197	211	218	233	214
22	183	193	209	230	236	249	263	272	289	255
23	239	245	259	280	282	288	305	311	322	280
24	282	278	294	310	310	317	334	333	338	292
25	302	293	306	323	325	330	341	338	343	296
26	307	296	307	329	323	330	330	332	337	293
27	298	286	300	327	317	330	326	326	329	286
28	283	269	287	314	311	320	321	317	319	278
29	266	252	275	299	299	311	309	305	312	270
30	245	236	259	288	285	297	298	293	300	265
31	227	221	242	271	272	279	285	282	286	257
32	211	206	228	253	260	269	270	272	278	245
33	194	188	210	237	245	258	263	260	270	230
34	175	172	195	222	230	244	254	246	258	218
35	159	160	184	205	218	229	239	230	243	213
36	148	146	175	192	204	214	223	216	226	200
37	133	132	162	183	188	198	205	204	210	182
38	118	119	141	167	174	181	188	189	191	162
39	107	105	123	146	156	165	170	168	167	144
40	91	90	108	124	131	144	149	143	144	124
41	73	74	90	102	110	118	122	118	120	99
42	56	59	70	83	88	96	94	96	95	77
43	41	44	49	61	64	73	71	70	69	57
44	28	29	35	39	43	48	48	46	47	41
45	17	18	24	24	28	31	31	30	33	0
46	12	12	15	17	18	19	−21	20	0	0
47	8	9	11	12	11	13	14	0	0	0
48	6	7	8	8	8	10	0	0	0	0
49	6	5	5	6	6	0	0	0	0	0
50	5	5	5	6	0	0	0	0	0	0
51	4	4	4	0	0	0	0	0	0	0
52	4	4	0	0	0	0	0	0	0	0
53	4	0	0	0	0	0	0	0	0	0
TOTAL	4511	4462	4921	5462	5559	5838	6006	5980	6121	5335

*Age at the time of birth

TABLE 7. Korea 1957-1966: Final estimates of fertility derived from the 1966 Census

Year	Total Fertility Rate	Age-specific fertility rate per 1000						
		15-19	20-24	25-29	30-34	35-39	40-44	45-49
1957	5342*	39	238	285	243	182	81	-
1958	6101*	40	266	329	279	210	97	-
1959	5941*	38	256	324	272	203	95	-
1960	5955*	35	251	326	275	206	97	-
1961	5786*	31	236	325	270	198	98	-
1962	5579	26	221	315	260	189	90	15
1963	5476	22	220	319	255	179	85	14
1964	4940	17	203	296	228	159	73	13
1965	4484	16	191	280	206	134	60	10
1966	4531	14	186	292	212	134	59	10

*Excludes the fertility rate of women 45-49 years of age.

TABLE 8. Estimated Total Fertility Rates[a] Based on the 1966 Census and the 1968 Fertility and Family Planning Survey, Korea: 1959-1966

Year	1966 Census	1968 Survey	Percent difference
1959	5,920	5,993	-1.2
1960	6,036	5,897	+2.3
1961	5,930	5,706	+3.9
1962	5,497	5,311	+3.5
1963	5,444	5,451	-0.1
1964	4,934	4,946	-0.2
1965	4,489	4,653	-3.5
1966	4,532	4,519	+0.3

[a]For the purpose of comparison, the total fertility rates are confined to the age range covered by the Fertility and Family Planning Survey; for example, the 1959 total fertility rate covers the age range from 15 to 39.

PGE STUDIES: COSTS AND EFFECTIVENESS

W. Seltzer, The Population Council

1. The PGE technique

The Population Growth Estimation (PGE) technique has been employed in many countries either to estimate fertility, mortality, and natural increase where these key demographic variables are not reliably known (e.g., India, Liberia, Pakistan, Thailand, and Turkey) or to estimate the completeness of the civil registration system (e.g., Canada, Chile, Commonwealth West Indies, Tunisia, USSR, and the United States).

Nearly all of the work on the PGE technique has proceeded as if it were a technique unique to demography. Actually, an identical procedure has been used in a number of studies of response error in a variety of fields. Moreover, it is also closely related to the "capture-tag-recapture" technique used in estimating the abundance of animal populations. Note that in the capture-tag-recapture technique one estimates population size. In PGE studies we estimate live births and deaths -- the two components of natural increase -- while population size is usually measured by some form of direct enumeration.1/ Even more broadly, the PGE procedure is related to any situation where a statistic, known to be incomplete, is adjusted by a ratio reflecting the estimated completeness of the statistic.

As used in vital statistics estimation, the PGE technique involves (1) collection of reports of vital events by two quasi-independent data-gathering procedures; (2) case-by-case matching of the reports from these two systems; and (3) the preparation of an estimate of the number of events adjusted for omissions or of the relative completeness of either system on the basis of the obtained matched rates.2/

Specifically, $\hat{N}$, the PGE estimate of the total number of births or deaths occurring in a given area during a given time period, can be expressed as

$$\hat{N} = \frac{N_1 N_2}{M} \qquad [1a]$$

or

$$\hat{N} = M + U_1 + U_2 + \frac{U_1 U_2}{M} , \qquad [1b]$$

where

$N_1 = M + U_1 =$ total number of births or deaths in source 1;

$N_2 = M + U_2 =$ total number of source 2 reports;

$M =$ number of reports in each source identified (by some matching procedure) as referring to the same event, i.e., the number of matches;

$U_1 =$ number of source 1 reports identified as referring to events not reported in source 2; and

$U_2 =$ number of source 2 reports identified as referring to events not reported in source 1.

Equations 1a and 1b yield identical estimates of N. The latter equation was given by Chandrasekaran and Deming in their 1949 article [1] which recommended the use of this approach in countries lacking adequate vital registration data.

One may also form an estimate of the completeness of either source 1 or source 2. Using equation 1a, we have for the estimated completeness of source 1

$$\hat{W}_1 = \frac{N_1}{\hat{N}} = \frac{M}{N_2} . \qquad [2]$$

Equation 2 has been used in a number of countries, starting with Canada in 1931, to estimate the completeness of civil registration.

As presented in equations 1 and 2 these estimates refer to sources which attempt to gather all the events occurring in the area under study. For example, infants enumerated in a census may be searched for in a civil registration system and the completeness of the latter may be estimated using equation 2. However, one may also restrict the collection of reports to a probability sample in either or both systems. The only restriction to the use of sampling unique to PGE estimation is that the samples used in each source must be identical or one must be a subsample of the other. If sampling is employed, equations given earlier can be rewritten in terms of the sample estimates. The expected value of the estimate is not changed by the use of sampling, i.e., $E(\hat{N}) = E(\hat{n})$, but its variance is usually larger.3/

Furthermore, these equations are quite general as to the method of data collection to be used. In practice, because the PGE estimate assumes that the probability of an event reported by one source is independent of its being reported in the other source, one tries to employ two methods of data collection as dissimilar as possible. This frequently leads to the use of some type of registration approach in one source, i.e., an effort to obtain reports of events as they occur, and some form of survey approach in the other source. The survey approach collects reports of events either by asking about events retrospectively or, in multi-round surveys, by obtaining a partial count of events by accounting for changes in the household composition recorded in consecutive survey enumerations.4/

This, in brief, is the PGE technique. Before trying to examine its strengths and limitations, let us look quickly at some of the possible alternatives.

2. Alternative approaches

Faced with the problem of using and interpreting those statistics of demographic change which, in Morgernstern's language, "simply accrue without any overall design or plan" [8] -- for example, most census or civil registration data -- demographers have tried three basic approaches: they made use of the available statistics, they tried to improve the methods of data collection, or they introduced new techniques for the analysis of data.

The first of these alternatives -- the uncritical use of whatever statistics are available -- continues to be a source of mininformation to the policy maker and confusion to the social scientist. Certainly policy makers and social scientists will not stop their activities until "good" data become available. Statisticians should respond to data requests from such sources as constructively as possible, attempting to guide the uninformed user so as to avoid the pitfalls of flawed data. However, not all of the requests for poor data come from uninformed users. The practice of various international statistical offices of requesting from each country, and then publishing, a single estimate for a long list of fertility and mortality variables is a powerful force influencing both producers and consumers of demographic statistics to act as if all such estimates were of equal substantive value.

Improvements in data collection techniques and methods of data analysis may be interrelated. However, reliance has often been placed on either alone to do the job. Thus, efforts are often made to improve questionnaire wording, training, and supervision so as to improve demographic estimates in censuses or household surveys. (The enormity of the task of improving civil registration has usually discouraged any attempts to improve this source.) In addition, more radical approaches have been tried: dual collection, pregnancy history, chemical pregnancy tests, randomized response, Sirkin's multiple respondent approach, etc.

Approaches involving improved data analysis -- which I shall refer to collectively as demographic analysis -- attempt to adjust deficient data on the basis of assumptions either about the nature of the population being studied (e.g., stable population analysis), or about the regularity of reporting errors (e.g., Brass fertility estimates, Som's recall lapse adjustments, and the Grabill-Cho method of estimating fertility from census data on own children), or both (e.g., Brass childhood mortality estimates). Some of these estimates are described in some detail in U.N. Manual IV [12] and in the papers by Page, Cho, and Zachariah presented at this session.

3. Choosing between alternatives

It is clear that a number of different methods for measuring demographic change exist. A decision to use one particular method should be based on a rational review of the alternative methods. This review should be based on a close look at the actual data needs and the resources available to carry out the measurements, rather than on one or more abstract imperatives. Three factors are needed to keep the review from being merely a formal exercise buttressing our methodological prejudices. They are: (1) comparable experience in the use of alternative methods; (2) specification of the measurement problem and the available resources; and (3) specification of one or more standards by which the choice is to be made.

Full knowledge about previous experience is required if we are to avoid past mistakes and benefit from past successes. The need for specification is of critical importance if we are to go beyond our preconceived notions. In brief, then, the need is for appropriate specification of the measurement problem and relevant knowledge about the means for its solution.

At this point let me specify five criteria for assessing the adequacy of basic demographic estimates such as the population growth rate, the crude birth rate, and the crude death rate: (1) accuracy, (2) timeliness, (3) detail, (4) user confidence, and (5) the cost of producing the estimate.

Also, let me quickly add that we do not have the techniques or the experience to apply these criteria to the alternative estimation procedures and come up with an unambiguous answer as to which procedure is preferable in a specific case. However, it is helpful to review what we do know, or think we know, about the various procedures available to us in terms of these criteria. Such a review may also help to define the criteria somewhat.

4. Costs of production

One of the truisms about PGE estimates is that they are more expensive to produce than single-system estimates. Let us leave aside for a moment that cost in the abstract, without reference to value received (in this case some mix of accuracy, confidence, detail, and timeliness), has little meaning, and compare the budget of a PGE study with that of a household survey.

Unfortunately, the analysis of cost information about data collection and analysis is an undeveloped science. The problem is threefold: (a) lack of interest, since the size of budget and the type study are often decided independently of each other; (b) lack of information on study costs, particularly for studies conducted in the developing world; and (c) the complexity of any equation that attempts to describe fully the costs of various components of the data production process.5/

As soon as one tries to list all the factors that can affect study costs one quickly becomes discouraged by the length and varied nature of such a list. After a largely fruitless search of the literature on this topic for a scheme, or a methodology, or a notation, or a something that would both simplify the cost picture and yet preserve those features needed to contrast the costs of a single system measurement effort with those of a PGE study, I gave up.6/ All of this is by way of introduction to the inelegance of table 1 which is an unimaginative listing of all the types of activities that go into a PGE and a single-system study. Despite its awkward notation, I think the cost picture revealed by the table is helpful. If nothing else, it may stimulate others to do better.7/

First let me explain the notation used in table 1. Annual aggregate costs are indicated by upper-case "C's", while lower-case "c's" are used for unit costs. The first-level subscripts α, β, γ, and δ refer, respectively, to study activities associated with data collection, vital events processing, base population processing, and presentation of study results. The second-level subscripts refer to a particular phase within one of the four broad types of activities designated by the first-level subscripts. The prime symbol is used to distinguish costs associated with the second collection source in a PGE study from those associated with the first source. Similarly, p and p' refer to the proportion of vital event reports sent for field investigation in source 1 and source 2, respectively. The letter m refers to the number of clusters in the sample, n to the total number of persons in the sample, and $\bar{n}$ to the mean number of persons per cluster. Assuming a crude birth rate of about 50 per 1,000 and a crude death rate of about 20 per 1,000, that some events go unreported, and that some out-of-scope events are reported, the maximum number of events reported by each system is .07n (i.e., 50 per 1,000 plus 20 per 1,000).

The cost equations in table 1 give the total cost of each type of activity (i.e., C_α, C_β, C_γ, and C_δ) in terms of the sample size (n), or the sample design (n and m), and the appropriate unit costs. Cost equations of this form permit one to assess directly the efficiency of a given sample design in terms of sampling error. Unfortunately, the equations in table 1 make no explicit recognition of differences between studies attributable to the timeliness, detail, or the accuracy of the estimates. Differences between studies involving these factors may be reflected indirectly in the unit costs of various phases of the study, as well as in the choice between single or dual collection. From the rightmost column of table 1, one can see that cost differences are limited to data collection activities and certain phases of vital events processing. Clearly, the extent to which extra collection costs are associated with dual collection will depend on the costs of the two data gathering procedures used in the PGE study and the single-system collection procedure used as a standard. To facilitate cost comparisons we assume that the first source in the dual collection system is identical to the single system source and that the sample design remains constant. Thus, all the added collection costs are associated with the second source.

Theoretically the data collection costs of the second source C'_α can take on any value; but, in practice, C'_α is usually less than C_α so that the ratio $1 + C'_\alpha/C_\alpha$ is almost always less than 2. For example, the second source may be the civil registration system, so that collection costs for the second source need cover only the cost of office sampling and, possibly, the transcription of records. Alternatively, as in the Turkish Demographic Survey, the second source also serves as the supervisory control for the first source. At a minimum, one of the sources may use an already existing infrastructure of statistical administration (e.g., the same regional field offices). To my knowledge no PGE study yet undertaken has used two fully funded, new data collection sources.

Another major determinant of data collection costs, given that source 1 is a household survey, is the frequency of survey rounds. If it is assumed that the aggregate annual data collection costs of source 1 in table 1, C_α, refers to a one-round household survey and that r_s survey rounds are carried out annually in the single-system survey then the total annual cost for data collection activities in the single-system survey is $r_s C_\alpha$. Assuming r_p rounds are employed in the comparable survey conducted in a PGE study, then the data collection costs for this survey are

Table 1 - Comparison of Cost Components for a Single-system Study and a PGE Study

[Upper-case 'C' refers to an annual aggregate cost, lower-case 'c' to a unit cost; for assumptions used and details of notation, see footnotes and text.]

Type of activity1/	Single-system costs	PGE study costs	Ratio of PGE total costs to single-system costs
Data Collection -- total2/	$C_\alpha = m(c_{\alpha_1} + \bar{n}c_{\alpha_2})$	$C_\alpha + C'_\alpha$	$1 + (C'_\alpha/C_\alpha)$
(1) Related to number of clusters	mc_{α_1}	$m(c_{\alpha_1} + c'_{\alpha_1})$	$1 + (c'_{\alpha_1}/c_{\alpha_1})$
(2) Related to number of elements	$m\bar{n}c_{\alpha_2}$	$m\bar{n}(c_{\alpha_2} + c'_{\alpha_2})$	$1 + (c'_{\alpha_2}/c_{\alpha_2})$
Vital events processing -- total3/	$C_\beta = .07n(c_{\beta_4} + c_{\beta_5})$	7/	8/
(1) Prematching phase4/	0	$.07n(c_{\beta_1} + c'_{\beta_1})$	∞
(2) Matching phase	0	$.07n(c_{\beta_2} + c'_{\beta_2})$	∞
(3) Field follow-up5/	0	$.07n(pc_{\beta_3} + p'c'_{\beta_3})$	∞
(4) Pretabulation phase6/	$.07nc_{\beta_4}$	$.07nc_{\beta_4}$	1
(5) Tabulation phase	$.07nc_{\beta_5}$	$.07nc_{\beta_5}$	1
Base population processing -- total	$C_\gamma = n(c_{\gamma_1} + c_{\gamma_2})$	C_γ	1
(1) Pretabulation phase6/	nc_{γ_1}	nc_{γ_1}	1
(2) Tabulation phase	nc_{γ_2}	nc_{γ_2}	1
Presentation of results -- total	C_δ	C_δ	1
All activities -- total	$C = C_\alpha + C_\beta + C_\gamma + C_\delta$	$C = C_\alpha + C'_\alpha + C_\beta + C_\gamma + C_\delta$	9/

1/ Includes stated activity, plus proportional share of costs of supervision and overheads. The cost analysis presented here makes no explicit recognition of expenditures incurred to increase the accuracy, timeliness, or detail of the estimates.

2/ It is assumed that collection costs are related to the number of clusters (m) and their mean size ($\bar{n}$), rather than to the level of vital events. For simplicity, all collection costs are treated as if they were solely linear functions of m and $\bar{n}$.

3/ It is assumed that vital event processing costs are related to the number of birth and death reports obtained (i.e., a maximum of .07n).

4/ Editing and other processing necessary to make the documents ready to matching. The matching is assumed to be done manually.

5/ As a first approximation field follow-up costs are assumed to be related to the proportion of reports from each source (p or p′) sent for follow-up. Follow-up costs will usually be described more accurately by $mc_{\alpha_1} + .07n(p + p')c_{\alpha_2}$.

6/ Editing, coding, and punching required for purposes of tabulation.

7/ $C_\beta = .07n\ (c_{\beta_1} + c'_{\beta_1} + c_{\beta_2} + c'_{\beta_2} + pc_{\beta_3} + p'c'_{\beta_3} + c_{\beta_4} + c_{\beta_5})$.

8/ Ratio $= 1 + \dfrac{c_{\beta_1} + c'_{\beta_1} + c_{\beta_2} + c'_{\beta_2} + pc_{\beta_3} + p'c'_{\beta_3}}{c_{\beta_4} + c_{\beta_5}}$.

9/ Ratio $= \dfrac{m(c_{\alpha_1} + c'_{\alpha_1}) + n(c_{\alpha_2} + c'_{\alpha_2}) + .07n(c_{\beta_1} + c'_{\beta_1} + c_{\beta_2} + c'_{\beta_2} + pc_{\beta_3} + p'c'_{\beta_3}) + n(c_{\gamma_1} + c_{\gamma_2}) + c_\delta}{mc_{\alpha_1} + nc_{\alpha_2} + .07n(c_{\beta_4} + c_{\beta_5}) + n(c_{\gamma_1} + c_{\gamma_2}) + C_\delta}$.

$r_p C_\alpha$.[8] It follows that ratio of total costs for a PGE study to that for a single-system survey is approximately

$$\frac{r_p}{r_s} + \frac{C'_\alpha}{r_s C_\alpha} \qquad [3]$$

or, if C_α is assumed to equal C'_α,

$$\frac{r_p + 1}{r_s} \; . \qquad [4]$$

In other words, if more than three survey rounds are contemplated per year (i.e., $r_s > 3$) it is possible that meaningful savings in data collection costs can be achieved by reducing the number of survey rounds and employing an appropriate dual collection procedure.

The expense associated with matching and field follow-up is unique to a PGE study. Unless the unit costs of these operations are very high, the relatively small numbers of cases involved, .07n and (p + p′).07n, suggest that the aggregate costs of these operations are only moderate compared to the total cost of a multi-round survey. Nevertheless, there is an urgent need for additional data on the costs of matching and field follow-up.

To indicate some idea of the range of costs involved in demographic field studies in the developing world let me cite two figures. The annual cost of the Pakistan PGE study came to $6.50 per household. This estimate is based on aggregate cost data covering all aspects of the study and all sources of funding. In the Pakistan study data collection continued for four years and the sample involved some 20,000 households so that the impact of necessary overheads on the annual average is not large. The second figure comes from another country in the developing world where the cost of a 27,000 household multi-round demographic survey extending over two years came to $16 per household per year. This costs estimate does not make any provision for the costs of tabulation or the presentation of results. Another major difference between these two cost figures is that personnel costs were quite low in Pakistan relative to those in the country in which the multi-round survey was conducted.

5. User confidence

In our real world of uncertainties and mistakes, user-decisions on confidence involve processes that are far from being either rational or accurate. In fact, there are many instances where the most "rational" procedures for determining the confidence to be placed in a particular estimate are not the most accurate.

Certainly the establishment of a confidence interval around a crude birth rate estimate in order to reflect the uncertainties introduced by sampling is an objective and rational procedure. However, most demographers would consider that the 1962 crude birth rate for rural India lay well outside the range of 32.6 to 36.6 per 1,000, even though this is the 2σ confidence interval of this estimate from the Indian National Sample Survey [9]. The point here is the simple one that the probabilistically determined consequences of random errors are not the only factors which should affect the confidence we place in any demographic estimate.

Though an ideal procedure for ascertaining confidence is well beyond us, we might try to approximate some measure of this concept in terms of (a) the likely accuracy of a statistic given the various types of errors it may be subject to, and (b) the likelihood of these errors occurring. I realize that even this somewhat loose and limited goal will be hard to achieve. Quite often it is the figure with no error statement attached on which the user places his greatest confidence. However, even if the policy makers are slow to heed technically sound assessments of data quality our fellow scientists should not be.

6. Detail and timeliness

I shall touch only briefly on the detail and the timeliness criteria as the paper by Louwes [5], deals extensively with these two factors. Details may refer either to the types of variables covered (e.g., crude rates, characteristic-specific rates, life table rates), or to the extent to which estimates are made for various analytical or geographic subgroups. The effect of increased detail on the cost equations of table 1 varies somewhat with the type of detail under consideration. For example, added geographical detail generally will necessitate an increase in the number of clusters (m) and hence data collection costs, as well as increasing tabulation and data presentation costs.[9] On the other hand, increased analytical detail (e.g., obtaining estimates of fertility) often will have only a marginal impact on collection costs. However, beyond a certain point increased analytical detail can also affect the cost and quality of the collection operation.

While timeliness is usually thought of in terms of speed of production, frequency and regularity are also factors involved in the concept of timeliness. In general, there is a reciprocal relationship between speed and detail as well as speed and accuracy. On the other hand, frequency of data collection, and to a lesser extent regularity, tend to be directly related to accuracy. Similarly,

Table 2 - Mean and Range of Approximate Intraclass Correlation Coefficients for Crude Birth and Death Rates, by Type of Cluster, for Six Specified Studies: 1950-66.

[For full qualifications, see sources cited. Values of δ are approximate and are rounded to 3 places.]

Type of cluster	Number of domains 1/	Mean population per cluster 2/	Crude birth rate			Crude death rate		
			Mean δ 3/	Range of δ		Mean δ 3/	Range of δ	
				Low	High		Low	High
All Types	46	572	+.002	-.001	+.008	+.003	-.001	+.013
Region and country								
Africa	33	333	+.002	-.001	+.008	+.003	-.001	+.013
Cameroon, 1960-65	23	356	+.001	-.001	+.005	+.002	-.001	+.010
Chad, 1964	7	300	+.001	-.001	+.005	+.005	+.000	+.013
Nigeria, 1965-66	3	235	+.005	+.004	+.008	+.005	+.002	+.012
Asia	13	1,179	+.002	+.000	+.005	+.002	-.001	+.006
India, 1950-52	4	455	+.002	+.000	+.004	+.004	+.000	+.006
Pakistan, 1964-65	2 4/	5,000	+.002	+.001	+.002	+.001	+.001	+.002
Turkey, 1965-66	7	501	+.003	+.001	+.005	+.001	-.001	+.002
Type of residence								
Urban	6	498	+.002	+.001	+.005	+.001	-.001	+.005
Rural	38	351	+.002	-.001	+.008	+.003	-.001	+.013
Mixed	2	5,000	+.002	+.001	+.002	+.001	+.001	+.002
Cluster size 5/								
Under 300	11	275	+.001	-.001	+.008	+.005	+.001	+.013
300-349	19	323	+.002	-.001	+.005	+.003	-.001	+.009
350-649	12	457	+.003	+.000	+.005	+.001	-.001	+.010
650 and over	4	2,917	+.002	+.001	+.002	+.002	+.001	+.005

1/ A domain is a group of clusters for which the intraclass correlation coefficient is separately available. Domains often correspond to sample strata.

2/ Mean of average cluster size reported for each domain in original source. Cluster size shown is that prior to additional within-cluster sampling, if any.

3/ Mean of unrounded intraclass correlation coefficients for specified number of domains.

4/ Each province is treated as a domain, with intraclass correlation coefficient based on the average survey and registration values for 1964 and 1965.

5/ Reported mean population per cluster of each domain.

Sources:

(a) Cameroon:

Scott, Christopher, "Vital Rate Surveys in Tropical Africa," in The Population of Tropical Africa, edited by J. Caldwell and C. Okonjo, London, 1968, Chapter 15, table 1, pages 164-165.

(b) Chad and Nigeria:

Scott, Christopher and J.B. Coker, "Sample Design in Space and in Time for Vital Rate Surveys in Africa," paper presented at the International Union for the Scientific Study of Population, London, 1969, tables 1 and 2. For Nigeria, estimates are based on artificially constructed clusters of 50 consecutive household questionnaires completed by the same interviewer.

(c) India, Pakistan, and Turkey:

Intraclass correlation coefficients calculated from published variance estimates from the Mysore Population Study, the PGE Experiment in Pakistan, and the Turkish Demographic Survey, using Scott and Coker's binomial approximation.

the more detailed statistics one has available from a study the more confident one can usually be about assessing its quality.

7. Accuracy

Accuracy as used here is a synonym for data quality and is measured in terms of the difference between an estimate and the value one is trying to estimate. Defined in this way the accuracy of an estimate is affected by both random and nonrandom errors, whether arising in the collection, processing, estimation, or presentation process.

The special sources of error unique to the PGE technique are:

1. lack of independence between the two collection procedures which, except in rare circumstances, can lead to an underestimate of the number of events;
2. use of matching criteria which fail to distinguish between reports referring to different vital events, leading to erroneous matches and an underestimate of the number of events; and
3. use of data in the matching process containing reporting or recording errors so that reports referring to the same event are not linked, resulting in erroneous nonmatches and an overestimate of the number of events.

In addition, the existence of out-of-scope reports in one or both sources or the use of a deficient estimate of the base population can lead to an upward bias of the PGE vital rate estimates. However, these two sources of error also affect most types of single-system estimates. All of these sources of error in the PGE estimate are discussed in much greater depth in Seltzer and Adlakha [11], Marks [6], and Marks et al. [7].

The principal advantage of the PGE technique is that the PGE estimate is largely unaffected by the errors and the uncertainties encountered in the collection phase of many single-system surveys and registration systems. Whether a single system is used to provide a vital rate estimate directly or is used as a source of data for demographic analysis, the amount of information available about the population being studied is limited to that obtained from the single source. Dual collection and matching by its very nature provides more information than is available from a single source. While the amount of information from a single source may be stretched by the use of suitable assumptions, the accuracy of the estimates made are then subject to both data errors and errors arising from the failure to meet the assumptions made.

In addition to the sources of error listed above, most PGE estimates are also subject to sampling variability.10/ Estimates of sampling error from the Pakistan PGE experiment indicate coefficients of variation between 4 and 9 percent for the annual crude birth rate estimate, with values approximately twice this size for the crude death rate. Because PGE studies often involve some type of registration of vital events, clustering is often more pronounced than is traditionally employed in survey sampling. Clearly, we pay a price for this clustering as the estimates of intraclass correlation coefficients presented in table 2 indicate.

Even with a δ as small as those shown in table 2, very large clusters will have a major impact on sampling variance; but sampling error is only one component of our accuracy criteria. Indeed, one might hypothesize that, in general, the smaller the cluster for a fixed budget and fixed total sample size, the larger would be nonsampling error. In other words, one supposes that a given supervisory effort is spread more thinly when the sample is based on a large number of small clusters than when an equal-size sample is composed of a smaller number of larger clusters.

I know of no data available to test this hypothesis directly and even if the relationship is established in one instance, there is no guaranty that any observed relationship between cluster size and nonsampling error will remain constant from study to study. Nevertheless, the design implications are important enough that some effort to test this hypothesis should be made.

I do not doubt that an intensive, well-run, single-visit retrospective survey can come up with high-quality demographic estimates. The problem is how can we rely on its accuracy in any given instance? The evidence is not encouraging. One study found a median 33 percent undercount of the number of births reported in one-time retrospective surveys relative to that of comparable dual collection estimates [10].

Finally, in attempting to improve the accuracy of any collection procedure we are caught in the dilemma of how much effort to spend per household to better the quality of data collection versus how many households should be sampled; that is, the choice of allocating limited resources to reduce nonsampling or sampling errors. Basically, dual collection provides a highly effective means of spending more per household so as to concentrate on the reduction of nonsampling errors.

8. Summary: effectiveness

My objective in this paper has been to compare the relative effectiveness of the PGE technique with that of some alternative procedures for obtaining estimates of basic demographic variables. Based on present knowledge, it is not possible to construct utility functions to this end that are both meaningful and rigorous.11/ Nevertheless, by specifying five factors (i.e., accuracy, timeliness, detail, user confidence, and cost) that might ultimately compose such a utility function and by examining alternative estimates in light of these factors I believe the goal of rigor has been advanced somewhat. At the same time, the introduction of experience from actual studies has kept the discussion from wandering too far from reality.

Unless the purposes for which demographic estimates are prepared are also adequately specified the concept of effectiveness has limited meaning no matter how rigorously this concept is defined. The question of proposed uses of demographic data has not been dealt with explicitly in this paper. However, the range of possible uses for demographic estimates is broad enough to guarantee that no technique can be termed, "universally most effective."

Given this general limitation, the findings of this paper can be summarized in terms of the five criteria of effectiveness as follows:

1. Accuracy -- The PGE technique is as good as the best of the alternatives; nevertheless the precision of our measurement techniques is such that small year-to-year changes in fertility and mortality can not be measured accurately in countries without an effective civil registration system.

2. Detail -- The PGE technique is as good as the best of the alternatives and, except in the case of historical data, provides more extensive detail than demographic analysis.

3. Timeliness -- The PGE technique will usually provide estimates more slowly than a one-time retrospective survey, more quickly than the 6 to 7 year lag between the mid-decade reference point of an intercensal growth rate and the date that such a growth rate becomes available subsequent to the census, and at about the same time that a good-sized multi-round household survey produces comparable estimates.

4. User confidence -- The informed user will find the PGE technique far ahead of the other available alternatives with respect to the degree of confidence that can be placed in the estimates. The fact that the PGE technique provides a built-in self-evaluation device -- through dual collection and matching -- does not guarantee that PGE estimates will be correct, or that the user will realize that any given estimates are quite incorrect. There is with the PGE technique, however, a much greater likelihood of realizing something has gone wrong, if it has, as well as of producing estimates that are moderately robust to the variations in the quality of data collection. This is an important point because sharp variations in the quality of field work are frequently encountered where data collection experience is limited or where field conditions are particularly difficult.

5. Costs -- In terms of cash outlay the PGE technique is usually, though not necessarily, more expensive than alternative approaches using comparable-sized samples. Whether the additional cost is justified depends upon the uses to which the estimates will be put. However, the difference in cost between a multi-round survey and a PGE study is generally not that much, so that whenever a multi-round demographic survey is contemplated, very serious consideration should be given to conducting a PGE study.

In order that we proceed beyond the tentative formulations of this paper we will need additional data on the costs, the accuracy, and the uses of various types of demographic estimates. Therefore, I would like to close with the request that statisticians concerned with demographic measurement increasingly turn their attention to identifying the costs, accuracy, and ultimate uses of the estimates they produce. In this request I am merely echoing some of the recommendations made two years ago by Ross Eckler [4] in his Presidential Address to this association.

FOOTNOTES

1/ However, see [3].

2/ Estimating equations for use in a three-source PGE study have been given by Deming and Keyfitz [3] and, independently, for a k-source study by Das Gupta [2]. Because it will usually be more efficient to improve the quality and independence of two collection systems than to attempt to use a third collection procedure, this paper is confined to an examination of PGE studies using only two sources.

3/ If the completeness of a source using no sampling is very poor (say, less than 50 percent) it may be desirable from a variance viewpoint to use a source with sampling and higher completeness.

4/ Unfortunately, not all events can be identified by examining changes from survey round to round in the list of persons enumerated in the household. Particularly in countries where infant mortality is high, migration rates are high, or women spend long periods of time at their parents home after childbirth, reports about a considerable number of events can only be obtained retrospectively.

5/ The classical approach of allocating collection costs between those associated with the number of clusters and those associated with the number of elements does not really help to assess designs which have major differences subsequent to the data collection phase.

6/ The major exception to this bleak picture is a paper by Louwes [5]. While Louwes' paper deals with agricultural surveys in the European common market countries and is thus not directly relevant to the problem at hand, it does suggest a number of promising leads, one or two of which are used in this paper.

7/ The basic monthly salary of the survey interviewer, or its hourly equivalent, seems to be a very promising standard unit from which comprehensive cost function can be built, thus permitting the kind of cost comparison suggested in this section.

8/ This is almost certainly an overestimate of the costs of a multi-round survey in that it assumes that data collection activities involve only recurring costs. However, the effect of this overestimate on the cost comparison of a PGE study with a single-system survey may be at least partially offset by assuming $C_\alpha = C'_\alpha$ as in equation 4.

9/ Of course, if sampling is not involved -- as is the case with a census or a national civil registration system -- no additional data collection costs are associated with increased geographical detail.

10/ PGE studies and single-system estimates based on retrospective survey questions are not alone subject to sampling errors. It is often unrecognized that many forms of demographic analysis are also subject to the effect of sampling errors. For example, the parameters used to enter stable population tables may be subject to sampling variability, implying a range of possible stable population estimates.

11/ In this formulation, utility is considered to be a joint function of the uses to which the estimates will be put and the effectiveness (in terms of the five criteria described in the paper) of the collection and estimation effort.

REFERENCES

[1] Chandrasekaran, C. and Deming, W.E., "On a Method of Estimating Birth and Death Rates and the Extent of Registration", Journal of the American Statistical Association, Vol. 44, No. 245, 1949, pp. 101-115.

[2] Das Gupta, P., "On the Estimation of the Total Number of Events and of the Probabilities of Detecting an Event from Information Supplied by Several Agencies", Calcutta Statistical Association Bulletin, Vol. 13, No. 49 and 50, 1964. pp. 89-100.

[3] Deming, W.E. and Keyfitz, N., "Theory of Surveys to Estimate Total Population" in Proceedings of the World Population Conference, Belgrade,... 1965 -- Volume III, United Nations, New York: 1967, pp. 141-144.

[4] Eckler, A. Ross, "Statisticians and Shoemakers ('Who is Worse Shod than the Shoemaker's Wife' from Heywood's Proverbs)", Journal of the American Statistical Association, Vol. 65, No. 329, 1970, pp. 9-21.

[5] Louwes, S.L., "Cost Allocations in Agricultural Surveys", Review of the International Statistical Institute, Vol. 35, No. 3, 1967, pp. 264-290.

[6] Marks, Eli S., "Problems in Estimating Population Growth in Developing Areas", paper presented at the seminar on Statistical Problems in Population Research, The East-West Center, Honolulu, August, 1971.

[7] Marks, Eli S., Krotki, K., and Seltzer, W., Population Growth Estimation: A Handbook of Vital Statistics Measurement, The Population Council, New York, forthcoming.

[8] Morgernstern, Oskar, On the Accuracy of Economic Observations, Second Edition, Completely Revised, Princeton University Press, Princeton, 1963 (Princeton Paperback Edition, 1965).

[9] Ramabhadran, V.K., "Experience and Problems in the Creation of the Sample Registration System in India", paper presented at the seminar on Statistical Problems in Population Research, The East-West Center, Honolulu, August, 1971.

[10] Seltzer, William, "Some Results from Asian Population Growth Studies", Population Studies, Vol. 23, No. 3, 1969, pp. 395-406, table 3.

[11] Seltzer, W. and Adlakha, Arjun, "On the Effect of Errors in the Application of the Chandrasekaran-Deming Technique", paper presented at the seminar on Population Growth Estimation, The Population Council, New York, January, 1969.

[12] United Nations, Methods of Estimating Basic Demographic Measures from Incomplete Data, Manual IV, Population Studies No. 42, U.N. Department of Economic and Social Affairs, New York: 1967.

THE PROGRAM OF THE STATISTICS DIVISION OF THE LAW ENFORCEMENT ASSISTANCE ADMINISTRATION

George E. Hall, Law Enforcement Assistance Administration

The Law Enforcement Assistance Administration (LEAA) was established by the Omnibus Crime Control and Safe Streets Act of 1968. The primary purpose of LEAA is to provide assistance to state and local governments to reduce crime and delinquency. The Act also enpowered the agency to collect and disseminate statistics.

The Statistics Division of LEAA, formerly known as the National Criminal Justice Statistics Center has the responsibility of quantitatively describing and monitoring the criminal justice system and for providing timely information on crime and its impact on society. In order to achieve these dual goals, reliable statistical programs must be developed. The major tool for monitoring the criminal justice system is the division's offender based transaction statistics program. The task of measuring crime and assessing its impact is being centered around a national crime survey panel.

The crime panel is the largest single activity of the Division. Basically, the panel is an omnibus sample survey. The core questions of the panel will provide measures of the incidence of serious crime and the effect on its victims. Historically, crime incidence estimates have been derived from information known to the police. However, this type of measure cannot provide all the information needed for criminal justice planning and evaluation. 1/

The crime panel will be able to provide estimates of the incidence and the socio-economic and geographic distribution of crime by utilizing a general sample of households and businesses. By interviewing a general population sample, the shortcomings of police statistics can be overcome. There is no necessity for the respondent to "get involved," nor is there any reason to fear the police. On the other hand, there is the natural positive incentive to discuss the incident. This is not to suggest that there are no serious response problems, but these problems seem to be amenable to solution.

The crime panel will provide a variety of crime measures. First, the panel will provide a measure of criminal events. This event measure provides a single count for each event no matter how many different crimes were committed during the course of that event. Only the most serious crime is counted. Thus, if a rape occurs during the course of a robbery or burglary, only the rape is reported. If only an event measure were provided, the implications would be serious. Most authorities for example, feel that there are steps which householders can take to reduce the incidence of burglary. Simple rape on the other hand is a much more difficult problem. However, if a significant proportion of rape incidents are coincident with a burglary, it becomes possible to view the rape problem in a different perspective.

Similarly, an armed robbery with three victims is counted in the same way as a strongarm robbery with one victim. The more precise differentiation between crime sub-types which the panel will produce will provide the kinds of data recognized as needed to develop strategies for crime reduction and overall criminal justice planning. Moreover, this information on the incidence of crime is the basis for the evaluation of any crime reduction program.

In addition to providing fresh estimates of incidence, the nation crime pan will also focus on the cost of crime. The direct costs of crime are many and varied: They include money and property lost through theft, the cost of medical attention, time lost from work and replacement services for persons injured. The panel will provide a measure of all these direct costs but of equal or more importance, it will provide the means of estimating indirect crime costs such as the flight of citizens from the cities. Restaurants, theatres, retail stores, other businesses and residential properties in the downtown parts of major American cities are being hard hit by the overall suburbanization of the society. Tax bases are shrinking, revenues have been reduced and other problems have resulted from this phenomenon. It would appear that a significant portion of this movement has been triggered by a fear of crime.

Some of this effect is a result of hyper-reaction of the population. Whether the reactions are rational or irrational, however, the effect on the urban quality of life is a major contributor to the current crisis in our cities.

By utilizing the latest sociometric and psychometric techniques adapted to mass surveys, we would be able to provide insights into the source of this public reaction and produce data which will suggest solutions. These data will provide both local and Federal authorities with valuable planning tools to determine program priorities.

The panel will also provide more direct input to criminal justice agencies by determining the attitudes of various segments of the population toward those agencies and the specific problems which seem most acute. This will provide justice agencies with the information needed to improve their effectiveness and their image.

Work involved in the development of this panel has uncovered a number of problems both methodological and philosophical. Conceptually, it is easy to determine a mix of events, which taken together legally constitute a crime. However, in the deeper sense this is not sufficient. Many people, for example, do not report events to the police because they do not consider the event to be a crime. Many violent activities which occur between relatives or close friends are often not considered crimes by the participants unless serious injury results.

One of the activities related to the panel will be a general population survey to assess public attitudes concerning the relative seriousness of actual events. The project will also examine the circumstances surrounding the event to be able to quantify those situations which the public considers extenuating circumstances.

The panel will also provide the capacity to conduct *ad hoc* surveys as well as the regular means of conducting methodological research. The panel design calls for interviewing approximately 10,000 households and a smaller number of businesses each month. While these monthly samples will have to be aggregated to provide crime incidence information, each is large enough to provide a national sample to use for gaining other information of interest to the criminal justice community.

Research into related data collection problems began in the winter of 1970. Small research projects were conducted in Washington, D. C., and Baltimore, Maryland, to evaluate question wording, victim recall, and other response problems. 2/ The results of these experiments were sufficiently positive to warrant going ahead with a research study aimed at the general population. Three such studies were scheduled for January 1971. One was a national survey appended to the Bureau of the Census' Quarterly Household Survey Panel. The other two, conducted in the LEAA pilot cities San Jose and Dayton, had a two-fold purpose, one to further refine the survey techniques 3/ but more importantly to provide baseline data for the evaluation of the pilot cities program itself.

Because of the large sample size required to provide crime estimates, the panel will be used as an omnibus vehicle to provide *ad hoc* data as required for planning, research or evaluation.

Because of the long lead time, planning is already underway to develop methodology to increase the routine utility of the panel. Preliminary investigations are already underway to examine the feasibility of regularly or periodically measuring such things as drug use, activities related to organized crime, juvenile crime and white collar crime.

The division's methodological research for the next several years will emphasize the development of survey information which can be routinely collected utilizing the crime panel.

UNIFORM CRIME REPORTS

Through a discretionary grant program administered by the division, almost a million dollars has been distributed in the past two fiscal years to encourage the states to set up central state reporting of UCR data. These state agencies then become responsible for quality control of the data within the state, thus the state-level agency is able to provide far closer supervision of the individual police agencies than the FBI would be able to do. The division is also encouraging new developments in the collection of UCR data. For example, one state is experimenting with the idea of obtaining sample data from police agencies. By utilizing samples, more detailed information can be obtained on individual events and more sophisticated cross-classifications can be developed at a reasonable cost. The division will also be working with individual police agencies and through Project SEARCH to develop uniform event reporting forms in a machine-readable format. With the development of these forms again more sophisticated information about crimes will become available. Of more importance, uniform arrest forms are being developed so that arrest registers will become available. In addition to making the Uniform Crime Reports more useful, the standardized arrest registers will facilitate the development of statewide offender based statistics programs.

TRANSACTION STATISTICS

As we attempt to assess the quality of the administration of justice, we encounter an information void in justice processes as they relate to the suspect or offender. It is our intention through various projects planned in 1972 to have the data to answer such vital questions as the following: What is the time element between various transactions in the process? Which outputs of agencies are inputs to other agencies? What is the drop-out rate at various points in the process? What is the caseload at each point? What are the characteristics of the offenders who re-enter the system? Which offenders drop out at various points? Which offenders experience greater time delays? What are the offenses that cause the greatest delay from one point to another? These questions must be answered separately for adults and juveniles as well as comparatively. These data, which will be made available can be used to predict events within the system, evaluate programs and practices of various agencies and plan new programs.

In order to provide this kind of information, we have encouraged states to develop "transaction statistics" systems. Since the basic unit common to all justice processes is the offender or the suspected offender, the transaction statistics system tracks the offender as he passes through the system and records the pertinent data for each criminal justice transaction. Thus it becomes possible to examine the processes of the entire system.

In fiscal 1971, we directly funded a number of states to develop at least modules of such systems. Out long-range goal is to develop the software and procedures for a national system which will include summary information from the state systems and Federal records.

One of the major efforts funded by LEAA over the last several years is Project SEARCH. This project which now consists of a consortium of 20 states was established among other things, to develop a prototype state statistics system. As an outgrowth of this prototype statistics system, we have launched a five-state effort to implement these systems. The system will be refined in an operational mode where many problems and questions relating to statewide transaction statistics systems will be resolved. Through this relatively small-scale effort, standards and guidelines will be established so that other states may use these as models for their own state systems. 4/

Moreover, to aid in the development of transaction statistics by the states, we must be prepared to aid the states as they assume responsibility for their own criminal identification functions. This is essential in an offender based statistics system since rapid positive identification of an offender is a must if we are to be able to determine recidivism patterns. It is only through the examination of these patterns compared with prior "treatment" that success rates can be determined and predictors developed for the justice process. The conversion of records to machine readable format is also necessary in order to fulfill this need. This will be done in conjunction with the conversion of records for the exchange of criminal histories through Project SEARCH.

The SEARCH states will develop as a preliminary step standard event reporting forms and standard arrest reporting forms which can be adopted by those states who are in the early stages of developing a transaction statistics system.

CORRECTIONAL STATISTICS

Transaction statistics will not soon provide all of the information needed on the justice processes. For example, many states are not now involved in the system and there is excessive lead time from the planning stage until comprehensive data become available for all parts of the criminal justice system. In the foreseeable future, this system will also only handle transactions related to serious (fingerprintable) offenses. Therefore, the Division will have to develop interim programs and support existing activities to provide much needed data for planning and evaluation. A project is now in the advanced planning stage which will utilize a sample of the institutions included in the jail census we conducted last year. While the previous study concentrated on the physical aspect, this study will concentrate on inmates.

For every inmate in the jail on the "Day of record" such information as age, sex, race, offense, reason for incarceration, time already served, time remaining, limited criminal history, place of residence and employment status before incarceration, marital status, number of dependents, personal income in preceding year, family income, educational attainment and enrollment status will be recorded. With these data available for a probability sample of all persons in jails in the United States, researchers for the first time will be able to assess the impact of jail on inmates and to some extent, the socioeconomic consequences of incarceration.

To increase further the understanding of local incarceration, we are conducting a study of juvenile detention centers to look at the physical aspects of juvenile facilities as well as study inmates of these institutions to focus on the same kind of characteristics for juveniles as in the follow-up Jail Survey.

Also in the development stage is a census of state correctional facilities. This census will provide much needed information on the programs, descriptions of the facilities including age, capacity, population and such characteristics of the staff as number, educational background, professional skills, and racial composition. We expect this program to go into the field by early 1972 with results becoming available by year's end.

Another project is National Prisoner Statistics. In fiscal year 1971 work was begun to rejuvenate this venerable system which in recent years was the responsibility of the Bureau of Prisons. Funds were provided to the Bureau to publish data collected through 1970. In FY 72 we will begin to publish data from the system.

The NPS program will involve the establishment of a data base of all inmates in state adult correctional institutions. Each agency will report admissions and releases on a monthly basis. As the states create their own central statistics bureaus, the states will be asked to provide summary data rather than individual records from each of the institutions.

Another interim program of interest is the Uniform Parole Reports. This work is coordinated by the National Council on Crime and Delinquency. We expect that beginning in 1972 funds will be provided to keep the program viable. Moreover, we expect that modifications will be made to make it compatible with the National Prisoner Statistics program.

Another pressing need is to coordinate this program with statistics concerning probation, but at this time there is no structured system for reporting probation information. A project now being funded will be a start in this direction. The NCCD has proposed to conduct a study directed toward filling the need for administrative control and research investigation for all types of probation programs. The proposed project will be a model using the San Francisco Bay Area Probation Departments. The project will develop a research system for describing and evaluating a wide variety of program elements to which juvenile and adult offenders are assigned. The system will be geared to meet the needs of adult offenders are assigned. The system will be geared to meet the needs of adult and juvenile probation with a general research methodology that is applicable on a nationwide basis. It will be designed with a goal of eventual integration into a nationwide correctional statistics program. This correctional system will be the final segment of the total offender based transaction system.

COURT STATISTICS

While there is a paucity of data concerning corrections at the national level, there is absolutely no data concerning the courts. In fiscal 1971, the first step was taken toward a national court statistics program. A Court Organization Study is now underway to examine the organization, jurisdiction, manpower, caseloads and practices of all state and local courts - civil, criminal, juvenile, and other courts of specialized jurisdictions. Such a study is requisite for a national court statistics program; first, proper analysis of data collected in a national court statistics effort can only proceed from a detailed understanding of the court system; and secondly, information collected in the system study would meet the methodological need of providing some of the necessary parameters for designing and optimum, stratified sample of the courts. In FY 1972 we will begin conceptualizing a national data collection program including coverage of trial courts of general jurisdiction and state appellate courts. Coverage will provide information about the most serious stage in an offender's contact with the judicial process - the felony trial. However, the overwhelming majority of criminal cases never reach the felony trial stage, thus for most offenders, the ultimate stage of contact with the criminal justice system is the lower court. Such courts will also be included on a representative sample basis.

With this program we will provide information on the number of criminal cases presented; the number of dispositions without trial: Number of trials with and without juries: and some information on trial outcomes and sentences. All of this will be presented by type of offense, type of court and possibly by a limited number of offender characteristics - age, sex, race. Information will be presented on an annual or perhaps quarterly basis at the U. S. level, and by city size groupings. Concurrently, work will begin to attempt to develop information on charge reductions, pleas, plea bargaining, type of defense and delay in judicial process.

EMPLOYMENT AND EXPENDITURES

The 1970 Omnibus Crime Control Act requires that beginning in FY 73 block grant funds to the states will have to be "Passed through" to units of local government on the basis of the proportion of funds spent by the local units of government. In order to ascertain this ratio, a Census of Criminal Justice Agencies will be undertaken to provide accurate data on expenditures by local and state governments for criminal justice purposes. As a side benefit of this census, we will get a more accurate picture of the manpower structure within the state and local criminal justice system. This information in the past two years has been gained partially on a sample basis and this will continue in the intervening years between censuses.

OTHER ACTIVITIES

In addition to its programs related to the collection, analysis and dissemination of data on crime and criminal justice, the Statistics Division has a number of other programs.

DICTIONARY OF CRIMINAL JUSTICE TERMS

Under the direction of the Division, a Dictionary of Criminal Justice Terms is being prepared. This dictionary will provide for the first time, definitions for criminal justice terms and concepts. This effort will involve isolating the various elements which taken together will reflect the meanings of a word or concept. For example, recidivist can have a number of different meanings. However, from a technical standpoint this is not acceptable. One definition of a recidivist is a person who has been in contact with criminal justice system who subsequently commits a crime. This definition of course does not permit accurate measurement since there is no definitive way of ascertaining that a subsequent crime has been committed unless the offender is apprehended. A more common working definition and one frequently used by law enforcement is someone who has been arrested for a crime and is subsequently arrested again. Correctional people often refer to recidivist as one who has been released from the prison system and was subsequently convicted of another crime and returned to prison. There are a number of elements then which go into the possible meaning of the word: arrest, conviction, imprisonment, and others. The dictionary will select a preferred meaning, utilizing combinations of those elements.

The dictionary will also provide the basis for the development of a standardized classification system for criminal justice statistics. There is a clear need for such a classification system. As the Division moves toward the Federal system, all states having statistics centers would report limited standardized information to the Division for comparative analysis and publication.

DIRECTORY OF CRIMINAL JUSTICE AGENCIES

The Division has also developed a directory of all criminal justice agencies. This directory, which is now available on computer tape, will be published as bound copy in fiscal 1972. The directory will be constantly updated to provide a current listing of agencies in the justice system. Moreover, additional information is being built into the directory which will identify with more precision the types of agencies. For example, information from our court survey will be coded into the directory so that we can look at courts by caseload, case mix, type of jurisdiction, etc.

STATISTICS DATA BASE

In order to derive maximum utilization of information collected by the Division and other statistical data relevant to criminal justice, the Division is establishing a computerized statistical data base. A prototype of this data base will be operational by the beginning of Fiscal 1972. By the middle of the fiscal year, the entire system should be in place with terminal access to the data base in all of the State Planning Agencies, large metropolitan police agencies, state correctional agencies, schools of criminal justice, as well as in LEAA. The data base will contain information from the decennial censuses, the Uniform Crime Reports, Survey of Employment and Expenditures in Criminal Justice and summaries of LEAA grants. As the statistics become available, the data base will be expanded to include information from the crime panel, and from the transaction statistics program.

The bulk of the staff time in the Division is devoted to the design of surveys, monitoring their execution, the analysis of resulting data and the preparation of publications from the data. The Division also develops standards for state statistical activities and provides technical assistance to the states. As the principal data collecting agency in criminal justice, the staff coordinates related statistical projects of other agencies. This coordinative function is of distinct advantage to the Statistics Division because it permits us to become aware of related projects and guide them in such a way that they will provide data of maximum utility to LEAA and the criminal justice community.

1/ For a description of methodology, see Dodge and Turner's paper "Methodological Foundations for Establishing a National Survey of Victimization," this section.

2/ See "Victim Recall Pretest (Washington, D. C.)" and "Household Survey of Victims of Crime Second Pretest (Baltimore, Md.)."

3/ See "The San Jose Methods Test of Known Crime Victims," LEAA Statistics Division, July 1971.

4/ A complete description of the program may be found in Wormeli and Kolodney's paper, "Computer-Linked Transactional Records for Criminal Justice Statistics.

METHODOLOGICAL FOUNDATIONS FOR ESTABLISHING A NATIONAL SURVEY OF VICTIMIZATION

Richard W. Dodge, U.S. Bureau of the Census
Anthony G. Turner, Law Enforcement Assistance Administration

One of the primary responsibilities of the Statistics Division of the Law Enforcement Assistance Administration is to provide timely statistical data on crime and its impact on society. Available statistics show counts of crimes which have been reported by citizens to the police and which the police, in turn, have reported in their statistics. However, evidence indicates a significant volume of crimes committed against citizens never become known to the police. In addition, administrative statistics cannot provide the demographic and socio-economic framework which is essential to understanding the broader impact of crime.

The Statistics Division of LEAA hopes to provide such data by establishing a National Crime Survey Panel which will be operated as a continuous national survey, administered by the Bureau of the Census to general probability samples of households, businesses, and institutions.

The core questions of the Crime Panel will provide measures of the incidence of serious crime and the effect on its victims. Data available from the survey will include national estimates of the number of crime events, the number of victims, the economic cost of crimes, multiple victimizations, characteristics of offenders, and victim-offender relationships. These data will be published to display the socio-economic and demographic distribution of crimes and victims, as well as the geographic distribution - that is, national and regional data, and data for some of the very large cities and states.

In its initial stages, for reasons to be described later in the paper, the Crime Panel will limit its focus to various forms of theft and interpersonal assaultive behavior. Later, as survey techniques are sufficiently developed and refined, we anticipate including the measurement of other types of crimes.

In planning for a national survey to measure victim experiences, a host of methodological problems must be addressed, evaluated, and documented. Since early 1970, the Bureau of the Census has launched a broad series of pilot studies for LEAA to ascertain the feasibility of measuring the total incidence of major crimes through the use of survey techniques.

Earlier attempts by other researchers were not only very promising in showing the analytical value of victim surveys, but they were invaluable as pioneering efforts from the standpoint of suggesting several methodological questions for Census and LEAA to address in their pilot tests. The only national survey ever undertaken was the National Opinion Research Center study of 1966. Criticism of this study pointed up the need to conduct further research on the differences in the amount of crime as estimated from questionnaires where the respondent reports for himself and from questionnaires where the respondent reports for others in the household.

Other surveys conducted for the National Crime Commission during the mid-60's were localized rather than national in scope. These studies, too, were useful in suggesting methodological problem areas, such as:

(1) What is the extent and nature of memory failure for victims of crime?

(2) What is the optimum length of the reference period for recalling crimes?

(3) What is the optimum mode of phrasing questions to avoid legal jargon for the answering public, yet to elicit responses which can be properly coded according to established standards for purposes of categorizing crimes?

This paper is devoted to a discussion of the methods tests conducted by Census and LEAA to focus on the aforementioned problem areas. In addition, we will also touch upon the topics of questionnaire format, use of telephone and mail survey techniques, and the use of business records to assess commercial victimization. Some of the results are presented, though a number of methodological inquiries are still in varying stages of completion and data for them are not yet available.

Victim Recall, Telescoping and Other Technical Problems Addressed Through Reverse Record Check Studies

A crucial issue in planning for a national household survey of victimization is the ability of respondents to recall incidents of victimization befalling them or other household members. Thorough study of this problem, and the related subject of telescoping, is needed in order to establish the optimum reference period to be used in the survey. Cost considerations become a significant element in this determination when it is recognized that cutting the reference period in half, from 6 months to 3, for example, necessitates a doubling of the sample size to achieve the same degree of reliability. Sample size is an especially critical parameter in setting up a crime incident survey since most major crimes, such as rape, robbery, or aggravated assault, are statistically rare phenomena. The recall problem has been more thoroughly studied by LEAA and Census than any of the other methodological problems being considered here. The studies have taken the form of a series of reverse record checks with samples of known victims drawn from police-maintained offense records. To date, these tests have been conducted in Washington, D.C. (March 1970), Baltimore, Maryland (July 1970), and in San Jose, California (January 1971). The San Jose test took place at the same time as the Pilot Cities Victimization Survey, conducted in both San Jose and Dayton, Ohio, which was designed to gather data on crime incidents from a general population sample.

There are certain difficulties in using police records as sources of samples. Only cases reported to the police are included. This leaves unstudied the large number of crimes which are not reported to the police and thus leaves unknown the degree to which recall problems for nonreported crimes differ from those that can be studied. A further problem in the use of police records involves sample selection. Our experience has been that although offense reports are public records, we have not been able to select a sample directly but have had to supply specifications to others. In general, the samples were quite satisfactory for our purposes, but errors in selection occurred which reduced the effective sample size. The most common of these were cases where the victim did not reside in the local metropolitan area or where the crime selected was directed against a commercial establishment or a person acting in a commercial capacity.

Crime victims seem to be more elusive than the general population, especially victims of personal crimes, and we have had great difficulty in locating our respondents. Only through exhaustive interviewer efforts were we able to achieve response rates in the three tests to date varying from 63 to 69 percent. This, of course, is separate from the ability or willingness of respondents to report crimes of which they were the victims once they have been located.

On the positive side, the advantages of using police records as a source for testing victim recall seem to us compelling. They provide a readily available sample of victims which, because victimization is a low incidence phenomenon, would be costly to identify in any other way. And, most importantly, they permit a direct comparison of a respondent report in a household interview situation, some time after the event, with the actual official report of the same event made when memory failure was at a minimum. Recognizing that the offense report is not the entire "truth" of the matter, it nontheless provides, at the very least, an anchor in time, not otherwise available, to which subsequent reports can be compared with a high degree of confidence.

The three pretests using samples of known victims had other purposes besides studying recall. The content of the questions, designed to screen for incidents, the order in which they were asked, and specific question wording were modified each time as a result of field experience. In Washington and Baltimore, victims of four major crimes were selected--robbery, assault, burglary and larceny. Cases of homicide and auto theft were not included because they are fairly well reported and not difficult to conceptualize. (In addition, victims of murder pose an obvious interviewing problem.)

Questions on theft of automobiles and other motor vehicles were included although no such cases were sampled from the police records. This was done to distinguish motor vehicle theft from other kinds of larcenies. Rape was excluded from the first two tests because of the sensitivity of the issue. In San Jose, however, a sample of rape cases (one-half the size of the samples for the other crimes) was selected for interview. The screen questions that had been used previously to elicit reports of assaults were left essentially the same to see if they would elicit reports of rape. More explicit wording was rejected as not appropriate for a federal agency to use and likely to be offensive to respondents.

In addition, revealed as a by-product of these tests was the problem of classification of crimes. Various inconsistencies were noted between the police classifications and those made as a result of the personal interviews. To some extent, these variations brought to light defects in the questionnaires which were subsequently corrected. Nevertheless, in the great majority of cases, there was sufficient detail obtained in the interview to enable a match to be made to the corresponding offense report.

The principal conclusions to emerge so far from these tests are these:

a. If the objective is to determine <u>whether</u> a crime occurred, as opposed to placing it in a more accurate time frame, then a 12-month reference period is as good as one of 6 months. This should be qualified by mentioning that two of these tests were anchored on the calendar year so that the furthest limit was one of the most salient of dates--New Year's Day. The recall bias which derives from time telescoping can be largely corrected by providing interviewers with bounding information, that is, the record of incidents from the previous interview. The plans for the National Crime Panel contemplate a substantial degree of overlap in sample addresses from one collection period to the next--in the neighborhood of 75-80 percent.

b. To the extent that it is desirable to place an incident in a specific time frame, greater accuracy is obtained from a shorter reference period. Thus, a 6-month reference period is better than 12, and a 3-month period is better than 6. As was mentioned earlier, cost constraints become increasingly important as the time reference is shortened.

c. Beyond the ability to locate and interview respondents is the probability of the respondent's recalling a specific act of victimization, which was determined in these studies by matching a respondent report with an incident selected from police records. This probability was very high for crimes involving theft of property (80 to 85 percent). With respect to personal crimes, robbery was well reported (75 percent and above), but rape and assault were less so (2/3 and 1/2, respectively). An important factor in the recall rates for cases of personal victimization is the relationship of the offender and victim. Recall rates vary directly with the nature of that relationship; that is, when victim and offender are strangers, recall rates are high (75 percent in San Jose). Acquaintance, and even more, kinship, result in lower reporting rates, as low as 22 percent for relatives in San Jose. Since assaults are more likely to occur between people who are at least known to each other, if not related, we would expect recall rates for assaults to be low. Robberies, on the other hand, tend to occur between strangers (70 percent of the cases selected in San Jose) and thus, recall rates are correspondingly high.

At the moment, our conclusion is, when considered in connection with a continuing survey, that a 6-month reference period is better than a 12-month period for producing calendar year data and for obtaining earlier and more timely results. With a 6-month rolling reference period, some data could theoretically be available after 12 months--assuming bounded interviews--and the data would be "centered" 3 months ago. For a 12-month reference period, 18 months would be required before data, comparably reliable, would be available and it would be centered 6 months ago. As was mentioned above, the sample size for a 6-month reference period is twice that for a 12-month period.

It is to be expected that any statistics which purport to measure the incidence of crime would inevitably be compared with crimes known to and reported by the police, issued regularly in the FBI's Uniform Crime Reports. For the victim surveys, therefore, considerable effort has been expended in developing the instruments so that certain major crimes elicited can be classified in accordance with the definitions used by UCR. This has been done in order to make comparisons between UCR and victim survey results meaningful. On the other hand, it should be noted that tabulation plans call for presenting victim-event data in sufficient detail to permit analysts who so desire to describe crimes in ways which may depart from the constraints imposed by UCR definitions.

Successive improvements in the survey questionnaires used in the three pretests have been made to the extent that we now feel our ability to classify crimes according to UCR standards cannot likely be improved further. We feel that any remaining inconsistencies that may show up between police and survey classifications would be due largely to normal response errors, legal differences in the definitions of crime from one jurisdiction to another, and variable police practices in recording crimes.

Screening for Incidents

In designing survey instruments for the various pretests and for the regular surveys to follow, it was decided to screen for all relevant incidents before obtaining details of any one incident. This was based on some experiences from previous surveys and also from our a priori judgment that better results would be obtained by letting the respondent remain in the incident-centered context while a series of specific questions attempted to elicit reports of victimization. This procedure has a very practical aspect, as noted by Biderman and Reiss, in that it takes advantage of the respondent's interest and freshness to establish the general victimization profile before proceeding to the specifics. The procedure of obtaining complete information about each incident at the time it is first mentioned, runs the risk of boring or tiring the respondent who can easily "forget" to report additional incidents. The screening procedure as adopted also has the added advantage of informing the interviewer of the total victimization picture so that she may be better able to assist the respondent in disentangling the facts of two similar larceny incidents, for example.

The content of the screening questionnaire itself poses crucial methodological problems. We have adopted what may be characterized as a "middle way" between a brief screen consisting of, say, one question concerned with each of the types of crimes in which we are interested and the alternative of compiling a lengthy list of very specific questions with which to bombard the respondent, explicitly mentioning a multitude of examples of the kinds of property that might have been stolen or the kinds of situations in which he might have been the victim of a personal crime.

We feel that the current version of the screen, while subject to further improvement, is a satisfactory compromise which achieves a reasonable measure of completeness of coverage without losing the respondent's attention. After each pretest we have modified the screen questions in order to overcome defects that have become evident. In the most recent version of the questionnaire, we have added two "catch-all" questions to the end of the screen in a final effort to elicit incidents that the more specific questions have not brought out. These questions ask the respondent if he called the police to report something that happened to him which he thought was a crime, and, second, if anything else happened to him which he thought was a crime but did not report to the police. As would be expected, these questions resulted in many reports of crimes other than those which are the focus of our studies--for example, vandalism, peeping toms, etc.--and also reports of non-crimes. However, they have also yielded descriptions of events which appear to qualify as one of the five major crimes. We use the word "appear" because the interviewer was asked to write as complete a description of the incident as possible, but did not fill a detailed incident report form. In a number of cases, the description of the event was too sketchy to permit conclusive determination of what kind of crime had occurred.

In a nationwide experimental survey conducted in July 1971 and utilizing the Census Bureau's Quarterly Household Survey, interviewers were instructed to fill an incident report on each situation where the crime reported in the two catch-all questions seemed to qualify as one that should have been mentioned in response to one of the earlier screen questions. We do not, as yet, have any results from this modification in procedure, but we do have some evidence from the surveys conducted in January 1971, on the kinds of events reported in these two final screen questions.

In the San Jose police sample, somewhat fewer than 3 percent of the successfully matched incidents were reported in the catch-all questions. However, there were a number of other reports of one of the five crimes which did not match the selected sample cases. Larcenies and assaults were most frequently picked up as a result of these additional probes. A hand tally of responses to these questions in the Pilot Cities Surveys indicated that as many as 5 percent of all incidents that qualified as one of the 5 crimes were reported in these two catch-all questions.

Self-Respondent vs. Household Respondent

Another methodological problem of significance in establishing a National Crime Panel is the choice of the respondent in a household. The most economical approach is to interview any responsible adult who is home when the interviewer calls--which means that the respondent will more often than not be the housewife. This respondent would report for himself and all other eligible household members. For crimes where the entire household can be considered the victim (i.e., burglary, auto theft, etc.), this procedure may produce satisfactory results. However, for those crimes where a person is the victim, there is evidence from the surveys conducted for the President's Commission on Law Enforcement and Administration of Justice that the household respondent reports other household members less frequently as victims than he reports himself, even though these persons are more likely to be exposed to crimes of this kind.

Interviewing all eligible household members individually is obviously a more expensive method. Less expensive would be the randomized pre-designation of household members based on household size. This has serious implications on the overall effective sample size, however, since for a fixed cost, it results in a sample size which is about 40 percent as large as if all household members had been included through the use of a household respondent. The decision as to which method to use has to balance the cost of the designated respondent procedure against the bias implicit in the household respondent approach.

A direct test of this problem was built into the Pilot Cities Victimization Survey. The sample households were divided equally in advance into those where a household respondent would be asked to report for himself and all other household members 16 years old and above; and those where each qualified household member would be interviewed individually.

At this time, only preliminary results are available based on hand tallies of raw data which have not been edited or weighted to allow for oversampling in the poverty areas of both cities. It is not known what effect, if any, editing and weighting will have on this comparison. The raw data indicate that the self-respondent households reported more incidents of crime than did those where the most available person responded for everyone. Although the interviewed households were almost equally divided, the self-respondent households reported 57 percent of all crimes. In addition, there was a tendency for certain crimes to be more frequently reported by persons in self-respondent households than the relative totals for all incidents would lead one to suspect. Petty larceny and assault were the principal examples of this. We would conjecture that petty larcenies are the most easily forgotten of all these crimes, but are likely to be better reported when each household member is interviewed for himself, including the owner of the particular item that was stolen. Assaults, on the other hand, may not be "forgotten" so much as they may not always be known to other family members, because of embarrassment, or if they occurred between family members or friends may be edited out by the respondent. Whatever the reason, the involvement of all family members as respondents has a better chance of bringing out these reports, especially if the interviews are conducted separately.

In contrast to petty larceny and assault, auto theft was reported at about the same rate, regardless of the interview method involved. However, it should be pointed out that even in those households where everyone eligible was personally interviewed, certain screening questions were asked only once in the household--and were asked of the first person interviewed, the equivalent of the household respondent in the other procedure. The screen questions that were deemed to fall into the category of household crimes that were to be asked only once were those concerned with burglary, larceny of household goods left outside, and theft of a motor vehicle or part of a motor vehicle. We would expect, therefore, that no significant difference would occur in the reporting rate for these crimes between the two procedures. If differences should appear, as in some kinds of larcenies, they might be attributable to another household member volunteering such information during the course of the interview, having been reminded of a "household" crime during the course of the individual screen questions. Obviously, the distinction between household and individual crimes is somewhat arbitrary and respondents cannot be expected to sort their reality out as neatly as researchers would like.

There is also a "fatigue" factor associated with the use of a household respondent who has to report for all household members. We have adopted the rule that once the household screen questions have been asked, that the individual screen questions must be asked about each household member in turn. Many respondents, especially when there are a number of other eligible household members, rapidly become conditioned and say something to the effect that the answer is "No" for everyone else, too. Interviewers find it difficult, under these circumstances, to follow the correct procedures and ask all questions, in turn, for each person--especially if it risks antagonizing the respondent. And, even if they persist, it is likely that the respondent, having decided that the answers are all "No," will not be giving any further thought to the matter. Our feeling is that this is a compromise procedure and, although it annoys some respondents, it probably evokes further reports of victimization which we would otherwise miss altogether. (See Reference 3.)

Age of Respondent

A problem which we feel is related to the type of respondent is that of the appropriate minimum age. The LEAA surveys to date have used age 16 as the minimum age for which victim data are sought. Sixteen is the age now used to designate the lower end of the labor force. The decision as to what age is appropriate for the study of crime victims is, to some extent, arbitrary. Serious crimes can and do occur to younger people (robberies of newsboys, to cite a well known example). On the other hand, threats, fights, and other "events" that would qualify, at least at the field collection stage, as crimes are common occurrences for many youth. Are these "crimes" of sufficient significance to warrant increased costs in the field only to be subsequently winnowed out at the processing stage?

To gain some insight into this problem, an experiment was conducted in five major cities in conjunction with the July 1971 Quarterly Household Survey of Victims of Crime. In New York, Chicago, Los Angeles, Detroit, and Washington, interviewers were instructed to obtain information for all household members 12 years and above. Since all these interviews used a household respondent, we have not studied the problems of interviewing these young people themselves. Nevertheless, we expect to accumulate a body of useful information on this age group which will have a bearing on the selection of the type of respondent for the National Crime Survey.

Mail Feasibility Test

Mail as an alternate data collection technique offers obvious economies. If the expensive process of screening for instances of victimization could be conducted by mail, field costs could be cut drastically. Our assumption is that the details of reported incidents would then be collected by personal interviews. For the moment, at least, we feel that mail would not be appropriate as an initial contact, but could be utilized in a sample design that provided for multiple interviews over time with persons residing at designated addresses.

As previously noted, preparations for the inauguration of the National Crime Panel have included the use of the Census Bureau's Quarterly Household Survey as a vehicle for testing questionnaire design and for collecting preliminary national data. The sample design of the QHS enabled us to conduct a mail feasibility test to run parallel with the personal interview survey in July 1971. The QHS sample is divided into six groups, each of which constitutes a national sample of approximately 3,000 occupied households. Each quarter a new group enters the sample and an old one completes its stay. The crime victim survey is being added to the QHS every six months. Thus, in the July 1971 survey, two-thirds of the addresses had been in sample for the previous survey in January. The other one-third, which had left the sample since January, was used for the mail test.

A mail questionnaire was designed containing a letter from the Director of the Census Bureau on the front and the screening questions, plus a few demographic items, on the inside. These questionnaires were mailed to coincide with the start of the regular personal interviewing for the July QHS. In August, a sample of nonrespondents to the mailing phase was followed up in the field. At the same time, interviewers were to collect details of incidents reported on the mail screening questionnaire. For all addresses in the sample in January, interviewers were supplied with information as to their earlier report--either a brief summary of any incidents reported, an indication that there were no incidents or that the household was not interviewed in January. One-half the households reporting incidents were designated for interview by personal visit, while the other half were to be obtained, insofar as possible, by telephone.

A comparison of the incident reporting rates for the mail survey with those obtained by personal interviews will indicate whether or to what extent, mail can be used in collecting these kinds of data. The results of this experiment will be available sometime next spring.

Commercial and Other Institutional Victimization

In addition to the methods testing that has been going on in the household sector for crime measurement, some work has also been undertaken to assess the feasibility of using crime victimization surveys in commercial establishments and other institutions.

In late spring of 1970 a commercial victimization pretest was conducted for LEAA by the Census Bureau. The survey took place in Cleveland and Akron, Ohio, with a general probability sample of about 500 business establishments. The methodology employed for the Cleveland-Akron test could be the subject of a paper in its own right. Briefly, however, the objectives were to determine the degree to which businesses keep written records of crime incidence and their losses due to crime, to test questionnaire wording and format, and to examine alternative reference periods for recalling crimes.

One of the findings of the Cleveland-Akron experiment indicates that it is not feasible to rely on existing written records maintained by businesses for estimating crime incidence. Only about half the businesses that were crime victims stated they kept written records of those crimes. Curiously, a higher proportion of non-victims stated they would keep records if victimized.

Another significant methodological finding was that commercial establishments have very little documentation on the amount of inventory shrinkage due to employee theft or to shoplifting, both of which are forms of larceny.

The evidence on reference period matches the findings of other researchers as well as other Census-LEAA efforts, namely that proportionately more incidents are reported for a recent period than for a distant one; and more so than can reasonably be accounted for by seasonal fluctuation. Additional information on the problem of reference period will be available, however, from a carefully designed reverse record check study in Dayton, Ohio. This study used a sample of several hundred known commercial victims taken from police reports. These victims were subsequently followed up for personal interview. The results are being compiled and should be available before the end of 1971.

Besides the commercial victimization methods tests, we have also conducted some research on record-keeping practices in governmental institutions and offices. There again, the findings indicate that except for public school systems, government organizations do not keep adequate records for survey uses. Another experiment is underway to test the feasibility of using a diary approach for recording crimes among a sample of government offices.

Recommendations for Future Methods Tests

In the course of working with the various test efforts to date, a number of methodological studies suggested themselves for the future. Some such studies might be undertaken prior to the establishment of the National Crime Panel, others in conjunction with the Panel, and still others independently of the Panel. Some of the possible methods tests under consideration are as follows:

(1) A test of the effects on reporting frequencies under varying reference periods (e.g., within the past 3 months, within the past 6 months, within the past year), utilizing a general population sample with a multiple split-sample approach.

(2) A test of whether the Warner randomized response technique is better than conventional questioning methods for eliciting reports of assaults (and perhaps rapes and robberies).

(3) An experiment designed to compare the categories into which various police agencies would classify crimes on the basis of data elements determined from an interview survey.

(4) A test of whether proxy-respondent reporting of crimes is different in amount and type from self-respondent reporting, utilizing a sample of known crimes from police files.

(5) A test of whether the measure of change in crime incidence between two periods differs by type of respondent (self versus proxy).

(6) Further exploratory work associated with the measurement problem of assessing the amount of certain types of commercial crime, such as employee theft, shoplifting, shipping fraud, embezzlement, vandalism, arson, and bomb damage.

We end this progress report on a tentative note. That is to say, we feel we have made a beginning in studying the methodological foundations for establishing a recurring national crime panel, but in so doing, we recognize that much remains to be learned.

REFERENCES

(1) Biderman, Albert D., "Surveys of Population Samples for Estimating Crime Incidence," The Annals of the American Academy of Political and Social Science, 374: 16-33, November, 1967.

(2) Biderman, Albert D.; Johnson, Louise A.; McIntyre, Jennie; Weir, Adrianne W., "Report on a Pilot Study in the District of Columbia on Victimization and Attitudes Toward Law Enforcement," U.S. President's Commission on Law Enforcement and Administration of Justice: Field Survey I (Washington, D.C.: U.S. Government Printing Office, 1967).

(3) Biderman, Albert D., "Time Distortions of Victimization Data and Mnemonic Effects," Bureau of Social Science Research, Inc., July, 1970.

(4) Ennis, Philip H., "Criminal Victimization in the United States: A Report of a National Survey," U.S. President's Commission on Law Enforcement and Administration of Justice: Field Survey II (Washington, D.C.: U.S. Government Printing Office, 1967).

(5) Law Enforcement Assistance Administration, "The San Jose Methods Test of Known Crime Victims," Statistics Division Technical Series, Report No. 1, 1971 (unpublished).

(6) Reiss, Albert J. (ed.), "Studies in Crime and Law Enforcement in Major Metropolitan Areas," Vols. I and II, U.S. President's Commission on Law Enforcement and Administration of Justice: Field Survey III (Washington, D.C.: U.S. Government Printing Office, 1967).

(7) U.S. Bureau of the Census, "Victim Recall Pretest (Washington, D.C.)," Demographic Surveys Division, June 10, 1970, (unpublished).

(8) U.S. Bureau of the Census, "Household Survey of Victims of Crime Second Pretest (Baltimore, Maryland)," Demographic Surveys Division, November 30, 1970 (unpublished).

(9) Warner, S. L., "Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias," Journal of the American Statistical Association, Vol. 60, (1965).

COMPUTER-LINKED TRANSACTIONAL RECORDS FOR CRIMINAL JUSTICE STATISTICS

Steve E. Kolodney and Paul K. Wormeli
Public Systems incorporated

For years, national and state authorities, commissions and hearings have recognized the need for and generally agreed upon the goals of criminal justice statistics systems. Studies and reports have focused both on the uses of such statistics and on the design features that would assure that these systems supply necessary criminal justice information.

The President's Crime Commission Report of 1967 [1, p. 123] summed up a comprehensive statement of user need by stating:

> "Adequate statistical programs are of enormous importance. If a serious effort to control crime is to be made, a serious effort must be made to obtain the facts about crime."

Likewise, the Science and Technology Task Force saw better information about crime and the criminal justice system as essential for both research and immediate operational improvements:

> "Information about the consequences of actions by the criminal justice system is essential for improving those actions" [2, p. 2].

Later work, notably the *Report on National Needs for Criminal Justice Statistics* by the Bureau of the Census in August 1968, and the *Hearings before the Subcommittee on Census and Statistics*, House of Representatives, March and May 1969, endorsed the same general goals, and agreed that summary statistics from separate agencies cannot provide a basis for any detailed analysis.

The major purposes of improved statistics systems [3, p. 1-4] can be summarized as follows:

- Better statistics are needed to determine the impact of crime; to determine the effects of criminal justice system policies and operations upon individual citizens and social groups, and to forecast the results of changes in penal policy or the redefinition of agency roles and responsibilities.
- Cost and effect data must be generated in order to allocate resources to the most efficient existing techniques, procedures and programs; to provide comparable agencies or personnel with standards of performance; to identify areas where increased expenditures will bring maximum benefits; and to ascertain that the use of the most basic criminal justice resources, both legal and fiscal, is generally adjusted to social priorities.
- The directors of operations must also use statistical methods to predict agency workloads in relation to both crime incidence and internal system factors such as changes in arrest policies, criminal procedures, or sentencing policies.
- Varying portions of this planning, evaluation, and daily decision making information are needed by legislators and administrators at all levels of government.

With so much agreement about goals and emphasis on the immediacy of these needs, can present criminal justice statistics systems provide the necessary information? Again, using the President's Crime Commission Task Force Report: *Crime and Its Impact--An Assessment* [4, pp. 190-199] to describe the state of the art:

Police Statistics

"The area of police statistics in this country is the area in which there is available the most highly developed reporting system--the Uniform Crime Reports prepared by the FBI with the cooperation of the International Association of Chiefs of Police. . . which is steadily increasing its coverage of arrest data.

Prosecution Statistics

". . . statistics pertaining to this area of law enforcement activities are not available not only on a national scale, but by and large, also not on a state or local level . . . after police report the arrests, a total statistical black out sets in . . .

Jail Statistics

"In the sense of either their total absence or their extremely low level of development, jail statistics are unquestionably next to the prosecution statistics . . .

Judicial Statistics

"There are no national judicial criminal statistics in the United States . . . Their absence is responsible for a major portion of a most serious gap in the total picture of criminality which consists in the absence of any data on crime between arrest statistics and the statistics of offenders committed to state and federal penal and correctional institutions . . .

Probation Statistics

"There are no national probation statistics in this country. . . Probation. . .offers a special difficulty in developing national or even statewide compilations, because the probation departments are frequently attached to the individual courts and thus are not subject to statewide administration. . .

Penal and Correctional Institution Statistics

"This country has National Prisoners Statistics. . . There is, however, one serious weakness. . . That is, the existence of local variations in the policies governing which institutions are classified as State institutions and which are treated as county or city jails or workhouses, etc., and also the policies concerning the kinds of sentences and the offenders to be sent to the State and local institutions.

Parole Statistics

"Presently there are no national parole statistics in this country. There is, however, a very promising effort to develop such a program. . ."

This review of the status of criminal justice statistics highlights the incapacities of present systems; incapacities which are a direct result of our traditional concepts of the administration of justice.

Traditionally, local agencies have been tasked with the responsibility for defining crime and developing a response to it. Different approaches are exhibited in the variety of administrative structures and policies that translate penal code and criminal procedures into actions, and that allocate funds to what is viewed as a serious crime or a serious offender.

More importantly, the criminal justice system is loosely structured and poorly defined: it is not a true system, but by law a set of systems with different aims and contradictory goals. The result is a network of agencies ill-suited to develop comparable or consistent statistics on crime, processes or persons. The structural problems are compounded by the fact that agency officials do not include statistical training in their backgrounds. Even now, when administrators are realizing the larger significance of their workload statistics, the ability to state criminal justice information requirements, design data systems, collect data, and interpret for diverse users is rare.

Faced with the foregoing realities, Project SEARCH, a federally funded multistate effort to develop a prototype computerized criminal justice information system, set about accomplishing its second of two major objectives--to design and demonstrate a computerized statistics system based on an accounting of individual offenders proceeding through the criminal justice system.

Sets of annual, single-agency process counts were immediately rejected as an adequate description of criminal justice activity.

Under the direction of a Statistical Advisory Committee, a new approach was developed. This approach focuses on the individual person, whether suspect or offender, and traces his movement through the agencies of the criminal justice system. The individual is the thread that holds the system together, for he is the common element that all agencies process and it is his experience that describes criminal justice functioning.

The approach, termed offender-based transaction statistics, loses none of the advantages of older systems because traditional summary data can be produced by analyzing cross-sections of the longitudinal files. In addition, the design provides new kinds of information that are needed for uses ranging from daily decision-making to long-range planning.

First, the passage of time is accounted for. Speedy prosecution and judicial processing are required for justice and economy. The evaluation of penalty levels and correctional programs and the effects of more or less punishment necessitates that time values be known.

Second, the relationship of agencies to one another, particularly the inputs of agencies related to the output of agencies preceding them in sequence of criminal justice processing, is described. The consequent data on the "fall-out" of offenders as they move through the system presents a structure which can be used as a model to reveal underlying assumptions and to provide a vehicle for simulated experimentation and calculation of the consequences of proposed changes.

Finally, multiple actions toward the same offender (the offender "recirculation") can be accounted for. This information has great implications regarding true arrest and conviction rates, and the extent to which prison input-output includes recirculation of the same people.

The basic concept was demonstrated by Project SEARCH in ten states which experimentally traced offenders step-by-step through the entire criminal justice process. Because of time and resource limits, an arbitrary group of 250 offenders arrested in 1968 was tracked in each state. Facts that were found scattered throughout the files of local police, county prosecutors, different levels of courts, and various state and local correctional agencies, were linked to show how each states' administration of criminal justice and adult criminal defendant processing could be analyzed.

Information about the following four stages of offender-system interaction was collected:

Stage 1 - Police Action
Stage 2 - Lower Court (Pre-Felony Trial) Action
Stage 3 - Felony Trial
Stage 4 - Corrections Action

The arresting agency gave Stage 1 detail. The personal characteristics and criminal history of the individual were recorded, along with other information about the offense and the police disposition of the arrest. Defendants who remained in the system entered Stage 2, where all data relating to lower court processing was secured. This included information on arraignments, hearings, and misdemeanor trials. Stage 3 described the processes and results of felony trials. Finally, for those who remained in the system, corrections action, Stage 4 was recorded.

The differences in the number of possible routes within stages were allowed for. Police and felony trial actions normally occur in only one sequence, although the offender may exit at any point. In the lower court and corrections stages an offender can follow several different routes without exiting from the system stage. That is, one proceeding produces another proceeding, which may produce still another proceeding before the individual can exit; all of these "cycles" were recorded.

Data elements were developed and data collection forms designed. To describe the project, explain data collection forms and define data elements, a data collection manual was compiled. The manual carefully defined the sample unit, the person-arrest, to assure that comparable cases were tracked in each state.

Upon execution of the experiment, technical shortcomings of present data collection structures became apparent:

- Data collection was irregular and incomplete, even within the limits of single agency annual workload concept.
- Stable offender identification codes were lacking and other information which would permit continuous tracking was also absent.
- The meaning of basic criminal justice terms were unstable across jurisdictions.

The data collected by each of the participating states was processed through a computer system. Tables were compiled separately for each state in the same analytical format.

Computer software was developed to reduce process and analyze the information from offender-based records to demonstrate the production of summary statistics describing each level or stage in the criminal justice process. Exhibits 1 through 3 are examples of the data content of this prototype system.

Experience with this demonstration project has made it clear that useful criminal justice statistics cannot be developed by linking the summary workload data collected by operating agencies. As management information is developed for particular segments of the system, however, design of the state-level statistical systems must include methods of integrating these data subsystems.

A number of guidelines for the organization and operation of state statistical systems were identified as a result of this experimental project.

Some of the major requirements for sound data collection design are:

- Offenders must be traceable throughout their processing, generally by the assignment of a unique identification number.
- Offender characteristics require careful definition and coding, and should not be repeated at various process levels.
- Agency reporting which must be exhaustive and consistent should be reinforced through training and quality control procedures.
- All transaction reports must account for the passage of time and be reconcilable between agencies.

From an organizational perspective, the ideal is a single agency in the state responsible for collection, analysis, interpretation, and dissemination of criminal justice statistics. The character and authority of the statistical center is extremely important. It must be staffed with professionals empowered to determine basic data needs and interpretations. Most importantly, the director of the statistics center must possess statutory authority to require all persons or agencies dealing with crime or criminals to report all requested data in a specified form that guarantees completeness and uniform quality of response. Most of the desirable and necessary conditions appear in the Uniform Criminal Statistics Act which was drafted in 1946 by the National Conference of

Commissioners on Uniform State Laws. The needs expressed in the Act have not changed in the 24 years since the model was published.

The findings from the SEARCH prototype statistical system experimentation led generally to the conclusion that this was the nature of the information required for the upgrading of state and national level statistical capability, and the next step should be taken to begin to create this capability. Consequently, the Law Enforcement Assistance Administration (LEAA) funded a first phase implementation effort involving five states: California, Florida, Michigan, Minnesota, and New Jersey.

The primary objective of this effort is to begin this implementation on as nearly a statewide basis as possible. An important element of the current effort is the intention to provide each of these states with a state level capability for the collection and preparation of statistical data of use to decision-makers within the state, and extract from the state systems the national level data required by LEAA and other federal agencies.

This project is a beginning toward a national system for the collection and dissemination of criminal justice statistics. The project is being coordinated by the State of California through the California Crime Technological Research Foundation. Public Systems inc. provides the technical staff for project coordination. Each state has a project coordinator. The Statistical Steering Committee for the project was appointed by the SEARCH Project Group to direct the project.

From a technical standpoint, the experience of the experiment originally conducted showed rather dramatically that a critical component of an ongoing national system is the establishment of routine methods for the collection of data from the various criminal justice agencies in such a way as to permit the linking of data on offenders as they pass from agency to agency. Although the post-facto collection of data for research purposes is certainly possible, it is far too costly for an ongoing statistical system. The main thrust, therefore, at this point in time is to establish the routine reporting techniques along with the quality control procedures needed to insure the accuracy of the data.

One of the continual problems in this field has been the difficulty of deciding on data elements and data element coding which would be consistent across jurisdictional boundaries. It is likely that this will continue to be a problem, and the present SEARCH effort is exploring ways to keep the flexibility required within jurisdictions and at the same time produce comparable national data. Because of these varying needs, the choice of data elements and coding is likely to vary and be changed considerably over time as more knowledge is obtained and more applications for the data are identified.

A long-range goal of this effort is to create a consistent, though rather loosely defined, "transaction space," having the general dimensions of: (1) the agency involved, (2) the offender and his characteristics, (3) the criminal justice process involved, (4) the event and its associated characteristics. The purpose of having this transaction space is to enable both routine and one-time research questions to be asked along one or more of these dimensions, with the capability of linking all subsets of the transaction space.

There is a fairly wide spread acceptance of this conceptual approach. There are many differences of viewpoints and needs in actually defining this concept in sufficient detail to make it operational. It should also be pointed out that the state of the art in data collection and analysis in criminal justice lags far behind this conceptualization. For example, only last year the Bureau of Census attempted to count the number of jails in this country. Not being able to know even the number of agencies involved in criminal justice makes it somewhat difficult to structure the transaction space.

If the present five-state implementation effort succeeds, these states will begin to have available for their own use and for reporting to the National Statistics Center, the beginnings of a transaction-based system. Hopefully, the calendar year 1972 will find the state of the art advanced by the routine operations of statistical centers in the states, producing data for improved decision-making in their own criminal justice systems. The experience of these beginning states will provide considerable assistance to other states attempting to begin to create a useful statistical system.

REFERENCES

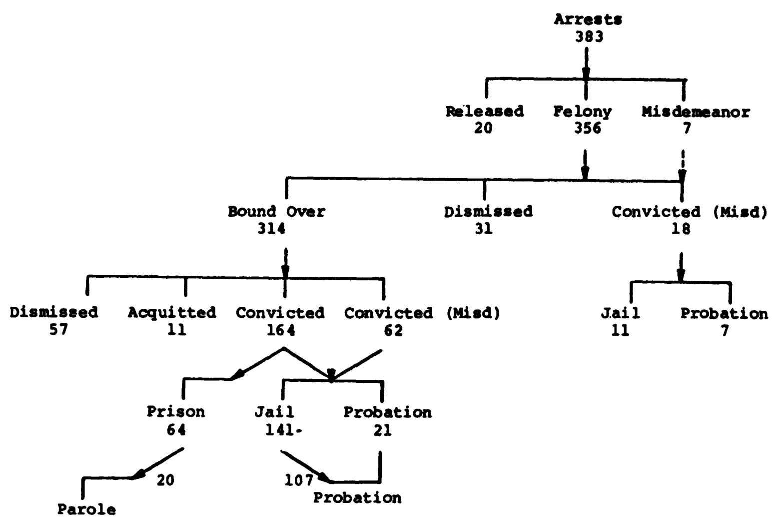
[1] President's Commission on Law Enforcement and Administration of Justice, "Crime and Its Impact--An Assessment," Task Force Report (1967).

[2] ________, "Science and Technology," Task Force Report (1967).

[3] Project SEARCH Statistical Advisory Committee, "Designing Statewide Criminal Justice Statistics Systems--The Demonstration of a Prototype," Technical Report No. 3 (November 1970).

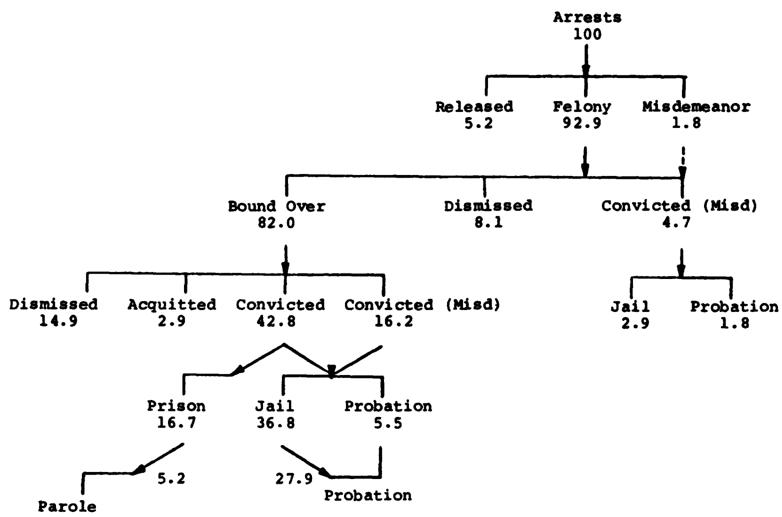
[4] Op. Cit. [1] , pp. 190-199.

(Not available from existing statistical systems)



a. Number of Arrests

(Not available from existing statistical systems)



b. Percent of Total Arrests

Exhibit 1: Flow of Arrestees Through Criminal Justice System (Example for One State)

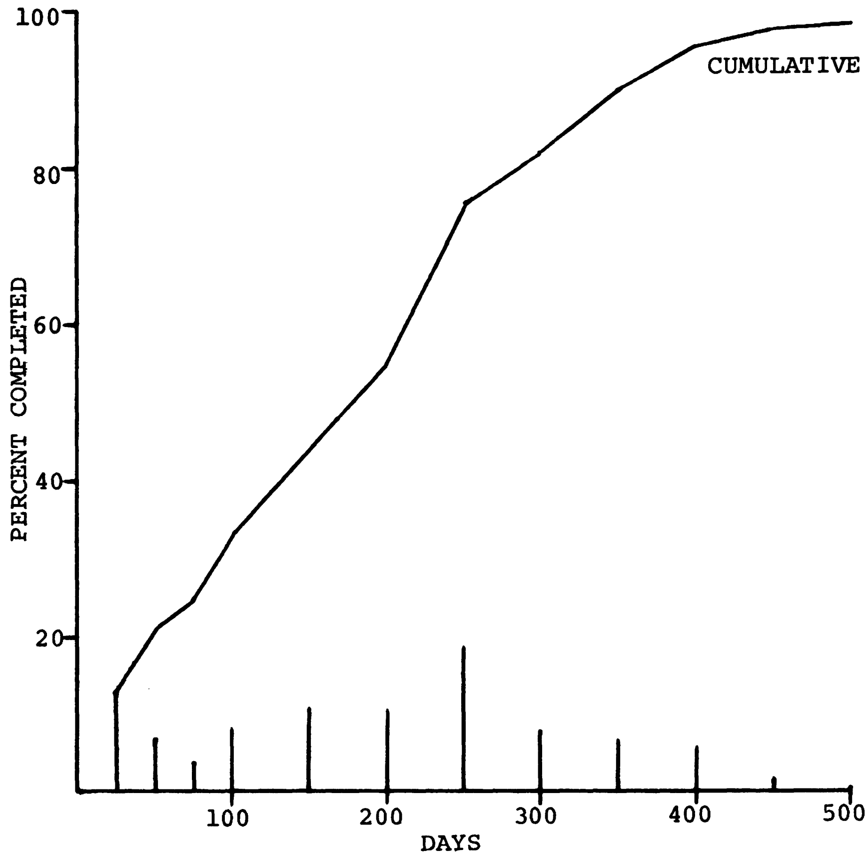


Exhibit 2: Time Lapse Between Filing and Disposition--Felony Trial

(Not available from existing statistical systems)

COMPARISON OF OFFENDERS ENTERING
CORRECTIONS RECEIVING AGENCY FROM FELONY TRIAL

(PERCENT OF ARRESTEES)

CORRECTIONS RECEIVING AGENCY	STATE A (%)	STATE B (%)	STATE C (%)	STATE D (%)	STATE E (%)	STATE F (%)	STATE G (%)	STATE H (%)
STATE INSTITUTION	4.1	1.8	6.0	16.7	37.7	19.9	13.4	10.5
PROBATION AGENCY	4.6		19.2	5.5	6.6	16.6	17.2	7.4
JAIL	6.6		1.8	36.8	13.9	1.1	4.2	12.1
OTHER							1.7	
TOTAL	15.3	1.8	27.0	57.0	58.3	37.6	36.5	30.0

Exhibit 3: Comparison of Offenders Entering Corrections Receiving Agency From Felony Trial (Percent of Arrestees)

DISCUSSION

John P. Conrad, National Institute of Law Enforcement and Criminal Justice

Social scientists have institutionalized a number of clichés about the decline of kinship in an industrial society. The values of clan and tribe have disintegrated; our bulwark against the storms of mass living is the nuclear family. I am reminded of these truths this afternoon because I now see the Law Enforcement Assistance Administration as a large tribe. My nuclear family is the National Institute of Law Enforcement and Criminal Justice; we have been hearing from the members of the family known as the National Criminal Justice Information and Statistics Service. We seem all to have been so busy that I have to come to a reunion like this to find out what my cousins have been up to. As we have heard, they have been planning and executing a statistical revolution. I am a fairly distant relative, but I take a sort of tribal pride in their accomplishments.

For many years we have been listening to laments about the lack of criminal justice statistics. Indeed, I have contributed a lament or two of my own. We have heard today that dearth is to be replaced by plenitude. Perhaps it is ungracious to be uneasy about this prospect, but I have been complaining for so long that I must persist for a while longer. I hope my colleagues will not construe these comments as in any way constituting a deprecation of their efforts. Sociologists make part of their living from a natural and enduring dissatisfaction.

Let me begin with an enormous question, one which no reasonable critic could have advanced before the refinement of statistical apparatus reached its present level of productivity. It runs like this: How Badly Does Who Need What Information For What Purpose and At What Time? All three of our speakers have discussed statistical programs which will produce unprecedented quantities of data about the volume of crime; the rates at which it increases or, please God, declines; its distribution in cities and census tracts; and the disposition of the criminals who commit it. To ask my enormous question is not to imply an answer that huge masses of useless data are to be accumulated. Rather, it is to ask how are those who must administer these systems of the criminal justice systems to use this unprecedented bounty?

I would like to demonstrate the significance of this question with a practical example drawn from experience unrelated to LEAA. Several years ago I was Chief of Research of the California Department of Corrections. This was in many ways a favored position. I was the beneficiary of the incomparable statistical zeal of Ronald Beattie, then and now the Chief of the California Bureau of Criminal Statistics. Using his data and those of our own statistics section, we used to conduct an annual audit of the decisions of the California Adult Authority, as that state's parole board is known. It was a solemn rite in which we addressed ourselves to the impact of parole decisions on prison population, on recidivism and on correctional costs. We prepared carefully and communicated clearly--or so we were told--and received the spellbound attention of the parole board members. One of the principal themes of our findings was that severe decision patterns had no apparent effect on recidivism but did, of course, force prison costs upward. I think our discussion was collaborative and, in a sense, heuristic. Our data were not challenged and our inferences were not contested. Yet I don't think these reviews resulted in any changes one way or another. We might have said that they were an entertaining diversion for the parole board members except that they had unintended consequences.

Eventually the legislature got wind of these findings. New legislation is now under consideration which will require the parole board to show cause when an inmate's sentence is to be extended beyond the minimum term. Our statistics provided finely tuned needle readings for a machine which lacked the capability of adjusting to them. The unexpected and undesired consequence is the radical alteration of the machine by external direction. But what gauges should be installed on the new model? How should they be calibrated? And how can the new machine be fitted to adjust to feedback more or less automatically?

This story illustrates my first point. We must bear in mind <u>who</u> it is who needs the data the statistician <u>can</u> supply and how he is going to use it. Administrators and statisticians must learn to collaborate. This state of affairs will not be easy to bring about. Last night we heard a distinguished statistician remind us that those who need help the most are the last to seek it. So it is in the criminal justice systems. Administrators are uneasy with figures and uncomfortable with the mysterious symbols and formulas of the statistician. They must learn to collaborate, and the statistician must be patient and respectful. There are exceptions. I have heard police chiefs say that they would not wish to work in a department which did not have first rate statistical service. They know how desperately they need data for the decisions they must make about the deployment of their undermanned forces, and to know their effectiveness in planning and budgeting. I have also heard police chiefs say that they see no point at all in statistics of any kind; money spent on statisticians would be better spent on more cops on the streets. Mr. Kolodney has noted that judges take a negligible interest in statistics and there are few signs that this apathy is changing. Correctional administrators have been exposed to statistics for longer than any of their counterparts, but methods for using correctional data effectively are far from standardization.

In the collaboration for which I have been pleading, we need a joint consideration by statistician and administrator of the specific questions on which the administrator needs light and of the power of statistics to shed that light. Who needs information? The administrator needs it, for one, but often he doesn't know it. The statistician must educate him by the dialectic of choices, not by presenting him with what the statistician thinks he needs and can most easily furnish him.

My second issue concerns the costs of crime. I am delighted to hear from Mr. Hall that we are to learn something substantial about this subject. I suppose most researchers in criminal justice have been subjected to disconcerting enquiries in this murky area. Most of us have to plead guilty to having contrived some sort of estimate to cope with importunities which could not be rejected. The resulting volume of misinformation is only beginning to be appreciated. The recent article by Max Singer in Public Interest concerning the persistence of mythical data in the assessment of the impact of narcotics on crime is one outstanding example. Any review of the speculative numbers related to organized crime will uncover another mine of misinformation. These distortions are not harmless; we can be sure that they have an unhealthy impact on our national priorities in acting for crime control. Nearly always they over-simplify extremely complicated problems.

What is needed is not so much the aggregated costs of all crime as the accurate costing of some kinds of crimes. If we had accurate knowledge of the volume and cost of drug-related property crimes we could plan our anti-robbery and anti-burglary efforts much more intelligently. The general public needs information about the costs of crime and the costs of reducing it so that it can vote its priorities with its eye open. The costs of crime control are going to rise along with the costs of other public services. Taxpayers should have factual rather than mythical guidance on the values to be obtained by increasing investments in crime control. They should have a fair idea as to whether the cost of crime will be a reduced burden if the investment in crime control is increased.

Third, I want to touch on Turner's thought-provoking discussion of the difficulties and methodological choices which must be borne in mind in devising victimization studies. In a study which our Institute is sponsoring, we are discovering that victimization and the fear of crime are not nearly as well correlated as we might have thought, at least in Philadelphia where the study is under way. There are, however, indications that the fear of crime is closely correlated to the numbers of delinquents residing in the neighborhood studied. Where the fear of crime is most intense--in some cases amounting to a fear of being on the streets in broad daylight--the incidence of delinquents seems to be highest even though the resident delinquents are not victimizing their neighbors.

I don't know what these preliminary results mean, but I think they already demonstrate the need for designing victimization surveys in forms which permit us to relate the data collected to other parameters. We need to consider not only the essentially criminological variables but also housing, public health, levels of public assistance and a whole gamut of the gauges of urban strain.

The plan for a National Crime Panel described by Mr. Hall will delight criminological researchers throughout the nation. We need the reports on prisons, jails, probation and parole which are promised us, and we need them yesterday or more accurately several decades ago. I am concerned that these reports should tell us what we need to know. The United States is a nation of compulsive consumers of statistics, as any sports page can tell us. We will absorb and consider and even memorize statistics of no conceivable relevance to anything of practical moment. I am sure that we will happily digest reports which authoritatively tell us how many prisoners per 100,000 population each state has, and, state by state, how many probationers per probation officer. We really ought to know these things, even though it's hard to see how that knowledge contributes to a solution of real life problems. What we really need is information which can link decisions to actions to consequences in a chain which can be tested by data. Such a chain calls for enumeration of programs and program consequences. If this sort of information can emerge from the National Crime Panel, we will have the statistical foundation for wise national policy making and effective local decision-making. And in that kind of outcome we can discern the most valuable service of the statistical community to the nation.

DIMENSIONS OF HOUSEHOLD FERTILITY: AN ECONOMIC ANALYSIS

Robert T. Michael, UCLA and NBER*

The study of population has had a long tradition in economics dating at least from the time of Thomas Malthus. Yet until quite recently very little attempt was made to apply the current tools of economic analysis to the area of human fertility. Gary Becker's paper[1] delivered in 1958 marked the beginning of a serious attempt to render fertility behavior comprehensible in the context of an economic framework where households exercise volition over their family size. Currently a number of economists are engaged in research in this area, perhaps partly in response to the current national interest in population growth and its relation to our environment;[2] but partly, too, because the analysis of fertility is an excellent application of recent advances in the theory of household decision making.[3]

The purpose of the present paper is to summarize some of the information available from the NBER-Census Bureau's Consumer Anticipation Survey (CAS)[4] on economic factors related to dimensions of household fertility behavior. The paper first sketches the broad outlines of the analytical framework currently being applied to household decision making with respect to fertility. It then discusses some of the implications of this theory and relates them to empirical findings from the CAS data. The fertility dimensions considered are (1) the number of children in the household; (2) the age interval in which the wife engages in childbearing; (3) the spacing interval between children; and (4) the expected educational attainment of the children. The empirical results are indicated in cross-tabulations by age, education, income and number of children (where appropriate) and in multiple regressions with three to five explanatory economic variables.

The dimension of fertility behavior most extensively analyzed by economists is the household's completed fertility. Very little work has been done, by contrast, on the spacing of children. So this paper in part tests relatively well established hypotheses, while with respect to some other aspects of fertility it is essentially descriptive. It approaches the household's fertility behavior from the perspective of a single-equation model instead of in the context of a model which emphasizes interdependencies related to fertility decisions.[5]

The Analytical Framework

The broad outlines of an economic analysis of fertility may be summarized as follows.[6] The household is the decision making unit which attempts to maximize an objective function subject to constraints on its available resources and on its capacity to convert these resources into the arguments of its objective function. The household is analogous to a small multi-product firm. It is endowed with time, a rate of conversion of time into money (a wage rate), and perhaps non-human wealth which yields property income.[7] It converts these resources into "commodities" through production functions. These commodities are the arguments in its utility function.

Assume one of these commodities to be "childservices," defined as a quality adjusted flow of services from children. The household produces childservices by spending time and money maintaining the child and effecting the child's quality. A given level of childservices can be produced from various combinations of number and quality of children. Economic theory suggests that the factors which might affect the household's demand for childservices include income and relative prices.

Since the production of childservices uses the household's own time as well as purchased goods and services, the opportunity cost of time becomes an important determinant of relative prices and hence of demand. It is usually assumed that the production of childservices is relative intensive in the wife's time. So the relative price of childservices is positively related to the opportunity cost of the wife's time. It is also generally assumed that the production of childservices uses relatively little of the husband's time, so the relative price of childservices is negatively related to his opportunity cost of time. In the cross-section, market prices of goods and services are assumed to be the same for all households. So differences in relative prices of commodities across households are determined by relative time intensities and differences in the opportunity cost of time.

Since the opportunity cost of time is integrally related to earnings, the effects of income on demand for childservices depend critically upon the source of the income.[8] Failure to contend with subtle changes in relative prices which accompany changes in income probably accounts for the widely different estimates of the effect of income on the demand for children in the literature. A final economic variable which may affect both relative prices and real income is the level of "technology" utilized in household production. Increases in the couple's level of education may raise the household's commodity output per unit of input and thereby raise its real wealth; if education affects productivity differently across commodities, relative prices will also be affected.[9]

Since it is not assumed that childservices are proportional to the household's stock of children across households, the implications with respect to childservices do not directly apply to the household's derived demand for children. Indeed substitution between number of children and quality of children is one of the most interesting, albeit difficult, aspects of this framework. To obtain implications related to the number of children, assume that their production is relatively intensive in the wife's time. Then increases in her opportunity cost of time raise the relative price of children as well as the relative price of childservices. So through substitution in production and substitution in consumption, the theory implies a negative relationship between her time value and the household's demand for children.

If we assume the husband's time is used relatively more extensively in the production of child quality, increases in his time value, holding income fixed, induce substitution toward quantity of children and away from higher quality children, while through substitution in consumption the demand for childservices rises. So the model predicts a positive effect of his time value on quantity of children and the effect on quality of children depends upon the strength of the effects of substitution in production (away from quality) and in consumption (toward more childservices and therefore toward higher quality). Increases in income which are not related to the value of time may still affect the relative price of quantity to quality if expenditures on quality are complementary with expenditures on "luxuries."[10] So without additional restrictive assumptions the predictions about the direction of effect of income on the derived demand for children are ambiguous.

The effect of the couple's education level on the relative price of quantity to quality operates through its effects on relative proficiencies of producing quantity of children and quality of children. With respect to quantity the dominant effect is presumed to be through contraceptive efficiency. Expenditures of time and money on contraception are made to prevent the acquisition of additional children. The higher the costs of contraception (for any given level of exposure to the risk of pregnancy) the lower the cost of acquiring an additional child. Said differently, the more it costs to avoid having a child the greater the economic incentive to have the child. Therefore if increases in education lower the expenditure on contraception necessary to achieve any given level of risk of pregnancy, increases in education will, in effect, raise the shadow price of having children.[11] This will induce a shift in the production of childservices toward fewer, higher quality children.

Likewise, if increases in education have a disproportionate, positive effect on the couple's capacity to produce quality in children, this too will raise the relative price of quantity to quality. This latter effect is difficult to substantiate directly without more definitive work on the characteristics of "quality" in children. If we assume "quality" to be positively monotonically related to the child's ultimate level of schooling, the hypothesis could then be more explicit: the higher the couple's education level the greater the marginal product of the couple's time spent with the child on the child's ultimate schooling level.

The previous paragraphs set out a conceptual framework in which the effects of changes in economic factors on the household's demand for childservices and derived demand for children and their quality have been analyzed. In some cases unambiguous predictions about the direction of effects have been made; in other instances opposing influences are identified. The focus has been upon the demand for number of children or quality of children while the empirical results also look at the timing and spacing of children. The framework discussed here will be of use in interpreting these latter results as well.

Household Fertility Patterns from the CAS Data

The household's fertility behavior was not a primary concern of the Consumer Anticipation Survey. Yet in the context of a model which views children as a consumer durable, the present analysis seems to fit comfortably with this survey which does emphasize the ownership and acquisition of durable goods. Since the survey obtained information on only the children under the age of twenty-two, the empirical analysis is restricted to households in which the wife is under the age of 40. It is further limited to husband-wife families in which the husband was not self-employed and was working at a full-time job for 50-52 weeks in 1967. The sample of 1711 households (from the approximately 4600 observations in the first wave of the survey conducted in May 1968) is a relatively high income, well educated group of suburban families. The means (and standard deviations) of key variables from the sample are indicated in the accompanying table.

Summary Statistics, Sample of 1711 Households
Consumer Anticipation Survey

variable	mean	(standard deviation)
Age of husband	34.9	(5.65)
Age of wife	31.8	(4.71)
Education of husband	15.4	(2.34)
Education of wife	13.9	(2.06)
Wage rate of husband	$5.73	($2.14)
Income of husband*	$16,455.	($6207.)
Number of children	2.44	(1.34)

*The "income of husband" variable used throughout this paper is age adjusted. The husband's observed full-time current income (not earnings) is used to project to age 40 his full-time income

(a) The Quality of Children

The CAS data contain information on the number of children in the household (under the age of 22) rather than the more frequently used variable, the number of children ever born to the woman. In a vast majority of cases the two variables will be the same although discrepancies may exist as a result of infant or child mortality, children from previous marriages, adoptions and children over the age of twenty-two.[12]

Table 1 indicates the average number of children in the household at specified age intervals for couples who indicated their fertility was completed.[13] Panel A reflects a negative relationship between the wife's education level and the number of children. Unless the youngest cohort in this table experiences an appreciably reduced completed fertility the table suggests that several of those who indicate their fertility completed will in fact have additional children in the future. Panels B and C suggest that the effect of the wife's education remains negative when either the husband's income or his education level is held constant. Since the opportunity cost of her time is positively related to the wife's education level, this negative relationship is consistent with the model's prediction. The partial effect of income in Panel B appears to be positive while the husband's education level (Panel C) has no clear systematic effect when the wife's education is held constant.

Multiple regressions within age groups are shown in Table 2. Again the wife's education has a negative partial effect on the number of children although it is statistically significant only at the higher age interval. The husband's wage rate has a positive effect on the number of children as predicted. The income variable also has a positive coefficient although its t-value is about 1.75 for both age groups. When both the husband's wage rate and age-adjusted income are included, the coefficients are positive for each in both age groups but neither exhibits statistical significance.[14] For the women aged 35-39 the regression which includes the income variable implies an income elasticity at the point of means of .086 which is not unlike other estimates in the literature. But when the husband's wage is held constant the implied elasticity falls to .058 and the income coefficient's t-value is 0.68. So in studies where a positive and statistically significant income elasticity is observed, if the husband's opportunity cost of time is not held constant, the observed income elasticity may simply reflect this substitution effect.

The effect of the husband's education level on the number of children in these regression is consistently negative although not statistically significant. The two education variables reflect the partial effects of increases in each separately. An alternative interpretation rests on the result from human capital theory that absolute differences in education levels are proportionate to relative differences in full-time earnings.[15] The coefficient on the wife's education level reflects an increase in the difference between her education level and her husband's, so it reflects the effect of an increase in her relative earning-power (or the opportunity cost of time). Summing the coefficients on the wife's education and husband's education reflects an increase in their education level holding her relative time value fixed.[16] Since his wage rate is also held constant in the regression the sum of the two coefficients is interpreted as the effect of education through nonmarket productivity. It is negative and statistically significant in the four regressions in Table 2.[17]

The third regression in Table 2 indicates that an increase in the wife's relative time value lowers the couples fertility (one more year of schooling lowers fertility by six-hundredths of a child on the average). An increase in the couple's level of education holding their market time value constant lowers fertility (one more year of schooling lowers fertility by seven-hundredths of a child). Increases in the husband's wage rate raise fertility (a dollar increase in his hourly wage rate raises fertility by four-hundredths of a child). So these results closely conform to the implications from the theoretical model. The magnitude of the effects though is small -- the residual variation is only slightly reduced and the slope coefficients are quite low. Economic theory predicts the direction of effects of factors which influence relative prices and real wealth but does not imply that these effects will dominate at the individual household level. Yet small effects if systematic and predictable may be relatively important to an understanding of aggregate behavior across cities, regions, or nations or over periods of time, since in the aggregate many of the individual idiosyncrasies which pervade micro data sets will cancel out.

(b) <u>The Timing of Children</u>

The timing and spacing of children are dimensions of fertility which have not received much attention by economists. The theory discussed above focuses on the household's demand for a stock of children and does not directly yield implications regarding the optimal age interval for the production of children or the optimal spacing of a given stock of children. These factors are of more than passing interest however since for a given number of children, changes in the average age at which childbearing takes place can have an appreciable effect on the long run rate of population growth through its effect on the length of a generation.

This paper does not develop a theory of the timing of childbearing although the household production function model would seem a logical framework for such a theory. Instead the same economic variables used in the previous section are again considered in looking at evidence related to the age at which childbearing began and the probability of additional children at specified age intervals.

Tables 3 and 4 deal with the age of the oldest child in the household. For age specific groups, an increase in the age of the oldest child reflects an equal decrease in the age at which childbearing began.[18] So the very consistent negative relationship between the wife's education level and the mean age of the oldest child suggests that more educated women begin their childbearing at a later age. Table 3 also indicates that they do so when the husband's income level or the number of children in the family is held constant as well. The strong negative effect of the wife's education on the age of her oldest child (with the wife's age given) is again seen in the regressions in Table 4 which hold constant several other variables as indicated. (The slope coefficient suggests that an extra year of schooling for the wife postpones the first child by about one-half a year). The regression also indicates that the higher the education level of the husband the later childbearing begins and the effect of increases in the husband's wage rate is to lower the wife's age at the first birth.

Despite the strong statistical relationship indicated in these two tables it is not suggested that the increase in education induced the woman to choose to begin her childbearing later. Instead, both acquiring formal education and raising children are presumably relatively time intensive activities. So while engaged in education, the woman's time value is relatively high which effectively precludes her simultaneously choosing to engage in childbearing. Thus the results probably reflect the sequential nature of the optimal strategy for acquiring education and children. Whether the possession of human capital has any independent effect on the age at which childbearing commences is not clear.[19]

In addition to looking at the age at which childbearing begins, the CAS data also permit us to look at the termination of the childbearing period. The variable is the household's own estimate of its chances of acquiring additional children within the next three years. The presumption is that differences in the household's expectations about its future fertility convey some information about its ultimate fertility behavior.[20] More certainly, the responses convey information about the couple's intentions, ex ante.

Tables 5 and 6 suggest that (holding her age constant) below age 35, increases in the wife's education raise the probability of additional births in the next few years, ceteris paribus. The effect is no longer present at the age interval 35-39 years, where the slope coefficient in the regression is in fact negative. These results along with the information in Tables 3 and 4 are consistent with the more educated woman beginning her childbearing at a later age; being more likely to have another child within any given short interval of time (i.e. three years) from the end of her schooling to, say, age 35; and ending her childbearing no later than her less educated counterpart. That is, the more educated woman may begin childbearing later, space her children closer together and end her childbearing at least no later than less educated women. The following section discusses the spacing of children more directly.

(c) The Spacing of Children

Child rearing is presumed to be relatively time intensive in the wife's time. So the higher the opportunity cost of her time the greater the incentive to compress the time interval in which childbearing takes place, in order to economize on the use of the wife's time. Not only does the wife forego other uses of her time during the period she spends with her children, she may also be disinvesting in her own marketable human capital. If so the more educated woman has an added economic incentive to concentrate her childbearing in a relatively short period of time.

Improvements in contraceptive efficiency may also shorten the time interval of childbearing. With only moderately effective contraception the couple may postpone births to achieve a smaller total family size.[21] But as contraceptive efficiency improves, the incentive to postpone for this purpose is reduced. So if the couple's contraceptive efficiency is positively related to its level of education, as hypothesized above, then aside from effects through the value of time and depreciation of human capital, the more educated may be expected to space their children closer together.

Tables 7 and 8 suggest that more educated women do in fact space their children closer together. For women aged 35-39 with three children, those with four or more years of college space their children about one-half a year closer together on the average than women with no college; for the women with four children the average difference is about two years. The regressions in Table 8 are run across households with the same number of children and the effect of the wife's education is negative though somewhat erratic. Other things held constant the effect of the husband's education is also seen to be negative as predicted although the coefficients do not exhibit much stability in this table.

Table 9 indicates the absolute and relative variation in the spacing of children within the household. The absolute variation is the standard deviation of the time intervals between successive children in the household. The relative variation in this standard deviation divided by the average spacing of the household's children. For the three columns in Table 9 in which all three cells have at least thirty observations, the figures imply that as the wife's education rises the variation in the spacing of children declines.[22] That is, more educated women tend to space their children more evenly. This may be related to contraceptive efficiency if outlying observations reflect contraceptive failures.

(d) The Quality of Children

The theory emphasized the possibility of substituting between quantity of children and quality of children in meeting the household's demand for childservices. It was indicated that the relative price of number of children to quality of children is, under the assumptions of the model, positively related to the wife's opportunity cost of time, to the education level of the husband (through technological effects which lower the cost of contraception and lower the cost of quality production) and to the household's income level (if luxuries are complementary with direct expenditures on quality of children). The relative price is negatively related to the husband's opportunity cost of time (if quality production is more time intensive in his time). So holding the household's demand for childservices fixed the theory implies a shift toward quality as the wife's opportunity cost, the couple's education level, and income rise and as the husband's opportunity cost of time falls.

The difficulty in testing these predicted effects is in holding the demand for childservices constant. Without this, one observes that the net effect of substitution in production between quality and quantity and substitution in consumption toward or away from childservices as discussed above in an earlier section. In the case of the husband's time value, for example, the former implies a negative effect on quality of children while substitution in consumption implies a positive effect. Only in the case of income, with the income elasticity of childservices positive, do these two effects move in the same direction yielding an unambiguous prediction of a positive effect on quality of children. What can be determined from the empirical results, then, is whether the substitution in production or consumption appears to dominate.

Table 10 suggests that the effect of the wife's education on the quality of children -- measured by the level of education the couple expects its oldest child to attain[23]-- is positive. The effect persists when the household's income or the number of children is held constant. Income itself appears to have a positive effect on the child's education, ceteris paribus. If one looks at only those cells in Panel B with at least thirty observations, there appears to be an erratic but negative relationship between the number and education level of children.

In the regressions in Table 11, in most cases the direction of effects of the four variables, although not statistically significant, are consistent with the predicted effects through substitution in production between quantity and quality of children. The wife's relative education level, the couple's education and the husband's income appear to raise the child's education level. The effect of the husband's wage is less clear. On the whole these regressions seem to suggest that the substitution between quantity and quality is of greater magnitude than the substitution between commodities toward or away from childservices.

FOOTNOTES

***This study was funded by NBER but has not been reviewed by the Directors of the National Bureau of Economic Research. I would like to thank Ates Dagli and Charles Lieberman for helpful research assistance.**

[1]Gary S. Becker, "An Economic Analysis of Fertility," in Demographic and Economic Change in Developed Countries, NBER, Princeton University Press, 1960.

[2]See for example the papers delivered at the 1970 American Economic Association meetings in the session "Population and Environment in the United States," American Economic Review, LXI (May 1971), pp. 392-421.

[3]In particular see Gary S. Becker, "A Theory of the Allocation of Time," Economic Journal, 75 (September 1965).

[4]This survey is a panel survey of some 4600 households living in suburban Boston, Minneapolis and San Jose. Its basic objective was an analysis of consumer anticipations regarding durable goods purchases and components of savings. In the first visit of the survey, conducted in May 1968, considerable information was obtained about the household's children under the age of 22 and about expectations regarding the education of the children and the couple's future fertility. Only the information from this first visit is utilized in this paper.

[5]For an example of a simultaneous equations model which treats fertility decisions along with marriage, migration and women's labor force participation decisions, see Marc Nerlove and T. Paul Schultz, Love and Life Between the Censuses: A Model of Family Decision Making in Puerto Rico, 1950-1960, RM-6322-AID, The Rand Corporation, Santa Monica, Calif., September 1970.

[6]There are a number of ongoing research projects, especially PhD. dissertations, which use a similar framework. For a thorough discussion of one such model see, Robert J. Willis, A New Approach to the Economic Theory of Fertility Behavior, mimeo, 1969.

[7]Of course, the amount of human and nonhuman wealth the household has is a function of its previous consumption-investment decisions. These are taken as already determined for the discussion here.

[8]On this point, see Robert J. Willis and Warren Sanderson, "Is Economics Relevant to Fertility Behavior?" presented at the AEA meetings, December 1970.

[9]See my The Effect of Education on Efficiency in Consumption, NBER, 1971 (forthcoming).

[10]If childquality, K, is a function of expenditures of time and money on the child's quality, k, and of the environment to which the child is exposed, e,

$$K = f(k,e)$$

then if $f_{ke} > o$ and if e is positively related to the couple's level of income, the higher their income the greater their incentive to produce childservices with relatively fewer, higher quality children.

Verbally the argument is this: by definition wealthier households spend proportionately more on "luxuries" and empirically these include travel, durable goods and housing. These expenditures for the adults' own use may complement the couple's direct expenditure on the child's quality. If so, they will lower the relative price of quality to quantity of children and hence that relative price will be negatively related to the couple's income.

[11]Considerable evidence exists that more educated couples use contraceptive techniques more extensively, adopt them at an earlier birth interval and tend to select more effective methods. See P.K. Whelpton, A.A. Campbell and J.E. Patterson, Fertility and Family Planning in the United States, Princeton University Press, 1966 for evidence based on a national survey conducted in 1960. Similar surveys conducted in 1955 and 1965 offer additional supporting evidence.

[12]Neither variable precisely reflects the couples derived demand for children since the one excludes children already grown up while the other ignores children not born of the mother. It is also not clear how child deaths or children from previous marriages should be considered in a study of demand. The date of birth of each child ever born is clearly preferable for the study of the timing and spacing of children. This information is not available in the CAS data so the ages of the children in the household have been used.

A later interview in the CAS did ask the respondent "How many children have you ever had." A check has been made for the women age 35-39 to see how closely the response to this question conforms to the information on the children in the household from the survey a year earlier. Adjusting for new births, on which information also exists, of the 425 households for which the comparison was possible about 5 per cent indicated having had a greater number of children than the number currently in their household. In half of these cases the oldest child listed was below the age of 12 and for these it would seem more likely that the discrepancy resulted from a child death than a child over the age of 22. It was also found that in about 1 per cent of the cases the number of children in the household exceeded the number ever born. This could result from adoptions, children born of a previous marriage by the husband or, of course, tabulation error. Some of these issues could be resolved if information were available on the age at marriage or the existence of previous marriages but neither was available in the CAS data.

[13] Couples were asked "Do you think you are likely to have one or more (additional) children at some time in the future?" Only those who answered "No" are designated as having completed fertility. While the usual practice is to consider women of a given age as having completed their fertility, the responses to this question reveal that about 14 per cent of the women 35-39 did not consider their family size completed.

[14] The simple correlation between the two variables is .825 and .847 in the two age subsets so the problem of multicollinearity may explain the relatively high standard errors. A more adequate test of the implication that the husband's price of time and the household's income have separate, distinct effects on fertility could be made using non-earnings income although even then the value of nonmarket time is affected by the amount of market purchases with which the time is employed.

[15] One of the fundamental equations in human capital theory states

$$\ln Y_s \cong \ln Y_o + rS$$

where Y is income, r is the rate of return to years of schooling, S is the number of years of schooling and the subscript designates the number of years of schooling. (See Gary S. Becker, Economic Theory, A.A. Knopf, 1971, p. 180.) So for the wife (w) and her husband (h), assuming $Y_{ow} = Y_{oh}$ and both have the same rate of return:

$$\ln \left(\frac{Y_{sw}}{Y_{sh}}\right) = r(S_w - S_h).$$

[16] If the two terms in the regression should represent the wife's relative time value and the husband's education level:

$$b_1(S_w - S_h) + c_1 (S_h)$$

and the regression run includes instead:

$$b_2(S_w) + C_2(S_h)$$

then the relevant coefficients are estimated as $b_1 = b_2$ and $c_1 = c_2 + b_2$.

[17] The values for the four regressions in Table 2 are: -.075 (.034) and -.078 (.035) for the regressions on women age 30-34 and -.072 (.028) and -.074 (.028) for the women age 35-39. The standard errors can be calculated from the variances and the covariance of the regression coefficients. In all four cases the slope coefficient is statistically significant at conventional levels of confidence.

[18] Two deficiencies exist in the data used here. First, the information does not relate to the woman's first live birth but rather the oldest child in the household. Second, the wife's age is indicated only by five year intervals. Since a principal interest here is in the education levels of the wife and husband it is important to consider whether the upward trend in education is sufficient to tend to place the more educated individuals in an age interval at the lower ages in the interval. It does not appear to be the case. For white women in the United States in March 1970 the median level of schooling was 12.5 for those in the 25-29 age interval as well as for those in the 30-34 age interval; the figure was 12.4 for those in the 35-44 age interval. (See Current Population Reports, P-20, November 30, 1970.)

[19] Another aspect of this issue is the age at marriage. Although this information was not available in the CAS data, the 1960 Growth of American Families (GAF) survey contained evidence on the wife's age at marriage and the time span between marriage and the first birth. For the 1931-1935 cohort, the average of the wife at the time of marriage rose from 18.5 for the grade school group to 19.7 for the group with four years of high school to 21.6 for those with four or more years of college. [It is interesting to note that Census data (CPR, Series P-20, No. 198, March 25, 1970) suggests a weaker relationship between husband's age at marriage and his education than the relationship seen here between wife's age at marriage and her education. This difference is consistent with an early marriage adversly affecting the wife's relative educational attainment. Another explanation for the observed difference is that the GAF data refer to means while the Census data use medians and the distribution of age of marriage is probably more negatively skewed at lower levels of education.]

The average number of months between marriage and first birth in the GAF data rose with the wife's education from 21 months for the grade school group to 28 months for the group with any college training (the comparable figures for couples using contraception during that birth interval were 24 months to 33 months). Similarly the percentages of couples with a first birth prior to their first wedding anniversary fell with education from 41% of the grade school group to 24% of the college group (see Whelpton, P.K., A.A. Campbell and J.E. Patterson, Fertility and Family Planning, 1966, pp. 320-329).

[20] From subsequent waves of this panel survey one can determine how accurately the household predicted its fertility in the next few years. This analysis has not been completed at this time.

[21] Keyfitz points out that the level of contraceptive efficiency required to assure no accidental pregnancies over a long period of time is surprisingly high. For a couple employing a contraceptive technique with 99 per cent effectiveness over a twenty-year period from say age 20 to 40, the chances of a pregnancy are about 40 per cent (for a ten-year period the probability falls to about 21 per cent). At a 99.9 per cent level of effectiveness the probability of a pregnancy in twenty years is about 5 per cent. (These figures are based on an assumed constant per month probability of conception of .2 for women exposed to the risk of conception.) See Nathan Keyfitz, "How Birth Control Affects Births," Social Biology, June 1971. Not all the above calculations are taken from Keyfitz's paper.

[22] In regressions not shown here on the standard deviation (which follow the format of the regressions in Table 8) the wife's education variable was always negative althouth significant only for the 30-34 age women with three children.

TABLE 1

NUMBER OF CHILDREN

Table includes only households not likely to have (additional) children at any time in the future.

A

Education of Wife	Age of Wife < 30	30-34	35-39	
≤ 12	2.45	3.06	3.13	2.98
13-15	2.21	2.83	2.85	2.75
≥ 16	2.32d	2.83	2.77	2.75
	2.36	2.93	2.94	2.85 (1033 obs)

B

Education of Wife	Income of Husband < 12,000	12-16,000	≥16,000	
Age of Wife: 30-34				
≤ 12	3.04	2.82	3.41	3.06
13-15	2.88d	2.82	2.82	2.83
≥ 16	2.17f	2.74	2.98	2.83
	2.92	2.80	3.06	2.93 (366 obs)
Age of Wife: 35-39				
≤ 12	3.06	3.03	3.32	3.13
13-15	2.59	2.86	2.97	2.85
≥ 16	2.22d	3.06	2.78	2.77
	2.81	2.98	3.01	2.94 (513 obs)

C

Education of Wife	Education of Husband ≤ 12	13-15	16	≥17	
Age of Wife: 30-34					
≤ 12	3.02	2.89	3.10	3.27	3.06
13-15	3.44f	2.86d	2.96	2.53	2.83
≥ 16	3.00f	3.20f	2.82	2.80	2.83
	3.09	2.90	2.96	2.84	2.93 (366 obs)
Age of Wife: 35-39					
≤ 12	3.00	3.26	3.31	2.86d	3.13
13-15	2.67d	3.00	3.08	2.54	2.85
≥ 16	1.40f	2.77e	2.84	2.81	2.77
	2.87	3.09	3.07	2.73	2.94 (513 obs)

Code: f indicates cell size 1-9
e indicates cell size 10-19
d indicates cell size 20-29
All others have 30 or more observations.

TABLE 3

AGE OF THE OLDEST CHILD

Table includes only households with at least one child.
(in years)

A

Education of Wife	Age of Wife ≤ 29	30-34	35-39	
≤ 12	4.79	9.55	12.99	9.51
13-15	3.77	8.43	11.76	8.17
≥ 16	2.91	6.57	10.09	6.83
	3.93	8.26	11.77	8.30 (1591 obs)

B

Panel B includes only households not likely to have additional children at any time in the future.

Education of Wife	Income of Husband < 12,000	12-16,000	≥16,000	
Age of Wife: 30-34				
≤ 12	10.48	9.15	10.41	9.94
13-15	9.09d	9.49	8.72	9.06
≥ 16	8.50f	7.68	7.94	7.88
	9.94	8.84	8.97	9.13 (364 obs)
Age of Wife: 35-39				
≤ 12	13.51	13.48	13.31	13.44
13-15	11.84	11.89	12.25	12.03
≥ 16	9.93	10.22	10.79	10.43
	12.50	11.96	12.08	12.16 (506 obs)

C

Education of Wife	Number of Children 1	2	3	4	
Age of Wife: 25-29					
≤ 12	3.06	5.36	7.50	8.35e	
13-15	1.60	4.86	6.35d	6.00f	
≥ 16	1.40	3.94	5.40e	6.33f	
	1.86	4.74	6.63	8.00d	4.29 (372 obs)
Age of Wife: 30-34					
≤ 12	7.27e	8.18	9.73	10.97	
13-15	6.08e	7.24	9.07	9.85d	
≥ 16	2.50e	5.52	8.07	9.27e	
	5.06	6.84	9.03	10.27	8.01 (503 obs)
Age of Wife: 35-39					
≤ 12	9.00e	11.69	13.20	14.65	
13-15	8.43f	10.77	11.70	13.47	
≥ 16	8.38e	8.68	10.82	10.91	
	8.65	10.50	12.02	13.24	11.49 (519 obs)

TABLE 5

PROBABILITY OF ACQUIRING A CHILD WITHIN THE NEXT THREE YEARS*

(in per cent; 100 = certainty)

A

Education of Wife	Age of Wife < 25	25-29	30-34	35-39	
≤ 12	62	37	11	8	20
13-15	76	57	20	3	29
≥ 16	84[d]	66	29	7	34
	70	53	19	6	27
					(1711 obs)

B

Panel B includes only households with one or more children.

Education of Wife	Income of Husband < 12,000	12-16,000	≥16,000	
Age of Wife: under 30				
≤ 12	33	44	43	40
13-15	51[d]	56	59	57
≥ 16	59[e]	55	66	63
	42	50	57	52
				(458 obs)
Age of Wife: 30-34				
≤ 12	12	7	11	10
13-15	21	19	21	21
≥ 16	44[e]	22	29	28
	19	15	22	19
				(550 obs)
Age of Wife: 35-39				
≤ 12	11	3	6	7
13-15	4	4	3	4
≥ 16	12	2	9	7
	10	3	6	6
				(583 obs)

C

Education of Wife	Education of Husband ≤12	13-15	16	≥17	
Age of Wife: under 30					
≤ 12	44	42	47	49	45
13-15	68[e]	53	68	54	61
≥ 16	37[f]	77[f]	72	64	68
	48	50	65	58	57
					(551 obs)
Age of Wife: 30-34					
≤ 12	15	11	10	5	11
13-15	16[e]	18	19	23	20
≥ 16	43[f]	35[e]	31	26	29
	16	16	22	20	19
					(563 obs)
Age of Wife: 35-39					
≤ 12	5	8	11	9[d]	8
13-15	2[d]	0	3	7	3
≥ 16	17[f]	4[e]	7	8	7
	5	4	7	8	6
					(597 obs)

TABLE 7

TIME BETWEEN BIRTH OF OLDEST AND YOUNGEST CHILD

(in years)

Education of Wife	Number of Children 2	3	4	5	
Age of Wife: 25-29					
≤ 12	2.73	5.10			
13-15	2.77	4.54[d]			
≥ 16	2.25	4.33[e]			
	2.59	4.73			(223 obs)
Age of Wife: 30-34					
≤ 12	3.03	5.86	7.28	8.69[e]	
13-15	2.70	5.53	6.96[d]	8.50[f]	
≥ 16	2.68	4.95	6.91[e]	7.67[f]	
	2.80	5.48	7.10	8.41[d]	
					(484 obs)
Age of Wife: 35-39					
≤ 12	3.81	6.04	8.79	10.78[d]	
13-15	3.03	5.85	8.07	9.78[f]	
≥ 16	3.02	5.52	6.81	9.67[f]	
	3.31	5.82	8.02	10.37	
					(523 obs)

TABLE 9

VARIATION IN THE SPACING OF CHILDREN

Standard deviation* (in years) and coefficient of variation (%)

Table includes only households with a positive time span between the ages of the oldest and youngest child.

Education of Wife	Number of Children 3	4	5	
Age of Wife: 30-34				
≤ 12	1.08	0.90	1.06[e]	
	(32.0)	(34.7)	(42.3)	
13-15	0.92	0.99[d]	0.74[f]	
	(29.3)	(36.7)	(30.0)	
≥ 16	0.72	0.95[e]	0.75[f]	
	(27.7)	(38.2)	(35.0)	
	0.93	0.94	0.90[d]	
	(29.9)	(36.0)	(37.0)	(293 obs)
Age of Wife: 35-39				
≤ 12	1.05	1.28	1.55[d]	
	(34.7)	(39.2)	(51.3)	
13-15	0.99	1.30	0.84[f]	
	(28.9)	(43.3)	(30.0)	
≥ 16	0.71	0.84	0.92[f]	
	(23.8)	(34.4)	(30.0)	
	0.92	1.16	1.28	
	(29.6	(38.9)	(42.9)	(328 obs)

TABLE 10

THE NUMBER OF YEARS OF SCHOOLING EXPECTED TO BE COMPLETED BY THE OLDEST CHILD*
(in years)

A

Panel A includes only households with one or more children.

Education of Wife	Income of Head: <12,000	12-16,000	≥16,000	
Age of Wife: under 30				
≤ 12	14.8	15.2	15.2	15.1
13-15	15.4[d]	15.9	15.5	15.6
≥ 16	16.2[e]	16.3	16.2	16.2
	15.2	15.7	15.7	15.6
				(458 obs)
Age of Wife: 30-34				
≤ 12	15.5	15.5	15.8	15.6
13-15	15.5	15.7	16.1	15.9
≥ 16	16.1[e]	16.2	15.7	15.9
	15.6	15.8	15.8	15.8
				(550 obs)
Age of Wife: 35-39				
≤ 12	15.2	15.3	15.8	15.4
13-15	15.5	15.8	16.1	15.8
≥ 16	15.6	15.8	16.2	16.0
	15.3	15.6	16.0	15.7
				(583 obs)

B

Education of Wife	Number of Children: 1	2	3	4	5	
Age of Wife: 25-29						
≤ 12	15.5	15.2	15.0	15.1[e]	16.0[f]	15.2
13-15	15.4	15.8	16.0[d]	16.0[f]	16.0[f]	15.7
≥ 16	16.2	16.2	16.1[e]	16.0[f]	16.0[f]	16.2
	15.8	15.7	15.6	15.2[e]	16.0[f]	15.7
						(376 obs)
Age of Wife: 30-34						
≤ 12	15.2[e]	15.7	15.6	15.6	15.5[e]	15.6
13-15	15.5[e]	15.8	16.0	16.0[d]	15.8[f]	15.9
≥ 16	15.1[e]	16.2	15.7	16.3[e]	16.2[f]	15.9
	15.3	15.9	15.7	15.8	15.7[d]	15.8
						(530 obs)
Age of Wife: 35-39						
≤ 12	15.6[e]	15.6	15.5	15.3	14.6[d]	15.4
13-15	16.3[f]	15.9	15.9	15.4	16.0[f]	15.8
≥ 16	16.2[e]	15.7	16.1	16.1	15.7[f]	15.9
	15.9	15.7	15.8	15.5	15.1	15.7
						(557 obs)

*The figures are obtained from the response to the question: what grade or year of schooling is the child expected to eventually complete?

TABLE 2

Number of Children

Regressions include only households not likely to have (additional) children at any time in the future.

Mean and (st. deviation) of dep. var.	Education of husband	Education of wife	Wage rate of husband	Income of husband	R^2 s.e.e.
	Age of wife: 30-34 (366 observations)				
2.93	**-.023**	**-.052**	**.057**		**.020**
(1.084)	**(.028)**	**(.032)**	**(.028)***		**1.078**
	-.026	**-.053**		**.002**	**.017**
	(.029)	**(.032)**		**(.001)**	**1.080**
	Age of wife: 35-39 (513 observations)				
2.94	**-.012**	**-.059**	**.042**		**.017**
(1.196)	**(.026)**	**(.028)***	**(.025)**		**1.189**
	-.012	**-.062**		**.002**	**.018**
	(.026)	**(.028)***		**(.001)**	**1.189**

* implies statistical significance at 95% level of confidence (two-tail test)

TABLE 4: Age of the Oldest Child [1]

Mean and (st. deviation) of dep. var.		Education of husband	Education of wife	Wage rate of husband	Income of husband	R^2 s.e.e.
		Age of wife: under 25 (80 observations)				
1.94	(1)	-.058	-.251	.194		.108
(1.487)		(.075)	(.120)*	(.151)		1.432
	(2)	-.017	-.279		-.000	.089
		(.074)	(.120)*		(.003)	1.447
		Age of wife: 25-29 (378 observations)				
4.35	(3)	-.103	-.523	.048		.133
(3.007)		(.076)	(.087)*	(.086)		2.811
	(4)	-.067	-.517		-.002	.134
		(.079)	(.088)*		(.003)	2.809
		Age of wife: 30-34 (550 observations)				
8.26	(5)	-.147	-.640	.180		.187
(3.353)		(.066)*	(.070)*	(.063)*		3.031
	(6)	-.141	-.639		.004	.179
		(.068)*	(.070)*		(.002)	3.046
		Age of wife: 35-39 (583 observations)				
11.77	(7)	-.195	-.449	.213		.098
(4.001)		(.079)*	(.083)*	(.074)*		3.810
	(8)	-.154	-.453		.003	.088
		(.079)	(.083)*		(.003)	3.832

1. Regressions include only observations with one or more children.

TABLE 6: Probability of Additional Children within the Next Three Years

Mean[1] and (st. deviation) of dep. var.		Education of husband	Education of wife	Wage rate of husband	Income of husband	R^2 s.e.e.
		Age of wife: under 25 (124 observations)				
7.0	(1)	.207	.410	-.372		.071
(4.00)		(.172)	(.241)	(.344)		3.91
	(2)	.145	.450		-.001	.062
		(.172)	(.241)		(.006)	3.92
		Age of wife: 25-29 (427 observations)				
5.3	(3)	.136	.543	-.110		.070
(4.50)		(.113)	(.126)*	(.128)		4.36
	(4)	.051	.535		.006	.073
		(.117)	(.126)*		(.004)	4.35
		Age of wife: 30-34 (563 observations)				
1.9	(5)	-.083	.423	-.055		.050
(3.61)		(.076)	(.080)*	(.073)		3.53
	(6)	-.079	.425		-.002	.049
		(.078)	(.080)*		(.003)	3.53
		Age of wife: 35-39 (597 observations)				
0.6	(7)	.121	-.057	-.123		.018
(2.30)		(.047)*	(.049)	(.044)*		2.28
	(8)	.122	-.052		-.005	.019
		(.047)*	(.049)		(.002)*	2.28

1. The probability is the mean times ten (7.0 implies a mean of 70%).

TABLE 8: Time (in years) Between Birth of Oldest and Youngest Child

Mean and (st. deviation) of dep. var.		Education of husband	Education of wife	Wage rate of husband	Income of husband	R^2 s.e.e.
		Age of wife: 30-34				
		Households with three children (196 obs.)				
5.5	(1)	-.006	-.195	-.027		.031
(2.270)		(.080)	(.090)*	(.082)		2.252
	(2)	.007	-.187		-.002	.033
		(.081)	(.090)*		(.003)	2.249
		Households with four children (70 obs.)				
7.1	(3)	-.040	-.099	.044		.010
(2.079)		(.134)	(.183)	(.129)		2.115
	(4)	-.020	-.084		-.002	.010
		(.139)	(.181)		(.004)	2.115
		Age of wife: 35-39				
		Households with three children (180 obs.)				
5.8	(5)	-.191	-.005	.028		.035
(2.263)		(.091)*	(.099)	(.081)		2.242
	(6)	-.172	.000		-.001	.036
		(.090)	(.100)		(.003)	2.241
		Households with four children (110 obs.)				
8.0	(7)	-.001	-.395	.046		.091
(2.825)		(.149)	(.151)*	(.126)		2.731
	(8)	.046	-.403		-.003	.094
		(.146)	(.150)*		(.004)	2.727

TABLE 11: Number of Years of Schooling Expected to be Completed by the Oldest Child [1]

Mean and (st. deviation) of dep. var.		Education of husband	Education of wife	Wage rate of husband	Income of husband	R^2 s.e.e.
		Age of wife: 30-34 (550 observations)				
15.8	(1)	.074	.035			.014
(1.728)		(.037)*	(.039)			1.718
	(2)	.068	.033	.031		.015
		(.037)	(.040)	(.036)		1.719
	(3)	.070	.034		.000	.014
		(.038)	(.040)		(.001)	1.720
		Age of wife: 35-39 (583 observations)				
15.7	(4)	.151	.033			.052
(1.790)		(.034)*	(.038)			1.746
	(5)	.132	.033	.057		.056
		(.036)*	(.038)	(.034)		1.743
	(6)	.127	.031		.003	.059
		(.036)*	(.038)		(.001)*	1.741

1. Regressions include only observations with one or more children.

AN INTEGRATED MODEL OF CONSUMPTION AND MARKET ACTIVITY: THE CHILDREN EFFECT*

Michael Landsberger, National Bureau of Economic Research, and the Technion - Israel Institute of Technology.

The usual approach to the theory of the consumption function is to assume a one-sided relation: family's income is supposed to determine its consumption. To be sure the relevant concept of income was, and still is, subject to a considerable disagreement among economists (see Mayer) but it seems that the one-sided relation is widely and may be even generally accepted. This approach ignores the importance of family members' time as a scarce productive factor.

Becker (1965) and Mincer (1960, 1963) were apparently the first to recognize the importance of this factor and have incorporated it into economic theory. This ingenous idea opened new avenues for investigations into the subject of the consumption function. A consistent approach to the problem along the ideas of Becker and Mincer suggests a simultaneous determination of consumption and earnings of family members. In view of the fact that earnings comprise the major portion of family's income, it may be argued that family's income and consumption are determined simultaneously. This is in line with an assessment made long ago by Mincer in which he expressed the desirability of a "set of simultaneous relations which undoubtedly exist between income, consumption, labor supply, and population" (Mincer, 1963, p. 26).

An attempt was made in this paper to show how the simultaneous determination of family's consumption and earnings is systematically affected by the presence of children of various ages. The outline of the paper is as follows: In section (a) a static version of the model was presented. Then in section (b) it was shown how the children effect was derived from the equilibrium conditions. This effect reflects a systematic relation between children's age and family's consumption and market activity. The propositions emanating from the analysis imply a negative effect on consumption by young children, and this effect declines systematically (in absolute terms) and ultimately in the case of older children it becomes positive. A similar effect was obtained with respect to the female's market activity.

*This paper was written while I was a research associate at the National Bureau of Economic Research. I owe a debt of gratitude to many of my colleagues at the NBER for their useful comments on various drafts of this paper. In particular I have to mention Jacob Mincer who initiated me into the subject and was very patient listening to my reservations. Discussions with John Hause, Thomas Juster, and Robert Willis were also very useful and I wish to thank them. I solely remain responsible for any remaining errors.

Empirical analysis of the theoretical propositions were presented in section (c). The findings do show an unequivocal support of the theoretical propositions.

(a) The Model

Following Becker's seminal paper (1965) it can be assumed that the family acts like a small factory where family members' combine their time and goods bought on the market in order to produce what Becker calls "commodities". These commodities enter directly into the utility function that the family tries to maximize. To make things easier all the goods and commodities were aggregated. Thus, the aggregate commodity is produced by the aggregate good bought on the market (which is what is usually referred to as consumption) combined with husband's and wife's time as inputs (a similar approach was taken up by Ghez, 1970. Gronau 1970 too followed the approach of regarding the household as a unit which combines goods and family members' time to produce various activities. But Gronau dealt with a very different subject than that discussed in this paper).

$$X = f(Y_1, Y_2, Y_3) \tag{1}$$

Y_1 = the aggregate good
Y_2 = husband's time devoted to nonmarket activities
Y_3 = wife's time devoted to nonmarket activities
X = the aggregate commodity.

Maximizing $U(X)$ is equivalent to maximizing X and therefore it may be assumed that the family tries to maximize

$$X = f(Y_1, Y_2, Y_3) \tag{2a}$$

$$\text{S.T.}\quad P_1Y_1 = V + (\bar{Y}-Y_2)P_2 + (\bar{Y}-Y_2)P_3$$

$$\text{or}\quad \sum_{i=1}^{3} P_iY_i = V + (P_2+P_3)\bar{Y} \tag{2b}$$

P_i = price of the i-th productive factor, i=1,2,3
P_2,P_3 = the market wage rates of the husband and wife
$\bar{Y}$ = total time available for market and nonmarket activities, like twenty-four hours a day, seven days a week, etc.
V = nonearning income.
P_1 = price of the aggregate good.

The left-hand side of (2a) is the cost of buying the aggregate good Y_1 on the market, and the right-hand side represents the resources available for that purpose. In (2b) the left-hand side is the cost to the family of the productive factors used in the process of producing the aggregate commodity X. The family buys its members' (husband and wife) time at the market wage rates. The right-hand side of (2b) is the maximum income available had family members' spent their total available time in market activity.[1]

The model as presented in (1) and (2) has no reference to savings nor to any planning for the future. This is of course a simplification but it was felt that at this stage it was justified in view of the high payoff of the model in its present form in terms of theoretical propositions which can be easily subjected to an empirical test. It can be shown that the conclusions reached here remain intact under a dynamic setup where savings are included (see Landsberger 1972). Maximizing (1) subject to (2b) results in the necessary conditions for an extremum.[2]

$$\frac{\partial f}{\partial Y_i} - \lambda P_i = 0 \qquad i=1,2,3 \qquad (3)$$

$$\sum_{i=1}^{3} Y_i P_i - V - (P_2+P_3)\bar{Y} = 0$$

where λ is the Lagrange multiplier.

Under the usual assumptions of convexity the equations in (3) are sufficient for determining the optimal values of the unknowns Y_1, Y_2, Y_3, λ, in terms of the parameters of the system P_1, P_2, P_3, V, $\bar{Y}$.

$$Y_i^* = g^{(i)}\ (P_1,P_2,P_3,V,\bar{Y}) \qquad (4)$$

for $i=1,2,3$ and

$$\lambda^* = g(P_1,P_2,P_3,V,\bar{Y})$$ the asterisks denote optimal values of the variables.

Remembering that Y_1 can be referred to as consumption, $g^{(1)}$ in (4) is the consumption function and $g^{(2)}$, $g^{(3)}$ are the demand functions for the male's and female's home services. Thus, consumption and earnings of the family are determined simultaneously as part of the optimal time allocation of family members.

What has been done heretofore is merely a convenient way of formalizing the ideas of Becker and Mincer, although the presentation of the problem is more in spirit with the approach taken up by Becker. As noted before Ghez, 1970 has followed a very similar approach. Relying on the first order conditions for an extremum an attempt was made in the next section to derive a set of propositions which relate in a systematic fashion children's age with the level of consumption and market activity of the male and female.

(b) The Children Effect

The effect of children on consumption has been largely ignored in the various theories of the consumption function. In empirical studies the number of children has been sometimes indirectly inserted as a variable in regression analysis. Most frequently a variable indicating the number of persons in the family (family size) was inserted and no distinction was made between adults and children, not to speak about a distinction between children of different ages. As far as I recall this has always been done with no particular theoretical reference, with the exception of Ghez, 1970, Ch. IV. However, Ghez did not make any distinction between children of different ages which is the main issue in this paper.

Children's age as a factor affecting the level of market activity of the male has been also ignored in the literature. But as opposed to consumption this subject as a whole was not discussed in a systematic way.

Market activity of the female was intensively covered in the literature (Mincer 1962, Cain 1966, Bowen and Finegan 1969). It appears that Mincer and Cain were mostly interested in wage rate effects. Mincer made a remark about the interaction between young children and wage rate effect (Mincer 1962, p. 92). Bowen and Finegan have discussed the effect of children's age on the female's market activity. It appears that the findings presented in this paper with respect to the relation between children's age and the female's market activity are very much in accord with the conclusions reached by Bowen and Finegan whose study had a predominant empirical orientation. Rewriting (3), the conditions for an extremum can be presented as

$$\frac{P_1}{P_2} = \frac{MP_1}{MP_2}, \quad \frac{P_1}{P_3} = \frac{MP_1}{MP_3}, \quad \frac{P_2}{P_3} = \frac{MP_2}{MP_3} \qquad (5)$$

where MP_i stands for the marginal product of the i-th factor.

It can be assumed that an increase in the number of young children in the family causes an increase in the productivity of home activity of the female. Her home activities are regarded now as being more important and the demand curve for her home services shifts to the right. As a result of it, at the former equilibrium level of $Y_3 = Y_3^*$ the marginal productivity - MP_3 is higher than the wage rate, $MP_3 > P_3$. It can also be argued that as a result of having young children not only is there an increase in MP_3 but also a decrease in MP_1, namely, a decrease in the marginal productivity of the aggregate good (consumption). Although there is an increase in the demand for goods which are related to the "needs" of the child, but this can be more than offset by a decrease in the demand for goods by the parents. This is because having young children implies usually a decrease in demand for such commodities as trips abroad, meals in restaurants and other kinds of entertainments which are very good intensive. Thus the demand curve for the aggregate good shifts to the left, and at the former equilibrium point $Y_1 = Y_1^*$ $MP_1 < P_1$. In addition it may be assumed that in view of the new composition of activities within the household there is a decrease in the demand for home activities of the husband which means a decrease in the MP_2 at $Y_2 = Y_2^*$. The rationale for this assumption is explained later on, however, it should be remarked that this assumption can be dropped without detracting from the significance of the main propositions of this study.

As it is accepted in economic theory the marginal productivity of one factor depends on the amounts of the other factors employed in the productive process, thus it may be claimed that a movement of one curve causes shifts of the other two. However, it may be assumed that this chain reaction is of secondary importance and the shifts discussed above reflect already these effects. As a result of these shifts in the demand curves the equilibrium conditions (5) are replaced by

$$\frac{P_1}{P_3} > \frac{MP_1}{MP_3}, \quad \frac{P_2}{P_3} > \frac{MP_2}{MP_3}, \quad \frac{P_1}{P_2} \gtreqless \frac{MP_1}{MP_2}. \qquad (6)$$

If remembered that in the context of the model presented in this study Y_1, Y_2, and Y_3 are inputs in a production function, it becomes clear that the change in the marginal productivities of the inputs will cause substitution effects. Y_1 and Y_2 will be substituted by Y_3 which became more productive. The meaning of these movements is an increase in home activity of the female, a decrease in that activity by the male and a decrease in consumption of the family. As the reader can verify it to himself by drawing the appropriate diagrams (see Landsberger 1972) these movements tend to restore equilibrium in (6). Recalling that home and market activities Y_i and $\hat{Y}_i$, respectively add up to a constant $\bar{Y}$, it comes out that young children tend to reduce the level of market activity of the female, increase the level of this activity in the case of the male and decrease the consumption of the family.

The effect of older children differs mainly because the change in the importance of woman's home activity is much smaller than before and on the other hand the importance of goods increases. Older children have a higher demand for commodities which are good (or money) intensive (like summer camps, clothes, and other expenditures related to education of the children). On the other hand, older children require less mother's care. The presence of older children does not impose on the parents various restrictions which young children (babies) do impose, and thus the parents may renew their stronger demand for good intensive commodities (such as noted before). Thus, the older the children the smaller are the shifts in the demand curves. Eventually, in the case of children of, say, thirteen years old or more the predominant factor is the increased demand for goods (consumption). The demand curve for the female's home services may not shift at all or even moves to the left with the result of older children having a positive effect on the level of market activity of the female.

It is imperative to the analysis here to realize that the movements of the three demand curves (for Y_1, Y_2 and Y_3) are interrelated. The position of the demand curve for the aggregate good $-Y_1$ determines the money expenditures of the family. The position of the demand curves for Y_2 and Y_3 determine the earnings of the family and its income for a given level of non-earning income. Thus, the shifts of the curves have to be simultaneously determined by the family.[3]

In terms of the equilibrium conditions (5) the effect of children of say 13 years old or more gives rise to the inequality.

$$\frac{P_1}{P_2} < \frac{MP_1}{MP_2}, \quad \frac{P_1}{P_3} < \frac{MP_1}{MP_3}, \quad \frac{P_2}{P_3} \gtreqless \frac{MP_2}{MP_3} \qquad (7)$$

The way equilibrium is restored can be described as follows: A movement to the right along the new (higher) demand curve for Y_1 results in a lower MP_1. A movement to the left along the new (lower) demand curves for the male's and female's home services results in higher values of MP_2 and MP_3. As the reader can verify to himself, these are movements in the right direction in terms of restoring equilibrium in (7). This equilibrium is being reached at a higher level of family's consumption and market activity of the male and female.

Before concluding this section a short summary of the children effect may be useful. Young children have a negative effect on the consumption of the family while at the same time cause an increase in the level of market activity of the male and a decrease in the market activity of the female. The effect on consumption increases (in absolute terms) with the age of the children and eventually becomes positive. At the same time the older the children the lower the decrease in the market activity of the female and finally this effect may become positive. Children of all ages have a positive effect on the level of market activity of the male, and it is very likely (although not necessary) that this effect is higher the younger the children.

(d) *Empirical Test of the Theoretical Propositions Concerning the Children Effect*

The empirical analysis presented in this study is based on cross-section data, composed of two surveys: Consumer Anticipation Survey (first visit) conducted by the National Bureau of Economic Research (henceforth, the NBER Survey), and the 1969 Survey of Consumer Finances, conducted by the Institute for Social Research at the University of Michigan (henceforth, the Michigan Survey). A detailed explanation of the data can be found in Landsberger 1972.

As shown in (4) the independent variables in the behavioral equations are P_1, P_2, P_3, V, and $\bar{Y}$. The first and last independent variables P_1 and $\bar{Y}$ can be eliminated because it is clear that in a cross-section their value are the same for all observations. Data for P_3 was available only for about 30-40 per cent of the observations because most women (60-70 per cent) did not have jobs on the market at the survey period. Therefore, it was decided, in this case, to use the female's education ED(W) as a proxy for her potential wage rate. The female's wage rate P_3 was inserted into the estimating equation only in the case where the estimation was done for those families where the woman had a job on the market. This calculation was based of course on a much smaller number of observations.

The fact that various families have children of different ages was utilized to evaluate the children effect. Having children of different ages can be interpreted as being on different demand curves for Y_1, Y_2, and Y_3, and therefore having different equilibrium values of the Y's.

It was decided to distinguish between three age groups of children: below six, six to twelve, and thirteen plus. The considerations which affected this division are related to the theoretical arguments as presented in section (b). It seems plausible that children below six cause the most significant increase in the demand for female's home services and decrease in the demand for consumption. On the other hand, children thirteen plus years old can be regarded as being able to provide themselves with various home services and therefore do not restrict, or only very little, the female's market activity. On the other hand, they cause an increase in the demand for consumption, because of their strong demand for good intensive commodities (see section (b)). Children in this age group are expected to have a positive effect on family's consumption.

With respect to children in the intermediate group six to twelve the effect on consumption can be positive or negative. But if positive, it should be lower than for children thirteen plus years old, and if negative, it should be higher (algebraically) than the effect of children below six. The effect on the level of the female's market activity is expected to be negative but smaller (in absolute terms) than that of children below six. To economize space the description of the statistical problems of estimation was omitted, however, this material can be found in Landsberger 1972.

An analysis of the empirical findings presented in Table 1 shows that the results follow strongly the theoretical propositions. In all cases the effect of children under six on consumption expenditure is strongly negative. On the other hand, older children, thirteen plus years old, increase consumption systematically. For children in the middle group the effects are negative in three cases and positive (but very low) in one. These results too are significantly in accord with the theory. The reader may recall that with respect to children in the middle group the theory did not predict the signs of the coefficients. This would be impossible because this is the range where the negative effect declines (in absolute terms) and eventually should become positive. However, the coefficients reflect the average effect of the whole range and therefore it is impossible to predict its value. But as it was pointed out the effects if negative, should be lower (in absolute terms) than the coefficients of NC_1, and if positive, should be lower than the coefficients of NC_3. And indeed the results in column (2) follow strictly these propositions.

The low t values for the children effects in the Michigan Survey are due to the fact that the calculations here are based on about 500 observations only (as compared to 2,400 in the NBER survey). What is even more important is the fact that information on age distribution of children in that survey was very inefficient for the purpose of this study. The results presented in Table 1 besides their support to the theory concerning the children effect provide useful information from a more general viewpoint. It appears that a child below six years old decreases family's yearly consumption by a few hundred dollars. The figures in column 1 run between \$400-800 which amounts to about 5 to 10 per cent of the yearly consumption of the families involved. The effect of children 6-12 years old is much less but still significantly negative. On the other hand children 13+ years old increase the yearly consumption expenditure of the family by about \$300-500 which is about 4-8 per cent. These results are well related to the effect of children of different ages on earnings.

These results cast a serious doubt on the procedure to lump together children of all ages into a single variable referred to as family size.[4] Essentially, the coefficients of such a variable are of dubious meaning. Usually, coefficient of such a variable is interpreted as reflecting the marginal contribution to the dependent variable as a result of a change in family size by one unit (see Ghez 1970, p. 113). But the results in Table 1 do suggest that such a change depends significantly on whether the change in family size reflects a change in the number of young or older children. If the increase happened to be in the number of children below six years old the results suggest that family's yearly consumption may decrease by some \$700 whereas in the case of an increase in the number of children 13+ years old it may increase family's yearly consumption by some \$400.

TABLE 1

The Effect of Children on Consumption and Female's Market Activity[a]

	Source of the Data	Estimating Equation	Children Effects: Under 6 (NC_1) (1)	6-12 Years Old (NC_2) (2)	13+ Years Old (NC_3) (3)
Consumption - Y_1	NBER Survey	D	-413.2 (-4.3)	-108.9 (-1.8)	389.3 (5.7)
		Q	-724.6 (-3.5)	-252.5 (-2.2)	378.1 (3.6)
	Michigan Survey	D	-696.9 (-2.4)	3.0 (.1)	311.7 (.9)
		Q	-783.1 (-2.5)	-254.7 (-.9)	516.7 (1.5)
Female's Market Activity - $\tilde{Y}_3$	NBER Survey	D	-.1257 (-11.1)	-.0560 (-7.5)	.0192 (2.3)
		Q	-7.1620 (-6.8)	-4.2650 (-6.9)	-.8123 (-1.5)
	Michigan Survey	D	-.1351 (-4.2)	-.0044 (-.2)	.0284 (.7)
		Q	-5.8330 (-3.5)	-4.9112 (-3.4)	-1.2126 (-.7)

[a]The numbers in parenthesis are the t values. The dependent variable - Y_1 was measured in terms of dollars per year.

$\tilde{Y}_3$ in equation Q was defined in weeks per year. Equation Q was estimated only for those observations where the wife had a job on the market. Equation D was estimated for all observations and here the dependent variable for female's market activity was defined in terms of 0 and 1.

In the last four lines of Table 1 the reader can find the estimates with respect to the female's market activity. Here again the results are in strong accord with the theoretical propositions. Young children (under six) have a negative effect on the female's market activity and this effect declines systematically, and eventually for children 13+ years old it becomes sometimes positive. (Similar results with unimportant deviations were obtained by Bowen and Finegan 1969, although their method of estimation was different from that reported here.) These findings do come out from both equations. The reader should recall that in D the coefficients should be interpreted as probabilities because the dependent variable was dichotomous. On the other hand in Q the dependent variable was defined in terms of weeks per year where a week was defined as 40 working hours. Thus, the coefficients obtained from Q can be interpreted as indicating that as a result of having a young child the female decreases her market activity by some 15 per cent (6-7 out of 40). This effect declines to about 10 per cent (4-5 out of 40) for children 6-12 years old. The female's market activity is decreased by only about 2-3 per cent by having children 13+ years old.

As explained before in order to evaluate the effect of children of different ages on female's market activity one has to combine the results obtained from Q with those obtained from D where the coefficients express the effect of children on the probability of having a job on the market. An addition of a young child in the family decreases by about 12-14 per cent the probability of the female to be engaged in market activity. This effect strongly decreases in the case of children 6-12. Older children tend to increase that probability by about 2-3 per cent only.

In Table 2 the results of the children effect on husband's market activity are presented, suggesting in a very clear fashion that children do have a positive effect on the level of husband's market activity and that this effect declines the older the children are. These findings fit well with the other results as well as with the theory. The fact that the children effect is stronger the younger the children fits well with the fact that the same is true for the female but in the opposite direction. Namely, while the female is decreasing the level of her market activity, the opposite happens with the male. Generally speaking, in the case of young children it appears that the decreases in the female's market activity is stronger than the increase in husband's market activity, which fits well with the decrease in consumption.[5] The same is true for children in the 6-12 age group. For older children there is too an increase in husband's market activity whereas the effect on the female's market activity is not clear. The probability of those who do not work to have a job on the market increases but on the other hand those already working tend to work a little bit less. This behavior of the parent's market activity (which apparently causes an increase in their earnings) is folowed by an increase in consumption.

TABLE 2

Children Effect on Husband's Market Activity*

Source of the Data	The Effect of Children by Age Groups		
	Under 6 (NC_1)	6-12 (NC_2)	13+ (NC_3)
NBER Survey	.4918 (3.1)	.3093 (3.0)	.1804 (1.6)
Michigan Survey	2.0580 (2.7)	.9757 (1.4)	.4826 (.5)

*The figures in parentheses are the t values.

REFERENCES

1. Becker, Gary, S., "A Theory of the Allocation of Time," The Economic Journal, September 1965.

2. Bowen, W.G. and Finegan, T.A., The Economics of Labor Force Participation, Princeton, N.J., Princeton Univ. Press, 1969.

3. Cain, G., Married Women in Labor Force, Univ. of Chicago Press, 1966.

4. Ghez, Gilbert R., A Theory of Life Cycle Consumption, Ph.D. Thesis, Faculty of Political Science, Columbia Univ., 1970.

5. Landsberger, M., Children's Age as a Factor Affecting the Simultaneous Determination of Consumption and Earnings, NBER, forthcoming 1972.

6. Mayer, T., Permanent Income, Wealth and Consumption (forthcoming).

7. Mincer, J., "Employment and Consumption," The Review of Economics and Statistics, Vol. XLII, No. 1, February 1960.

8. ________, "Labor Force Participation of Married Women," published in Aspects of Labor Economics, Princeton Univ. Press, Princeton, 1962.

9. ________, "Market Prices, Opportunity Costs, and Income Effects," published in Measurement in Economics, Stanford Univ. Press, Stanford, California, 1963.

FOOTNOTES

[1]It is assumed here that the nonnegativity restrictions on the Y's are ineffective. Discussion of this subject can be found in Landsberger 1972.

[2]Second order conditions for a maximum were developed in Landsberger 1972.

[3]This is the argument to be used to justify the shift to the left of the demand curve for the husband's home services in the analysis of the effect of young children.

[4]Although the variable "family size" includes all members of the family, it can still be regarded as reflecting the number of children because the different between the two is a constant - the parents.

[5]This development is not a necessary one because among other reasons husband's wage rate is much higher, and there is always the possibility of savings.

THE DETERMINANTS OF HOUSEHOLD EXPENDITURES ON TRAVEL

Thomas Stoterau, Bureau of the Census

I. Introduction

Most of the research concerning ex-ante consumer spending has centered on the consumption of autos, housing, and major household durables. This is largely due to the notion that most of the variance in consumer outlays is attributable to changes in expenditures on durable goods. One suspects that in the future, however, spending on such services and nondurable goods as vacation travel will become increasingly important, both absolutely and cyclically.1/ This paper is predicated on the belief that the need to forecast vacation travel expenditures is becoming increasingly salient. In the past the subject has received limited attention, partly because it has been overshadowed by the greater need to predict the consumption of durables. It is also partly due to the difficulty in measuring spending on travel.2/ The only source for this data is the household itself, and no survey collects travel information on a frequent and regular basis. Thus, there is presently no acceptable series to use as a dependent variable in a vacation expenditures forecast equation.

The study presented here is based upon data collected in the Census-NBER Consumer Anticipations Survey (CAS). Briefly, this research oriented survey was a non-random, convenience sample drawn from Census tracts in the suburban areas of Boston, Minneapolis, and San Jose. Each household was interviewed five times at roughly six-month intervals between May 1968 and September 1970. A general profile of the 4,000 households in the sample would be: family income -- $10,000 to $25,000; age of head -- 30 to 50 years; and value of property -- $20,000 to $40,000. The surveys collected a rather broad range of economic and demographic data from these households.

One of the primary objectives of the CAS research project is to investigate new areas for which consumer anticipations could be worthwhile. From the very beginning, those of us involved in the design of this survey felt that travel expenditures might be an area in which anticipations could be useful. Accordingly, it was decided that household expenditures on travel would be collected in all five interviews. Aggregate expenditures were obtained by summing the amount spent per vacation trip (on up to three trips per household) for each time period. The convention was adopted that a vacation trip counted only if it cost the household $200 or more. Travel paid for by an employer or someone else outside the household was excluded. Additional data were gathered on the duration, distance, method of travel for each trip. Questions on the subjective probability of someone taking a trip costing $200 or more in the next twelve months and the likely expenditure per trip were asked in the first and third visits. This study also makes use of the data on type of family, age of head, education of head, and other objective and anticipatory variables.

It should be stressed that the study presented here is a preliminary report. This paper does not utilize all of the information available and the scope is somewhat limited. It should also be noted that since the CAS is in no sense a random sample, no attempt will be made to estimate sampling errors. Results are reported simply as the findings of a biased, although hopefully useful sample.

II. The rational for Consumer Anticipations Data

Consumer anticipations surveys are premised on the failure of traditional, non-survey variables to adequately forecast outlays for durables. There seems, however, to be little evidence that this is also the case for non-durables and services. In fact, econometric models and related research have found, on the whole, that consumption of non-durables and services is highly dependent upon personal disposable income.3/ If this were true for travel expenditures as well, there would be little reason to entertain the notion of collecting travel anticipations data. In short, a survey measure of anticipated spending on travel (PPT*) cannot be tested by simply discerning whether PPT* explains a significant proportion of the variance in travel expenditures. It must also be able to explain variation net of income and other non-anticipatory variables. If anticipations are able to do this, it is possible to make a case for collecting this kind of data on a regular national sample. The Census Bureau's success over the last several years with anticipations to buy cars, houses, and household durables has been modest. It is reasonable to suppose, therefore, that the case for expanding the present Survey of Consumer Buying Expectations to include travel would probably have to be most compelling.

III. Analysis of the CAS Travel Data

Some years ago, Lansing and Blood wrote about the determinants of non-business air travel in the Journal of the American Statistical Association (JASA). 4/ Although the intent of their article was quite different from this one, several of their findings are useful as a point of departure. Among their conclusions were: (1) the probability that an adult will take one or more non-business air trips increases with his income; (2) the stage in the life cycle of a person's family is important in understanding the tendency to travel by air; and (3) attitudes toward, and experience with air travel are useful in explaining air travel behavior. Since the focus here is on cyclical fluctuations in expenditures on travel, we can seemingly ignore conclusion "(3)" above. Certainly attitudes and experience affect travel, but these are much more likely to cause secular rather than cyclical changes.

Table 1A provides an indication of the relationship between income and household expenditures on travel in the CAS study.

Table 1A.--DISTRIBUTION OF CAS HOUSEHOLDS BY INCOME AND EXPENDITURES ON VACATION TRAVEL - NOVEMBER 1968 TO JUNE 1969

Characteristics	Percent not taking a trip	Percent spending \$200 to \$299 on trips	Percent spending \$300 to \$599 on trips	Percent spending \$600 or more on trips
INCOME 1/				
Under \$10,000............	72.1	14.0	11.2	2.7
\$10,000 to \$14,999.......	60.8	18.8	14.4	6.0
\$15,000 to \$19,999.......	51.3	18.4	17.9	12.3
\$20,000 to \$24,999.......	48.1	12.7	19.9	19.3
\$25,000 and over.........	43.0	13.0	20.0	23.9

1/ Total annual income in 1967.

As in the Lansing and Blood paper, there appears to be a fairly good relationship between income and travel expenditures. Except for trips costing \$600 or more, income above \$20,000 doesn't seem to influence vacation expenditures greatly. This may be partially due to the increased likelihood of families with large incomes owning vacation homes, and thus utilizing these facilities in lieu of taking a vacation trip.

Attempts to relate travel expenditures to the stage in the life cycle were less successful. There are several possible explanations for this difference in findings. The Lansing life cycle variable was almost a dummy variable for having children. It equalled zero if children were present, 1 if married with no children or if over 45 and single, and 2 if under 45 and single. These scaling procedures would be meaningless in the CAS study because such a large proportion of the sample would be assigned a value of zero. Instead, marital status and number of children under six were used as proxies for a life cycle variable. Neither, surprisingly, were found to shed much light on travel expenditures. Age and education of head were found to be more useful. (See Table 1B, and 1C.)

Table 1B.--DISTRIBUTION OF CAS HOUSEHOLDS BY AGE OF HEAD AND EXPENDITURES ON TRAVEL - NOVEMBER 1968 TO JUNE 1969

Characteristics	Did not take a trip 1/	Percent spending \$200 to \$299 on trips	Percent spending \$300 to \$599 on trips	Percent spending \$600 or more on trips
AGE OF HEAD 2/				
Under 30 years.............	68.9	14.7	7.9	8.4
30 to 34 years.............	61.5	16.3	14.1	8.1
35 to 39 years.............	52.0	20.7	17.6	9.6
40 to 44 years.............	53.7	17.0	17.8	11.5
45 to 54 years.............	54.2	13.4	18.9	13.6
55 years and over..........	53.8	19.6	14.7	11.9

1/ The convention was adopted that a trip counted only if the household spent \$200 or more.
2/ Age as reported in the first CAS interview (May 1968).

Table 1C.--DISTRIBUTION OF CAS HOUSEHOLDS BY EDUCATION OF HEAD AND EXPENDITURES ON TRAVEL - NOVEMBER 1968 TO JUNE 1969

Characteristics	Did not take a trip 1/	Percent spending \$200 to \$299 on trips	Percent spending \$300 to \$599 on trips	Percent spending \$600 or more on trips
EDUCATION OF HEAD 2/				
High school or less.........	63.4	15.6	14.7	6.3
1 to 3 years of college.....	59.0	16.1	13.7	11.2
4 years of college..........	56.3	17.5	15.2	11.0
5 years of college or more..	46.6	17.1	20.8	15.5

1/ The convention was adopted that a trip counted only if the household spent \$200 or more.
2/ Education of head as reported in the first CAS interview (May 1968).

In general, the probability of a family taking a vacation trip and their chances of spending large sums on trips increase with age and education. These tables, however, do not separate the tangled influence of income, education, and age. Nevertheless, it is interesting to observe the nearly monotonic relationship between education and travel in Table 1C.

Attention is now focused on the subjective travel anticipations variables. In the past we have mainly concerned ourselves with relating the consumer's expected behavior in the period with his actual behavior in a subsequent period. Increasingly, however, we are inclined to think that the first differences in reported anticipations are an important consideration. In the CAS survey we asked respondents, "What are the chances that you will take a vacation trip costing $200 or more during the next 12 months?" Respondents answered by giving one of the numbers on an eleven point scale card (0, 10, 20, ... 100). The first difference is simply the remainder of the reported anticipation in the first period minus the reported anticipation in the second period. Table 1D shows the average change in travel expenditures by income and the first difference in reported subjective probability of taking a trip costing $200 or more. The relationship is not perfect, but the numbers do move in the right direction for the most part.

Table 1D.--AVERAGE CHANGE IN EXPENDITURES ON TRAVEL BY INCOME AND CHANGE IN PROBABILITY OF TAKING A TRIP

(Total number of households in each cell shown in parentheses)

Characteristic	Change in reported probability of taking a trip						
	-100 to -80	-70 to -50	-40 to -20	-10 to +10	+20 to +40	+50 to +70	+80 to +100
INCOME 1/							
All households.	-$68 (3,527)	-$73 (3,527)	$12 (3,527)	$39 (3,527)	$153 (3,527)	$38 (3,527)	$152 (3,527)
Under $5,000........	-$305 (27)	$375 (4)	$118 (8)	$28 (46)	$20 (5)	$250 (4)	$203 (16)
$5,000 to $9,999....	-$94 (84)	-$69 (31)	-$103 (38)	$22 (280)	$110 (14)	$41 (12)	$85 (44)
$10,000 to $14,999..	-$37 (369)	$51 (83)	$25 (108)	$57 (529)	$23 (67)	-$77 (38)	$159 (100)
$15,000 to $19,999..	-$70 (295)	$35 (35)	$102 (79)	-$3 (293)	$256 (36)	$96 (27)	$105 (81)
$20,000 to $29,999..	-$38 (190)	-$60 (29)	$14 (46)	$110 (166)	$267 (20)	$88 (18)	$251 (49)
$30,000 and over....	-$133 (112)	-$303 (15)	-$336 (15)	-$6 (85)	$513 (11)	$250 (5)	$169 (13)

1/ Total annual income in 1967.

Table 3 displays the R-squares, regression coefficients, α's, and Sy.x for nine regression equations. (See Table 2 for the definition and scaling of individual variables.) In all of the equations, the dependent variable is the actual expenditures on vacation travel from June 1968 to November 1968 ($EXPV_2$). Expected expenditures on travel (EEV_1*) completely dominates all the equations in which it is included. Only total income (I_{1967}) is also clearly significant in all of the equations in which it appears. In fact as long as EEV_1* and I_{1967} are included, no other variables improve the regression equation. In most of the equations, however, education of head (EH_1) explains a significant portion of the variance. Other objective variables such as marital status, number of children, and number of children under six years are seldom significant. A priori, it would have seemed likely that the number of children under six would have proved more useful. Liquid assets (LA_1) is significant both times it appears in the equation. But LA_1 does not reduce Sy.x or improve R^2.

The performance of the attitudinal variables is also somewhat disappointing. Neither expected business conditions (EBC_1*) nor good or bad time to buy durable goods (TB_1*) explain a significant portion of the variance in household vacation expenditures. Attitudinal variables, of course, have never performed particularly well in cross-section studies. Consequently, this does not necessarily mean that attitudinal variables would also perform poorly in a time series study of vacation travel expenditures.5/

The five regression equations in Table 4 relate changes in the objective and subjective variables with the change in expenditures on vacation trips. The results in general are quite similar to those reported above. The major exception is that the only significant first difference is expected vacation expenditures (Δ EEV_1*). None of the other first differences make any contribution to the regression equation. It is also interesting to note that actual expenditures on vacations in the previous period ($EXPV_2$) is highly significant. In a sense, $EXPV_2$ is a stock variable.

Summary

A number of variables which are worthy of study do not appear in these equations, but will be examined in the future. Among these are spending on durable goods, changes in assets and debts, permanent income, and ownership of vacation homes. The analysis presented above, however, indicates that anticipated expenditures on travel is a powerful variable in a cross-section study of household expenditures on travel. Income, although much weaker, also appears to be fairly important. Other objective variables such as the number of children under six, age of head, education, and liquid assets seem to be rather weak determinants of household spending on vacation travel.

Table 2.--CAS VARIABLES INCLUDED IN REGRESSION ANALYSIS

AH_1 = Age of head (as reported in the first CAS interview)

1 = Under 25 years
2 = 25 to 29 years
3 = 30 to 34 years
4 = 35 to 39 years
5 = 40 to 44 years
6 = 45 to 54 years
7 = 55 to 64 years
8 = 65 years or older

AW_1 = Age of wife

Scaling is same as for AH_1

EH_1 = Education of head

1 = no education
2 = 1 to 8 years of elementary school
3 = 1 to 3 years of high school
4 = 4 years of high school
5 = 1 to 3 years of college
6 = 4 years of college
7 = 5 or more years of college

EW_1 = Education of wife

Scaling is same as for EH_1

NC_1 = Number of children

1 = no children
2 = 1 child
3 = married, 1 child, and head under 45
4 = married, 1 child, and head over 45
5 = married, no children
6 = single, no children

$NC6_1$ = Number of children under six years

Same as NC_1

LA_t = Amount in liquid assets

Actual dollar amounts (5 digits)

ΔLA_{-1} = Amount in liquid assets in period t minus amount in liquid assets in period t -1

EBC_t = Expected business conditions

1 = much better
2 = better
3 = about the same, don't know
4 = worse
5 = much worse

$EXPV_2$ = Actual expenditures on vacation travel expenditures in the second interview

Actual dollar amounts (5 digits)

ΔEBC_{-1}= Expected business conditions in period t minus expected business conditions in period t -1

TB_t = Good or bad time to buy large durable goods

1 = very good
2 = good
3 = partly good, partly bad; don't know
4 = bad
5 = very bad

ΔTB_{-1} = Good or bad time to buy in period t minus good or bad time to buy in period t -1

VX_t = Vacation expenditures by household since last visit

Actual dollar amounts (5 digits)

ΔVX_{-1} = Vacation expenditures in period t minus vacation expenditures in period t -1

EEV_t* = Expected expenditures on vacations (the product of a household's probability of taking a trip and their likely expenditures if they do take a trip)

Actual dollar amount (5 digits)

Δ EEV*= Expected expenditures on vacations in period t minus expected expenditures in period t -1

I_y = Total annual income from all sources

Actual dollar amount (5 digits)

ΔI_{-1} = Total annual income for year y minus total annual income for year y-1

Table 3.--REGRESSION EQUATIONS WITH OBJECTIVE AND ANTICIPATORY VARIABLES, DEPENDENT VARIABLE IS HOUSEHOLD VACATION EXPENDITURES

(t values are shown in parentheses)

Equation number	R^2	REGRESSION COEFFICIENTS										α	Sy.x
		I_{1967}	AH_1	EH_1	$NC6_1$	NC	LA_1	MS_1	EEV_1*	EBC_1*	TB_1*		
I..........	.24	0.003 (3.03)	15.81 (2.72)	19.1 (2.68)	--	--	--	--	0.35 (18.55)	--	--	-120.71	355.22
II..........	.24	0.004 (3.96)	--	--	-11.94 (-1.05)	--	--	--	0.35 (18.52)	--	--	50.28	356.16
III.........	.24	0.004 (4.06)	--	--	--	--	--	--	0.35 (18.67)	15.74 (1.42)	--	0.553	356.07
IV..........	.14	0.008 (7.94)	17.43 (2.26)	2.51 (0.64)	--	--	--	--	--	--	--	-96.97	400.72
V...........	.25	0.004 (3.67)	17.37 (2.93)	18.73 (2.62)	--	--	-0.002 (-2.28)	--	0.36 (18.67)	11.33 (1.02)	--	-149.38	354.79
VI..........	.25	0.004 (3.54)	7.56 (0.63)	19.08 (2.64)	-0.002 (-2.27)	--	--	29.31 (0.64)	0.36 (18.58)	11.92 (1.07)	-6.12 (-0.59)	-163.91	354.94
VII.........	.24	0.004 (4.07)	--	--	--	--	--	--	0.35 (18.67)	15.66 (1.42)	--	0.94	356.14
VIII........	.24	0.004 (4.05)	--	--	--	--	--	--	0.35 (18.63)	--	--	42.6	356.17
IX..........	.25	0.004 (3.41)	7.06 (0.59)	19.91 (2.42)	-2.31 (-0.17)	3.40 (0.58)	-0.002 (02.20)	30.39 (0.67)	0.36 (18.49)	12.04 (1.08)	-6.11 (-0.59)	159.77	355.20

NOTE: In order to test several hypotheses the CAS sample was split into A and B segments. The regression equation shown above are based on the 1747 households included in the A segment. The numbers shown above are coefficients for the variables included in each equation.

Table 4.--REGRESSION EQUATIONS WITH SELECTED OBJECTIVE AND ANTICIPATORY VARIABLES, DEPENDENT VARIABLE IS CHANGES IN VACATION EXPENDITURES

(Numbers in parentheses are t-values)

Equation number	R^2	REGRESSION COEFFICIENTS							α	$S_{y.x}$
		ΔEEV_{-1}*	ΔI_{-1}	ΔEBC_{-1}*	ΔTB_{-1}*	ΔLA_{-1}	EEV_{-1}*	$EXPV_2$		
I........	0.14	0.35 (16.56)	0.003 (1.68)	--	--	--	--	--	2.79	520.21
II........	0.14	0.35 (16.56)	0.003 (1.69)	-5.20 (-0.43)	--	--	--	--	2.41	520.33
III.......	0.14	0.35 (16.54)	0.003 (1.65)	-5.80 (-0.48)	6.00 (0.52)	--	--	--	2.95	520.44
IV........	0.14	0.35 (16.50)	0.003 (1.60)	5.83 (-0.48)	5.93 (0.51)	0.0002 (0.15)	--	--	2.53	520.59
V.........	0.45	0.35 (18.38)	--	-5.88 (-0.61)	--	--	0.30 (12.52)	-0.88 (-31.6)	100.56	414.98

NOTE: In order to test several hypotheses the CAS sample was split into A and B segments. The regression equations shown above are based on the 1747 households included in the A segment. The numbers shown above are coefficients for the variables included in each equation.

FOOTNOTES

1/ Detailed household travel expenditures will be collected in the 1972 Consumer Expenditures Survey sponsored by the Bureau of Labor Statistics. It will then be possible to compare them with 1961 expenditure data.

2/ Part of the difficulty lies in the fact that travel expenditures are spread out over a rather large number of non-durables and services.

3/ See Michael K. Evans and Lawrence R. Klein, The Wharton Econometric Forecasting Model (2nd edition, University of Pennsylvania, Economics Research Unit, 1968).

4/ John B. Lansing and Dwight M. Blood, "A Cross-Section Analysis of Non-business Air Travel," Journal of American Statistical Association, December 1958.

5/ See F. Gerard Adams, "Prediction With Consumer Attitudes: The Time Series-Cross Section Paradox," Review of Economics and Statistics, November 1965.

SOME RELATIONSHIPS BETWEEN CONSUMER ATTITUDES AND SPENDING

John McNeil, Bureau of the Census

Introduction

A number of organizations are engaged in the collection of periodic survey data on consumer anticipations. These data are usually thought of as falling into two categories: attitudes and intentions. The former category includes those questions which seek to reveal the respondents feelings about his personal economic situation or about the economic situation in general. Intentions questions are concerned with expected purchases of specific items or groups of items.

Consumer anticipations data have been collected in the United States at least since the 1946 Survey of Consumer Finances. Among the organizations now engaged in collecting this information are the Survey Research Center of the University of Michigan, the Conference Board, the Albert Sindlinger Co., and the U.S. Bureau of the Census.

These data are collected with the expectation that they will make a net contribution to forecasts of consumer spending. The importance and difficulty of forecasting changes in consumer spending can hardly be exaggerated. Changes in consumer spending on durable goods, especially automobiles, are probably the most important source of cyclical instability,1/ and most forecasters regard consumer spending as the most intractable of sectors. Anticipations surveys owe their existence to the generally poor record of forecast equations which contain only the traditional stock, income, and price variables. The man who is most responsible for the initiation of consumer anticipations surveys, George Katona, states flatly that "attitudes matter" and "willingness to spend" must be given equal consideration with "ability to spend."

The failure of traditional equations and the acceptance of the proposition that attitudes are important does not necessarily lead to the conclusion that forecasts of consumer spending can be improved by using any of the survey results currently available. Twenty five years have passed since consumer anticipations data were first collected on a national basis, and the issue of predictive value is still alive. There have been, of course, a number of studies designed to measure the explanatory and predictive power of anticipations data. Before we consider the implications of some of the new results from the experimental Consumer Anticipations Survey, it might be useful to summarize some of these earlier studies.

A Review of Selected Studies

(1) In 1955, Klein and Lansing2/ offered this conclusion after a cross-section study of 1,000 households interviewed initially in early 1952 and again in early 1953:

> "In working with the attitudinal variables, we were particularly impressed with the importance of buying plans. The coefficient for this term in the equation was highly reliable, amounting to almost 4½ times its own standard error. Fluctuations from year to year in the estimated proportion who buy turned out to be dominated by plans to buy and by the feeling of financial well being. Nevertheless the whole analysis shows that buying plans alone are not adequate to discriminate between purchasers and nonpurchasers."

> "In addition to plans to buy, the question on feeling of financial well being stood up well. In each of our calculations, those who felt 'better off' were more likely to buy even after taking plans to buy and the other variables into account."

(2) In 1955, the report of the Consultant Committee on Consumer Survey Statistics3/ (organized by the Federal Reserve Board at the request of the Joint Committee on the Economic Report) contained this summary:

(a) "Buying intentions, properly interpreted, appear to have predictive value. The extent of their predictive usefulness and the optimal way of combining them with other information are still to be determined by further research and experience.

(b) Other attitudes are highly correlated with buying intentions, both over time and as among spending units; and there is so far no convincing evidence that they make an independent contribution to ability to predict, however interesting these attitudes may be for other purposes."

(3) In a paper published in 1960, Mueller4/ reported on the results of a panel study which involved four interviews with 800 households at 6 month intervals over the period June 1954 - December 1955. Mueller offers this conclusion:

> "The results of the tests described here are not yet conclusive. On the positive side were (1) the strong relationship between attitudes and purchases obtained in the aggregative test over the short period for which data were available, and (2) the findings that at the individual level attitudes exhibited a pronounced influence on purchases in two of the three periods studied (as long as buying plans were disregarded). On the negative side is the finding that data on consumer attitudes consistently made only a small net contribution to forecasts of consumer spending at the individual level, when income, age, and buying plans were also taken into account. However, theoretical considerations suggest that a small net contribution by the attitudinal data at the individual level is not inconsistent with a considerably greater contribution to forecasting at the aggregative level."

Mueller's paper presents an interesting discussion of the validity of using cross-section results to make judgements about probable time-series performance, she states:

"More important, some variables vary more over time than others. There are variables such as age or thrift which vary considerably between individuals, producing correlations with spending behavior in a cross-section without varying appreciably over time; their value to business cycle analysts or forecasters is almost nil. Hence any conclusions about the significance and relative importance of attitudes and buying plans must be drawn in the light of their cross-section relation to behavior and their variation over time.

Consideration of the bias imparted by omitted variables leads to similar conclusions. Kuh has demonstrated that 'the biases from excluded variables are likely to be of strikingly different nature in the two cases, time series and cross sections. Therefore, the propriety of applying estimated behavior relations for prediction purposes in one context that were estimated in another context is highly questionable.' The time-series error is likely to be caused by dynamic excluded variables, which vary to some extent with the business cycle; the cross-section error primarily to static excluded variables such as demographic characteristics, personality traits, stocks of durable goods owned. Conceivably consumer attitudes, having a clear business cycle reference, are correlated with the dynamic excluded variables and reflect some of their impact on spending, while buying plans may be more closely related to the static excluded variables. If this assumption is correct, time-series tests would have a tendency to overestimate the influence of attitudes and cross-section tests a tendency to overestimate the influence of buying plans."

In a comment on Mueller's paper, Eisner makes this statement:

"To summarize, I think the weight of evidence including the new data presented by Miss Mueller suggest that consumer-intentions data in the major household expenditures on durable goods area do have predictive value, whereas the evidence for consumer attitudes as distinct from intentions is mostly negative though not conclusive."

(4) In a paper published in 1960, Arthur Okun[5/] examined the time-series performance of the anticipations data collected by the Survey Research Center during the period 1949-1955. His results showed that buying plans made a net contribution to the explanation of expenditures on cars, but were not useful in explaining expenditures on other durables. No other SRC attitudinal measure made a contribution to the explanation of expenditures net of buying plans.

Okun also demonstrated mathematically that if intentions have predictive value in the cross-section, they will also have, except under certain unrealistic conditions, predictive value over time. Unfortunately, the demonstration does not tell us anything about the amount of variance intentions are likely to explain, nor does it tell us whether predictive value will exist net of other variables.

(5) In 1959, Tobin[6/] offered these conclusions from a cross-section study:

"Buying intentions have predictive value; other attitudinal questions do not. This conclusion is the inescapable testimony of this analysis of the evidence of this re-interview sample."

(6) In 1963, Mueller[7/] presented the results of a time-series study, of the explanatory value of attitudes and intentions. Her regression results indicated that attitudes, but not intentions, had predictive value. The analysis did not attempt to prove that attitudinal variables had predictive usefulness net of all "objective" variables, but several income variables as well as lagged depedent variables were included to test their effect on the estimated coefficients of the attitudinal variables.

(7) In an exhaustive cross-section study published in 1964, Juster[8/] found that intentions to buy were highly significant in explaining actual purchases. Only income and a question asking "whether there is a good or bad time to buy durables" are consistently significant when intentions are present in the equation.

Juster also states: "The most important finding is that consumer buying intentions essentially reflect judgements by respondents about their probability of purchasing a particular commodity. It follows as a matter of course that surveys should attempt to estimate mean purchase probability in the population, not the proportion with sufficiently high probabilities to report that they 'intend to buy'."

(8) In a 1964 paper, Adams[9/] investigated the time series performance of the SRC measures of attitudes and intentions to buy, using 24 observations covering the period 1952-62. His conclusion:

"Regression analysis of attitudes and of buying plans as predictors of consumer durable expenditures show that attitudes made a significant contribution to forecasting durable expenditures. Buying plans do not improve the correlation once the income and attitudes are present in the equation..."

(9) In a 1966 paper, Juster[10/] reported on a test of a "new" method of measuring expected purchases. The test was the basis for the Census Bureau's decision to move from an "intentions" format to a "probability" format. The intentions question asked respondents if they "expected to buy." The probability question asks respondents about their "chances (in 100) of buying." According to Juster:

"A number of points stand out. First, it is clear that households classified as nonintenders have been successfully distributed into more homogeneous subgroups by the probability survey.... It is not so clear that the probability scale works as well among the straightforward intender classes.... On the other hand, the intentions classes do not generally appear to be effective discriminators within probability classes.... Finally, it should be noted that the vast majority of purchases are made by households that report non-zero purchase probabilities."

(10) In a paper prepared in 1967 but published in 1969, Juster[11] examined a number of models, based primarily on anticipatory variables. This was one of the first studies to include an analysis of the time-series performance of the Census Bureau series on buying intentions. The models were estimated over several time periods, including the period 1953-1967. The buying intentions series used for the longer time periods linked the SRC series on intentions to the Census Bureau series to obtain pre-1959 values (the Census Bureau survey began in 1959). Juster concludes:

"On the whole, this examination of anticipatory demand models brings out two clear-cut conclusions. First, the anticipations series themselves are strong cyclical indicators; both consumer attitudes and consumer buying intentions have cyclical turning points which precede those in durable goods and automobile expenditures by about six months. The attitude index appears to be a bit better at reflecting turning points than buying intentions, partly because the series itself is considerably reduced since the initiation of the large sample Census Bureau survey in 1959.

Although both anticipations series contain pronounced cyclical movements, only buying intentions appear to have a distinct trend component. This factor works to the comparative disadvantage of the attitude variable in regression models, since all of the trend influences on durable goods expenditures must be picked up by other variables. This difference in ability to measure trends is very probably the explanation for results obtained in Section III, where it was found that the attitude index was comparatively more useful in predicting changes in the purchase rate of nonintenders than in predicting changes in the population purchase rate."

(11) A December 1969 paper by Burch and Steckler[12] provided an analysis of the performance of the SRC Index of Consumer Sentiment in predicting turns in the consumption of real durables. Their conclusion:

"When the index is used as an indicator with every reversal of movement counted as a 'signal' the index correctly forecasts every major movement in durable consumption but also provides a number of false turns. When stringent minimum criteria are applied, the number of false leads declines, but the date at which turns in the index can be identified often lag the consumption movements."

(12) In a 1969 paper, Juster and Wachtel[13] introduced the hypothesis that attitudes are a measure of "uncertainty." According to this hypothesis, a survey measure of expected car purchases together with a variable intended to reflect the influence of unforeseen events are enough to provide unbiased forecasts, given the "typical" amount of consumer uncertainty. The variable selected to represent unforeseen events is the rate of unemployment during the forecast period: hence the model produces contingent forecasts. Forecasts based on these two variables must be corrected, however, by a measure of the extent of deterioration of improvement in attitudes during periods when uncertainty is changing. The third variable in the (contingent) forecast equation becomes, therefore, a "filtered" Index of Consumer Sentiment. During periods when the SRC Index is stable or moving randomly, the amount of uncertainty is considered "typical" and the sentiment variable takes on a value of zero. The sentiment variable takes on a non-zero value only when the change over two quarters is large or the change over three quarters is small but consistent.

The paper presents regression results using the three variable model described above. The results show that the "filtered" Index of Consumer Sentiment is superior to the "continuous" Index in terms of significance and parameter stability (the model was tested for various time periods).

(13) In a 1970 paper, Hymans[14] tested both the continuous and "filtered" SRC Index of Consumer Sentiment in a stock-adjustment automobile equation. He found that the "filtered" variable significantly improved the explanatory power of the equation. Attempts to include the continuous variable proved fruitless. Hymans concluded:

"Economists who make substantial use of sentiment variables, stock market changes, and other such non-real (as distinct from unreal) quantities in their own forecasts of consumer spending tend to shun the structural stock-adjustment framework preferred by the majority to econometric forecasters. The latter, in turn, have tended to reject the complex of stock market-sentiment-expectational variables as of dubious value and in any case unpredictable. There no longer appear to be many good reasons to maintain this dichotomy of approaches.

Changes in consumer sentiment -- if properly filtered -- do improve the forecasting accuracy of a stock-adjustment model of automobile expenditures. It is apparently possible to forecast ahead at least one quarter (and perhaps further investigation will suggest still longer) on the basis of the current quarter's sentiment index. It is also possible to forecast the systematic component of the sentiment index one quarter ahead with the aid of current stock market prices, thus permitting an auto forecast at least two quarters ahead without a forecast of stock market prices. Beyond this, the need to forecast the stock market may well establish the practical limit of the usefulness of the sentiment index in auto forecasting, except for conditional projections of the kind undertaken in the previous section. Nonetheless, the potential of meaningful improvement in forecasting accuracy for two quarters into the future is not to be taken lightly. Many four-quarter forecasts would have been much more accurate if only the errors present in the first quarter or two of the forecast could have been measurably reduced."

Some Comments on Previous Studies

Studies of the predictive (or explanatory) usefulness of anticipations data fall into three categories; (1) cross-section, (2) panel, and (3) time-series. Some of the early studies illustrate an apparent paradox which has received widespread publicity: in cross-section tests, intentions to buy are significant but attitudes are not; in time-series tests, attitudes are significant but intentions are not.

Juster has shown that at least a portion of the paradox is illusory. Analysts who claimed that intentions data had no signficance in a time-series test made that judgement after examining the performance of a SRC series on intentions which was based on a quarterly sample of fewer than 3,000 households and had not been adjusted for seasonal variation. Regression studies of the series produced by the Census Bureau indicate that intentions to buy are significant if adjusted for seasonal variation and collected from a sample of sufficient size (in this case, about 12,000 households per quarter).

A more basic question is whether cross-section and time-series results should necessarily be consistent. Mueller's (and Kuh's) conclusion that they need not be seems well taken. In a cross-section test (involving a single observation on attitudes), attitudes are basically a reflection of interpersonal differences in optimism. Intentions reflect the age and condition of the present car, and such considerations as whether a son or daughter is about to reach the driving age. Consider the case of two neighbors involved in a cross-section study. Neighbor A has been having mechanical difficulties with his 3 year old car and says the probability of his buying a car is positive. He thinks that business conditions will be "about the same" next year as they are now. Neighbor B purchased a car last month and reports a zero probability of buying. He thinks that business conditions will be "better" next year. When these two are visited six months later we find that Neighbor A did buy a car but Neighbor B did not. The conclusion: intentions have predictive power; attitudes do not.

The limitations of cross-section tests are clear. The intentions data may and often do reflect variables which explain individual behavior but which have no importance in explaining aggregate behavior over time. Such important explanations of individual behavior as accidents, mechanical difficulties, and a son or daughter reaching the age of 16 are of no interest to the forecaster. It seems equally true that attitudes reflect variables which are important cyclically, e.g., consumer reaction to news about unemployment, prices, and income, but which should not be expected to explain differences in individual behavior over a single time period.

Judgements concerning the predictive power of anticipations data must rest on time-series and panel evidence.

The time-series evidence presented in the above studies can be summarized as follows:

1. The SRC Index of Consumer Sentiment has little or no net explanatory power when continuous values of the index are tested in a relatively sophisticated forecast equation. The "filtered" version of the Index is consistently significant in such equations.

2. Buying intentions are consistently significant in equations seeking to explain the variation in new car sales since 1960. The contribution of intentions is weakened when the period of fit is expanded to include years when Census Bureau intentions data are not available, and the series must be taken from the smaller sample SRC. A qualification to this time-series evidence is the presence of trend in both car sales and intentions during the 1960's.

There is no published panel evidence on the predictive value of anticipations data although Mueller's 1960 paper was based on data collected from a panel (800 households were visited four times at six month intervals). Her analysis involved the classification of households by their attitude score at the beginning of a period and by their change of attitude during the period. She then computed each group's "expected" purchase rate for major durable goods based on the income level of the families in the group. She then examined the ratio of actual to expected purchases for each group to measure the net influence of attitude change. The analysis was repeated for four time periods. The results showed a rather weak net relationship between attitude change and purchases of major durables. Households with improving attitudes had the highest relative purchase rate twice in the four tests; households with no change in attitudes and households with deteriorating attitudes had the highest rate once each.

Some Evidence from the Consumer Anticipations Survey

The experimental Consumer Anticipations Survey (CAS), conducted by the Bureau of the Census, collected data on both purchase probabilities and attitudes several times on a panel of approximately 3,500 households. The CAS data on attitudes are rather limited; five attitudinal questions were asked in the first visit, but only two each in the second and third and none in the fourth and fifth.

Tables 1 through 3 show that results usually shown for cross-section studies of anticipations data. They are based on data collected in the first two CAS visits; May 1968 and November 1968. The tables are based on data for 3,527 CAS respondents and show average household expenditures on a collection of major items including appliances, television sets, hi-fi equipment, furniture, home improvements, cars and light trucks (less trade-ins), and vacation trips, cross-classified by income level and responses to questions on attitudes and expected purchases. Tables 4 through 6 are based on data collected in four of the CAS visits and show changes in actual expenditures by changes in attitudes and expected purchases. The changes in expenditures are from the six month period May 1968-November 1968; to the six month period May 1969-November 1969. The changes in attitudes and expected purchases are from May 1968 to May 1969.

The CAS method of asking attitudinal questions required interviewers to probe in an effort to distinguish between "very good" and "good" and between "very bad" and "bad." For example, if a respondent said he expected business conditions to be better a year from now, he was asked if he thought conditions would be "much better" or just "better." These "very good" and "very bad" categories are usually so small so as to be of little analytical use.

Tables 1 and 2 show that there is a tendency for optimists to spend more than pessimists, and this tendency persists when households are classified by income. The relationship is not particularly strong, however. It tends to break down at the "much better" and "much worse" categories, and there are a number of exceptions throughout the income categories. Table 3 shows expenditures on cars and trucks by responses to a question on the chances of buying a car within 6 months. There is a fairly strong relationship between the expected and actual measures, but the table also illustrates the old problem of the "nonintenders purchase rate." In this instance, households with a reported zero probability of buying actually spent an average of $251 during the subsequent 6 months, and accounted for 35 percent of the total expenditures of the group.

Tables 4 through 6 show changes in expenditures by changes in responses to questions on attitudes and chances of buying. The following scales were used to code answers to questions on attitudes:

Expected business conditions	Good/bad time to buy
1. Much better	1. Very good
2. Better	2. Good
3. Same	3. Partly good/ partly bad
4. Worse	4. Bad
5. Much worse	5. Very bad

The changes in attitudes shown in tables 4 and 5 are calculated by substracting the May 1969 code from the May 1968 code. For example, a change from "good time to buy" to "partly good/partly bad" would equal minus one.

Tables 4 and 5 show almost no relationship between changes in attitudes and changes in spending. Households reporting a one or two point decline in their responses to the question on expected business conditions increased their spending more than those with no change, and those reporting no change had a larger spending increase than those with a one or two point improvement. The question on good/bad time to buy produced no better results. Households with one or two point declines in attitudes did have a relatively small spending increase, but households with no change in attitudes had a much larger increase than those with an improvement in attitudes.

Table 6 shows the relationship between changes in expected car purchases and changes in actual spending. Households reporting lower purchase probabilities tend to reduce their spending; households reporting higher probabilities tend to increase their spending.

Tables A and B present selected regression results based on 1,747 observations, or about one-half of the households represented in tables 1 through 6. Table A shows results using expenditures on household durables, home improvements, cars and light trucks, and vacation trips as the dependent variable and as the independent variable (1) annual levels of income, and (2) amount in liquid assets (savings accounts, bonds, and stocks), (3) probability of buying a car within 6 months, (4) expected expenditures on appliances, entertainment items, furniture, and home improvements, (5) probability of buying a house within 12 months, (6) expected business conditions, and (7) good or bad time to buy. Table B shows the results of using changes in these items.

Table A shows that four of the seven independent variables are significant in explaining the level of aggregate expenditures. In descending order of significance, they are, (1) the probability of buying a car within 6 months, (2) expected expenditures on appliances, entertainment items, furniture and home improvements, (3) income, and (4) liquid assets. Both attitudinal measures have the right sign (scaled from 1 "very good" to 5 "very bad," but neither is significant.

Table B shows that the income and assets variables lose their significance when first differences are taken. The car probability variable is just as powerful as in the levels regression and the significance of the expected expenditures variable is only slightly diminished. Neither attitudinal measure is significant but the change in expected business conditions has the right sign and a "t" ratio of over 1.

Conclusion

There is time-series evidence that anticipations data, in the form of both attitudes and intentions, have net predictive value. The evidence must be qualified by two considerations: (1) attitudes appear to be important at some but not all, points in time, and (2) the relationship between intentions and actual car purchases is strongly influenced by trends in both series.

If the evidence that both attitudes and intentions provide unique information which is helpful in explaining and predicting consumer behavior over time is accepted, it should be possible to demonstrate the usefulness of such data on the individual level. By individual level, we do not mean the sort of cross-section test reviewed above. Comparing the purchase rates of optimists and pessimists is not a very useful exercise. But if individual changes in attitudes and intentions are not related to individual changes in spending, the time-series evidence should be called into question.

The CAS results shown above are not inconsistent with the time-series evidence concerning intentions. There appears to be a reasonably strong first difference relationship on the individual level. The CAS results on attitudes are less favorable. Changes in attitudes were not associated with subsequent changes in spending. Even this result is not necessarily inconsistent with the Juster-Wachtel hypothesis that only large and/or persistent attitudinal changes matter.

The failure of attitudes in this panel test suggests that changes in attitudes are important only if they reflect widespread changes in other economic phenomena. The contribution of attitudes in the Juster-Wachtel and Hymans studies suggest either that these "other" variables have yet to be identified or that their relationships to spending are not simple and linear. Until these "other" variables are identified and correctly specified, forecast equations can be improved by including an attitudinal variable.

FOOTNOTES

1/ E. Scott Maynes has calculated that expenditures on cars account for 51 percent of the time-series variance in total consumption. See E. Scott Maynes, "Consumer Attitudes and Buying Intentions: Retrospect and Prospect," 1966 (Mimeographed).

2/ "Decisions to Purchase Consumer Durable Goods," L.R. Klein and J.B. Lansing, The Journal of Marketing, October, 1955.

3/ Consumer Survey Statistics, Report of the Consultant Committee on Consumer Survey Statistics, 1955 (Organized by the Federal Reserve Board at the request of the Joint Committee on the Economic Report).

4/ E. Mueller, "Consumer Attitudes: Their Influence and Forecasting Value." In The Quality and Significance of Anticipations Data. A Report of the National Bureau of Economic Research. Princeton University Press, 1960.

5/ Arthur M. Okun, "The Value of Anticipations Data in Forecasting National Product," In The Quality and Economic Significance of Anticipations Data.

6/ James Tobin, "On the Predictive Value of Consumer Intentions and Attitudes," The Review of Economics and Statistics, Feb., 1959.

7/ Eva Mueller, "Ten Years of Consumer Attitude Surveys: Their Forecasting Record," Journal of the American Statistical Association, Dec., 1963.

8/ F. Thomas Juster, Anticipations and Purchases, National Bureau of Economic Research, 1964.

9/ F. Gerard Adams, "Consumer Attitudes, Buying Plans, and Purchases of Durable Goods: A Principal Components, Time Series Approach," Review of Economics and Statistics, November, 1964.

10/ F. Thomas Juster, "Consumer Buying Intentions and Purchase Probability: An Experiment in Survey Design," Journal of the American Statistical Association, September, 1966.

11/ F. Thomas Juster, "Consumer Anticipations and Models of Durable Goods Demand," Economic Forecasts and Expectations, Jacob Mincer, Editor, National Bureau of Economic Research, 1969.

12/ S.W. Burch and H.O. Stekler, "The Forecasting Accuracy of Consumer Attitude Data." Journal of the American Statistical Association, December, 1969.

13/ F. Thomas Juster and Paul Wachtel, "A Note on Uncertainty, Expectations and Durable Goods Demand Models," 1969 (Mimeographed).

14/ Saul H. Hymans, "Consumer Durable Spending: Explanation and Prediction," Brookings Papers on Economic Activity (2:1970).

Table 1.--EXPENDITURES ON HOUSEHOLD DURABLES, CARS AND LIGHT TRUCKS, AND VACATIONS BY LEVEL OF ANNUAL INCOME AND RESPONSE TO QUESTION ON EXPECTED BUSINESS CONDITIONS ONE YEAR HENCE

(Average reported expenditure during the period May 1968-November 1968)

1967 Income	Total	May 1968 response to question on expected business conditions one year hence					
		Much better	Better	Same	Worse	Much worse	Don't know
All households...	$1,100 (3,527)	$930 (73)	$1,171 (1,606)	$1,063 (997)	$970 (509)	$1,148 (30)	$1,106 (312)
Under $5,000..........	$813 (124)	$1,033 (3)	$897 (58)	$915 (31)	$831 (18)	$425 (1)	$145 (13)
$5,000 to $9,999......	$632 (310)	$270 (9)	$700 (115)	$704 (107)	$449 (52)	$238 (1)	$542 (26)
$10,000 to $14,999....	$860 (1,121)	$265 (12)	$849 (511)	$883 (308)	$767 (170)	$1,659 (12)	$975 (108)
$15,000 to $19,999....	$1,046 (946)	$413 (21)	$1,125 (429)	$1,025 (277)	$842 (120)	$432 (9)	$1,210 (90)
$20,000 to $29,999....	$1,384 (449)	$1,903 (8)	$1,538 (208)	$1,239 (128)	$1,118 (70)	$1,774 (1)	$1,397 (34)
$30,000 and over......	$1,749 (577)	$1,763 (20)	$1,793 (285)	$1,651 (146)	$1,845 (79)	$1,370 (6)	$1,645 (41)

NOTE: Household durables include kitchen range, washing machine, clothes dryer, refrigerator, freezer, dishwasher, television set, hi-fi equipment, musical instrument, room air conditioner, furniture, floor coverings, and home improvements.

Table 2.--EXPENDITURES ON HOUSEHOLD DURABLES, CARS AND LIGHT TRUCKS, AND VACATIONS BY LEVEL OF ANNUAL INCOME AND RESPONSE TO QUESTION ON GOOD/BAD TIME TO BUY

(Average reported expenditure during the period May 1968-November 1968)

1967 Income	Total	May 1968 response to good/bad time to buy question					
		Very good	Good	Partly good, partly bad	Bad	Very bad	Don't know
All households.....	$1,100 (3,527)	$1,059 (48)	$1,176 (1,611)	$1,098 (859)	$903 (605)	$856 (61)	$1,149 (343)
Under $5,000............	$813 (124)	$204 (7)	$852 (50)	$256 (24)	$1,050 (27)	- (2)	$1,111 (14)
$5,000 to $9,999........	$632 (310)	$666 (6)	$737 (110)	$594 (76)	$527 (79)	$533 (7)	$636 (32)
$10,000 to $14,999......	$860 (1,121)	$810 (6)	$894 (480)	$868 (283)	$709 (208)	$1,146 (25)	$913 (119)
$15,000 to $19,999......	$1,046 (946)	$617 (11)	$1,056 (437)	$1,059 (248)	$984 (145)	$640 (16)	$1,186 (89)
$20,000 to $29,999......	$1,384 (449)	$1,910 (5)	$1,467 (239)	$1,489 (92)	$1,164 (63)	$279 (6)	$1,117 (44)
$30,000 and over........	$1,749 (577)	$1,863 (13)	$1,794 (295)	$1,767 (136)	$1,357 (83)	$1,580 (5)	$2,106 (45)

Table 3.--NET EXPENDITURES ON CARS AND LIGHT TRUCKS BY LEVEL OF ANNUAL INCOME AND RESPONSE TO QUESTION ON PROBABILITY OF BUYING A CAR WITHIN 6 MONTHS

(Average reported expenditure during the period May 1968-November 1968)

1967 Income	Total	May 1968 response to question on probability of buying a car within 6 months (Number of chances in 100)				
		0	10 to 30	40 to 60	70 to 90	100
All households.....	$463 (3,527)	$251 (2,286)	$443 (439)	$740 (177)	$938 (332)	$1,435 (293)
Under $5,000...........	$308 (124)	$165 (89)	$105 (11)	$292 (6)	$925 (10)	$1,425 (8)
$5,000 to $9,999........	$338 (310)	$263 (219)	$110 (34)	$1,089 (14)	$426 (23)	$913 (20)
$10,000 to $14,999......	$405 (1,121)	$231 (761)	$465 (138)	$580 (46)	$978 (95)	$1,155 (81)
$15,000 to $19,999......	$409 (946)	$228 (614)	$428 (123)	$498 (53)	$909 (85)	$1,277 (71)
$20,000 to $29,999......	$569 (449)	$277 (268)	$369 (69)	$725 (20)	$890 (45)	$2,157 (47)
$30,000 and over........	$680 (577)	$333 (335)	$742 (64)	$1,221 (38)	$1,110 (74)	$1,592 (66)

Table 4.--CHANGE IN EXPENDITURES ON HOUSEHOLD DURABLES, CARS AND LIGHT TRUCKS, AND VACATIONS BY LEVEL OF ANNUAL INCOME AND CHANGE IN RESPONSE TO QUESTION ON EXPECTED BUSINESS CONDITIONS

(Change in expenditures from May 68-Nov. 68 to May 69-Nov. 69: Change in attitudes towards expected business conditions from May 68 to May 69)

1967 Income	Total	Change in attitude towards expected business conditions					
		-3 to -4	-1 to -2	0	+1 to +2	+3 to +4	Don't know
All households....	$129 (3,527)	$101 (22)	$151 (892)	$142 (1,449)	$137 (765)	$873 (14)	-$12 (385)
Under $5,000..........	$163 (124)	$3,838 (2)	$98 (35)	-$106 (48)	$170 (19)	- -	$550 (20)
$5,000 to $9,999.......	$40 (310)	-$273 (5)	$1 (74)	-$98 (113)	$245 (79)	$925 (1)	$99 (38)
$10,000 to $14,999.....	$56 (1,121)	$20 (6)	$65 (312)	$85 (442)	$104 (232)	$945 (5)	-$196 (124)
$15,000 to $19,999.....	$148 (946)	$163 (2)	$250 (228)	$105 (398)	$186 (199)	$822 (8)	-$28 (111)
$20,000 to $29,999.....	$256 (449)	-$1,400 (2)	$232 (111)	$418 (181)	$407 (113)	- -	-$704 (42)
$30,000 and over.......	$181 (577)	-$345 (5)	$212 (132)	$247 (267)	-$204 (123)	- -	$748 (50)

Table 5.--CHANGE IN EXPENDITURES ON HOUSEHOLD DURABLES, CARS AND LIGHT TRUCKS, AND VACATIONS BY LEVEL OF ANNUAL INCOME AND CHANGE IN RESPONSE TO QUESTION ON GOOD/BAD TIME TO BUY

(Change in expenditures from May 68-Nov. 68 to May 69-Nov. 69: Change in attitude towards good/bad time to buy from May 68 to May 69)

1967 Income	Total	Change in attitude towards good/bad time to buy					
		-3 to -4	-1 to -2	0	+1 to +2	+3 to +4	Don't know
All households....	$129 (3,527)	$376 (43)	$42 (878)	$246 (1,421)	$89 (727)	$801 (27)	-$82 (431)
Under $5,000...........	$163 (124)	$575 (6)	$645 (35)	$202 (45)	-$482 (21)	$1,325 (1)	-$381 (16)
$5,000 to $9,999.......	$40 (310)	$392 (3)	-$67 (80)	$69 (121)	$149 (61)	$1,192 (3)	-$106 (42)
$10,000 to $14,999.....	$56 (1,121)	$91 (12)	-$24 (302)	$207 (404)	$63 (245)	$181 (9)	-$192 (149)
$15,000 to $19,999.....	$148 (946)	-$23 (11)	$51 (236)	$265 (396)	$204 (188)	$539 (4)	-$157 (111)
$20,000 to $29,999.....	$256 (449)	$689 (5)	-$22 (102)	$354 (204)	-$19 (82)	$3,775 (3)	$601 (53)
$30,000 and over.......	$181 (577)	$1,211 (6)	$140 (123)	$286 (251)	$106 (130)	$698 (7)	-$175 (60)

Table 6.--CHANGE IN NET EXPENDITURES ON CARS AND LIGHT TRUCKS BY LEVEL OF ANNUAL INCOME AND CHANGE IN RESPONSE TO QUESTION ON PROBABILITY OF BUYING A CAR WITHIN 6 MONTHS

(Change in expenditures from May 68-Nov. 68 to May 69-Nov. 69: Change in response to question on car buying probability from May 68 to May 69)

1967 Income	Total	Change in reported probability of buying a car within 6 months (Number of chances in 100)				
		-70 to -100	-20 to -60	-10 to +10	+20 to +60	+70 to +100
All households...	$177 (3,527)	-$868 (409)	-$319 (370)	$57 (1,862)	$764 (507)	$1,592 (379)
Under $5,000..........	$286 (124)	-$350 (14)	-$261 (9)	$125 (81)	$1,250 (12)	$1,612 (8)
$5,000 to $9,999......	$31 (310)	-$360 (30)	-$245 (33)	-$137 (188)	$971 (33)	$852 (26)
$10,000 to $14,999....	$48 (1,121)	-$1,006 (115)	-$201 (114)	$16 (634)	$409 (158)	$1,175 (100)
$15,000 to $19,999....	$232 (946)	-$941 (105)	-$299 (94)	$93 (486)	$814 (158)	$1,672 (103)
$20,000 to $29,999....	$357 (449)	-$884 (54)	-$233 (58)	$164 (207)	$899 (70)	$2,079 (60)
$30,000 and over......	$251 (577)	-$848 (91)	-$772 (62)	$121 (266)	$1,106 (76)	$1,877 (82)

Table A.--SELECTED REGRESSION RESULTS USING MAY 1968 - NOVEMBER 1968 EXPENDITURES ON HOUSEHOLD DURABLES, CARS AND LIGHT TRUCKS, AND VACATIONS AS THE DEPENDENT VARIABLE

(Independent variables as measured in May 1968 survey, "t" ratios shown in parentheses)

Equation	Constant	1967 income	Amount in liquid assets	Probability of buying a car within 6 months	Expected expenditures on household durables and vacations	Probability of buying a house within 12 months	Expected business conditions 1 year hence	Good or bad time to buy large durable goods like cars and appliances	R^2	SEy.x
I......	570.5 (9.5)	.0315 (10.4)							.059	1314.4
II.....	548.3 (9.0)	.0278 (8.1)	.0074 (2.3)						.062	1312.8
III....	822.0 (23.0)			132.2 (15.1)					.116	1274.2
IV.....	261.9 (4.3)	.0171 (5.2)	.0089 (2.9)	115.5 (13.5)	.2512 (7.1)				.180	1228.0
V......	420.7 (3.5)	.0169 (5.1)	.0085 (2.8)	115.5 (13.5)	.2490 (7.0)			-54.9 (1.5)	.181	1227.5
VI.....	503.5 (3.3)	.0169 (5.1)	.0086 (2.8)	115.2 (13.4)	.2488 (7.0)		-33.6 (0.8)	-52.3 (1.5)	.181	1227.6
VII....	503.9 (3.3)	.0169 (5.1)	.0086 (2.8)	115.2 (13.4)	.2493 (7.0)	.1801 (0.1)	-33.6 (0.9)	-52.1 (1.4)	.181	1227.9

Table B.--SELECTED REGRESSION RESULTS USING CHANGE IN EXPENDITURES ON HOUSEHOLD DURABLES, CARS AND LIGHT TRUCKS, AND VACATIONS FROM MAY 1968 - NOVEMBER 1968 TO MAY 1969 - NOVEMBER 1969 AS THE DEPENDENT VARIABLE

Equation	Constant	Change in income	Change in liquid assets	Change in probability of buying a car within 6 months	Change in expected expenditures on durables and vacations	Change in probability of buying a house within 12 months	Change in expected business conditions	Change in attitude towards good/bad time to buy	R^2	SEy.x
I.....	127.5 (2.7)	-.0073 (1.2)							.030	1956.8
II....	118.8 (2.4)	-.0081 (1.4)	.0036 (0.9)						.036	1956.9
III...	132.5 (3.0)			139.1 (14.3)					.105	1852.4
IV....	174.2 (3.9)	-.0087 (1.6)		138.5 (14.3)	.2228 (4.9)				.118	1839.6
V.....	177.5 (3.9)	-.0089 (1.6)		138.7 (14.3)	.2228 (4.9)		46.7 (1.1)		.119	1839.5
VI....	170.0 (3.7)	-.0096 (1.7)	.0030 (0.8)	138.7 (14.3)	.2212 (4.9)		46.1 (1.1)		.119	1839.7
VII...	169.5 (3.6)	-.0095 (1.7)	.0030 (0.8)	138.7 (14.3)	.2214 (4.9)		46.6 (1.1)	-5.5 (0.1)	.119	1840.2
VIII..	169.5 (3.6)	-.0095 (1.7)	.0030 (0.8)	138.7 (14.3)	.2219 (4.9)	-.1949 (0.1)	46.7 (1.1)	-5.5 (0.1)	.119	1840.8

AN ANALYSIS OF EX-ANTE SAVINGS DATA: SOME PRELIMINARY RESULTS

F. Thomas Juster and Paul Wachtel
National Bureau of Economic Research*

INTRODUCTION

For the past several years the National Bureau of Economic Research, in conjunction with the U.S. Bureau of the Census, has been conducting an experimental survey designed to test a number of hypotheses about the possible usefulness of ex-ante data on consumer behavior. The experimental survey, known as the Consumer Anticipations Survey (CAS), began in May 1968 with a sample of roughly 4,400 households in three suburban areas of the country (San Jose, Minneapolis, and Boston). The sample selection was non-random and purposive, both features designed to reduce survey costs; neither constituting a serious interpretative or methodological shortcoming. Sampling was restricted to moderately high to high income Census tracts, in order to get a high frequency of "positive" readings on activities like saving, spending on durables, vacation outlays, etc. The survey design called for five waves of interviews, each six months apart. The fifth interview was not conducted precisely on schedule because of budgetary problems, but was carried out some four to five months later than originally planned. The final interview has not yet been completely processed, hence results in this paper consist of data from the first four waves.

The survey yielded information on a wide range of questions concerning household decision-making. One central element in the survey design was the testing of specific question forms about prospective expenditures on a wide range of discretionary outlays including automobiles, home appliances, furniture, home improvements, vacations, recreation, and housing. Another was a test of the usefulness of ex-ante data on savings. A second set of hypotheses concerned the effects of family income on spending and saving decisions, with special attention to the composition of family income between earnings of the household head, earnings of supplementary members of the labor force, non-wage income from a variety of sources, capital gains, and so forth. Thus we obtained data on annual earnings for a number of past years, hours worked, multiple job holdings, labor force participation on the part of the wife and other adult family members beside the principal earner, variations in hours for supplementary earners, and so on. The idea was to examine the effects on expenditure and savings patterns of both long run and more transitory aspects of family income.

A third component of the survey design focussed on a wide range of questions of peripheral interest to the analysis of cyclical variability in spending and saving, but of substantial and growing interest for analysis of household decision-making generally. Thus we collected extensive data on educational level of all household members, on schooling status and schooling plans for children, on family size and expected family size, and on a number of basic demographic characteristics of the household. In conjunction with the cyclically oriented analysis, we obtained data on a number of expectational and attitudinal variables similar to those used by the Survey Research Center to construct the Index of Consumer Sentiment. Other expectational variables include judgments about the probability of changes in family income and the likelihood of changes in earnings, multiple job holding, labor force participation, and so on.

Other papers at this session will concentrate on family decision-making models of a more general sort; both Michael and Landsberger rely extensively on data from the CAS for their empirical analysis. The usefulness of the CAS for examination of these questions represents an unexpected but sizeable bonus from the survey.

One interesting by-product of this paper, and perhaps it may represent more of an embarrassment than anything else, is the possibility of comparing the analysis of ex-ante durables expenditures in this paper and in the Stoterau and McNeil papers. The last two will focus entirely on the ex-ante discretionary expenditure variables for durables and vacations, and should in principal show identical results for identical empirical tests. The difference is that McNeil and Stoterau have based their results on a tape prepared at the Census Bureau, while the results in this paper are based on a different tape prepared at the NBER. We have not, unfortunately, had much chance to compare allegedly identical results, and thus all of us may learn more about the sensitivity of results to differences in editing and tape-making procedures than about the substantive questions about consumer anticipations that the CAS was designed to answer.

This paper will concentrate on the analysis of the ex-ante savings data and will attempt to assess its possible usefulness for short-run forecasting of consumer behavior. For a variety of ex-ante measures relating to discretionary expenditures, results are presented for comparison with already available data from operating Census Bureau and other surveys.

One fact that should be kept in mind in interpreting the results, and the survey design that produced them, is that the focus of the experimental CAS survey was on providing possible inputs into an operating survey. That is, the ex-ante measures from the CAS were specifically designed for their possible use as ex-ante variables on an operating Census Bureau survey such as the CBE. Thus we have constructed and tested variables which could be used within the framework of the Census Bureau's household

*This paper has neither been reviewed nor approved by the Board of Directors of the National Bureau of Economic Research, and represents only the personal judgments of the authors.

surveys which are subject to constraints on interview time and on total resources. In short, the experimental survey, so far as the ex-ante data are concerned, is not designed to answer the question: Is it possible to explain spending and saving behavior from ex-ante data reported by households? Rather, the question is: Is it possible to structure relatively simple and easily handled questions that can improve our understanding of the likely future course of household spending and saving? These are obviously not the same questions.

The available data tape contained information, as noted above, from the first four waves of the survey. A total of approximately 3,500 completed interviews are available for the full four waves, and these constitute the basic sample for the results presented here. Responses have been deleted for various reasons: families reporting no family income at all have been eliminated, as have those where the household head is past the age of 65 and those which are not husband and wife families. Other responses have been deleted because of errors or probable errors on the part of respondents, and one group of responses have been eliminated from the analysis based on a measure of "response quality" for the asset and asset change questions. Total deletions amount to about 23 per cent of the available sample. The error deletions were for households reporting ex post or ex-ante changes in savings of less than $-12,500 or more than $+37,500; both responses seemed more likely to represent errors than real changes. The quality code deletions represent families that, on the basis of responses to the asset questions, should have provided responses to other asset questions and failed to do so. Any household reporting less than 75 per cent of the number of responses that should have been reported has been eliminated from the sample. Some comparisons are made between the sample without quality deletions and the truncated quality control sample.

HYPOTHESES

Experience with the Consumer Buying Expectations survey now conducted by Census have been quite unsatisfactory as regards the usefulness of data on expected purchases of household durables. The CBE contains a single global expected purchase estimate for a category of products described as household durables and appliances, furniture, home improvements, and so forth. No specific quantity information is obtained in the survey, but only a single dollar value for total expected purchases. The CAS survey contains that version on the B half of the sample, but on the A half, respondents are asked about the likelihood of purchasing any one of a specified collection of household durables and appliances, about the likelihood of spending money on furniture, on home improvements, on vacations, and on recreation. The question posed by this design was whether or not disaggregation of the discretionary outlay variable results in better forecasts of total discretionary outlays. If there are offsetting errors in the forecasts for individual components of discretionary outlay, the aggregate forecast of outlay would be better than the sum of the component forecasts.

Relationships are estimated for both single-time cross sections, and from a first difference version of the cross section. The latter uses changes between second and fourth wave data to measure actual change, and between the first and third waves to measure the corresponding expected change. In most cross-section regressions, first wave ex-ante measures are associated with behavior variables measured from the second wave and thus corresponding to a six-month period. In first difference form, the dependent variable is the difference between the first and third six-month periods. In principal, it would be desirable to relate ex-ante first wave data to actual behavior obtained from the second and third waves, thus covering a twelve-month period, but the counterpart first difference comparisons would not be possible at this writing because we do not have actual behavior data from the fifth wave.

Regression estimates of the association between alternative ex-ante variables and the corresponding ex-post value use a standard set of supplemental variables. These include three family status and two education-level dummy variables, family income, expected and actual change in family income, an attitude variable comprising two of the relatively volatile components in the SRC Index of Consumer Sentiment (expectations about future business conditions and opinions about whether the present is a good or bad time to buy durables), and expected price change. To this standard collection of variables is added the test variable on expected purchases; the corresponding actual expenditure variable is regressed on the full set of independent variables.

The standard set of variables for the first difference equations are similar to those in the single-time cross sections. Exactly the same set of family status, education and income variables is used. The latter are unchanged because complete information on actual and expected family income were not collected in each wave. The attitude variables is the first difference of the corresponding variable, as is the price change variable. The level of expected price change is also included.

The first difference equations do not correspond exactly to the specification obtained when one cross-sectional equation is subtracted from the previous one. The expected and actual income change and actual income variables from the second single-time cross section are omitted. The remaining income variables appear with the opposite signs from those in the single-time cross section because of the differencing. Thus estimates of the first difference equations are subject to omitted variables bias if omitted and included variables are correlated, although the predictions are unbiased. If actual change reported on the fourth wave minus expected change reported on the third wave were uncorrelated with the actual and expected change variables included, or with the first difference of expected purchases, the estimated coefficients are unbiased estimates of the true coefficients.

For evaluation of the potential forecasting value of the ex-ante variables, it can be argued that the most relevant comparisons involve the first difference specification, where changes in the ex-ante variables are used to explain changes in the counterpart ex-post variables. In the past, analysis of micro relationships have often been limited to examination of single-time cross-sectional differences. A serious problem with such comparisons is that the observed differences among households in a cross section largely reflect differences in the permanent (structural) characteristics of the families involved, less so differences in transitory phenomena that are of major interest in the analysis of cyclical behavior. Thus we can be sure that expected purchases will have very strong correlations with actual purchases in single-time cross sections, as both are largely determined by the same structural factors. For example, families whose automobile has just been wrecked in an accident will report high probabilities of car purchases, and are quite likely to report having purchased a car, while families who bought cars the day before the survey are quite likely to report zero or low probabilities of future purchase and equally likely not to purchase. Hence the powerful cross-section association between ex-ante and ex-post behavior does not necessarily tell us very much about the potential usefulness of the data in time-series predictions. By contrast, the empirical relationships observed in the first difference comparisons come much closer to resembling the time-series world, and that is the world in which these survey data are designed to be used.

The set of supplementary variables are not without interest themselves. Some are included mainly to standardize for obvious and easily measurable influences in order to determine the net contribution of the ex-ante data--family income level, family structure, and the educational level variables fall in this category. Others are of potential forecasting use in an ex-ante model, since they represent information relevant to the interpretation of the ex-ante spending and saving measures. Variables in this category are expected income change, actual income change, the attitude variable, and expected price change. The analysis of the ex-ante data should be largely unaffected by the choice of other variables used; except for income the simple correlations with the ex-post and ex-ante data of the other variables is always small.

In the model, household investment is explained by the combination of expected investment, expected income change, and actual income change. Holding constant expected investment and actual income change during the purchase period, expected income change should be negatively associated with actual investment because it represents a favorable or unfavorable income surprise variable. That is, given expected investment and actual income change, the higher is expected income change the less favorably surprised the household or the more disappointed--and the lower investments should be. And conversely for families reporting low expected income change.

For the attitude variable, one expects either a nil or a net positive association with investment, since this variable may reflect an additional dimension of consumer optimism than the expected investment variable itself. For expected price change, we do not have an a priori conviction about the appropriate sign. One could argue that families would expand purchases relative to expected purchases if actual price change turned out to be larger than expected, since they would have misread the strength of inflationary forces and would expand purchases to protect themselves against the anticipated further price rise. On the other hand, some families might feel poorer as a result, causing a contraction of actual investment relative to expectations.

The equations are all estimated by Ordinary Least Squares. Heteroscedasticity in the error terms is possible because the majority of households report zero ex-post expenditures. Of course, the problem is not as serious as it is in many other cross-section demand studies because the ex-ante data may adequately explain the concentration of expenditures around zero.

Finally, we should note that the model specified for these empirical tests clearly represents a minimal exploitation of the available data in the CAS. Preparation of the data tape, as usual, took substantially longer than we had hoped or expected, and we have been forced to restrict the scope of the empirical analysis severely. Thus we are able to test only the simplest hypotheses regarding alternative versions of expectational variables, hypotheses which by and large have been built into the basic survey design. Even here, our results must be viewed as preliminary and exploratory, since the comparisons across expectational variables obviously could be influenced by the specification of the equations.

EMPIRICAL RESULTS

The basic empirical results are presented in Tables 1 through 10, below. Table 1 summarizes the means values of variables used in the analysis. Some of these values are worth comment. First, we have obviously selected a sample with well above average mean income, roughly $17,000 per family in 1967. Actual purchase levels for durables and actual savings amounts are correspondingly higher than one would find in a random population sample. Next, there are modest differences in the mean values of expected expenditures on durables when measured by a single global question than when measured by a series of questions about specific components. Building from the components arrives at an expected expenditure level larger than that yielded by the single variable, and probably closer to the level implied by the corresponding actual expenditure variable. In passing, we should note that there is a seasonal problem in comparing expected and actual levels; the actual expenditure or saving data cover a six months span, while the expected values are for a twelve-month period. Thus expected outlays for vacations are only slightly larger than actual outlays, but the six-month actual span covers the summer months.

Finally, there are marked differences in means for the two alternative versions of the expected savings variable. One of these, labeled S_1^*, is obtained from responses to a series of questions about expected changes in specified types of asset holdings. The second, S_2^*, is obtained from asking households whether their "expenses" (undefined), are likely to exceed or fall short of their income, and by how much. The levels of expected savings are about 50 per cent larger for the latter than the former; the level of actual savings (when annualized) is between the two. And the very high education level in the sample is worth noting--about half the respondents are college graduates or above.

Table 2 shows regression coefficients in a cross-section analysis for a variety of expected outlay variables for discretionary expenditure. The regressions relate ex-ante and ex-post data for individual expenditure components, and for an increasingly broad aggregate of total discretionary outlays. These data answer the question: How much ex-ante information must be included in a consumer survey in order to get the maximum usefulness in explaining variability in expenditure, and indirectly, in savings?

For individual expenditure components, the ex-ante variable does about as well in one category as another with the apparent exception of vacation outlays, where the contribution of the ex-ante variable is perceptibly larger. But there is no evidence that aggregation of individual variables for appliances, furniture and home improvements produces a better forecast of total outlays than implied by the simple addition of forecasts for the components. That is, there is no evidence of negative correlation in the error terms across components, hence no evidence of canceling out of errors as the ex-ante variable covers an increasing range of expenditures. This is a disappointing result, in a sense, since we already know (or think we know) from the CBE operating survey that there is very limited forecasting value in the household durables expenditure variable included on that survey.

One curious feature of these results, which is similar to those found in an experimental survey conducted three or four years ago by Census in conjunction with NBER, is the apparent lack of difference in the cross-sectional results between equations designed to explain expenditures for automobiles and those for household durables. Neither the general structure of the equations, the contribution of the ex-ante variables, nor the proportion of variance explained differ when cars or household durables are the dependent variable. Hence one would infer that time-series predictions with ex-ante household durables variables would be about as successful as with ex-ante automobile variables. But all of our experience with time-series data suggests that this is not the case: In the most clear-cut comparison, the simple time-series correlation between ex-ante household durables outlays and actual household durables outlays appears to be virtually nil, while ex-ante automobile expenditures have always been an important part of automobile demand models whenever they have been tested.

We had hoped that the data summarized in Table 3, where the first difference form of the equation is estimated for the same relationships as summarized in Table 2, would resolve that particular issue. But that turns out not to be the case either. Taking first differences in actual and expected outlays for either individual components of household durables outlays or various subaggregates, it continues to be true that aggregation has no apparent payoff and that the relation is about the same for automobiles as for any of the household durable categories or any of the subaggregates. Thus we are still left with a puzzle, at least on the level of the relatively simple specification of the Tables 2 and 3 equations.

One rather striking result in these data, and one that constitutes the most convincing evidence that we have yet seen on the point, concerns the behavior of the actual and expected income change variables. As argued earlier, a rational decision-making model calls for positive regression coefficients on actual income change and negative ones for expected income change. It ought also to be true that, excluding actual income change from the regression, one might get either positive or negative coefficients on expected income change; the result is not predictable a priori. While the latter test has not been carried out because of time pressure, the regression coefficients in Table 2 are systematically significant, with the alternate positive and negative signs predicted by the rational decision model. This is more true for the subaggregates than for the individual components on household durables, but it is consistently true throughout. Just on a signs test, for example, 9 of the 10 actual income change coefficients have positive signs in Table 2, all 10 expected income change coefficients have negative signs. In Table 3, the pattern of signs on actual and expected income change variables is exactly reversed. This is as expected, considering that the first difference version of the equation can be obtained by subtracting one cross-sectional equation from the preceding one. On balance, the results shown in Tables 2 and 3 are remarkably consistent with the hypothesis that the combination of expected purchases and an income surprise variable is the appropriate model with which to explain household investment decisions.

Before turning to the ex-ante savings data, some brief comment on Tables 4 and 5 is in order. A more specific test of the CBE version ([B] survey) of household durable expenditures with the experimental CAS version ([A] survey) is shown in Table 4. The results show little evidence of improvement with the more elaborate CAS version. However, the CAS regressions have all the correct signs and most of the t-ratios are larger than in the CBE version. In addition, Table 1 shows that the mean of the experimental ex-ante durables variable is of about the same magnitude as ex-post expenditures, whereas the CBE version substantially underpredicts expenditures on average.

Another test built into CAS was to determine whether the ex-ante automobile expenditure variable would be improved by attempting to get ex-ante information on multiple purchases of cars within a single period, and also to determine whether omission of vehicular purchases like trailers had any influence on the accuracy of the ex-ante, ex-post automobile comparisons. Thus we designed a question which asks about the probability of buying "more than one car" during a given time span. In Table 5, actual purchases of cars (which of course include multiple purchases) is regressed on ex-ante expenditures for one car (the present CBE version) and on ex-ante purchases for one or more cars (the experimental CAS version). The differences are quite noticeable, and they go in the appropriate direction both in the cross-sectional and first difference equations. Thus, even though ex-ante purchases of more than one car are a very small part of the total, their inclusion does make a contribution to explanation of the variance in observed automobile purchases, and inclusion of a multiple ex-ante question on the operating CBE survey would apparently represent an improvement in accuracy. Incidentally, in this sample about 10 per cent of the ex-ante purchases represent those reported by households under the "more than one" variable, as indicated by the mean values shown in Table 1.

Results from the ex-ante savings data are shown in Tables 6, 7, and 8. Table 6 tests alternative specifications of the ex-ante variable against alternative definitions of savings. S_1^* represents ex-ante asset change, while S_2^* represents the difference between income and expenses ex-ante. S represents changes in savings in the form of savings accounts, saving bonds, common stock equity (excluding capital gains), and investments in property and land, while L represents the first three categories only but excludes the fourth. Cross-section results are shown in Table 6, first difference equations in Table 7.

The cross-section results in Table 6 suggest that ex-ante savings questions are about as useful as the ex-ante questions on household durables and appliances. The ex-ante variables are always significant, actual and expected income change have the expected positive and negative signs, and income level is also significant. Explanations of S are a little better than explanations of L with either of the ex-ante variables, and the asset change form of the ex-ante variable looks to be slightly better than the income less expenses version. The regression coefficients of the ex-ante variables are significantly higher for the asset change version, and the t-ratios are also higher as is the explained variance. Hence the ex-ante variable looks promising; more precisely, it looks about as promising as the household durable expenditure ex-ante variables.

This promising look evaporates when we turn to Table 7, where the first difference form is shown. Here we appear to be looking at essentially random numbers except for the actual and expected income change variables, which have the same sign pattern as discussed earlier. The ex-ante variable, representing the difference between two consecutive ex-ante estimates, always has a negative sign and is never significant. On a preliminary view, therefore, the results are discouraging but not necessarily hopeless. That the ex-ante responses bears some resemblance to actual behavior is clearly shown by the cross-sectional results in Table 6, and the first difference form may be quite sensitive to equation specification.

Another test of the ex-ante savings variable is shown in Table 8, where we estimate a saving regression which include either ex-post or ex-ante expenditures on durables as an additional independent variable. It is a well documented fact that savings and at least some types of durable goods expenditures are close substitutes for each other, and that both are appropriately included in a household investment function. Thus the inadequacy of the savings function might be due to the effect on actual saving of variation in expenditures on durables.

The first two equations in Table 8 contain ex-post and ex-ante measures of total outlays for durables, including household durables, vacations, and automobiles. The last two equations contain only the automobile expenditure variable in ex-post and ex-ante form. We expect to find a negative correlation between actual savings and expenditures on durables.

The only equation form where the expected negative sign emerges is the last equation, which has actual expenditures on automobiles as an additional dependent variable. The effect is not very strong, since the regression coefficient is only .14; but in all other equations the durables variable has a positive rather than a negative sign. Thus total durables outlay appears to be complementary rather than competitive with savings. Even for cars, only the ex-post expenditure variable contributes significantly to the explanation of savings behavior. In forecasting equations, one would have to make use of either the ex-ante variable or of some kind of predicted value for automobile expenditures, hence even these results are not as encouraging as they might appear.

Two further tests of the savings data are shown in Tables 9 and 10. In Table 9, a typical equation is shown for the (A) and (B) samples. The variables are the same; the differences, due solely to sampling variability, are not small. In Table 10, a series of equations with alternative dependent variables are shown. S is the total savings variable used in the previous tests, respondents with poor financial quality information are eliminated. S' is total savings for 1,537 observations including 150 eliminated from the S regressions. The coefficients are essentially the same in the two sets of regressions. The probable reason is that "poor quality" relates mainly to refusal or non-response on particular questions about actual asset change. Non-response would be translated into zero asset change, and a similar non-response on expected asset change, which is not unlikely, would also be translated into zero expected asset change.

The coefficients on the standard collection of family structure, education, attitude and expected price change variables show few consistent and significant effects. This is not unexpected for the dummy variables representing the age of the head and the presence of children in the household and the education of the head. Net of the household's plans there are no significant shifts according to household type. The attitude and price variables appear significantly in some of the single cross-section equations but not in the first difference equations. The attitude variable has a positive sign in many of the expenditure equations but is insignificant in savings equations. The expected price change variable is positive in many savings equations. However, neither variable exceeds its standard error in the car expenditure equations.

SUMMARY

On the whole, we find some results in the experimental survey which are promising and warrant further examination, while others do not appear to be worth pursuing much further. Perhaps the most discouraging feature of the results is the apparent lack of difference between the current CBE version of expected household durable outlays and the much more precise and hopefully improved version of that variable on the CAS experimental survey. The regressions yield no evidence that the experimental version is better than the existing version, and since we know that the existing version doesn't help much in time-series predictions, that does not auger well for the experimental version. But perhaps the explanation lies in our failure to examine more carefully the role of the existing CBE version in time-series prediction models. There are only a limited number of observations available with the CBE durables expenditure variable, and the negative judgments about its value are largely a consequence of the casual observation that it clearly does not predict movements in actual outlays when taken by itself. In a more fully specified demand model, perhaps even the present version might make some contribution, and the CAS version should turn out to be a bit better since its mean is much closer to mean expenditures in the cross section.

On the encouraging side, one would have to put the results on ex-ante vacation outlays and on multiple plans for purchasing cars. Both in cross sections and first differences, the strongest ex-ante, ex-post relationship in this batch of survey results concern expenditures for vacation. This is clearly a major discretionary outlay, and it looks as if one might well be able to predict changes in its level from an ex-ante survey variable. On multiple plans to purchase cars, the results suggest that a consumer survey would be substantially improved simply by the addition of questions designed to find out if households expected to buy more than one car during the purchase period.

On the "in-between" side, we would put the savings results. This paper is, after all, a very preliminary report based on results which have been obtained within the last week. There are significant cross-sectional associations between ex-ante and ex-post saving, as measured on the survey. And there are clearly enormous measurement error problems when dealing with both ex-ante and ex-post savings. Despite these problems, there are some limited positive results, and the subject is worth pursuing further. The first difference results are discouraging, and it may turn out that those results are accurate. Finally, there are several additional variables concerned with expected savings that appear on the CAS questionnaire, and these have not been examined at all.

TABLE 1

Mean Value of Variables[a,b]

	Twelve-Month Expected (1A)	Twelve-Month Expected (1B)	Six-Month Actual (2)
Appliances = App	196.34		109.03
Home improvements = HI	242.92		119.75
Furniture = F	280.63		154.02
Vacations = V	352.85		235.67
Total cars = C	1090.94	1027.15	582.15
First car = C_1	999.57		
Household durables = D	719.89	337.20	382.80
D + V	1072.74		618.47
D + V + C	2163.68		1200.62
Actual savings = S			660.08
Actual savings, excluding land = L			526.86
Expected savings = S_1^*	1404.00	1372.87	
Expected income less expenditures = S_2^*	1990.37	1915.63	

First Differences

	Twelve-Month Expected (3A)-(1A)	Twelve-Month Expected (3B)-(1B)	Six-Month Actual (4)-(2)
App	-6.37		1.06
HI	-25.83		31.46
F	8.23		-1.03
V	54.66		21.14
C	-62.41	-50.33	83.05
C_1	-48.18		
D	-23.96		31.48
D + V	+30.76		+52.62
D + V + C	-31.65		+135.67
S			247.80
L			-158.76
S_1^*	151.21	130.69	
S_2^*	-178.48	-78.98	

Family Structure	Per Cent	Head's Education	Per Cent
Head 25, children	22.9%	0 - 12 years	25.1%
Head 45, children	26.7	13 - 15 years	20.5
Head 35-44, children	39.4	16 or more	54.4

	Mean of Attitude Variables (1)	Mean of Attitude Variables (3)		Income Variables[c] (1)	Income Variables[c] (3)
Attitude index	.661	.504	Y	17019.33	18623.98
Expected rate of price change	2.286	2.301	ΔY^a	.146	
			ΔY^e	.066	

[a]Data are in dollars, except as noted. Interview and sample are in parentheses. The (A) sample has 1,410 observations and (B) 1,312.

[b]Variable name is followed by the symbol used in following tables. An asterisk is added to the symbol to refer to ex-ante data.

[c] Y = Family income.
$\Delta Y^a = (Y(3) - Y(1))/Y(1)$ = actual income change.
$\Delta Y^e = (\text{Expected income} - Y(1))/Y(1)$ = expected income change.

TABLE 2
Aggregation Tests for Household Durables and Vacation Outlays, Cross-Sectional Data, A Sample

Dependent Variable	Y	ΔY^a	ΔY^e	Ex-Ante Variable	SE	R^2
App	.0016(2.5)	24.71(1.5)	-15.83 (.6)	.2298 (9.6)	203.4	.079
F	.0021(2.1)	19.47 (.8)	-14.24 (.4)	.2917(14.5)	314.6	.147
HI	.0024(2.3)	81.95(3.1)	-101.2 (2.5)	.2090(11.7)	329.9	.106
V	.0029(2.2)	-2.932 (.1)	-4.932 (.1)	.3842(17.6)	372.4	.255
C	.0108(2.7)	257.8(2.6)	-298.9(1.9)	.3248(13.2)	1,238.6	.137
App + HI	.0040(3.1)	107.0(3.3)	-117.7(2.4)	.2158(11.4)	400.9	.106
App + HI + F = D	.0057(3.3)	129.4(3.1)	-142.7(2.2)	.2741(15.2)	526.7	.171
D + V	.0104(4.6)	136.9(2.6)	-154.7(1.9)	.2971(15.8)	660.6	.225
D + V + C	.0202(4.2)	391.7(3.5)	-457.8(2.6)	.3220(15.3)	1,406.7	.202
App + C	.0127(3.1)	281.5(2.8)	-312.2(2.0)	.3125(12.9)	1,260.8	.134
HI + C	.0166(3.7)	386.1(3.6)	-440.7(2.6)	.3052(14.0)	1,343.3	.163

TABLE 3
Aggregation Tests for Household Durables and Vacation Outlays, First Difference Data, A Sample

Dependent Variable	Y	ΔY^a	ΔY^e	Ex-Ante Variable	SE	R^2
App	-.0016(1.6)	-20.50 (.8)	-20.4 (.5)	.2879 (9.7)	306.5	.074
F	.0015(1.1)	-7.415 (.2)	24.09 (.4)	.3294(12.8)	449.6	.114
HI	.0005 (.3)	-50.87(1.3)	110.5(1.9)	.2784(13.3)	481.2	.122
V	-.0062(3.5)	-51.93(1.2)	73.54(1.1)	.4154(16.9)	549.5	.179
C	.0063(1.0)	-356.8(2.3)	532.4(2.2)	.3936(13.8)	1,930.4	.128
App + HI	-.0011 (.6)	-73.88(1.6)	93.53(1.3)	.2962(13.3)	587.1	.123
App + HI + F = D	.0004 (.2)	-85.21(1.4)	123.2(1.3)	.3232(14.3)	778.1	.141
D + V	-.0055(1.8)	-134.0(1.7)	195.3(1.6)	.3304(14.2)	980.7	.141
D + V + C	.0006 (.1)	-485.5(2.8)	718.5(2.6)	.3572(13.4)	2,191.1	.122

TABLE 4
Tests of Alternative Household Durables Variables, (A) and (B) Surveys

Sample	Y	ΔY^a	ΔY^e	Ex-Ante Variable	SE	R^2
(A)	.0057(3.3)	129.4(3.1)	-142.7(2.2)	.2741(15.2)	526.7	.171
(B)	.0098(5.5)	45.84(1.2)	99.19(1.3)	.3458(14.6)	554.3	.197
(A)	.0166(3.7)	386.1(3.6)	-440.7(2.6)	.3052(14.0)	1,343.3	.163
(B)	.0215(4.9)	240.3(2.6)	-95.26 (.5)	.2871(11.3)	1,327.9	.137

TABLE 5
Tests of Alternative Expected Car Purchase Variables, A Sample

Ex-Ante Variable	Y	ΔY^a	ΔY^e	Ex-Ante Variable	SE	R^2
C*	.0108(2.7)	257.8(2.6)	-298.9(1.9)	.3248(13.2)	1,238.6	.137
C_1^*	.0152(3.7)	251.5(2.5)	-254.1(1.6)	.2865(10.7)	1,263.5	.102
			First Difference Data			
C*	.0063(1.0)	-356.8(2.3)	532.4(2.2)	.3936(13.8)	1,930.4	.128
C_1^*	.0033 (.5)	-361.8(2.3)	532.3(2.2)	.3331(10.9)	1,975.6	.086

TABLE 6
Alternative Ex-Ante Savings Functions, A Sample

Dependent Variable	Y	ΔY^a	ΔY^e	Ex-Ante S_1^*	Ex-Ante S_2^*	SE	R^2
S	.0347(3.3)	960.0(4.2)	-799.1(2.2)	.2759(6.8)		2,865.6	.081
L	.0319(3.1)	767.2(3.4)	-581.0(1.7)	.2701(6.8)		2,799.9	.077
S	.0455(4.3)	959.8(4.1)	-765(2.1)		.1040(4.1)	2,894.8	.062
L	.0436(4.2)	766.1(3.4)	-535.0(1.5)		.0962(3.9)	2,830.4	.057

TABLE 7
Alternative Ex-Ante Savings Functions, First Differences, A Sample

Dependent Variable	Y	ΔY^a	ΔY^e	Ex-Ante S_1^*	Ex-Ante S_2^*	SE	R^2
S	.0127 (.8)	-360.9 (.9)	404.6 (.7)	-.0182(.4)		4,795.1	.008
L	-.0138(1.1)	-738.1(2.4)	572.6(1.4)	-.0339(.9)		3,881.2	.012
S	.0098 (.6)	-365.0(1.0)	376.1 (.6)		-.0451(1.4)	4,792.0	.009
L	-.0147(1.2)	-767.5(2.5)	687.3(1.4)		-.0182 (.7)	3,881.8	.011

TABLE 8
Ex-Ante Savings Functions with Durable Goods Expenditure Variables, S is Dependent Variable

Y	ΔY^a	ΔY^e	S_1^*	D	D*	SE	R^2
.0316(3.0)	961.2(4.2)	-847.5(2.4)	.2700(6.6)		.2324(2.4)	2,860.9	.085
.0330(3.2)	937.6(4.1)	-784.7(2.2)	.2752(6.7)	.1750(1.3)		2,864.9	.082
				C	C*		
.0337(3.2)	962.4(4.2)	-812.5(2.3)	.2735(6.7)		.0383 (.7)	2,866.2	.081
.0375(3.6)	993.6(4.3)	-824.8(2.3)	.2782(6.8)	-.1424(2.4)		2,860.5	.085

TABLE 9
Survey Sampling Test, S Dependent Variable

Sample	Y	ΔY^a	ΔY^e	S_1^*	SE	R^2
(A)	.0347(3.3)	960.0(4.2)	-799.1(2.2)	.2759(6.8)	2,865.6	.081
(B)	.0428(4.1)	710.0(3.5)	-102.1 (.3)	.1094(2.8)	2,881.5	.045

TABLE 10
Test of Financial Asset Quality Editing in Savings Functions, A Sample

Dependent Variable*	Y	ΔY^a	ΔY^e	S_1^*	S_2^*	C	SE	R^2
S	.0347(3.3)	959.9(4.2)	-799.1(2.2)	.2759(6.8)			2,865.6	.081
S'	.0366(3.8)	809.1(4.0)	-509(1.6)	.2684(7.3)			2,794.5	.085
S	.0375(3.6)	993.6(4.3)	-824.8(2.3)	.2782(6.8)		-.1424(2.4)	2,860.5	.085
S'	.0392(4.0)	836.4(4.2)	-527.2(1.7)	.2702(7.4)		-.1390(2.5)	2,789.6	.089
S	.0455(4.3)	959.8(4.1)	-765.4(2.1)		.1040(4.1)		2,894.8	.062
S'	.0488(4.9)	813.1(4.0)	-463.7(1.5)		.0970(4.1)		2,827.8	.063
S	.0483(4.5)	991.4(4.3)	-787.9(2.2)		.1042(4.2)	-.1341(2.3)	2,890.5	.066
S'	.0514(5.1)	839.5(4.1)	-480.8(1.5)		.0977(4.1)	-.1341(2.4)	2,823.3	.066

*S = Total savings with observations eliminated if financial asset information of poor quality (1,387 observations).

S' = Total savings, no financial asset quality eliminations (1,537 observations).

DISCUSSION

Robert J. Willis, National Bureau of Economic Research
and Graduate Center, City University of New York

In addition to their common use of data from the Consumer Anticipation Survey, the papers by Landsberger and Michael share a common theoretical framework based on Gary Becker's theory of the allocation of time and the theoretical characterization of the household which stems from it. Becker's theory enables them to incorporate into economic models of the household forms of behavior such as the number, spacing and quality of children and the division of labor within the household between the spouses which are often considered to be more within the realm of sociology than economics. More generally, Becker's model provides a way by which economists may attempt to formulate a unified explanation of a large number of variables, such as those encountered in the CAS survey, which describes the household and its behavior. If such attempts are to be fruitful, certain variables must be measured and certain theoretical difficulties must be met. I think that the Becker model or some modification of it will increasingly be the framework within which economists organize their inquiries into household behavior and I think the two papers before us represent interesting, but flawed examples of how to use the model. Accordingly, I shall first briefly describe the theoretical structure of the Becker model to set the stage for evaluating the use made of it by Landsberger and Michael.

According to Becker, households do not obtain direct satisfaction or utility from goods and services purchased in the market. Rather, to obtain satisfaction a household must combine the purchased good -- say, soap -- with the time of one or more household members as inputs into a household production function whose output -- say, cleanliness -- is the quantity that directly affects utility. In effect, then, the household is both the demander and supplier of its final wants and its demands for produced goods and services, conventionally treated as final, are in this model derived demands analogous to the derived demands for labor or capital in standard production theory.

Becker likened the household to a small factory using inputs to produce outputs. A more apt analogy, I think, is to liken the household to a small socialistic economy in which the "planner's preference" is maximized by allocating resources physically according to "commands." The planner represented by the household decision maker(s) -- perhaps the husband and wife acting in concert -- must allocate many factors of production -- member's time and individual types of purchased goods -- between many alternative uses, e.g., current cleanliness, warmth and child services, future productivity via education and future purchasing power via saving. The characteristics of this model are formally identical to those of a dynamic Walrasian general equilibrium system in which the planner's utility function and the household production functions generate market clearing, utility maximizing shadow prices for n factors of production and m final consumer goods where n and m are large numbers and where production and consumption may occur at different dates. Economists have learned that in its full glory, the Walrasian system is intractable for practical problems. Thus the essence of applying the Becker or the Walrasian model is to choose the appropriate level of simplification or complication. It is at this stage that both papers before us are flawed.

The paper by Landsberger presents a model of the labor supply of husbands and wives and the amount of family consumption as functions of their wage rates and the number and ages of their children which is derived from a time allocation model of the type just described. Since the main focus of this paper is on the "children effects," it is rather surprising to find that children occur nowhere in Landsberger's mathematical model. Instead, the effect of children on labor supply and consumption is introduced from outside the model by means of shifts in the marginal product schedules of goods and the time of each spouse devoted to household production that the presence of children of certain ages are supposed to cause.

The reasons given for these shifts seem to be inconsistent with the aggregation of all household outputs into one aggregate commodity X and this inconsistency, in turn, may be the reason that the effects of children were introduced in an <u>ad hoc</u> fashion from outside the model. The children effects seem to derive mostly from the hypothesis, parts of which are stated in various parts of the paper, that young children are relatively time intensive users of the wife's time and that, as they age, children become progressively less time intensive. Added to this is the more tentative hypothesis that the husband's time is used more intensively in non-child oriented activities.

One of the main implications of Becker's model is that the "shadow prices" of household outputs of commodities which are relatively time intensive will tend to increase as the price of time measured by market wage rates rises. Thus, households whose wage rates differ will also face different shadow prices for commodities. The Hicksian composite commodity theorem which justifies the aggregation of all purchased market goods into Y, on the assumption that all households face same set of market prices cannot, therefore, be used to justify the aggregation of all commodities into X unless all commodities are assumed to have identical factor intensities which, of course, contradicts the hypothesis that leads us to expect children effects.

These difficulties could easily be overcome if Landsberger would specify a two sector model in which household outputs are aggregated into two commodities, child services and other activities, the former being assumed intensive in the wife's time relative to the latter but growing more like the latter as the children age. One implication of this two sector model is that the relative shadow price of child services will increase as the wife's market wage increases, but that the sensitivity of the price of child services to the wife's wage diminishes as they age and become less time intensive. Landsberger assumes fertility to be exogenous so that this particular implication is without behavioral significance in his model. However, it is also implied by the two sector model that a shift in the composition of household consumption toward (wife's) time intensive commodities would tend to raise the marginal product of the wife's time at home causing her to withdraw labor from the market until the marginal product of her time at home and in the market are equal. This effect will diminish as children age and may even disappear or reverse itself if children become equally or less time intensive than other household activities. Landsberger's empirical results support these implications and, therefore, support the intuition which lead him to the time intensity hypothesis but they contradict his formal model.

Michael's model of household fertility, child spacing and child quality, on the other hand, seems to me to be insufficiently aggregated or at least insufficiently specified to sustain the considerable explanatory burden he places upon it. Using the same triad of inputs as the Landsberger model (husband's time, wife's time and market goods), the household produces a large set of commodities whose quantities enter into household utility. Of these commodities, Michael focuses mainly on the household's consumption of the commodity child services which, unsurprisingly, requires inputs of actual children in addition to time and goods inputs. Household fertility behavior, therefore, will depend on the demand for the quantity of children considered as a factor of production which is derived from the final demand for the commodity child services. The demand for child quality which, for empirical purposes, is identified with the level of schooling the child is expected to complete, is also derived from the demand for child services.

The implications of Michael's model for fertility and child quality, therefore, depend on the properties of derived demand functions in the case in which there are four factors of production. A strong argument can be made for the proposition that such derived demand functions have no empirically refutable implications unless the structure of the model (i.e. the utility function and household production function) is severely restricted. Diewert[1/] has shown, for example, that the elasticity of a derived demand function in the three factor case depends on eight parameters which include the partial elasticities of substitution between factors, the supply elasticities of each factor, factor shares in total cost and the elasticity of demand for the final product. The four factor case involves still more parameters. Michael's assumption that child services are relatively intensive in the wife's time will suffice to establish that the shadow price of child services will be an increasing function of the wife's wage if the wife works, but far more must be assumed before we know what this implies for fertility or child quality. Until we do know these implications, it is difficult to know what to make of his empirical work in either a hypothesis testing or descriptive framework.

While the emphasis in my discussion has been to stress the theoretical difficulties that may be encountered in applying the Becker model, I think these papers also illustrate the exciting prospect that economics may provide a really unified account of many seemingly unconnected aspects of household behavior.

FOOTNOTES

1/ W. E. Diewert, "A Note on the Elasticity of Derived Demand in the N-Factor Case," Economica (New Series) May, 1971, pp. 192-7.

LABOR MOBILITY: AN ANALYSIS OF SOCIAL SECURITY DATA FOR ATLANTA[1]

Kathryn P. Nelson, Oak Ridge National Laboratory

Over 30 years ago, Dorothy Thomas recommended that data being collected by the Social Security Administration on all those covered by Old Age, Survivors, Disability and Health Insurance (OASDHI) program be used for migration and sociological research [1]. Ten years later, Donald Bogue and his associates used these data for a pioneering work on labor force mobility in Michigan and Ohio [2]. Though the number of papers based upon the Social Security data is finally growing, the materials still represent a largely unmined resource.

Using one city, Atlanta, as an example, this paper will illustrate the unique capability of the Social Security One Percent Current Work History Sample (CWHS) to indicate the direction and extent of both geographical and industrial mobility as they relate to aggregate changes in employment. At present, monthly employment and unemployment for states and large SMSAs by industry, by race, and by sex can be estimated from reports from a sample of establishments. Estimates of net migration for states--and with less accuracy for large SMSAs--are possible using census survival methods, and with each census the streams of migration over the previous five years can be more precisely delineated. But except for the census one percent public use sample, these are group statistics, and cannot be related to each other.

On the other hand, with the CWHS estimates of employment and labor force by industry can be combined with knowledge of the personal characteristics of the workers, including income, sex, and age; and with these data the same individuals can be followed over time. Thus we can study the personal characteristics of the mobile and nonmobile--including their origins and their destinations--by industry. This paper draws from a longer monograph on Atlanta to illustrate briefly some of these possibilities. First we compare establishment data with Social Security data, and then devote the remainder of the paper to a discussion of industrial and geographic mobility.

Data were drawn from the CWHS for every worker in the one percent sample who was in covered employment[2] in the Atlanta SMSA in 1962 or in 1967. The following items were extracted for each individual for both years: major industry of employment by SIC (defined as the single industry of highest earnings); location of employment by state and SMSA; total earnings for the year; and age, race, and sex.

Changes in Employment by Industry

Since the Social Security data constitute a one percent sample and do not cover the entire labor force, we are interested in ascertaining its biases when we seek to use it for small area analyses.[3] As Table 1 indicates, both the Social Security data and data derived from establishment reports of average yearly employment[4] show a 35% rate of growth for the Atlanta SMSA over the five year period, although the difference in magnitude between these two sets of data is roughly 20%. In a sense, this 20% disparity is a measure of turnover, since the Social Security data relate to the number of persons holding jobs in Atlanta at any time during the year, while the establishment data refer to the average number of jobs available during the year.

The final two columns of Table 1 show that the ratio between the two measures varies widely by industry--from 1.75 for services in 1967 to .65 for government in 1967. The latter ratio undoubtedly reflects the limited Social Security coverage of state and local government personnel, a gap that must make tentative any conclusions about government employment from these data. Equally, the disparities in the areas of services and construction suggest caution in interpretation. Over the five year period many more people were, in fact, employed in these industries, but the number of jobs did not grow as quickly as the Social Security data might indicate.

When the industries are ranked by rates of growth, as shown by the two data sources, the order is quite similar except for services and government. The range of rates of growth shown by Social Security data is greater--varying from 17% for finance, insurance, and real estate to 68% for services; while establishment data range only from 25 to 48%. Thus we must bear in mind that Social Security data appear to overstate the extent of change in employment.

At both ends of the period, trade accounted for the largest number of workers, more than a quarter of the total, a reflection of Atlanta's preeminence as a trade and distribution center. Manufacturing was second, but the proportion in this industry declined from nearly a quarter to little more than a fifth in five years. Although the high turnover rate in services results in some overstatement of growth, both in number and proportion, services was the growth industry for Atlanta during this period. Nearly a third of the total increase in workers was in services, and there were three service workers in 1967 for every two in 1962.

The lowest rates of growth, excluding the insignificant agriculture and mining categories, were in manufacturing and finance. Government and trade also grew at below average rates. The declining role of manufacturing mirrors the national trend, and is partly due to increases in productivity, and a change in the industrial mix toward larger and more efficient operations.

Change in the Atlanta Labor Force 1962 and 1967

Figure 1 shows that, of the 484,000 persons in covered employment in 1962, only 60% were still employed in Atlanta in 1967, and they constituted less than half of the city's labor force. One in five had left covered employment and one in five were working outside the SMSA.[5]

Blacks of both sexes were more likely to remain in Atlanta, and were less likely to migrate to other states. Males were more likely to move than females, and were more likely to migrate over long distances.

Net in-migration and 206,000 new entrants combined to produce a 1967 labor force in which only 44% were holdovers from 1962. It is obvious that the high percentages of female retires and new entrants reflect more mobility in and out of the labor force by women. Such high apparent turnover reminds us that the Social Security data are not the perfect registration system we might desire for studying labor mobility. Even among males, almost half of the retires were under age 45 and thus can be assumed not to have retired in the conventional sense. However, the extraordinarily high percentage of black female entrants undoubtedly reflects new opportunities for employment. It is also notable that these new entrants were not concentrated in personal services.

Considering this change in terms of age, we find that partly because of the large influx of young entrants, there were, percentage wise, more workers under age 24 and relatively fewer aged 25 through 44 in 1967 than in 1962. The percentage of workers aged 45 and over was almost exactly the same (27.7 and 27.8%, respectively), although the proportion of oldest (over 60 years) workers increased slightly. The relative decrease in the middle age-group may reflect the small cohorts of Depression babies; it is more likely that the decrease arises from the greater tendency of younger workers to change jobs and take advantage of a quickly growing area like Atlanta.

Labor Mobility of Atlanta Workers

The large number of entrances into and exits from the Atlanta labor force, its substantial growth during the five year period, and the degree of turnover that may be inferred from the disparities in level between Social Security and establishment data all combine to suggest high labor mobility and many job openings in Atlanta during this period. But since we have not considered changes in jobs for those who remained in the Atlanta labor force, we have thus far underestimated the true extent of mobility.

With the CWHS, it is possible to infer change of employers (job mobility) since all sources of income in covered employment are given, but the tabulations prepared for this report do not contain such information. For ease of data handling, employees were assigned to the industry and location of their job of highest earnings in each of the two years considered, 1962 and 1967. Thus we consider here only two kinds of mobility, industrial and geographic. We define industrial mobility to have occurred when the industry (as measured by one-digit SIC code) of employment in 1967 differs from that in 1962. A change in the place of employment across the boundaries of the Atlanta SMSA was taken to constitute geographic mobility.[6]

Table 2 presents the data on mobility for all workers who appear in tabulations for both 1962 and 1967. The upper portion of the table shows the destination of all in the Atlanta labor force in 1962 who were still in covered employment in 1967, and percentages have been calculated to indicate differences in mobility among the different sex-race groups. We observe again that whites are more geographically mobile, but we also note that blacks are much more likely to change industry. When only those who are geographically mobile are considered in the lower portion of the table, the differences in industrial mobility between blacks and whites appear to be less for Atlanta in-migrants than for the out-migrants.

Industrial Labor Mobility

Perhaps as well as migration, industrial mobility indicates how efficiently the economy is functioning to reallocate manpower and resources in spite of institutionalized hindrances. Indeed, as the number of intrafirm moves increases, industrial mobility (with its inferred abandonment of pension plans, seniority, etc.) may better measure the flexibility of the labor force. In this respect the Social Security data permit examination not only of the personal characteristics and industrial affiliation of industrially mobile workers, but also of the effects of such mobility on particular industries. Thus we can, for example, determine the attractiveness of given industries to experienced workers, the labor force components of their relative growth and decline, and what kinds of workers move in and out and at what rates of pay.

For the Atlanta SMSA, as has been observed more generally elsewhere [3], it is the young, the poorly paid, and the blacks who are more likely to change industry, regardless of geographical mobility. Figure 2, which shows the relative distribution of Atlanta stayers, out-migrants, and in-migrants by age and industrial mobility, graphically represents the greater likelihood of young workers to change industry. When stayers are compared with migrants, we see also that geographical mobility is associated with a greater likelihood of industrial change; 56% of those who changed location also changed industry, while only 38% of those who remained in Atlanta worked in a different industry in 1967. Those who changed industry also had greater increases in incomes than those who did not change industries.

To consider the impact of industrial shifts by those employed in both 1962 and 1967 (both stayers and migrants) upon industries in the Atlanta area, net industry change (the total of moves in and out of a particular industry) as a percentage of employment in 1962 was calculated. As would be expected, agriculture experienced net losses equivalent to 73% of its 1962 employment. Less obvious, perhaps, is the 20% loss of experienced workers by retail trade--an industry that grew by 30% in the five year period. The greatest gains of experienced workers through industry shifts occurred in the areas of personal services and construction.

Geographic Labor Mobility

Figures 3 and 4 illustrate two of the many ways in which the Social Security data can illuminate the processes of labor mobility in a metropolitan area. Streams of migration between Atlanta and other areas and the resulting net migration are shown in Fig. 3. As is usually the case, the total number of migrants between Atlanta and any other region was much greater than the net redistribution of population accomplished by such streams. The effectiveness of migration, defined as:

$$\frac{|\text{in} - \text{out}|}{\text{in} + \text{out}}$$

was greatest in redistributing people from the other Southern states to Atlanta. To the extent that Atlanta's work force grew through net migration, it can be seen that the favorable balance was supplied largely by interchange with the South, and only marginally by net in-migration from the Northeastern and Western states. Atlanta's status as a growth center relative to the whole country is shown, however, by the lack of significant net out-migration to any region. Net in-migration into Atlanta also characterized all sex, race, and age groups, except for nonwhite males aged 18 to 24.[7]

In the graph depicted in Fig. 4, we take advantage of our knowledge of the characteristics of the entire 1962 Atlanta work force to calculate out-migration as a rate of all employed for 12 different age groups, thus in a sense, measuring the probability of such migration. Rates are shown for the total work force, for white males and females, and for nonwhite males.[8] Since our sample is now quite small, the resulting curves are something less than smooth. However, comparison with Fig. 5, which shows the remarkably regular curves that approximate rates of migration over a five year period for white and nonwhite male by single years of age in 1960, reveals important similarities. In each case rates of migration rise from low levels for teenagers to peak in early adulthood and decline thereafter. As with the national sample, migration appears to be more selective by age for nonwhites than for whites, and for females than for males.

Finally when the economic consequences of migration are considered, we find that the migrant improves his level of income at a greater rate than the nonmigrant, even when we account for the different distributions of the two groups in terms of age, sex, race and industry change classifications.

Summary

This paper has illustrated how the Continuous Work History Sample can be used to provide insight into the relationships between labor mobility and changes in the level of employment for large SMSAs, portraying to a greater extent than is otherwise available the individual characteristics and work histories of those who are mobile. The examples given are scattered--for example, little has been said about income although this is one of our major interests and our data appear to indicate that the migrant improves his level of income at a greater rate than the nonmigrant even when we control for differences in age, sex, and race. Yet, even the few examples given here are suggestive--for trade, the largest industry in terms of size and one that grew at a close to average rate by both data sources, we find a net loss of experienced workers, such that growth in employment depended upon recruiting new entrants and same-state migrants.

For Atlanta, an SMSA whose central city is now 51% black, we find that in spite of respectable growth and high turnover during the five year period, blacks increased their share of the labor force only slightly from 18 to 19%. Regional migration flows show at the same time that Atlanta is attractive to migrants from all parts of the country.

To turn to the quality of the data for these purposes, it is evident that the usefulness of these data for small area analysis will be increased as we gain information about the significance of the "precise" estimates that can be drawn from the data. For Atlanta, estimates of employment level and growth by broad industrial classifications compare favorably with those derived from establishment data. Further work should include comparing results with those from other sources of data on the labor market and migration.

The imminent release of the 1970 First Quarter CWHS data will provide valuable opportunities for checking results for large SMSAs with census data, both the sample questions on migration, place of work, occupation, income and such, and the one percent public use sample for large SMSAs. Comparisons between different types of cities should also increase the utility of these data, and it is in this direction that our future work will head.

Footnotes

[1]Work supported by the Department of Housing and Urban Development and the U.S. Atomic Energy Commission under Interagency Agreement No. IAA-H-35-70 AEC 40-192-69, and conducted at Oak Ridge National Laboratory, Oak Ridge, Tennessee operated by Union Carbide Corporation for the U.S. Atomic Energy Commission. Data from Social Security Administration's Work History Sample tapes were processed by David Hirschberg, Regional Economics Division, Office of Business Economics. At no time were individual records made available to the author. Opinions expressed in this report are solely those of the author and do not necessarily represent the views of ORNL, AEC or HUD.

[2]Throughout the U.S. about 88% of workers are in covered employment and thus represented by the sample. Those workers who are most likely to be excluded from coverage are federal civilian employees, some state and local government employees, household and farm workers who do not work long enough or earn enough to meet the minimum requirements, and very low income self-employed persons (Current Population Survey P-23, No. 31). Therefore, reported figures for women and blacks are probably a lower percent of the actual employment.

[3]The Atlanta SMSA with a 1970 population of 1,390,164 is smaller than the smallest areas studied with Social Security data in the past, such as Michigan and Ohio [4]; North Carolina, South Carolina, and Georgia [5]; and the TVA region [6,7].

[4]Georgia Employment and Earnings: A selected sample of employers report the number of full and part-time workers during the second week of each month, and these figures are averaged for the year.

[5]The following terminology is used in this paper:

Retires = those who were in covered wage employment in 1962 but not in 1967.

New entrants = those who were in covered wage employment in 1967 but not in 1962.

Stayers = those who were in covered wage employment in Atlanta in both 1962 and 1967.

In-migrants = those who changed from covered employment outside the Atlanta SMSA in 1962 to covered employment in Atlanta in 1967.

Out-migrants = those who changed from covered employment in the Atlanta SMSA in 1962 to covered employment outside Atlanta in 1967.

[6]We cannot strictly equate such mobility with migration. Especially as Atlanta's highway system expands and commuting distances lengthen, such an assumption becomes less tenable. Because of Atlanta's location in the state, however, it seems safe to assume that interstate job changes represent migration. Such essentially arbitrary decisions must be made whenever Social Security data, showing change in county of employment, are used to study migration, which is usually defined as a change in county of residence.

[7]However, considering the use of Social Security's Continuous Work History Sample for measure of net migration by geographic area, Zitter and Nagy [8] show that it appears more likely that the CWHS will pick up in-migration to low coverage states from high coverage states, and less likely to reflect out-migration from low coverage states to high coverage states. Thus in-migration may be overestimated and out-migration underestimated by our data between Atlanta and the more industrialized states of the Northeast, especially for nonwhites for whose rates standard errors are higher.

[8]There were too few nonwhite female migrants to warrant this calculation.

References

[1] Dorothy S. Thomas, "Utilization of Social Security Data for Sociological Research," American Sociological Review, October 1938.

[2] Donald J. Bogue, An Exploratory Study of Migration and Labor Mobility Using Social Security Data, Scripps Foundation Studies in Population Distribution - No. 1, 1950.

[3] Lowell A. Gallaway, "Interindustry Labor Mobility in the U.S.--1957 to 1960," Research Report No. 18, Office of Research and Statistics, Social Security Administration, Washington, D. C., U. S. Government Printing Office, 1967.

[4] Donald J. Bogue, A Methodological Study of Migration and Labor Mobility in Michigan and Ohio in 1947, Scripps Foundation Studies in Population Distribution - No. 4, 1952.

[5] Robert L. Bunting, et al., "Labor Mobility in Three Southern States," Industrial and Labor Relations Review, Vol. 14, No. 3, April 1961, pp 432-45.

[6] Boonlert Leopropraі, Mobility and Economic Progress in the Tennessee Valley Region: 1957-63, University of Georgia, unpublished Ph. D. thesis, August 1970.

[7] Wesley G. Smith and Robert A. Matson, "The Mobility of the Tennessee Valley Employed Labor Force," presented at the Symposium on Assessing Social Costs sponsored by the Civil Service Commission and The Association for Public Program Analysis, Washington, D. C., June 16, 1969.

[8] Meyer Zitter and Elisabeth S. Nagy, "Use of Social Security's Continuous Work History Sample for Measurement of Net Migration by Geographic Area," American Statistical Association, Proceedings of the Social Statistics Section, 1969, pp 235-60.

Table 1. Growth in Employment by Industry, Atlanta SMSA Counties, 1962 and 1967

Industry of Employment	Establishment Data*			Social Security Data**			Ratio of Social Security to Establishment Data	
	1962	1967	Growth Rate (%)	1962	1967	Growth Rate (%)	1962	1967
Total Covered Employment				484,000	653,000	34.8		
Total Nonagriculture	397,250	534,000	34.3	478,200	647,100	34.8	1.20	1.21
Construction	22,500	33,200	48.0	31,800	49,900	57.0	1.41	1.50
Manufacturing	90,150	117,000	30.0	116,100	137,300	18.0	1.29	1.17
Transportation	37,200	51,100	37.0	37,000	55,100	49.0	.99	1.08
Trade	104,700	140,300	34.0	136,900	177,400	29.6	1.31	1.26
Finance, etc.	29,950	37,300	25.0	34,700	40,700	17.3	1.16	1.09
Services	55,750	77,300	39.0	80,600	135,600	68.2	1.45	1.75
Government	57,000	78,100	37.0	41,000	51,100	24.3	.72	.65

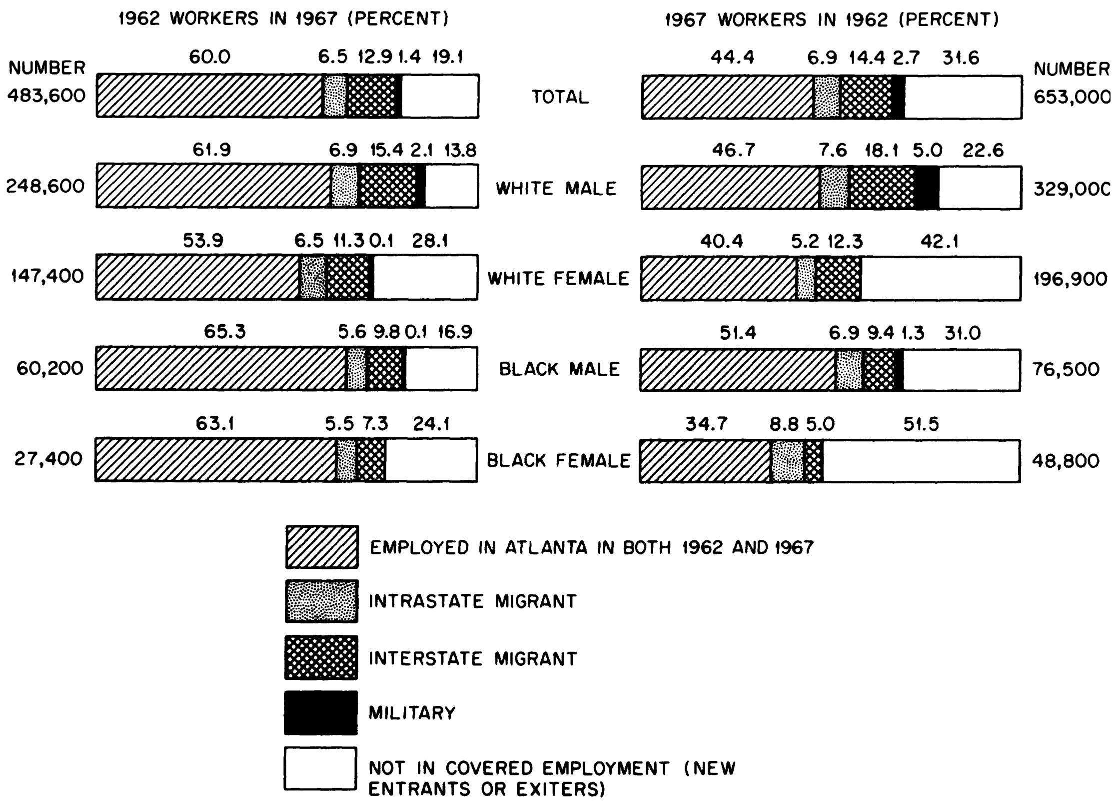
*Source: Georgia Department of Labor "Employment and Earnings."

**Source: Estimated from the One Percent Continuous Work History Sample, Social Security Administration.

Table 2. Mobility of Atlanta Workers

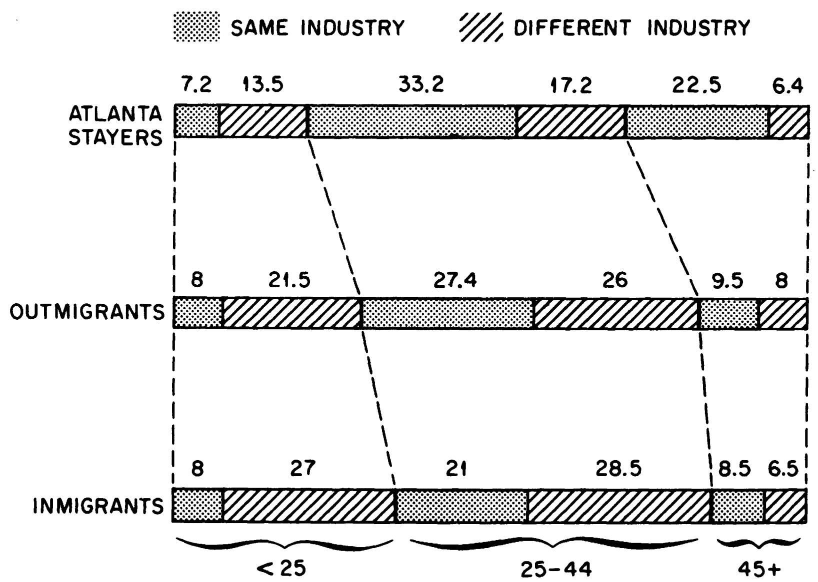
Mobility Status	Number of Workers (in thousands)					Percent				
	Total	Male		Female		Total	Male		Female	
		White	Black	White	Black		White	Black	White	Black
	Working in Atlanta in 1962 and in Covered Employment in 1967									
No mobility	182.2	101.1	18.6	53.1	9.4	47.2	48.2	38.9	50.2	44.3
Same SMSA Different industry	107.6	52.8	20.5	26.2	8.1	27.9	25.1	42.8	24.8	38.2
Geographical mobility from Atlanta										
Same industry	43.1	27.1	2.2	11.6	1.2	11.1	12.9	4.6	10.9	5.6
Different industry	52.4	28.7	6.5	14.7	2.5	13.5	13.6	13.5	13.9	11.7
Totals for all workers	385.3	209.7	47.8	105.6	21.2	100.0	100.0	100.0	100.0	100.0
	Working Elsewhere in 1962 and in Atlanta in 1967*									
Geographical mobility to Atlanta										
Same industry	52.5	33.0	3.4	13.8	2.3	38(45)	39(48.5)	31(24)	38.5(44)	33(32)
Different industry	86.4	52.1	7.7	22.0	4.6	62(55)	61(51.5)	69(76)	61.5(56)	67(68)

*Figures shown in parentheses below indicate percent of persons in Atlanta in 1962 but elsewhere in 1967 who remained in the same industry or changed industry.



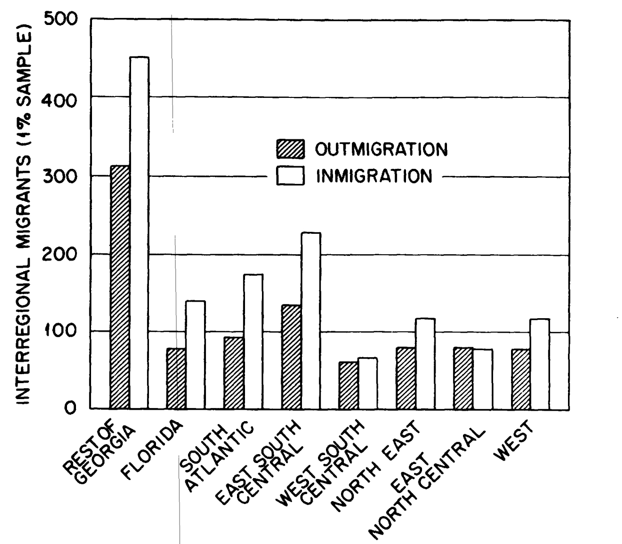
Migration and Labor Force Status of Atlanta.

Fig. 1



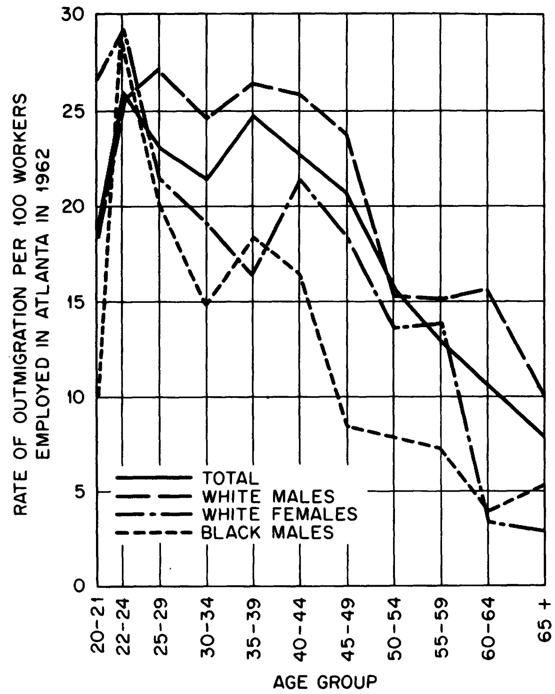
Atlanta Stayers, Outmigrants and Inmigrants by Age in 1962 and Industrial Mobility.

Fig. 2



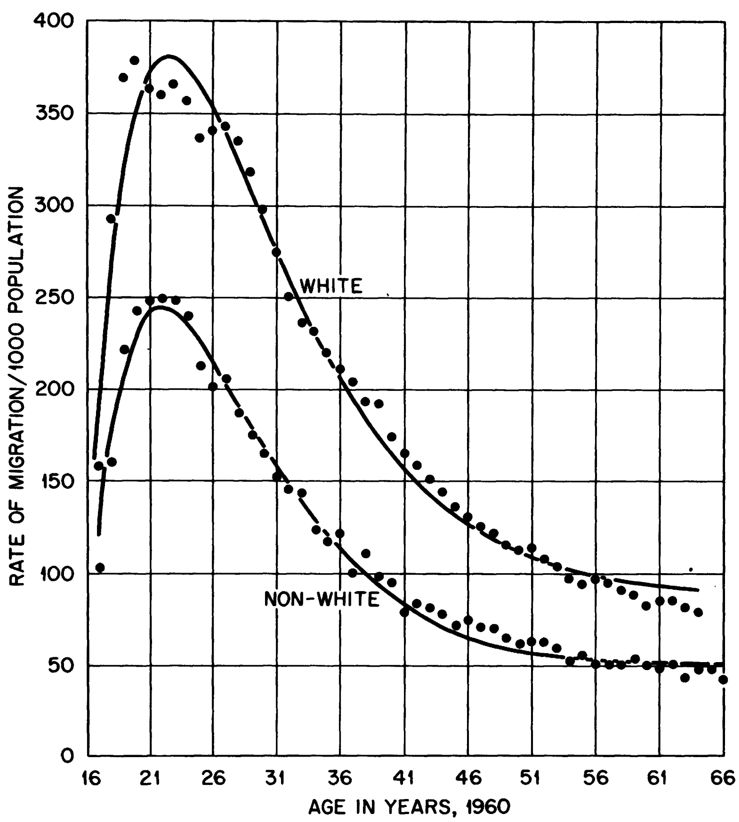
Regional Migration Streams to and from Atlanta 1962 and 1967.

Fig. 3



Rates of Outmigration from Atlanta in 1962 by Age, Sex, and Race.

Fig. 4



Total White Male Migration, U.S. 1955-60, as Compared with Total Non-White Male Migration.

Fig. 5

INTERSTATE HIGHWAYS AND THE GROWTH AND DISTRIBUTION OF POPULATION[1]

Robert A. Bohm[2] and David A. Patterson[3]

Introduction

The relationship between improved transportation facilities and regional economic growth is generally believed to be strong and positive. With the exception of some quantitative historians [5, 8], economists and geographers agree with the conventional wisdom that the canals, railroads and highways all contributed substantially to U. S. economic development. Further support for this view has been provided by a large volume of state and federal government studies that purport to show a positive relationship between new and better highways and economic growth [2, 3, 6, 7, 12, 15, 20].

In addition to the economic effects of improved transport systems, many observers also predict that population movements, and thus inter-regional differences in population growth, might also be influenced by more efficient transportation. Brian Berry argues that better roads, and then the automobile, enabled the rural population to bypass lower level market centers (small central places) and to journey to the larger centers where chain store services are available [1, pp 114-5]. The net result has been a population decline in the very small places beyond that associated with the mechanization of agriculture. "Before 1930 hamlets with populations of 100 or less were declining; thereafter, as centralization of functions in higher levels of the hierarchy progressed, the general decline embraced villages with population of less than 500" [1, p 115].

Of course, the relatively recent decentralization trend of major industries out of the central cities and into the suburbs and out of the older more mature northern states and into the previously less industrialized southern states, has also had an effect on relative rates of population growth, more or less reversing the previous out-migration from the southern region. Industrial location theorists as well as empiricists tend to place transportation costs well up on the list of important locational considerations [4, 13, 14]. Hence, improvements in highways in general, and the interstate highway system in particular, should have affected the location of economic activity and hence the distribution of population.

In spite of the weight of theory and of empirical evidence (albeit relatively unsophisticated in the case of the empirical support), there are some who have questioned the real value of transportation improvements as a stimulus to real economic growth. Fogel and Cautner [5, 8] have both presented arguments that minimize the influence of the early railroads and canal systems. With specific reference to the interstate highway system, Friedlaender argues that "since all of the centers of production ...are already connected by an extensive network of highways and rail facilities, it seems unlikely that the...system will trigger sizable investments that would not have occurred in its absence" [9, p 64]. These authors would appear to be supporting the contention that the interstate highway system merely serves to connect already growing urban centers. According to this line of reasoning, current rates of local economic and demographic change are determined in the main by past rates of change.

The impact of highway system improvements on economic development and on inter- and intra-regional population distribution is of more than academic interest. Federal Government policy makers and others are currently expressing an interest in a policy of urban decentralization. Under our political system a decentralization policy can only be accomplished indirectly through the manipulation of a relatively small number of policy variables. Thus the possible role of the interstate system in the concentration or dispersal of population and economic activity is an empirical question of policy significance, at least to the extent that highway location is a policy variable.

The interstate highway system was begun in 1957. By the end of 1968 nearly two-thirds of the systems' planned mileage was open to traffic. The effect of this new highway system on the population growth of counties and on the intra-regional distribution of population is the subject of our analysis.

In a general sense, any change that lowers the cost of producing and distributing a product can be a source of economic growth. If the cost reduction is differentially distributed geographically, then the growth effects should be likewise distributed. That is, the cost reduction will cause some geographic areas to grow more rapidly than those areas not sharing in the cost saving. We might expect this differential growth to be composed of two parts:

(1) Net new growth that would not have taken place in the absence of the significant cost saving, and

(2) Transfer effects which can be either:
   a. Replacement of already existing or planned economic activity; e.g., a shift in the locational pattern of industries as firms move from their previous locations or expand in different ones in an attempt to realize locationally determined cost savings.

   b. Use of resources otherwise employed; e.g., land shifted from agricultural to industrial use.

Thus, a portion of an area's growth can be described as "new," or growth that would not otherwise have taken place at that time, while the balance represents a "transplanted" growth, i.e., a redistribution of activity that would normally have taken place elsewhere.

The growth that results from a cost saving also may induce further growth to the extent that new concentrations of industry and population provide new supplies of material and labor as well as new markets for output. As before, a portion of this growth may be described as "net new growth," developing, in this case, out of the external economies associated with the first-round growth effects. And, a portion of the growing area's change will reflect transfers, for example, a relative decline elsewhere as population and industry shift to the places offering lowered cost or increased marketing opportunities.

An interstate highway should have the effect of lowering the cost of transportation, possibly changing the distribution of feasible locations. Whether or not the new set of locations is sufficiently attractive to encourage net new growth or to force a transfer depends on whether the cost reduction is sufficiently large to offset locational inertia. Only a small number of industries, such as textiles and some assembly operations are described as "footloose" [13]. For most industries, the perceived cost of a move is enough to yield a high degree of locational stability. The degree of competition perceived by the individual firm is also a factor conditioning its need to respond to marginal changes in locational advantage.

Another factor affecting the response of firms to lowered transportation cost must be the extent of external economies at various location alternatives. These may change over time due, for instance, to changes in the structure of the labor force, to the presence of complementary firms and to the availability of services. Thus, some firms will have a lagged response to changes in the optimal location due to changes in transport cost.

In general, we would expect regional changes in economic activity to be reflected in corresponding regional population changes. Studies have recorded instances of employees commuting to work very long distances, as far as 60 miles and more [3, pp 65-68]. However, other studies support the view that employees' transportation cost is an important variable in the choice of residence decision [16, 18]. Thus, while granting the possibility of a lagged response, we would expect regional shifts in industrial location to be accompanied by shifts in the distribution of population within and among regions.

If the interstate highway system has merely served to connect already growing places without markedly shifting the pattern of optimal industrial locations, the effects of interstate location on county population change should have been negligible over the 1960 to 1970 decade. On the other hand, if the interstates have lowered transport costs sufficiently to generate net new growth and/or transfer effects as defined above, population changes over the decade should reflect this phenomenon. That is, interstate highway location should result in changes in county population growth that are independent of past population changes.

The Model

In this analysis, the process of county population growth is assumed to be largely autoregressive in nature. In other words, county population change from 1960 to 1970 is primarily determined by population change in previous decades. Specific county characteristics, such as the existence or nonexistence of an interstate highway, will merely lead to deviations about the growth trend.

The model may be stated as follows:

$$\Delta P_t = \alpha + \beta_1 \Delta P_{t-1} + \beta_2 \Delta P_{t-2} + bX_t + \epsilon_t \quad (1)$$

where county population change during the current period, ΔP_t , is expressed as a function of population change in previous time periods, e.g., t-1, t-2, etc. A matrix of county specific parameters is shown as X_t, and ϵ_t is a random disturbance variable which is assumed to be [4] distributed independently of ΔP_{t-1}, ΔP_{t-2}, etc. [11, pp 272-4]. In order to capture the influence of interstate highway location on county population change 1960-70, four dummy variables are included in the analysis. These variables are:

IS = 1 for all counties in which an interstate highway was completed by 1968, 0 otherwise;
ISI = 1 for all counties containing an intersection of two or more interstate highways by 1968, 0 otherwise;
ISA = 1 for all counties adjacent to IS counties, 0 otherwise;
ISIA = 1 for all counties adjacent to ISI counties, 0 otherwise.

The expected sign for all four variables is positive. It is also expected that the IS and ISI variables will have a much stronger influence on population change than the two adjacent specifications. In some instances, both the IS and ISA variables and the ISI and ISIA variables are combined. Values for all of these variables were obtained by inspection of the 1969 Rand McNally Road Atlas which included the status of the interstate system at the end of 1968[5] [17]. Our assumption is that completions after that date would have little impact on the population changes shown in the 1970 Census.

Three additional variables are included in the analysis. These variables cover the degree of urbanization of a county (URBAN), the county's Standard Metropolitan Statistical Area designation (SMSA), and whether or not it is adjacent to such a county (ASMSA). The SMSA designation is used to control for metropolitan-nonmetropolitan differences among counties. Previous analysis of 1960 Census data revealed significant differences between these two groups in county population change for the period 1950-60 [25]. Except for counties composed entirely of large central cities, metropolitan county growth was considerably greater than that of nonmetropolitan counties. The SMSA dummy variable is included to pick up a possible continuation of this dichotomous growth pattern. A positive sign is expected.

Counties adjacent to SMSA counties might be expected to be affected by their proximity to these more heavily urbanized areas. Furthermore, the boundaries of SMSA areas have not been redefined since 1966. Hence the ASMSA designation is used to bring out county population change due to a spillover from SMSA areas. A positive sign is expected here also.

The degree to which a county was urbanized in 1960 was determined from 1960 Census data [23]. The percent of urban population in a county (URBAN) would be expected to affect county population growth in three ways: (1) Within metropolitan areas, the process of decentralization (i.e., urban sprawl) leads to the older, more urbanized areas having slower rates of population growth and perhaps population decline. (2) In rural areas, relatively high values for URBAN indicate the existence of local service centers, i.e., small towns. Due largely to the decline of traditional rural-farm markets, the majority of these small towns have been experiencing population decline since 1945. (3) Some threshold level or urban size is probably necessary for a place to begin to achieve self-sustaining growth [21, pp 15-60]. The first and second influences should have a negative effect, strong enough to offset the positive influence of urban places that have achieved a growth threshold. Thus we expect a negative sign for URBAN.

Regression Results

Regressions were run for each of the nine major census regions.[6] For the purpose of this analysis, ΔP_t in Eq. (1) is defined as the relative change in county population 1970/1960 (POP76), ΔP_{t-1} is defined as the relative change in county population 1960/1950 (POP65), and ΔP_{t-2} is defined as the relative change in county population 1950/1940 (POP54).[7] All other variables included in the regression equations are as defined above. Results are reported in Table 1. For each census region, only the best overall estimate is shown. In all cases the dependent variable is POP76.

With the exception of the Pacific equation, all coefficients reported in Table 1 are statistically significant at the .10 level or better. In the Pacific census region, several variables with rather large estimated coefficients have been retained even though they failed to pass the usual significance tests.

In the discussion to follow, the continuous variables included in the regression will be considered first. The significance of the various dummy variables on county population change 1960-70 will then be considered.

In all census regions except Mountain and Pacific, POP65 is a significant variable. This variable is strongest in the Middle Atlantic region where its coefficient is .595. By way of contrast, the coefficient of POP65 is .271 in the West South Central region and .167 in the West North Central region. Apparently, the influence of population change in the immediately preceding period on 1960-70 county population change is considerably less in the western portion of the country than in the East. The variable, POP54, appears in six of the nine equations. In five of these regions, the influence of POP54 is less than POP65 (East North Central, South Atlantic, East South Central, West South Central, and Mountain). In fact, in the West South Central and Mountain regions, the sign of POP54 is negative. The implication of these results is that fairly remote population changes have a negligible or negative influence on current developments in most of the country.

In New England, however, the relationship between the coefficients of POP65 and POP54 is reversed. The coefficient of POP65 is .348 while POP54 is .503. In New England, it seems clear that the influence of factors which resulted in high county population growth rates between 1940-50 are still being felt. Moreover, indications are that the influence of these 1940-50 growth factors exceeds that of more recent factors.

In Table 1, URBAN appears in five equations. It is significant in four equations. In the Pacific equation, URBAN has a "t" value of 1.44. In accord with our a priori expectation, the sign of URBAN is consistently negative, i.e., a relatively high percent of county population defined as urban in 1960 retards county population growth 1960-70. Note that this variable has a significant influence only in those regions which contain a substantial number of established large cities. It is interesting to note further that these are also the regions in which past population changes are important variables.

Table 2 has been constructed to facilitate the discussion of the dummy variables. To properly interpret the dummy variables, they must be considered in relation to the constant term. For example, in the Middle Atlantic region the constant is .456 and the coefficient of ISI is .048. In otherwords, for interstate intersection counties the regression plane shifts up .048. If recent population change (i.e., POP65, POP54) equaled zero, POP76 would equal .456 in all Middle Atlantic counties according to the estimate presented in Table 1. Under the same condition, POP76 would equal .504 in interstate intersection counties. The latter number is arrived at by adding the estimated coefficient of ISI to the constant. It is recorded in the appropriate cell in Table 2, Part A. In Part B of Table 2, the percentage effect of each dummy variable on county population growth is entered. For example, interstate intersection counties in the Middle Atlantic region grew (1960-70) 10.5% faster than all counties in the region.

The immediate impression created by Table 2 is that the interstate and SMSA dummy variables have quite different effects on county population change in different regions of the country. As Table 2, Part B shows, IS counties grew 6.4%, 5.1% and 4.4% faster than all counties in the South Atlantic, West South Central and West North Central regions respectively. In the

Mountain region, the comparable figure was 50.9%. The percentage growth differential for ISA was 4.9% in West South Central while that of IS + ISA was 34.4% in New England. The percentages were 10.5, 25.6, and -13.0 for ISI counties in the Middle Atlantic, West South Central and West North Central regions.

The variable ISIA has an effect in the largest number of regions (4). For counties adjacent to interstate intersections, POP76 would equal .656, .589, .995, and .602 as opposed to .534, .518, .824, and .522 for all counties (POP65 = 0, POP54 = 0) in the South Atlantic, East South Central, West South Central and West North Central regions respectively. The percentage figures were 22.8, 13.7, 28.8, and 15.3.

The SMSA counties had significantly different growth rates in three regions. In the South Atlantic region, they grew 7.7% faster than all counties. In the West North Central, the SMSA percentage differential was 26.1. The ASMSA counties had higher growth rates in the Middle Atlantic, West North Central and Mountain regions. The percentages were 11.0, 12.1 and 68.5.

The interpretation of the dummy variables included in the Pacific estimate is somewhat different than in the rest of the results. In the Pacific equation, the sole continuous variable is URBAN. If URBAN = 0, then county population change 1960-70 would be 1.942. The percentage effects were 121.9 and 65.9 for ISA and SMSA, respectively.

Finally, it should be noted that the level of R^2 reported in Table 1 indicates important variables have been omitted from the analysis in all regions. The highest R^2 obtained is .662 for the Middle Atlantic region. R^2 is .634 in New England, .603 in the East North Central region and .557 in the South Atlantic region. In the East South Central, West South Central and West North Central, R^2 equals .422, .191, and .394 respectively. The model exhibits especially poor performance in the Mountain and Pacific regions. The standard error of the estimate and R^2 obtained here require that the regression results for these regions be interpreted with care.[8]

Some Conclusions

The principal objective of this paper has been to investigate the impact of the interstate highway system on county population change. From the results presented in Tables 1 and 2, it is clear that this impact has been fairly substantial. Variables reflecting the influence of interstates are significant in all regions of the country except the East North Central.

The interstate variables have their strongest impact in the South Atlantic, East South Central, West South Central and West North Central regions. The influence of all dummy variables including SMSA and ASMSA, is most pervasive in South Atlantic, West South Central and West North Central. These are also the regions where POP65 and POP54 have relatively small coefficients. In New England, Middle Atlantic, and East North Central, the situation is reversed and past population change dominates. Although some large dummy coefficients are reported for the Mountain and Pacific regions, these results are considered somewhat suspect. As noted above, R^2 in these two regions is quite low.

The conclusion that the dummy variables have a greater impact in the South and Plains sections of the country is not especially surprising. These regions have been slower to industrialize than the North East and Middle West. Perhaps of greater importance, past investment in highways has lagged in the South and Plains. A past deficiency in highway development should serve to magnify the current influence of the interstate system on county population change.

Population Concentration

An additional aspect of this study is a detailed consideration of the determinants of intra-regional population concentration. To facilitate this analysis, a modified Lorenz Curve technique has been utilized. This technique permits us to calculate coefficients of population concentration. Preliminary results reveal considerably increased concentration between 1940 and 1970 in the South Atlantic, East South Central, West South Central, West North Central and Mountain regions. In New England, East North Central and the Pacific regions there was very little change. Population dispersal became apparent in the Middle Atlantic region during the 1960-70 period. Although our attempts to fully explain these phenomena have just begun, it is interesting to note that concentration is occurring in the rapidly industrializing states and stabilizing in the more mature states (where suburbanization may be having an important effect). Furthermore, many of the regions where increased concentration is evident have received more than their proportional share of interstate highway mileage in use by the end of 1968.

Footnotes

[1]Work supported by the Department of Housing and Urban Development and the U.S. Atomic Energy Commission under Interagency Agreement No. IAA-H-35-70 AEC 40-192-69, and conducted at Oak Ridge National Laboratory, Oak Ridge, Tennessee operated by Union Carbide Corporation for the U.S. Atomic Energy Commission. Opinions expressed in this report are solely those of the authors and do not necessarily represent the views of ORNL, AEC or HUD.

[2]University of Tennessee, Knoxville

[3]Tennessee Valley Authority, Knoxville

[4]Thus, the autoregressive linear regression model is applicable, since regressors and errors are contemporaneously uncorrelated. Ordinary least squares estimates of Eq. (1) will exhibit desirable asymptotic properties.

[5]The authors' judgement was used in some cases to exclude a county from a particular classification because of terrain or distance; for example, Grand, Park and Teller counties in Colorado were excluded from an ISA classification because of terrain. A list of all exceptions is available upon request.

[6]The regions are New England (NE), Middle Atlantic (MA), East North Central (ENC), South Atlantic (SA), East South Central (ESC), West South Central (WSC), West North Central (WNC), Mountain (Mt.), and Pacific (Pac.).

[7]A table showing average percentage county population change by Census Region 1940-50, 1950-60, 1960-70 is included as an Appendix.

[8]A partial explanation for the poor performance of the model in the Mountain and Pacific regions may lie in the extremely large population changes that have occurred there since 1940 (see Appendix).

References

[1] Brian J. L. Berry, Geography of Market Centers and Retail Distribution, Prentice-Hall, Inc., Englewood Cliffs, New Jersey (1966).

[2] Bureau of Business and Social Research, A "Before and After" Study of Effects of a Limited Access Highway Upon the Business Activity of By-Passed Communities and Upon Land Value and Land Use, The University of Denver, Denver, Colorado (1958).

[3] Bureau of Public Roads, Highways and Economic and Social Change, United States Government Printing Office, Washington, D. C. (1964).

[4] R. E. Carrier and W. R. Schriver, Plant Location Analysis: An Investigation of Plant Locations in Tennessee, Memphis State University, Memphis, Tennessee (1969).

[5] P. H. Cautner, "The Role of Railroads in U. S. Economic Growth," Journal of Economic History, 23 (December, 1963).

[6] Colorado Department of Highways, Colorado Land Economic Study, Denver, Colorado (1963).

[7] First Session, 87th Congress, Final Report of the Highway Cost Allocation Study, House Document No. 72., 1961.

[8] R. W. Fogel, "A Quantitative Approach to the Study of American Economic Growth," Journal of Economic History, 22 (June, 1962).

[9] A. F. Friedlaender, The Interstate Highway System: A Study in Public Investment, North Holland Publishing Co., Amsterdam, Holland (1965).

[10] V. R. Fuchs, Changes in the Location of Manufacturing in the United States Since 1929, Yale University Press, New Haven, Connecticut (1962).

[11] A. S. Goldberger, Econometric Methods, John Wiley and Sons, New York (1964).

[12] Highway Research Board, Indirect Effects of Highway Improvement, Bulletin 327, National Academy of Science, National Research Council, Washington, D. C. (1962).

[13] E. M. Hoover, The Location of Economic Activity, McGraw Hill Book Co., New York (1948).

[14] W. Isard, Location and Space Economy, The Massachusetts Institute of Technology Press, Cambridge, Massachusetts (1956).

[15] Charles T. Moore, et al., The Expected Economic and Social Impact of Interstate Highways in the Industrial and Commercial Trading Area of Birmingham, Alabama, University of Alabama, Birmingham (1965).

[16] Richard F. Muth, "The Spatial Structure of the Housing Market," Papers and Proceedings of the Regional Science Association, 7 (1961) pp 207-20.

[17] Rand McNally Road Atlas, 45th Annual Edition, Chicago (1969).

[18] J. Siegal, Intrametropolitan Migration of White and Minority Group Households, reprinted for Manpower Administration, U. S. Department of Labor, PB 195410 (May, 1970).

[19] D. M. Smith, "A Theoretical Framework for Geographic Studies of Industrial Location," reprinted in: R. D. Dean, W. H. Leahy and D. L. McKee (eds.), Spatial Economic Theory, The Free Press, New York (1970) pp 73-99.

[20] The University of Connecticut, The Connecticut Turnpike: Ribbon of Hope, Storrs, Connecticut (1965).

[21] Wilbur R. Thompson, A Preface to Urban Economics, The John Hopkins Press, Baltimore, Md. (1965).

[22] U. S. Department of Commerce, Bureau of the Census, County and City Data Book, 1962, Washington, D. C. (1962).

[23] U. S. Department of Commerce, Bureau of the Census, County and City Data Book, 1967, Washington, D. C. (1967).

[24] R. L. Williams and R. K. Koshal, Highways and Economic Development, Vol 1, Ohio University, Athens, Ohio (1970).

[25] Hal H. Winsborough, "City Growth and City Structure," Journal of Regional Science (1962) pp 35-49.

Table 1. Regression Results By Region

Variable Name	Region								
	NE	MA	ENC	SA	ESC	WSC	WNC	Mt.	Pac.
Constant	.224	.456	.390	.534	.518	.824	.522	5.668	1.942
POP65	.348[a] (.124)	.595[a] (.038)	.370[a] (.034)	.411[b] (.023)	.519[a] (.035)	.271[a] (.042)	.167[a] (.025)	-	-
POP54	.503[a] (.164)	-	.306[a] (.048)	.102[b] (.040)	-	-.096[b] (.037)	.275[a] (.038)	-5.126[c] (2.698)	-
IS	-	-	-	.034[b] (.016)	-	.042[c] (.022)	.023[c] (.014)	2.887[c] (1.690)	-
ISA	-	-	-	-	-	.040[b] (.020)	-	-	2.367[b] (1.044)
IS+ISA	.077[b] (.031)	-	-	-	-	-	-	-	-
ISI	-	.048[b] (.022)	-	-	-	.211[a] (.057)	-.068[c] (.035)	-	-
ISIA	-	-	-	.122[a] (.027)	.071[a] (.017)	.131[a] (.034)	.080[a] (.027)	-	-
SMSA	-	-	-	.041[c] (.016)	-	-	.136[a] (.028)	-	1.279 (1.129)
ASMSA	-	.045[b] (.018)	-	-	-	-	.063[a] (.016)	3.880[c] (2.080)	-
URBAN	-.081[b] (.037)	-.119[a] (.033)	-.108[a] (.018)	-.117[a] (.034)	-	-	-	-	-2.547 (1.762)
R^2	.634	.662	.603	.557	.422	.191	.397	.037	.056
F	26.8[a]	69.0[a]	218.9[a]	113.9[a]	131.8[a]	18.2[a]	57.3[a]	3.48[b]	2.57[c]
SE	.080	.096	.080	.151	.105	.178	.118	12.55	4.25
SAMPLE SIZE	67	146	436	551	364	470	618	278	133

[a] Significantly different from zero at the .01 level.
[b] Significantly different from zero at the .05 level.
[c] Significantly different from zero at the .10 level.

Table 2. Impact Of Dummy Variables By Region

Variable Name	Region								
	NE	MA	ENC	SA	ESC	WSC	WNC	Mt.	Pac.
Constant	.224	.456	.390	.534	.518	.824	.522	5.668	1.942
A. Total Effect (coefficient + constant)									
IS	-	-	-	.568	-	.866	.545	8.555	-
ISA	-	-	-	-	-	.864	-	-	4.309
IS+ISA	.301	-	-	-	-	-	-	-	-
ISI	-	.504	-	-	-	1.035	.454	-	-
ISIA	-	-	-	.656	.589	.995	.602	-	-
SMSA	-	-	-	.575	-	-	.658	-	3.221
ASMSA	-	.501	-	-	-	-	.585	9.548	-
B. Percent Effect (coefficient ÷ constant)									
IS	-	-	-	6.4	-	5.1	4.4	50.9	-
ISA	-	-	-	-	-	4.9	-	-	121.9
IS+ISA	34.4	-	-	-	-	-	-	-	-
ISI	-	10.5	-	-	-	25.6	-13.0	-	-
ISIA	-	-	-	22.8	13.7	28.8	15.3	-	-
SMSA	-	-	-	7.7	-	-	26.1	-	65.9
ASMSA	-	11.0	-	-	-	-	12.1	68.5	-

Source: Table 1.

APPENDIX

Average Percentage County Population Change by Census Region 1940-50, 1950-60, 1960-70

Decade	Region									
	US	NE	MA	ENC	SA	ESC	WSC	WNC	Mt.	Pac.
1960-70	13.8	12.7	10.6	8.6	9.2	2.7	3.5	-2.6	81.9	54.5
1950-60	6.4	10.4	14.8	10.9	11.3	-3.5	0.9	-0.5	13.2	24.7
1940-50	4.7	9.4	9.5	7.0	8.3	-1.4	.010	-3.1	5.6	38.4

SOURCES: United States Department of Commerce, Bureau of the Census, 1970 Census of Population, Advance Report (Washington, D.C.: United States Government Printing Office, 1970); United States Department of Commerce, Bureau of the Census, Statistical Abstract of the United States 1970, (Washington, D.C.: United States Government Printing Office, 1970), Table 13.

DISCUSSION

Larry H. Long, U. S. Bureau of the Census

The three papers in this session all have an "area" focus in that the authors are primarily concerned with how migration affects areas and not how migration affects individuals. All three papers, therefore, have a strong "ecological" orientation. Both perspectives--the ecological and the individual--are needed in order to understand migration processes.

The first paper, by Bowles and Oh, concerns educational levels and migration to and from the South and is based on data from the 1967 Survey of Economic Opportunity, a nationwide sample of about 35,000 households. Their paper contains numerous good ideas and a great deal of enthusiasm, but it contains no data and hence no concrete findings. I'm sure they will produce the data in good form, and I hope that they will publish their cross-tabulated data and not just the summary measures (various contingency statistics) which they say they will produce. Publication of cross-tabulated data serves the dual purpose of letting people see what one has done and letting other researchers use for their own purposes the raw data one has produced.

But the most important omission from the paper is a statement of hypotheses. The authors state that "It is hoped that this technique (utilization of various contingency statistics) will provide insights into the relationships between education and migration beyond those which can be gained from the cross-tabulated data themselves." The question is, "How?" The authors can make a significant contribution by identifying what is not known about how education affects migration, what they expect to find from their study, and why they expect some patterns and not others.

The next paper, by Kathryn Nelson, is a study of migration to and from Atlanta between 1962 and 1967 and is based upon the 1% Continuous Work History Sample maintained by the Social Security Administration. The advantages of such longitudinal data are obvious and have been so for almost as long as there has been a social security system in this country. From this source, one obtains nearly continuous data on county of employment, amount of income subject to social security withholding, industry of employment, age, sex, and race of persons paying into the social security system each year. The limitation is that one does not know any more than that. One does not know county or city of residence, occupation of the person, his education, or his family status; and, of course, one does not know anything about people not covered by social security, including federal workers and many state and local government employees, some self-employed persons, and very low-wage workers mostly in agriculture and private household employment. Also troublesome are the numerous persons who enter and leave the system, leaving incomplete records.

I think that the most interesting conclusions reached by Nelson are: 1) movers are more likely to change industry than stayers, and 2) those who changed industry also had greater increases in incomes than those who did not change industry. With respect to the first finding, it would be very interesting to ask what is the strength of the relationship when type of move is controlled for. With the data, I believe that it would be possible to control for at least three types of moves: nonmetropolitan to metropolitan, metropolitan to nonmetropolitan, and inter-metropolitan. Is the relationship between probability of moving and changing industry equally strong for all three types of moves?

With respect to the second point above, it would be interesting to relate frequency of moving to income changes. At least one study based on tax records reported greater income increases for persons who changed occupation or city of residence [1]. At some point, however, frequency of moving probably comes to interfere with orderly income advancement, and it would be useful to utilize the longitudinal records to investigate the potentially negative effects of excessively frequent moving.

The third paper, by Bohm and Patterson, attempts to assess the effect of interstate highways on 1960-70 population change in counties. Their method is to use rate of 1960-70 population change in counties as the dependent variable in multiple regression equations for each of nine regions of the county. Each regression equation consists of ten independent variables, three of which are continuous (1940-50 rate of population change, 1950-60 rate of population change, and percent of the county that was urban in 1960), with the other independent variables being scored in dummy fashion to represent whether or not a county had an interstate highway completed as of 1968, whether or not a county had an intersection of interstate highways, whether or not a county was adjacent to either of the two preceding types, whether or not the county was in an SMSA, and whether or not the county was adjacent to an SMSA.

The authors show that interstate highways have their strongest effects in the South Atlantic, the West South Central, the West North Central, and Mountain regions (divisions) of the country. The authors say that the effect was greatest in these areas because they previously lagged behind the rest of the country in the development of highways. Had the regressions been run for the country as a whole and not for specific regions, the overall effect of interstate highways would probably have been muted. Their results show that the effect of interstate highways is significant and varies from region to region.

What is surprising to me is that when other things are held constant, percent urban has either a non-significant effect or actually has a negative effect on 1960-70 rate of population change; i.e., the higher the percent urban, the lower the rate of 1960-70 population change, other things being equal. What may be confounding the result is the behavior of counties that are 100 percent urban, many of which lost population between 1960 and 1970. It is quite possible that counties begin to behave quite differently as they approach 100 percent urban, thereby violating the assumption of linearity underlying the regression analysis.

In addition to investigating the effect of interstate highways, the authors also seem to be trying to maximize the amount of explained variance in the rate of 1960-70 population change, somewhat in the fashion of Tarver and Gurley [2] and Kariel [3]. If this is their purpose, they can, as they are aware, explain more variation simply by increasing the number of independent variables. With more than 3,000 counties in the United States, they are not likely to run out of degrees of freedom. But is 1960-70 population really what they want to explain? Shouldn't net migration be more sensitive to economic changes than total population change? Of course, when the study was being done, the authors did not have data on net migration, but they might consider using rate of net migration as the dependent variable now that the data are available. Otherwise, they might try using rate of change in population 20-29 years old rather than change in total population, in view of the high degree of age-selectivity of migration. The authors might even consider using data on commuting as their dependent variable. The 1970 Census asked each person working the address of his place of employment, and the authors might consider investigating how interstate highways affect not only place of residence but place of employment.

Finally, Bohm and Patterson might want to give greater attention to explaining why their model does a much better overall job of accounting for population change in the North and East than in the South and West. Their values of R^2 vary from .037 in the Mountain region to .662 in the Middle Atlantic region.

REFERENCES

[1] Schroeder, Larry and David, Martin, "Long- versus Short-Term Impacts of Mobility on Income and Earnings", in Proceedings of the Social Statistics Section (1970), American Statistical Association, pp. 354-360.

[2] Tarver, James D. and Gurley, William R., "The Relationship of Selected Variables with County Net Migration Rates in the United States, 1950 to 1960", Rural Sociology, 30 (March 1965), 3-13.

[3] Kariel, Herbert G., "Selected Factors Areally Associated with Population Growth due to Net Migration", Annals of the Association of American Geographers, 53 (June 1963), 210-223.

AFDC PAYMENT LEVELS AND NONWHITE MIGRATION TO CITIES

Gordon F. DeJong and William L. Donnelly
The Pennsylvania State University

The problem is public welfare, and to many governmental officials public welfare is a problem. This is particularly true of the Aid to Families with Dependent Children (AFDC) program which is at the heart of proposals for welfare reform. The problems attributed to AFDC are many, but three themes are repetitive: (1) Benefits are inadequate and hardly begin to care for the needs of the dependent; (2) Family break-ups are encouraged by rules which frequently make families with males present ineligible for assistance; and (3) Migration, usually to urban areas, is encouraged by regional differences in levels of payment. To quote President Nixon's message on welfare reform (1969; 2): "By breaking up homes, the present welfare system has added to social unrest and robbed millions of children the joys of childhood; by widely varying payments among regions, it has helped to draw millions into the slums of our cities." The interest here is in the migration thesis as it applies to nonwhites in the United States. Do differential AFDC payment levels predict, at least in part, nonwhite migration to U.S. cities?

AFDC (formerly ADC - Aid to Dependent Children) is a federal participation program through grant-in-aid to states. As a national average the federal government contributes nearly sixty percent, state governments about one-third, and local governments the remainder. The basic purpose of the program is to enable needy children who are deprived of parental support or care to have the economic support and services they need for health, education, and family-based development. The AFDC program was a product of the 1930's when its typical recipient was pictured as a West Virginia mother whose husband had died in a mine accident. Honest, hard-working, rural, God-fearing, white Protestant folk. Gradually the typical recipient has become an urban Negro or a member of some other minority group (Moynihan, 1967; 11). As a result of the changing clientele, many policy makers feel that the nonwhite population is the key sub-population for a test of whether or not different AFDC payment levels affect migration.

The rapid increase in the AFDC program is notable. In 1936, the first year of its operation, there were about one-half million recipients. By 1960 this figure increased to more than three million and to nearly seven and a half million children and adults by 1970. The number of children aided per 1,000 under the age of 18 years also increased from 20 in 1940, to 35 in 1960, and 85 in 1970. Not only have the number of participants increased but also payment levels. Using 1957-59 purchasing power as a base, average AFDC payment per month per recipient increased from $20.05 in 1940, to $27.25 in 1960, to $39.00 in 1970 (U.S. Department of Health, Education, and Welfare, 1966; and 1970). However, national averages in payment levels mask state variations. For example, in 1967 the average amount paid per recipient in Louisiana was $24.00, in Georgia $25.35, and in North Carolina $25.40, compared with $43.50 in Illinois, $44.85 in California, and $59.70 in New York (Bureau of Social Science Research, 1968; 27).

Frame of Reference

The unit of analysis in this study is population aggregates - Standard Metropolitan Statistical Area (SMSA) counties of the contiguous United States. The more usual framework for welfare research is the analysis of case materials with individual records as the basis for generalizations. The perspective here is that area variations in AFDC payment levels can be viewed, along with other social aggregate and demographic characteristics, as indicators of community structure which influence human behavior. These data do not, however, provide a direct test of the motivations of welfare clients.

The frame of reference is applicable to an analysis of social and economic conditions that stimulate or retard migration to or from an area. Viewed behaviorally, an index of migration (in this study the net migration rate of nonwhites aged 25-29) is influenced by actual or perceived differences in the social and economic conditions and services of areas. As Bogue notes (1959; 501), there have been relatively few opportunities to study how net migration rates of given age and color groups of the population are related to social and economic conditions of ecological areas. This type of research provides an ecological complement to studies of the differential migration in that migration forces are sensitive to the social and demographic characteristics of the migrant.

Since most SMSA counties have experienced net migration gains in nonwhites during recent decades the primary link to migration research concerns the forces which "pull" nonwhite migrants to cities. Largely ignored by the definition of the problem are areas of origin "push" factors in nonwhite migration (Myrdal, 1944; Bogue and Hagood, 1953; Bowles, 1956; Ginzberg, 1956; Hamilton, 1959 and 1964; Cowhig, 1964; Taeuber and Taeuber, 1964; Stinner and DeJong, 1969). While the "push-pull" distinction may be somewhat artificial, it is a useful one when assessing the attracting forces of areas largely on the receiving end of net migration flows. Particularly in view of the finding that the pull of better conditions in the city is quite influential as a capturing mechanism, while the push of poor conditions at home is less effective in encouraging departure (Lowry, 1966; Lansing and Mueller, 1967; and Morrison, 1970).

Development of the Model

The basic goal of the study is to test the level of AFDC payment per family in a model which includes "pull" factors which are most applicable to nonwhite migration to metropolitan United States counties. The dependent variable is the 1950-60 county net migration rate for nonwhites aged 25-29 developed by Bowles and Tarver (1965). The choice of this age group is on the basis of our regression analyses of other age categories and the findings of case study materials which indicate that the 25-29 age group includes adults family members most involved with the AFDC program. An alternate would be the 20-24 age group; however, this cohort was only 10-14 years of age at the beginning of the decade, and it is likely that they were relatively nonmigratory as an independent family unit, during a better part of the period. Although the research is based on evidence from the 1950-60 period the continuation of heavy nonwhite migration to metropolitan areas since 1960 as well as upward trends in AFDC payment levels and numbers of clients argue for the importance of available evidence in testing the migration thesis.

AFDC payment levels are measured here by the average county payment per family in 1960. Unpublished statistical data of county AFDC programs was kindly provided by the Welfare Administration, U.S. Department of Health, Education, and Welfare.

From the literature it is clear that a most important attracting force in migration is superior opportunities for employment which reflect differences in economic conditions. It is true for nonwhites as well as whites (Tilly, 1968; 141), and has been confirmed in different research models. When asked why they move, Lansing and Mueller (1967) found that nearly three-fourths mentioned economic or occupational reasons. Some migrants already had a job while other migrants, and nonwhites are more likely to be in this group, were looking for work. Low levels of employment opportunity or low income levels in an area did not stimulate out-migration, but high levels of employment opportunity attracted in-migrants (Lansing and Mueller, 1967; 89-123). Using SMSA data Lowry (1967) came to a similar conclusion that in-migration is a function of the characteristics and conditions of the area's labor market. And Negroes were found to respond more sharply than whites to changes and regional variations in economic opportunity during the period from 1870 to 1950 (Eldridge and Thomas, 1964).

Closely related to the attraction of superior opportunities for employment in one's preferred occupation is the opportunity to earn a larger income. Migrants tend to be attracted to areas with populations of higher socio-economic status (Rogers, 1969; Blevins, 1969). In addition to the broader aspects of social status, the potential for more satisfactory family income itself may be a "pull" factor, particularly for nonwhite migrants who frequently must work at low-paying jobs. A complement to level of family income is the relative gap between nonwhite and white income levels. Following the general rationale of the relative deprivation thesis, one would expect a higher migration rate for nonwhites to areas where family income levels for whites and nonwhites were similar.

Numerous indicators of employment opportunities and socio-economic structure of an area's population have been used by researchers. Some of the more frequent include change in civilian nonagricultural employment; unemployment rate; change in civilian labor force employment in various occupational groups such as white collar, professional, laborer; family income and wage structure, and levels of educational attainment (Anderson, 1956; ter Heide, 1963; Blanco, 1964; Lowry, 1966; Tarver and Beale, 1968; Blevins, 1969; Rogers, 1969; Stinner and DeJong, 1969; Zuches, 1970; Greenwood and Gormely, 1971). Through empirical tests with the above indicators (Donnelly, 1970), three variables were selected as most sensitive to employment opportunities and socio-economic structure in relation to AFDC and nonwhite migration to cities. First, the indicator of employment opportunities is the percent change in employed persons, 1950-60, adjusted to exclude estimated employment change attributable to the net in-migration of nonwhites 25-29 years of age.[1] Second, median income in 1959 for all families is used as an indicator of income levels and the socio-economic structure of the area's population. Median family income is highly interrelated with occupation and education status indicators. The third variable is nonwhite median family income as a percent of median family income of all families. Data for all measures were derived from 1950 and 1960 U.S. Bureau of the Census publications.

From Bogue's (1969; 754) summary of "pull" factors in migration, a final variable which seems particularly applicable to the migration of nonwhites in the 25-29 age category is the lure of new or different activities, environments, and people in the city. Such activities - cultural, recreational, and intellectual - are captured in the life of the larger metropolitan areas more than in the life of smaller cities or rural areas. Undoubtedly included is the lure of perceived freedoms which also tend to be identified with larger cities. Total county population in 1960 is used as an indicator of this factor.

The sample, described in Table 1, is composed of all 185 SMSA counties for which age-color specific 1950-60 net migration rates are available. Further specification of net migration rates by sex added no new findings to the analysis. Bowles and Tarver (1965) calculated net migration data by age and color only for counties which had at least 5,000 nonwhites in 1960. A multiple regression statistical analysis is employed with a two-fold format. First the model is tested for all 185 counties.

Then the sample is divided into three sub-samples: southern SMSA counties, larger northern and western SMSA counties, and smaller northern and western SMSA counties (Table 1). The purpose of testing the model for these divisions of the sample is to assess the significance of disaggregating the relationships by regional and metropolitan character.

Table 2 presents the means and standard deviations for each sample and sub-sample and Tables 3 and 4 give zero-order correlations. As expected the net in-migration of nonwhites aged 25-29 to southern metropolitan counties was very low - average rate of 12.26 - while the average net migration rate for larger and smaller northern and western SMSA counties was much higher, 94.59 and 101.33, respectively (Table 2). Average AFDC payment level per family also varied considerably by region with southern counties averaging nearly $80 as compared with $146 for larger and $140 for smaller northern and western metropolitan counties.

Findings

Looking at the results for all SMSA counties in Table 5, the nonwhite total income ratio, family income level, and AFDC payment per family appear as the most significant factors in the model. All three factors either directly or indirectly tap differentials in income potential and perhaps indicate the importance (for this age cohort of young nonwhites) of moving to areas most indicative of the affluent society. Population size, as an indicator of the range of experiences and services available in the city, was of less significance than income and AFDC indicators in predicting nonwhite migration, and changes in employment opportunities was not a significant "pull" factor for this migratory age group.

The total model accounted for 39 percent of the variance explained in net migration rates for all SMSA counties (Table 5). However, the predictive value of the model is sharply differentiated by region. The model has the highest predictive value for nonwhite migration to southern SMSA counties with 44.9 percent of the variance explained (Table 6). This compared with 33.8 percent of the variance explained for migration to larger northern and western cities (Table 7) and 12.5 percent for smaller northern and western cities (Table 8).

City size is the strongest attracting force in the migration of young nonwhites in the South (Table 6), while a second highly significant factor is a more equal ratio between nonwhite and total income levels. Change in employed persons is also a significant factor, perhaps because of proximity to and information about actual developments in employment opportunities in southern cities. AFDC payment level is not a "pull" factor, and this is not attributable to a lack of variation in payment levels between southern metropolitan counties (Table 2).

The crux of the migration thesis is usually considered to be the northern and western cities, and for the larger SMSA counties AFDC payment level is a significant component of the model (Table 7). None of the other variables, except perhaps median family income level, even approaches statistical significance. However, for the smaller northern and western cities AFDC is not a "pull" factor and, as before, only median family income level approaches statistical significance. In summary, then, AFDC is a significant factor in nonwhite net in-migration to larger but not smaller northern and western cities, and not important in nonwhite migration to southern cities where the chance to earn larger incomes, more equal to those of whites, and the size of the SMSA are more significant attracting factors.

Discussion

There may be several ways to view the findings, albeit we choose to generalize at the aggregate level, that is the urban community, rather than at the individual level. However, findings from this perspective are seen as relevant to the more social psychological formulations involving individual motivations and migratory behavior, in that predictive results contrary to the findings of aggregate models would be open to question.

In terms of migration theory, the model has minimal predictive power, particularly for nonwhite net migration to northern and western SMSA counties. Perhaps "push" factors at areas of origin are more important in the urbanward migration of young nonwhites than the literature for all migrants would seem to indicate (Stinner and DeJong, 1969). Much of the past research on "pull" factors has not explicitly considered attracting factors for racial groups, particularly in regard to the consistently reported influence of employment opportunities.

Not tapped in this model but inexorably intertwined with employment opportunities and higher income potentials for nonwhite migrants are variables which represent distance between origin and destination and the existence of friends and relatives in the region of origin and/or destination (Barth, 1970; 188-189). Friends and relatives in a metropolitan area of destination often provide information concerning perceived and/or potential job opportunities, and this helps determine why migrants choose one destination rather than another (Blumberg and Bell, 1959; Rubin, 1960; MacDonald and MacDonald, 1964). Among groups subject to discrimination, the support of friends and family may be quite important (Lurie and Rayack, 1966). Even though opportunities for employment may be relatively near the area of origin, economic and social costs of migration may be less if a migrant goes to a distant place where initial accommodations, job information, and primary group social relationships are available through friends and relatives. Important as distance and friends and relatives may be in explaining nonwhite migration to cities, they cannot be adequately operationalized in a model based on net instead of stream migration data where counties rather than individuals or families are the unit of analysis.

Turning to the findings concerning AFDC payment level as a "pull" factor in nonwhite migration, the hypothesis receives some support, although the picture is complex and inconsistent. First, AFDC is related to net in-migration to larger but not smaller northern and western SMSA counties, yet payment structures are basically the same for all metropolitan counties within a given state. In other words, despite within-state similarities in payment levels, AFDC appears to "pull" migrants to the New Yorks, the Philadelphias, the Detroits, and the San Franciscos but not to the Albanies, the Harrisburgs, the Flints, and the Fresnos. Second, AFDC is not a "pull" factor in the migration of young nonwhites to southern SMSA counties. Third, level of AFDC payment and median family income are interrelated which may suggest that income level has both a direct and indirect affect on nonwhite migration, with the indirect affect being through higher AFDC payment.

Accepting the migration hypothesis for larger northern and western cities, the frame of reference here would suggest that AFDC is a community resource for income support which migrants and nonmigrants can fall back on if necessary. But to interpret this resource, greater in some cities than others, as a primary cause of net in-migration seems questionable at best. Not all nonwhite migrants are the "welfare poor." In fact, perhaps only a small number are. Characteristics of migrants studies (Hamilton, 1964; Suval and Hamilton, 1965) indicate that nonwhite migrants to the north and west have a higher level of years of school completed than the nonwhite population in the areas of destination. If education is related to employment, migrants may not be disproportionally unemployed. And since the measure here is net migration, it must be recognized that some migrants who "didn't make it" return to their area of origin rather than go on welfare. Finally, the 25-29 age cohort was 15-19 years of age at the beginning of the decade, and it seems reasonable that many migrated before they entered the life-cycle stage that necessitated AFDC assistance.

Lending substantiation to the argument that AFDC is not a primary cause of migration is a New York City study of welfare clients which found that a maximum of 14 percent of AFDC cases were people who had migrated to New York in a 23-month period prior to the study. "In spite of the coping problems facing the newly arrived migrant family, they appeared to be less likely to use public assistance than long-term migrant families" (U.S. Department of Health, Education, and Welfare and the New York State Department of Social Services, 1969; 39). Supporting evidence also comes from another study of welfare families in New York City (Podell, 1967) which reported that 75 percent of the mothers responding to the survey had either been raised in New York City or were ten-year residents and had not migrated solely for the purpose of becoming AFDC clients.

Perhaps it can be concluded that once the decision is made to migrate to a northern or western SMSA, the area of destination for young nonwhites is more likely to be a large city in a state that provides a higher AFDC benefit than one that provides a lower benefit. The point to be made is that nonwhite migrants tend to move to areas which appear to offer them greater civil and economic opportunities. It is, however, these same areas (especially large cities in the northern and western states) which also provide the most liberal AFDC programs and payments. From this conclusion we find agreement with Steiner (1970; 13) that welfare reform which standardizes AFDC benefits throughout the nation, admirable as it may be from a humanitarian viewpoint, is unlikely to make Mississippi as attractive as New York to Blacks in search of greater freedom and opportunity.

FOOTNOTE

1. To eliminate the contribution of nonwhite in-migrants aged 25-29 from the 1950-1960 percent change in employed persons in the civilian labor force, the following procedure was applied to data for SMSA counties with a 1950-60 net in-migration of nonwhites aged 25-29.

$$\frac{\Delta E - M}{E_{1950}}$$

where:

ΔE = change in the number of employed persons in the civilian labor force, 1950-1960.

M = number of nonwhite net in-migrants aged 25-29, 1950-1960.

E_{1950} = number of employed persons in the civilian labor force, 1950.

REFERENCES

1. Anderson, Theodore R.
1956 "Intermetropolitan Migration: A Correlation Analysis. American Journal of Sociology 61:459-462.
2. Barth, Michael C.
1970 "Migration and Income Maintenance." In The President's Commission on Income Maintenance Programs, Technical Studies, 187-206.
3. Blanco, Cicely
1964 "Prospective Unemployment and Interstate Population Movements." Review of Economics and Statistics 46:221-222.
4. Blevins, Audie L.
1969 "Migration Rates in Twelve Southern Metropolitan Areas: A 'Push-Pull'

Analysis." Social Science Quarterly 50:337-353.
5. Blumberg, Leonard and Robert Bell
1959 "Urban Migration and Kinship Ties." Social Problems 6:328-333.
6. Bogue, Donald J.
1959 "Internal Migration." In Philip M. Hauser and Otis Dudley Duncan (eds.), The Study of Population: An Inventory and Appraisal. Chicago: The University of Chicago Press.
7. Bogue, Donald J.
1969 Principles of Demography. New York: Wiley and Sons.
8. Bogue, Donald J. and Margaret M. Hagood
1953 Subregional Migration in the United States, 1936-1940, Vol. II: Differential Migration in the Corn and Cotton Belts. Miami, Ohio: Scripps Foundation Studies in Population Distribution, No. 6.
9. Bowles, Gladys K.
1956 Net Migration from the Rural-Farm Population, 1940-1950. Agricultural Marketing Service. Statistical Bulletin No. 176.
10. Bowles, Gladys K. and James D. Tarver
1965 Net Migration of the Population, 1950-60, by Age, Sex, and Color. Volume I, Parts 1-6. In cooperation with Oklahoma State University Research Foundation and United States Area Redevelopment Administration. Washington, United States Government Printing Office.
11. Bureau of Social Science Research
1968 A National Study of the AID to Families with Dependent Children Program. Report No. II for the Social and Rehabilitation Service, Department of Health, Education and Welfare.
12. Cowhig, James D., and Calvin L. Beale
1964 "Socio-economic Differences Between White and Nonwhite Farm Populations of the South." Social Forces 42:354-362.
13. Donnelly, William L.
1970 Demographic Correlates of the Aid to Families with Dependent Children Program. Unpublished Master's Thesis, The Pennsylvania State University.
14. Eldridge, Hope J. and Dorothy Swaine Thomas
1964 Demographic Analyses and Interrelations. Philadelphia: American Philosophical Society, Vol. III of Population Redistribution and Economic Growth, United States, 1870-1950.
15. Greenwood, Michael J. and Patrick J. Gormely
1971 "A Comparison of the Determinants of White and Nonwhite Interstate Migration." Demography 8:141-155.
16. Ginzberg, Eli
1956 The Negro Potential. New York: Columbia University Press.
17. Hamilton, C. Horace
1959 "Educational Selectivity of Net Migration from the South." Social Forces 38:33-42.
1964 "The Negro Leaves the South." Demography 1:273-295.
18. Lansing, John B. and Eva Mueller
1967 The Geographic Mobility of Labor. Ann Arbor: Survey Research Center Institute for Social Research.
19. Lowry, Ira S.
1966 Migration and Metropolitan Growth: Two Analytical Models. San Francisco: Chandler Publishing Co.
20. Lurie, Melvin and Elton Rayack
1966 "Racial Differences in Migration and Job Search: A Case Study." Southern Economic Journal 33:81-95.
21. MacDonald, John S. and Leatrice MacDonald
1964 "Chain Migration, Ethnic Neighborhood Formation, and Social Networks." Milbank Memorial Fund Quarterly 42:82-97.
22. Morrison, Peter A.
1970 Urban Growth, New Cities, and "The Population Problem." The Rand Corporation, P-4515-1.
23. Moynihan, Daniel P.
1967 "The Crisis in Welfare: A View from New York. Working paper for a conference on public welfare. Harriman, New York.
24. Myrdal, Gunnar
1944 An American Dilema. New York: Harper and Brothers.
25. Nixon, Richard
1969 "Proposals for Welfare Reform." U.S. House of Representatives, 91st Congress, 1st Session, Document No. 91-146.
26. Podell, Lawrence
1967 "Mothers' Nativity and Immigration." Preliminary Report No. 4, Families on Welfare in New York City. Center for Social Research, Graduate Center, City University of New York.
27. Rogers, Tommy W.
1969 "Migration Attractiveness of Southern Metropolitan Areas." Social Science Quarterly 50:325-336.
28. Rubin, Morton
1960 "Migration Patterns of Negroes from a Rural Northeastern Mississippi Community." Social Forces 39:59-66.
29. Steiner, Gilbert Y.
1970 "Reform of Welfare Reform." (Mimeograph)
30. Stinner, William F. and Gordon F. DeJong
1969 "Southern Negro Migration: Social and Economic Components of an Ecological Model." Demography 6: 455-471.
31. Suval, Elizabeth M. and C. Horace Hamilton
1965 "Some New Evidence on Educational Selectivity in Migration to and From the South." Social Forces 43:536-547.
32. Taeuber, Karl E. and Alma F. Taeuber
1965 "The Changing Character of Negro Migration." American Journal of

Sociology 70:429-441.
33. Tarver, James D. and Calvin Beale
1968 "Population Trends of Southern Nonmetropolitan Towns, 1950-1960." Rural Sociology 33:19-29.
34. ter Heide, H.
1963 "Migration Models and Their Significance for Population Forecasts." Milbank Memorial Fund Quarterly 41:56-76.
35. Tilly, Charles
1968 "Race and Migration to the American City." In James Q. Wilson (ed.), The Metropolitan Enigma. Cambridge, Massachusetts, Harvard University Press.
36. United States Department of Health, Education and Welfare
1966 Graphic Presentation of Public Assistance and Related Data: Trend Report. Social and Rehabilitation Service, Assistance Payments Administration.
37. United States Department of Health, Education and Welfare and the New York State Department of Social Services
1969 "Aid to Families with Dependent Children in New York City." Committee on Ways and Means, U.S. House of Representatives, 91st Congress, 1st Session, Document No. 34-557.
38. Zuiches, James J.
1970 "In-migration and Growth of Nonmetropolitan Urban Places." Rural Sociology 35:410-420.

Table 1. Description of Total Sample and Sub-Samples

Area	Sample Size
Counties composing the 211 SMSA's of the contiguous United States, 1960	346
SMSA counties for which age-sex-color specific 1950-60 net migration rates are available;* > 5,000 nonwhites, 1960	185
Southern SMSA counties**	85
Northern and western SMSA counties	100
Larger SMSA counties with populations of 500,000 or more and their suburban counties	64
Smaller SMSA counties with populations of less than 500,000	36

*Source: Bowles and Tarver, Net Migration of Population, 1950-60 by Age, Sex, and Color.

**South as defined by the U.S. Bureau of the Census.

Table 2. Means and Standard Deviation of Each Variable

Variables*	All SMSA Counties		Southern SMSA Counties		Northern and Western SMSA Counties: LARGER CITIES		Northern and Western SMSA Counties: SMALLER CITIES	
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
X_1	29.97	35.26	36.30	41.35	26.70	29.87	20.84	25.00
X_2	$5,956	$1,015	$5,332	$989	$6,572	$742	$6,337	$541
X_3	0.62	0.11	0.55	0.09	0.67	0.11	0.68	0.06
X_4	500,393	691,267	279,940	255,425	823,054	785,321	446,772	96,475
X_5	$114.46	$40.62	$79.93	$25.64	$146.09	$26.68	$139.79	$20.76
X_6	58.08	86.29	12.26	49.24	94.59	84.34	101.33	105.17

* X_1 = Percent change in employed persons, 1950-60.
X_2 = Median family income, 1959.
X_3 = Nonwhite/total income ratio, 1959.
X_4 = Total population, 1960.
X_5 = AFDC payment per family, 1960.
X_6 = Net migration rate, 1950-60, for nonwhites aged 25-29.

Table 3. Matrix of Zero-order Correlations; All SMSA Counties Above the Diagonal and Southern SMSA Counties Below the Diagonal

Variables	X_1	X_2	X_3	X_4	X_5	X_6
X_1 Percent change in employed persons, 1950-60	--	0.156	-0.030	0.006	-0.008	0.029
X_2 Median family income, 1959	0.388	--	0.380	0.306	0.688	0.512
X_3 Nonwhite/total income ratio, 1959	0.136	-0.021	--	0.241	0.515	0.483
X_4 Total population, 1960	0.042	0.290	0.043	--	0.340	0.332
X_5 AFDC payment per family, 1960	0.218	0.575	0.023	0.078	--	0.563
X_6 Net migration rate; 1950-60, for nonwhites aged 25-29	0.325	0.337	0.375	0.528	0.171	--

Table 4. Matrix of Zero-order Correlations: Larger Northern and Western SMSA Counties Above the Diagonal and Smaller Northern and Western SMSA Counties Below the Diagonal

Variables	X_1	X_2	X_3	X_4	X_5	X_6
X_1 Percent change in employed persons, 1950-60	--	0.231	0.055	-0.150	0.174	0.073
X_2 Median family income, 1959	-0.077	--	0.189	0.126	0.372	0.379
X_3 Nonwhite/total income ratio, 1959	-0.094	0.276	--	0.124	0.287	0.226
X_4 Total population, 1960	0.447	0.278	0.143	--	0.269	0.199
X_5 AFDC payment per family, 1960	0.177	0.270	-0.072	0.152	--	0.591
X_6 Net migration rate; 1950-60, for nonwhites aged 25-29	-0.049	0.461	0.278	0.204	0.175	--

Table 5. Measures of Relationship Between Model Components and the 1950-60 Net Migration Rate for Nonwhites Aged 25-29 for All SMSA Counties

Model Components	Standardized Regression Coefficient	Standard Error	Student t
X_1 Percent change in employed persons, 1950-60	0.005	0.059	0.088
X_2 Median family income, 1959	0.207	0.082	2.524*
X_3 Nonwhite/total income ratio, 1959	0.244	0.067	3.624**
X_4 Total population, 1960	0.124	0.062	2.008*
X_5 AFDC Payment per family, 1960	0.253	0.087	2.898**
Multiple Coefficient of Determination (R^2) = 0.391; F (5,179) = 24.83			

*p < .05
**p < .01

Table 6. Measures of Relationship Between Model Components and the 1950-60 Net Migration Rate for Nonwhites Aged 25-29 for Southern SMSA Counties

Model Components	Standardized Regression Coefficient	Standard Error	Student t
X_1 Percent change in employed persons, 1950-60	0.211	0.089	2.363*
X_2 Median family income, 1959	0.119	0.111	1.075
X_3 Nonwhite/total income ratio, 1959	0.329	0.082	3.993**
X_4 Total population, 1960	0.469	0.086	5.480**
X_5 AFDC payment per family, 1960	0.012	0.100	0.124
Multiple Coefficient of Determination (R^2) = 0.449; F (5,79) = 14.70			

*p < .05
**p < .01

Table 7. Measures of Relationship Between Model Components and the 1950-60 Net Migration Rate for Nonwhites Aged 25-29 for Larger Northern and Western SMSA Counties

Model Components	Standardized Regression Coefficient	Standard Error	Student t
X_1 Percent change in employed persons, 1950-60	-0.059	0.108	0.544
X_2 Median family income, 1959	0.187	0.113	1.658
X_3 Nonwhite/total income ratio, 1959	0.090	0.108	0.836
X_4 Total population, 1960	0.021	0.109	0.193
X_5 AFDC payment per family, 1960	0.500	0.118	4.242**
Multiple Coefficient of Determination (R^2) = 0.338; F (5,58) = 7.44			

**p < .01

Table 8. Measures of Relationship Between Model Components and the 1950-60 Net Migration Rate for Nonwhites Aged 25-29 for Smaller Northern and Western SMSA Counties

Model Components	Standardized Regression Coefficient	Standard Error	Student t
X_1 Percent change in employed persons, 1950-60	-0.065	0.186	0.347
X_2 Median family income, 1959	0.360	0.182	1.980
X_3 Nonwhite/total income ratio, 1959	0.165	0.168	0.984
X_4 Total population, 1960	0.096	0.190	0.507
X_5 AFDC payment per family, 1960	0.086	0.170	0.510
Multiple Coefficient of Determination (R^2) = 0.125; F (5,30) = 2.00			

RACIAL DIFFERENTIALS IN THE DETERMINANTS OF PROSPECTIVE MOBILITY

Jeanne C. Biggar, University of Virginia

Because traditional definitions of migration included only intercommunity spatial movement, generalizations on geographic mobility are applicable primarily to the white middle-class families predominant in these long-distance moves (Lansing and Mueller, 1968: 263). Furthermore, most residential mobility studies have centered on the factors associated with local moves for white families. This emphasis on whites has been due in part to the nature of sample areas selected in those studies. Peter Rossi's Potential Mobility Index was based on four white census tracts selected to represent upper and lower SES levels and high and low mobility areas in Philadelphia (Rossi, 1955). Wendell Bell's analysis of familism, consumerism, and career mobility drew from families in two Chicago suburbs (Bell, 1958). Leslie and Richardson's life-cycle determinants were based on a sample from a LaFayette white middle-class housing development (Leslie and Richardson, 1961). Despite the fact that the frequency of Negro intracommunity moves exceeds that of whites and that local moves account for 80 to 85 per cent of residential exchanges in the United States, little attention has been given to the relevance of existing theoretical mobility explanations for Negro moves.[1]

Implicit in most general discussions of geographic mobility is the idea that similar causal factors operate in the decision to move for Negroes as for whites but that racial mobility differentials are a consequence of the Negro's less advantaged socioeconomic position and the "hedonistic" values associated with inadequate short-run decision-making which Beshers imputes to this social strata (Beshers, 1967: 135). If this is true, it is expected that the pattern of mobility determinants for Negro families would be similar, if not identical, to that for whites after the effects of SES variables are controlled. Such is the speculation which directed the primary purpose of this paper to test whether a theoretical prospective mobility model synthesizing the findings of earlier mobility research will apply to moving propensity for urban Negro families. More specifically, the task outlined is to discern significant differences in the patterns of prospective mobility determinants among white and Negro families when SES levels are held constant.

The Prospective Mobility Model

Previous studies of residential mobility have shown the family's propensity to move to be a function of previous moving experience, duration of residence, housing tenure, relative dissatisfaction with dwelling unit and neighborhood, size of family - particularly in the early stages of the family life-cycle -- and the age and career mobility of the household head.[2] Drawing on the findings of these studies, the theoretical prospective mobility model included four sets of independent variables: retrospective mobility, i.e., characteristics of the last move; dissatisfaction with present dwelling unit and neighborhood; family status variables; and dwelling status variables (Figure 1). The dependent variable, prospective mobility, conceptualized as the propensity to move, is indicated by the prefer-

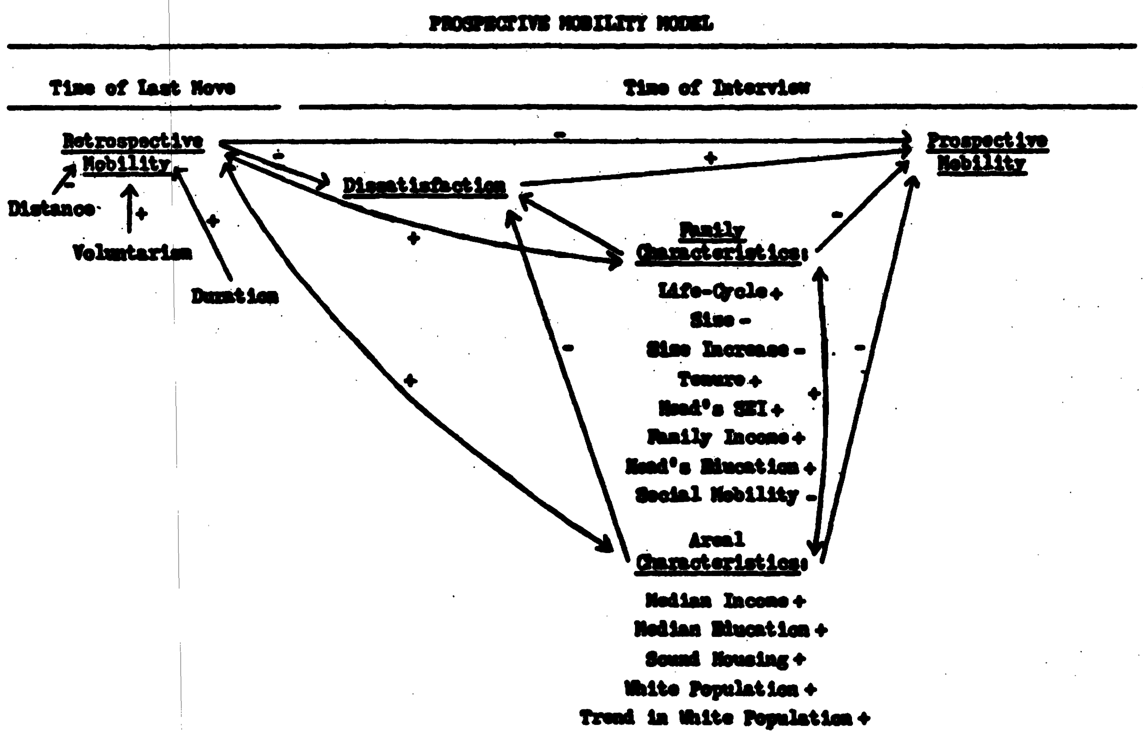


Figure 1. Hypothesized Relationships For Retrospective and Prospective Mobility With Dissatisfaction With Present Location, Family and Areal Characteristics.

ences and/or intentions to move in the near future. Consideration of exchange of dwelling unit, actual search for alternative locations, the desire if given the choice, and actual plans to move are all dimensions of this moving propensity.

The structure of relationships outlined in this model proposes first that prospective mobility depends upon retrospective mobility. Voluntarism, the degree of control the family had in the last decision to make the last move and in the selection of the present dwelling unit, will decrease the degree of moving propensity (Rossi, 1955). The distance over which the family traveled in the last move was expected to constrain the amount and accuracy of knowledge of destination opportunities, housing types, and neighborhood qualities and hence increase prospective mobility (Westefeld, 1947; Ladinsky, 1967). Duration of residence is expected to reduce moving propensity as in Land's axiom of cummulative inertia, "The probability of an individual continuing in a state -- residential area -- increases with increasing length of previous residence" (Land, 1969 : 133). Second, prospective mobility is expected to depend upon the degree of dissatisfaction with present location, the variable hypothesized to intervene in the relationships of retrospective mobility, family and dwelling status variables to prospective mobility. Third, characteristics of the family unit, such as stage in family life-cycle, size, income, education and occupation of the household head, tenure and social mobility commitment were expected to determine prospective mobility both indirectly through the degree of dissatisfaction, and directly or independent of the degree of dissatisfaction. Finally, the qualities of the urban environment were expected to influence the degree of moving propensity, again, indirectly through dissatisfaction, and directly despite the degree of dissatisfaction of the family unit.

Methodology

A national survey of metropolitan households (N=1476) in 1966 provided the data for testing the prospective mobility model.[3] Interviews centered on objective and attitudinal dimmensions of the present and previous dwelling units and neighborhoods, on consideration, choice and plans to move in the future, and characteristics of the members of the household unit. The responses of a subsample of recent urban movers (237 white and 117 Negro households in Central City Tracts who had exchanged dwelling units since 1960 and whose head was a full-time worker are analyzed in the following sections to provide the comparison of white and Negro prospective mobility determinants. Because the original question posed for research required an examination of net relationships, path analysis was employed to permit examination of the nature of relationships between any two variables while controlling simulaneously for the effects of all other variables in the model.

Racial Differentials in Prospective Mobility

The responses of recent movers into the urban segments of the metropolitan areas to the four questions indicating moving propensity showed significant racial differentials (Table 1).

Table 1. Per Cent Distribution of Urban Respondents on Mobility Criteria for White, Negro, and All.

Mobility Criteria	White	Negro	All
(N)	(237)	(117)	(354)
Have considered moving again:			
No	63.6	51.4	59.5
Yes	36.4	48.6	40.5
. . . and looked for another place:			
No	15.4	20.6	17.1
Yes	21.0	28.0	23.4
If had choice, would move again:			
No	53.2	35.0	47.2
Don't Know	2.1	2.6	2.3
Yes	44.7	62.4	60.6
Actually plan to move in next year:			
No	67.5	57.3	64.1
Don't Know	8.0	4.3	6.8
Yes	24.5	38.5	29.1

Almost half (48.6%) of the Negro respondents, as compared to a third (36.4%) of the whites had considered moving again. Furthermore, more Negroes had actually searched for alternative housing (Negro, 28.0%; white 21.0%). The restlessness of the urban Negro is apparent on the last two question as well. Almost two-thirds said they would move if they had their choice. Not quite half of the whites expressed this desire. As to actual plans to move -- the most reliable indicator of moving behavior -- thirty-eight per cent, almost double the national moving rate, said they expected to relocate in the year following interview (Van Arsdol, Sabagh, Butler, 1968). In contrast about one-fourth (24.5%) of the whites planned to move. The Negro-white differential in moving propensity over all criteria is clearly apparent in the comparison of distributions of respondents on the summary prospective mobility score (Table 2).

Table 2. Per Cent Distribution of Urban Respondents on Prospective Mobility Score for White, Negro, and All.

Prospective Mobility Score	White	Negro	All
(N)	(237)	(117)	(354)
0	42.2	30.8	38.3
1 - 12	28.2	30.8	38.3
13 - 24	18.9	23.0	20.5
25 - 36	10.5	19.7	13.5
(Mean)	(9.17)	(13.15)	(10.49)
(Standard Deviation)	(11.04)	(13.22)	(12.16)
(Median)	(3.40)	(6.62)	(3.73)

Negro and White Prospective Mobility Models

Submission of the data gathered from these metropolitan households to path analysis permitted the empirical test of the theoretical prospective mobility model deduced from previous studies of white residential mobility.[4] As expected, among urban white families, the pattern of relationships followed rather closely the hypothesized model (Figure 2). In terms of net relation ships, the propensity to move again was determined by the degree of dissatisfaction with the present dwelling unit and neighborhood. Dissatisfaction was highest in areas of lower family incomes with larger shares of non-white population for households in the early stages of family life-cycle who were committed to upward social mobility. Socioeconomic variables influenced this scheme only to the degree that they inhibited the acquisition of ownership of the dwelling unit.

The findings above are congruent with those of previous studies of white mobility. Moving propensity for white families was primarily due to the dissatisfaction with present dwelling unit and urban enviroment, and despite the relative level of satisfaction cr dissatisfaction,their desire to own a home of their own. Prospective mobility appears to be a function of upward career mobility for those families in the expansion stages when they are expecting to achieve new dwelling status to match anticipated higher social status.

The empirical prospective mobility model for urban Negro families shows several points of contrast to that for whites (Figure 3). First, only one of the three paths of direct influence to prospective mobility, dissatisfaction with present location stemmed from the family unit itself. The other paths were from areal characteristics such that prospective mobility was more probable from census tracts with higher education levels and from those which showed the least invasion of non-whites between 1950 and 1960.

Retrospective measures showed fewer but clearer indirect paths through dissatisfaction and family characteristics for Negro families than for white families. Distance influenced only through its correlation with voluntarism and so directly through greater dissatisfaction to greater prospective mobility. Voluntarism linked directly to dissatisfaction and on to prospective mobility. In addition, it was correlated with tenure and so co-varied with dissatisfaction and prospective mobility in the same direction. Duration had a direct path to dissatisfaction and an indirect one as well through size of family. For this subsample, the relationship of duration is a clear negation of the axiom of cummulative inertia. The longer a Negro family had lived at location, the more dissatisfied they tended to be, and hence the more inclined to move again.

The pattern of influence of family characteristics on dissatisfaction -- both direct and indirect -- was identical to that for urban white families. Attention should be called to the inconsistent roles of income, however. Note that one path through its correlation with tenure tended to decrease the level of dissatisfaction and so moving propensity as well, but its second path through areal educational level actually increased the probability of prospective mobility.

Unlike the white model which indicated prospective mobility is associated with anticipated social mobility, the Negro model suggests that anticipated residential mobility is associated with achieved social mobility. When the effects

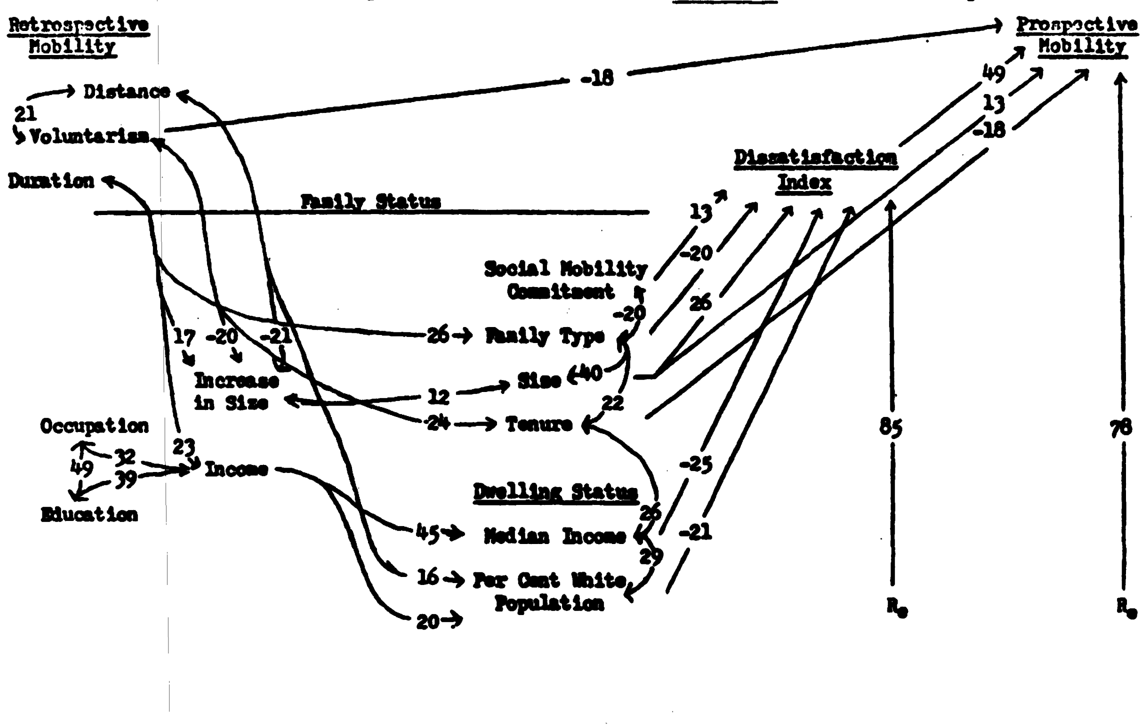


Figure 2. Empirical Model of Retrospective Mobility, Family and Dwelling Status, and Dissatisfaction Determinants of Prospective Mobility for Urban White Families

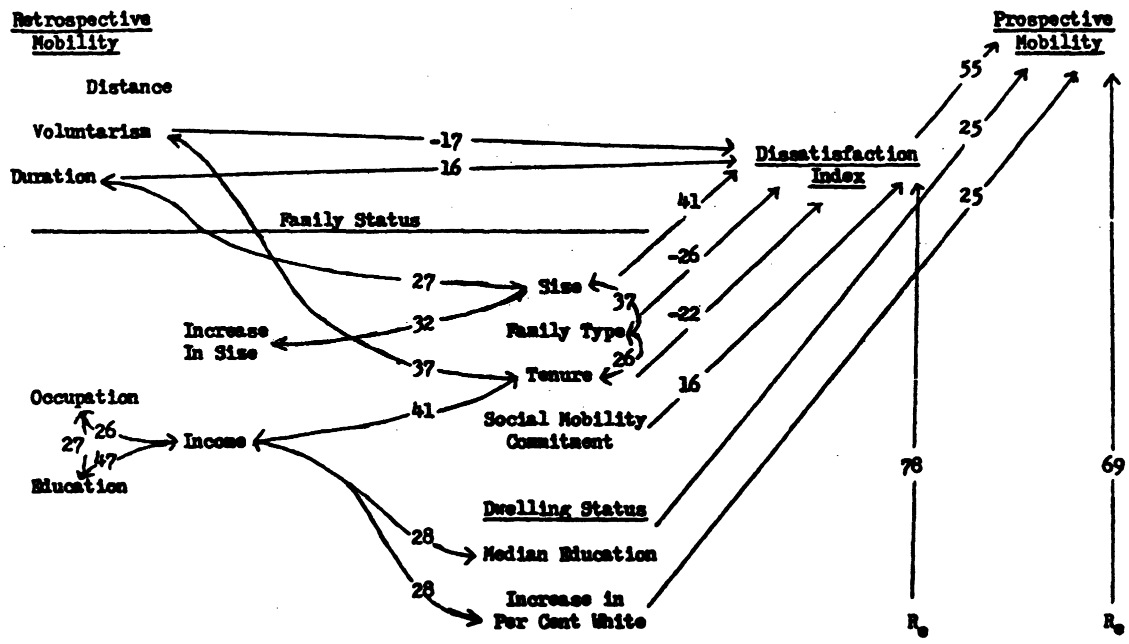


Figure 3. Empirical Model of Retrospective Mobility, Family and Dwelling Status, and Dissatisfaction Determinants of Prospective Mobility for Urban Negro Families

of retrospective mobility, dissatisfaction and family characteristics are held constant, it is the Negroes who have already achieved dwelling status in the more favorable urban enviroments in which Negroes reside that show the higher degree of propensity to move again. Therefore, it is inferred from this analysis that prospective mobility for Negroes is more often a consequence than an antecedent of upward social mobility. Further, it is speculated that the lag in match of dwelling status with other social status dimensions -- occupation , education, and income -- for the socially mobile Negro family is due to the constriction of urban housing opportunities from discriminatory practices such as restrictive housing covenants.

Conclusions

The hypothesis that racial differentials in prospective mobility determinants would be minimized when socioeconomic levels were controlled found little support in this analysis. Two points of contrast should be emphasized in comparing mobility determinants for white and Negro families living in the urban area. First, while like white families, dissatisfaction with present location is related to characteristics of the family and less voluntarism in the last move, unlike white families, dissatisfaction for Negro families stems from longer duration of residence but not directly from the attributes of the area in which they live. Second, for Negro families the influence of areal characteristics is directly to prospective mobility where it seems the more desirable the tract -- in terms of societal evaluations -- the greater the probability of subsequent moves. This is in contrast to white families where the pattern indicates the less desirable the tract, the greater the family's dissatisfaction and hence propensity to move again.

It is argued here that these differences were due to the constraints of discriminatory housing practices. Where families were free to move from area to area in order to afford a fit between the desired and their actual dwelling status, the absence of fit was expressed in dissatisfaction among families who reside in the less desirable census tracts. However, when available alternatives were restricted by color considerations, as in the case of Negro families here, after SES characteristics and the degree of dissatisfaction had been controlled, higher degrees of moving propensity could be expected in areas of greater relative rather than greater absolute deprivation. In this study Negro families with greater relative deprivation were those with higher socioeconomic levels who had invaded some of the relatively more desirable central city tracts.

This multivariate analysis of white and Negro moving propensity indicated that racial differentials in mobility patterns cannot be adequately explained solely in terms of the more advantageous position of white middle-class families to utilize occupational and housing opportunity structures available. Furthermore, the implications of the white-Negro contrasts are that decision-making for Negro families is less a matter of "inadequacy" and more a matter of paucity of opportunities to select freely among housing alternatives suitable to their family needs.

Until the constraints of discriminatory housing structures are relaxed, Negro families can be expected to continue their pattern of shelter opportunism.

FOOTNOTES

[1]U.S. Department of Commerce, Bureau of Census. 1966. Current Population Reports, Series P-20, Number 156, Washington, D.C.

[2]See Kenkel, 1965; Beshers, 1967; Westefeld, 1947; Ladinsky, 1967; Butler, Chapin, et.al., 1968; Land, 1969; Mangalam, 1968; Rossi, 1955; Burchinal and Bauder, 1965; Butler, Sabagh, and Van Arsdol, 1964; Leslie and Richardson, 1961.

[3]In 1965, the Center for Urban and Regional Studies, University of North Carolina, undertook a study of moving behavior and residential choice in metropolitan areas under the auspices of the National Cooperative Highway Research Program of the Highway Research Board. An extensive interview was conducted in 1966 by NORC on a standard multi-stage area probability sample to the "block or segment" level. The universe sampled was composed of all non-institutional metropolitan population of the U.S., 21 years or older. The Primary Sampling Units were 1960 Standard Metropolitan Areas; the Sampling Units, localities or Census Tracts, drawn at random from listings stratified for size and urban type.

[4]Operational measures for the concepts included in the model are as follows: Prospective Mobility Index: score summing responses on consideration, search, choice and plans to move; Dissatisfaction Index: score summing responses on dissatisfactions with number of rooms, number of bedrooms, size of rooms, inside appearance, outside appearance, age of the dwelling unit, and with the kind of people, cleanliness, condition of houses and apartments, and reputation of the neighborhood; Distance of last move; score based on seven "lengths" ranging from moves within same neighborhood to migrations from another state; Voluntarism: sum of scores on a twenty-five point scale based on reasons for last move ordered according to the degree of push from previous origin to degree of pull from destination; Duration of Residence: number of years of residence at present dwelling unit; Stage in Family Life-Cycle: ten categories combining seven family types with age of household head; Size of family: number living in the household; Increase in family size -- difference (if positive) between present number and number in household at time of last move; Tenure: four categories: own, buying, renting, other; Education and Income: U.S. Census classification for household head; Occupation: Duncan's SEI; Social Mobility Commitment: score summing three-point scale for last four of Westoff's items; Dwelling Status: five Characteristics of Census Tracts in sample as obtained from 1960 Census.

REFERENCES

Bell, Wendell. 1958. "Social Class, Life Styles, and Suburban Residence", in The Suburban Community, edited by William K. Dobriner, New York, G.P. Putnam's Sons.

Beshers, James M. 1967. Population Processes in Social Systems, New York, The Free Press.

Burchinal, Lee G., and Ward W. Bauder. 1965. "Adjustments to the New Institutional Enviroment", in Family Mobility In Our Dynamic Society, Iowa State University Center for Agricultural and Economic Development, Ames, Iowa.

Butler, Edgar W., and F. Stuart Chapin, Jr., George C. Hemmens, Edgar J. Kaiser, Michael A. Stegman, Shirley F. Weiss, 1968. Moving Behavior and Residential Choice: A National Survey, Review Draft. University of North Carolina, Mimeo.

Butler, Edgar W., George S. Sabagh, and Maurice D. Van Arsdol, Jr. 1964. "Demographic and Social Psychological Factors in Residential Mobility," Sociology and Social Research 48: 129-154.

Kenkel, William F. 1965. "The Family Moving Decision Process," in Family Mobility In Our Dynamic Society, Iowa State Center for Agricultural and Economic Development, Ames, Iowa: 180-193.

Ladinsky, Jack. 1967. "Sources of Geographic Mobility Among Professional Workers; American Sociological" Demography 4: 293-309.

Land, Kenneth C. 1969. "Duration of Residence and Prospective Migration: Further Evidence", Demography 6: 133-140 .

Lansing, John B., and Eva Mueller. 1967. The Geographic Mobility of Labor, Institute for Social Research, Ann Arbor, Michigan.

Leslie, Gerald R., and Arthur G. Richardson. 1961. "Life-Cycle, Career Pattern, and the Decision to Move", American Sociological Review: 26: 894-901

Mangalam, J.J. 1968. Human Migration, University of Kentucky Press, Lexington, Kentucky.

Rossi, Peter. 1955. Why Families Move, New York: The Free Press.

U.S. Department of Commerce, Bureau of the Census. 1966. Current Population Reports, Series P-20, Number 156, Washington, D.C.

Van Arsdol, Maurice D., Jr., Georges Sabagh, and Edward W. Butler. 1968 "Retrospective and Subsequent Metropolitan Residential Mobility," Demography 5: 249-266.

Westefeld, Albert. 1947. "The Distance Factor in Migration", Social Forces 19: 210-218.

Homeownership, Race, and Housing Inequality in Metropolitan U.S.

Ernest B. Attah, Brown University

A substantial amount of research has been conducted on the general subject of residence, housing inequality, and race in cities of the U.S. Examples in the sociological tradition include the series of studies undertaken for the Commission on Race and Housing(summarized in McEntire 1960), the literature on racial residential succession(e.g. Duncan and Duncan 1957, and Taeuber and Taeuber 1965), the more general "housing" literature (e.g. Duncan and Hauser 1960), and the extensive literature dealing generally with the overall social and economic condition of blacks as compared to the rest of the society. While this literature has documented extensively the general patterns of housing inequality between the races, and over different areas of the city, there has not been much focusing on specific factors, to search for the ways in which they may influence what emerges as the general pattern. It should be noted, so as not to misrepresent the literature, that some attempts have been made to bring several demographic and socioeconomic characteristics of population and some basic characteristics of housing to bear on analysis of overall patterns of housing inequality between races. However, the need still exists for more in-depth exploration of specific factors, and for attempts to explicate the role which these play in the complex of factors and processes involved in the ecological dynamics of the metropolitan community. To paraphrase for the present context a recent statement(Taeuber 1969, p.146) of continuing research needs in the general area of "race relations:" statistical documentation of white-black differences in total housing and changing patterns of residential distribution is ample; what is now needed is increasingly detailed analysis of identifiable central factors having far-reaching implications in this overall complex.

The factor chosen for investigation here is homeownership. The supporting argument for this choice runs as follows: in the aggregate, the quality of housing obtained by whites is better than the quality of housing obtained by blacks; in terms of occupancy of the total housing inventory, the housing obtained by blacks in metropolitan areas has for the most part come to them via a turnover process from previously white-occupied housing; this turnover process occurs gradually on a neighborhood basis; the blacks who have pioneered in this process of neighborhood change have been of generally high socioeconomic status - clearly higher, also, than the rest of the black population; homeownership has been found to be positively associated with socioeconomic status. As specific questions to explore, therefore, the following might arise: (a) what role does homeownership play in this process of neighborhood change? and (b) what identifiable role does homeownership thereby play in upgrading the quality of housing obtained by blacks? It is the purpose of this paper to explore these questions. Stated otherwise, the research intent here is to analyze the implications of patterns of homeownership for the trends in racial residential distribution, and the corresponding distribution of housing and neighborhood amenities.

The plan of analysis will be first to review the recent trends in homeownership and in some general factors that have some bearing on the homeownership trends and on the overall role of homeownership: trends in the representation of blacks in the local population, changes in family income, and changes in the overall housing inventory. Then, after a review of the general trends in the quality of housing obtained by blacks and whites, the analysis will be carried out on a neighborhood basis: given a classification of local areas according to stage in the turnover process from white to black occupancy, trends in levels of homeownership will be examined for the different stages, and then the analysis will focus on the areas undergoing the first stages to review the trends in family income and in housing quality. Finally, this area-wise analysis will be articulated with an aggregate-level examination of the implications of homeownership for the quality of housing obtained by blacks and whites.

The review of trends in homeownership will include data for the U.S. - total and nonfarm - and for the nonfarm areas of the four regions, to provide a general overview of the national and regional context. The local-area analysis will deal with four major cities, each in one of the four regions of the country: Atlanta, Ga., Boston, Mass., Cleveland, Ohio, and San Francisco, Cal. The data will refer to the city proper only.

Note on Data

The data for this study were drawn entirely from publications of the U.S. Bureau of the Census from the decennial censuses of population and housing. Thus their character and scope are circumscribed by the limitations inherent in census data. Especially problematic here is the issue of reliability and comparability in the measurement of quality of housing. The categorization of housing quality on the basis of condition and plumbing has been different for each of the three census years to be dealt with in the analysis here(1940, 1950, 1960), with the result that the level of comparability of data between censuses is indeterminate. Furthermore, the rating of each unit has rested solely with the census enumerator with the result, in this case, that the level of scorer reliability is largely indeterminate. The many problems of measurement involved in the assessment of quality of housing are discussed at length in a Census Bureau Working Paper(No. 25, 1967).

As a conventional measure widely used in the literature, the condition and plumbing measure will also be used in this paper subject, of course, to the limitations discussed. In an attempt to amelicrate(to the extent possible) some of these problems, another measure will also be used as an indicator of housing quality. If the argument is made that for such various reasons as normal wear, disrepair, over-use, etc., housing tends on the whole to deteriorate in quality with time, then the age of the structure may serve as at least a rough indication in the aggregate of quality.

Indeed, the tabulations of age of structure against the corresponding condition and plumbing measure in census reports from the three census years supports this conjecture.

Yet a third measure of housing quality will be used in this paper. Whereas the condition and plumbing measure and the age of structure refer to the physical characteristics of the housing unit itself, the aim this time is to find a measure which reflects the pattern of living implied by occupancy of the unit by the household involved. The extent of room crowding is one such measure, and an indicator is available for it in the form of persons per room in census reports. This measure also has limitations(e.g. it ignores possible differences in the sizes of the rooms), but it should reasonably serve as at least a rough indicator.

Where the data are available, all three measures will be used in the analysis below. The hope is that the different perspectives which they represent concerning the general issue of quality of housing should make for a more rounded view than any one of them could provide singly. The attendant risk of compounding errors must of course be kept in mind.

A final note on data concerns notational convention. The bulk of the census data relevant to the analysis below is tabulated by color ("white" vs. "nonwhite"), but the orientation of the discussion leans towards interpretations for races("white" vs. "black" or "Negro"). The problem of equivalence which arises is negligible for the purposes of analysis at the level of aggregation involved here in areas where the representation of "Other Nonwhite" in the total nonwhite population is small. Atlanta, Boston, and Cleveland fall in this category. In San Francisco, However, "Other Nonwhites" constitute a substantial proportion of the total nonwhite population. Thus whereas it is legitimate to use "black" and "nonwhite" interchangably for the other three cities, the equivalence breaks down for San Francisco. This distinction should be borne in mind in reading through the discussion below. Parenthetically, however, it should be interesting as a subsidiary research question to see whether or not the white-nonwhite differentials observed in San Francisco are similar enough to the white-nonwhite differentials observed in the other cities to warrant the conclusion that for the purposes of the kinds of issues dealt with in the analysis below, the "Other Nonwhite" population of San Francisco may as well be considered "black."

Review of Broad Trends: 1940-1960

Homeownership levels are measured as the percentage of occupied units which are owner-occupied. Table 1A shows trends in homeownership between 1940 and 1960 for the U.S.(total and nonfarm) and for the four regions(nonfarm only).

On the whole, rates of homeownership increased between 1940 and 1960. This was true for the country as a whole, for nonfarm areas of the country, and for the nonfarm areas of each region of the country, but the specific patterns differed markedly between the two decades spanned by the total period of study. The decade of the 1940's marked a sharp rise in the levels of homeownership, but these levels actually declined slightly between 1950 and 1960. For the U.S. total, for instance, the level increased by eleven percentage-points between 1940 and 1950, but decreased by six-tenths of a percentage-point in the 1950's. Perhaps the most pronounced rise-and-fall pattern was evidenced by the West with changes of 10.5 and 4.3 percentage-points respectively.

To compare the different areas, ownership of nonfarm housing, first, is understandably at somewhat lower levels than ownership of total housing. Over the regions, the pattern of ownership(nonfarm housing only) falls in the order of lowest levels in the Northeast, followed by the South and West, with levels in the North Central region being the highest. This pattern is consistent through the three census years reviewed, but the clustering of regions by levels changes over the period: in 1940 the Northeast and the South fall relatively close together, at some remove from the West and North Central which are close together. In 1950 the pattern is of the Northeast separated from the South which is in turn separated from the West and North Central which fall together. In 1960 the South and West are clustered together with the Northeast falling below at some remove, and the North Central above.

Table 1B shows the trends in homeownership, by color, between 1940 and 1960 for the four cities under study. There are parallels between the homeownership trends in these four cities and the larger nationwide and regional trends, but there are also some departures. Generally, the levels rose between 1940 and 1960, but there were clear differences between the two decades involved. Between 1940 and 1950 there were relatively substantial increases in levels in Atlanta and Cleveland, and somewhat smaller increases in Boston and San Francisco. The levels in Atlanta, Boston and Cleveland continued to rise in the 1950's, although at a much reduced pace than in the preceding period. Over this period, the increase in Atlanta was somewhat larger than the increase in the other two cities. But in San Francisco the level declined slightly between 1950 and 1960. Thus the ranking pattern wherein overall levels of homeownership were highest in Cleveland and lowest in Boston clearly held true for 1940 and 1950, but was modified in 1960 by the emergence of Atlanta with a slightly higher level than Cleveland.

The data by color reveal first that the trends in white ownership levels paralleled quite closely the trends for the total population: the levels increased over the total period in all four cities, but the increase over the 1940's was generally greater than the increase between 1950 and 1960, and in San Francisco the level actually fell slightly in the latter period. For the nonwhite population the pattern of substantial increase in levels between 1940 and 1950, followed by a reduced increase over the 1950's also occurred in Atlanta, Boston, and Cleveland, but the distinction of San Francisco this time is that the percentage-point increase between 1950 and 1960 actually exceeded the comparable figure for the first decade of study. Thus although in 1940 the level of nonwhite homeownership in San Francisco was distinctly below the comparable levels in the other cities, by 1960 it was clearly higher than the comparable level in Boston, and not much lower than the levels in Cleveland and Atlanta.

As a final note on homeownerhip trends by color, mention should be made of the fact that there have been marked and persistent differentials in levels between whites and nonwhites, with nonwhites owning their own homes much less than whites. In 1940 the differences were of the order of twenty to thirty percentage-points, and the highest nonwhite level(Atlanta) barely exceeded 10%. Given the different patterns of change in levels over the period 1940 to 1960, the differentials in 1960 were narrowed somewhat to the range of between ten and twenty-five percentage-points, but the highest nonwhite levels(Atlanta and Cleveland) were still under 30%.

Tables 2 and 3 show data relevant for a review of overall trends in some major population and housing characteristics. As the significant units competing on the market for housing, households are used here as the basis for indicating trends in the representation of nonwhites in the local population. Concerning the total number of households, it should be noted that the first impression given by the trends in Atlanta are misleading: the apparent large increase between 1950 and 1960 resulted from the annexation of territory during that period. The actual patterns of growth in numbers of households were quite similar among all four cities. In Atlanta, Boston and Cleveland, the overall increase was of the order of between 10% and 15% between 1940 and 1960, and the overwhelming bulk of this growth took place in the first of the two decades involved. San Francisco displayed a generally similar pattern, but the levels involved were higher: there was a 25% increase in total number of households between 1950 and 1960, and an overall increase of about 40% between 1940 and 1960.

The trends in representation of nonwhites in the local population seem to reflect in these four cities the population shifts which took place on a larger scale. Specifically, with the great migration of blacks from predominantly the rural South into the urban North, coupled with the continuing suburbanization of the white population, the proportion black of the population of central cities rose steadily between 1940 and 1960 in the North and West, but did not change much in the South, showing only a slight upturn in the 1950's (see Farley 1968). Thus the proportion nonwhite of households in Boston, Cleveland and San Francisco increased substantially between 1940 and 1960, and decreased by one percentage-point in Atlanta over the same period(N.B. the annexed territory in Atlanta had a smaller proportion nonwhite than the rest of the city). Data not reported here indicate that the representation of blacks among the nonwhite population of San Francisco also increased over the period: in 1940 the overall representations of Negro and Other Nonwhite households among all households were 0.7% and 3.1% respectively, and in 1960 the comparable representations among total population were 10.0% and 8.3% respectively.

The rather sketchy data available indicate that family income(in current dollars) generally rose for both whites and nonwhites between 1940 and 1960. In raw percentage terms, the increase in median family income seems to have been greater for nonwhites than for whites. However, given the extremely low levels(relatively and absolutely) from which the nonwhite figures started at the beginning of the period, very substantial differentials persisted through the end of the period between the white and nonwhite levels. These differentials seem to have been at their worst in Atlanta: whereas the median family income for whites fell within a relatively narrow range across cities, the nonwhite level in Atlanta was substantially lower than the corresponding levels in Cleveland and San Francisco. Comparable data by color are not available for Boston.

Trends in characteristics of housing are reported in Table 3. The measure of proportion of all units which are in one-unit structures attempts to indicate the general representation in the overall housing inventory of units suitable for owner-occupancy. The levels involved there are distinctly lower in Boston than in the other three cities. Part of the explanation for Boston's generally lower levels of homeownership may lie in this fact. Further, the increase in the levels involved was very slight for Boston and San Francisco between 1940 and 1960, but was quite substantial in Atlanta and Cleveland. Thus by 1960 somewhat over half of the housing units in Atlanta were in one-unit structures, and the corresponding levels for Cleveland, San Francisco and Boston followed at intervals of roughly fifteen, ten, and twenty percentage-points respectively.

Construction of new units took place at a fairly vigorous pace over the two decades of study in San Francisco(roughly 61,000 units authorized), somewhat less in Atlanta and Cleveland(47,000 and 39,000 units authorized, respectively), and relatively much less in Boston(21,000 units authorized). Vacancy rates in all four cities fell to below 2% in 1950 from levels of roughly between 3% and 7% in 1940, and then rose again to between 3% and 5% in 1960. These trends may perhaps partially reflect pressures on housing resulting from the trends in numbers of households in the cities.

The final item of review in this section will be the general trends in housing quality in the four cities of study between 1940 and 1960, on the basis of three indicators(see Table 4). Consider first the age of structure. There was not much change in the proportion of all units which were in structures ten years old or older in Boston over the period 1940-60. In Cleveland the corresponding proportion declined somewhat - more so between 1940 and 1950 than in the 1950's. In San Francisco it fell slightly over the first decade of study and rose again in the second, and in Atlanta it fell steadily over the entire period. Concerning racial differentials, it seems from the data available for 1960 that there really was not much difference on the whole in the age of housing occupied by whites and nonwhites.

Since the condition and plumbing measures are not comparable between census years, it is not possible to discuss the time-trend in quality of housing as indexed by that measure. Looking within census years, however, it becomes quite clear that in the aggregate the quality of housing occupied by nonwhites has been distinctly poorer than the quality of housing occupied by whites. This conclusion holds true for all four cities of study, and for all dates in the period of study for which data are available. And to compare cities, the gap between whites and nonwhites in occupancy of basically sound housing was largest in Atlanta(with more than a twenty-five percentage-point difference on both the 1940 and 1960 measures), and was progressively less in San Francisco and Cleveland. From the data available, it seems that the nonwhite population of Boston was relatively the least disadvantaged of the four.

The trends in room crowding paralleled closely the trends in occupancy of sound housing: in Atlanta, San Francisco and Cleveland the proportions of nonwhite-occupied units housing more than one person per room were clearly larger than the corresponding proportions of white-occupied units. And this was true for the two census years between 1940 and 1960 for which data are available. This differential between whites and nonwhites persisted even as overall levels of room crowding decreased for both groups over the period of study. Boston also emerges as an exception to the patterns evidenced by the other three cities: there was hardly any difference between the white and nonwhite levels of occupancy of units with more than one person per room in 1940. In fact, the differential that did exist(six-tenths of one percentage-point) was in the direction of whites being at a disadvantage. Unfortunately, no more color-specific data are available for Boston, and statements about the rest of the 1940-60 period are therefore precluded.

In summary, therefore, the review undertaken in this section has revealed the following: first, homeownership levels have generally increased over the period 1940-60. The specific patterns have differed between regions and between the four cities studied, but the general statement holds true for the total population in each area, as well as for the white and nonwhite segments of the population. Differentials were observed between whites and nonwhites in initial levels of homeownership and in amounts of increase over the period, and consequently in final levels. The figures for nonwhites in all cases were lower than the corresponding figures for whites, except for amounts of increase in homeownership level over the period for San Francisco.

These trends in homeownership took place in the face of continued growth in the numbers of households in the four cities, and also of both an increasing representation of nonwhite households in the local population(Boston, Cleveland, San Francisco), and of no particular change in representation of nonwhite households(Atlanta). At the same time, income levels generally rose, but the available data indicate that although the percentage increase between 1940 and 1960 in median family income was greater for nonwhites than for whites, the dollar-difference between white and nonwhite levels of median family income actually widened over the period.

Construction rates varied between the four cities over the two decades, the proportion of all units which were in one-unit structures increased appreciably for two of the four cities(Atlanta and Cleveland) but did not change much for the other two, and vacancy rates fell somewhat in the first of the two decades, then rose again(still without attaining high levels) in the second.

Finally, clear and persistent differentials in general quality of housing obtained have been observed between whites and nonwhites over the period of study for three of the four cities. The scanty data available for the fourth city(Boston) suggest that the differential has perhaps not been as marked there as in the other cities.

All the processes so far reviewed in broad terms(trends in homeownership levels, in general housing and population characteristics, and in quality of housing) necessarily had parallel manifestations on an areal basis: the housing inventory as well as the population resident in it was necessarily distributed in some way over the area of the city. Thus for the purposes of fuller explication of these trends it should be interesting to study their ecological parallels. In addition, there are aspects of the trends in the ecological pattern which are not shown by aggregate data of the sort so far reviewed, but which are important for shedding meaningful light on issues relevant to the central concerns of this paper. Therefore the analysis will now turn to examination of areal data pertaining to these issues for the four cities involved, over the period of study.

Spatiotemporal Patterns

The analysis in this section will be based on a grouping of census tracts according to stage in the process of racial residential succession. The overall thrust of the analysis will be to start by taking the process of neighborhood change as given, and then to study the changes that occur during this process in the character and color-tenure distribution of the housing inventory, and in the socioeconomic profile of the resident population. Specifically, the plan is to review changes in homeownership over all stages, and then to study closely the changes in nature and character of housing, and in family income, in areas undergoing the first stages of the process.

The scheme by which census tracts are classified is designed to distinguish between areas of established and unchanging white and nonwhite residence, stable interracial areas(where they exist), areas undergoing a general loss or gain of population of both races, and areas undergoing the classical racial change-over process(white to black, or black to white where such is the case). Details of the classification scheme are shown in

Figure 1. Scheme for Classification of Census Tracts According to Succession Stage.

Classification	Initial Year of Period: NW pop	Initial Year of Period: %NW	Terminal Year: NW pop	Terminal Year: %NW	Terminal Year: W pop
I. Early Integration	<250	<2 %	≥250	increased	decr or stable
II. Integration		2-49.9%	≥250 incr or stable	incr	decr or stable
III. Succession	≥250	50-89.9%	incr or stable	incr	decr or stable
IV. Stable Interracial	≥250		stable		stable
V. Growing			incr		incr
VI. Declining			decr		decr
VII. Re-segregation			decr or stable	<90% decr	incr or stable
VIII. Segregated black		≥90 %		≥90%	
IX. Segregated white	<250	<2 %	<250	<2%	
X. Penetration	<250		incr <250	incr	decr

Notes:

1. The population - white or nonwhite - of a census tract is defined as stable over the period if it changes by less than 100 persons and less than 10 per cent of its original level. Conversely, it is defined as increased or decreased if it changes in excess of any of these two criteria in the applicable direction.
2. On the basis of this definition, it is possible that some of the tracts designated "segregated white," or "segregated black," for example, may also be growing or declining. In a strict sense, therefore, the "growing" and "declining" tracts are mixed growing and declining.
3. The category labelled "re-segregation" may also include three different alternatives: "pure" re-segregation would involve the situation wherein the nonwhite population decreased over the period, and the white population increased. The two other possibilities are that the nonwhite population may remain stable while the white population increased, or the white population may remain stable while the nonwhite population decreased. The nonwhite percentage would decrease in all three cases.
4. In view of the many possible questions that may arise regarding the total set of logically possible combinations of patterns of change, the order in which the factors are checked for assignment of census tracts to their appropriate categories is: per cent nonwhite first, nonwhite population second, and white population third.
5. The development of this classification scheme clearly owes much to the work of Duncan and Duncan (1957) and of Taeuber and Taeuber (1965).

Figure 1. It should perhaps be mentioned explicitly that the specific cut-off points used in this scheme are essentially arbitrary. However, their choice has been informed not only by some generally similar precedents in the literature(Duncan and Duncan 1957, and Taeuber and Taeuber 1965), but particularly also by the constraints inherent in the nature of the available data: data by color were only published separately for census tracts having at least 250 nonwhites in 1940 and 1950, and having at least 400 nonwhites in 1960. Thus the "Penetration" category, for instance, holds particular interest as the very earliest identifiable stage in racial residential succession, but it unfortunately cannot be studied in depth owing to the unavailability of separate tabulations of data by color for such tracts.

Some further comments should be made on methodological issues. First of all, to minimize classification errors census tracts are excluded from the analysis if changes in their boundaries alone accounted for a change of fifty or more dwelling units and/or ten per cent of the total number of units in the tract. These cut-offs have been chosen to articulate with the cut-offs associated with whether or not the population of a census tract will be considered stable or changed(see Fig. 1, Note 1). Secondly, census tracts are excluded from the analysis if 10% or more of their population resides in group quarters. The basic intent here is to concentrate the analysis on the population actively in the general housing market, and to eliminate areas in which that part of the population not actively in the market might distort significantly the factors under study. Finally the two periods 1940-50 and 1950-60 will be treated separately for the analysis below since the differences between them in the processes reviewed above may have resulted in different forms of neighborhood change in any given area.

Tables 5A and 5B show the trends in homeownership over the periods 1940-50 and 1950-60 for the different groupings of census tracts by stage in racial residential succession. Comparisons among classes within census years reveal the following general patterns: the areas of generally highest homeownership levels(for both whites and nonwhites) were predominantly the areas of segregated white residence. At the least, the homeownership levels in those areas were consistently relatively high through both decades. But there were two outstanding exceptions to this general pattern: for whites in Boston at the end of the decade of the 1940's, and also for blacks in Atlanta at the beginning and at the end of the 1950's, homeownership levels were lowest in these segregated white areas than in any other grouping of census tracts. A further exception is that in San Francisco the highest homeownership levels for whites at the beginning and at the end of the 1950's, and for nonwhites at the end of that decade fell in areas that underwent re-segregation over the period. For nonwhites in San Francisco at the beginning of the period, the highest level of homeownership occurred in stable interracial areas.

A second general pattern that emerges from time-constant inter-class comparisons is that homeownership levels for both whites and nonwhites in segregated black areas were generally low in relative as well as absolute terms. In San Francisco, in fact, they were the lowest among all classes of census tracts, at all date-points of the two decades of study, and for both whites and nonwhites, except for the nonwhite population at the beginning of the 1950-60 period. An outstanding exception to this general pattern of homeownership levels being lowest in segregated black areas is in Cleveland where, between 1940 and 1950, homeownership levels were consistently lowest in re-segregating tracts than anywhere else for both whites and nonwhites, and for whites in 1940 the level in segregated black areas was actually higher than anywhere else.

The general patterns of change in homeownership levels may be summarized briefly as follows: between 1940 and 1950, first, the general increase in homeownership levels in these four cities(see discussion above) was closely paralleled by the trends in virtually all areas of the cities. Of the sixty-eight color-by-succession-stage categories in Table 5A, the level of homeownership failed to increase over the decade in only twelve. These departures from the overall trend occurred with the highest frequency(three and four respectively) in succession tracts and in growing areas. Otherwise they were evenly distributed over all the other classes of census tracts excluding early integration and integration tracts, stable interracial areas, and declining areas. The range in the overall magnitudes of the percentage-point changes in homeownership is such that the simple arithmetic average of these percentage-point changes, computed within classes of census tracts (disregarding city and color), would vary between a low level of 0.0% for re-segregation tracts to a high of 15.7% for integration tracts. The broad racial differentials over these four cities in the patterning of change in homeownership levels between 1940 and 1950 were such that the percentage-point increase in levels for nonwhites was generally higher than the increase for whites in all classes of census tracts except growing areas and re-segregation tracts.

Over the decade of the 1950's, the patterning of change in homeownership levels within groupings of census tracts by stage in racial residential succession over the decade differed substantially from the corresponding patterns for the decade of the 1940's. First of all, there were many more instances of declines in homeownership percentages: in all, twenty-nine of the sixty-eight color-by-succession-stage categories in Table 5B showed decreases in level of homeownership over the period, and in two more the ownership level at the end of the period remained the same as at the beginning. In Atlanta in fact, homeownership percentages declined over the decade in all classes of census tracts, for both whites and nonwhites, with only two exceptions: there were very slight increases for whites in segregated black areas, and for blacks in declining areas.(N.B. In comparing the city-wide trend implied by these changes with the data reported in Table 1B, it should be noted that Table 5B excludes the annexed territory which is included for Table 1B, and in which homeownership levels were higher than in the rest of the city).

In terms of relative frequency, these declines in homeownership were fairly evenly spread out over all the classes of census tracts, with the sole exception of the one set of stable interracial areas which exists in the table. However, there were substantial differences between classes of census tracts in the magnitudes of the changes involved, with the result that the simple arithmetic average of percentage-point changes in homeownership levels, within groupings of census tracts, (disregarding city and color), would range from -5.4% for succession tracts up to 14.3% for growing areas. These patterns of change in ownership percentages within classes of census tracts do not fall into any clear pattern of systematic differentials or regularities by race, except perhaps for the observation that in growing areas the direction of the change in levels(increase or decrease) was consistently the same in each city for both whites and nonwhites.

It is interesting to note, parenthetically, that among the various color-by-succession-stage categories of Tables 5A and 5B, there are scattered cases of quite dramatic changes in homeownership levels within a decade, some of which represent strikingly anomalous departures from predominating trends. For example, it was noted above that homeownership levels generally tend to be highest in segregated white areas than in other areas of the city. But for the white population of such areas in Boston between 1940 and 1950, the change in homeownership levels over the decade took the form of a sharp decline that resulted in the ownership level in that category being the lowest of all at the close of the decade. A closely comparable decrease in ownership levels also took place between 1940 and 1950 for the nonwhite population of growing areas in Cleveland - and this happened in a period when homeownership levels(particularly among blacks) were generally rising. There are also examples of outstandingly high increases in the level of homeownership in a particular class of census tracts as, for instance, for both the white and nonwhite populations of growing areas in Cleveland between 1950 and 1960. Indeed, it seems that growing areas have exhibited at least one instance of each of a range of possible patterns of change in homeownership levels. In addition to the rapid increase and drastic decrease already mentioned, there are cases where the level remained high (relative to other areas of the same city) over the entire period(San Francisco, white, 1940-50 and 1950-60; Atlanta, nonwhite and white, 1940-50, and nonwhite, 1950-60), and there are cases where the level remained low over the period(Boston, white, 1940-50, and nonwhite, 1950-60).

On the whole, however, a general conclusion which emerges concerning growing areas is that they quite clearly evidenced the largest overall percentage-point increases in homeownership levels between 1950 and 1960, for both whites and nonwhites. But between 1940 and 1950 the pattern was less distinctive. A large increase in levels for nonwhites in Atlanta, for instance, was countered by an equally large decrease for nonwhites in Cleveland. For whites, a large increase in Atlanta and a more modest increase in Cleveland were countered by slight decreases in Boston and San Francisco.

The intergroup patterning of change in homeownership levels among census tracts grouped according to stage of racial residential succession was such that the highest levels of increase for both whites and nonwhites in the 1940's occurred in the areas which underwent early integration and integration over the decade. But in the 1950's the general trends in such areas were not particularly distinct from the trends in other areas. For succession tracts, the general patterns were of modest increases in levels in the 1940's for both whites and nonwhites, but of a general increase for whites and a general decrease for nonwhites in the 1950's. Stable interracial areas, to the extent that they existed, evidenced modest but consistent increases in both periods. Declining areas also evidenced modest but consistent increases on the whole over both periods, and their 1940-50 levels of increase in fact ranked quite close to the highest levels for that period. Finally, except in Boston, the homeownership levels for nonwhites in penetration tracts generally increased over both decades of the period between 1940 and 1960. But for whites, although the same general statement can be made for the decade of the 1940's, the overall pattern in the 1950's was of a slight decrease in levels over the period.

Turning now to focus the analysis on the areas which were undergoing the early stages of racial residential succession, time-constant interclass comparisons, first of all, indicate that at the beginning of the 1940-50 period, homeownership levels over the four cities of study were generally highest(disregarding color) in penetration tracts. The levels in early integration and integration tracts were alternately higher and lower than each other an equal number of times. At the end of the period, the predominant pattern was of levels in early integration tracts being the highest of the three, followed by levels in integration tracts, and the levels in penetration tracts were the lowest. For the decade of the 1950's, the most prevalent pattern was for the levels of homeownership in these areas to fall in the same rank order at the beginning and at the end of the period: highest in early integration areas, next in penetration areas, and lowest among the three in integration areas.

If distinctions are made among these levels of homeownership on the basis of the color of the occupants, the observed pattern for the decade of the 1950's is not altered. The predominant ranking of the three types of areas at the beginning and at the end of the period, for both whites and nonwhites, is still: early integration-penetration-integration. For the decade of the 1940's, however, there are substantial differences. At the beginning of the period, the prevalent ranking of levels for whites is: penetration-integration-early integration; but for nonwhites, no clear pattern is noticable. At the end of the period, one clear observation emerges: the levels for whites were generally lowest in integration tracts, but the nonwhite levels were generally highest in those areas.

The patterning of change in homeownership levels over the periods of study also differed somewhat among these three classes of census tracts. Direct comparisons of the actual percentage-point differences from Tables 5A and 5B would yield only two specifically observable rankings: first, for nonwhites, the percentage-point increases over the 1940's were highest in early integration tracts, next highest in integration tracts, and lowest in penetration tracts; secondly, the same ranking of areas would be obtained for the percentage-point increases in white homeownership levels between 1950 and 1960. On the basis of these direct comparisons of the changes in homeownership levels, no clear or consistent patterns of change can be said to have occurred for whites in the decade of the 1940's, or for nonwhites between 1950 and 1960. However, as manifested through the simple averages of the percentage-point changes in levels, the increases for whites between 1940 and 1950 were highest in integration tracts, next highest in penetration tracts, and lowest in early integration tracts. And for nonwhites between 1950 and 1960, comparisons of the simple averages of the percentage-point changes in levels would yield the ranking: penetration-early integration-integration.

Some further data are shown in Tables 6A and 6B for the trends in selected characteristics of population and housing in areas undergoing the early stages of racial residential succession. For reasons noted above, these data are available for only the early integration and integration tracts.

In terms of the representation of one-unit structures, the housing inventory of these areas did not undergo very pronounced changes between 1940 and 1950: except for integration tracts in Atlanta, the changes involved were of the order of less than five percentage-points in any direction. It seems to be the case that the proportions of units which were in one-unit structures were generally lower in early integration tracts than in integration tracts. In the 1950's, however, these patterns were different: the proportional representation of one-unit structures in the housing inventory was generally higher in early integration tracts than in integration tracts in all cities except Boston, the changes in these levels over the decade were on the whole somewhat more substantial than in the previous decade, and in every case except one(early integration, Atlanta) these changes involved increases.

Vacancy rates generally fell in the 1940's and rose again in the 1950's. The rates in early integration tracts were lower than the rates in integration tracts for Boston and San Francisco in the 1940's, and for Boston, Cleveland and San Francisco in the 1950's. But the reverse was true for Cleveland in the 1940's and for Atlanta between 1950 and 1960.

Concerning quality of housing, the overall conclusion is that housing located in early integration tracts was of generally higher quality than housing in integration tracts. The specific patterns differed in details between whites and nonwhites, for the two decades of study, and somewhat also with the particular indicator of housing quality involved.

At the close of the 1940-50 period, the proportions of units which were in old structures were actually higher in early integration tracts than in integration tracts, although the differences did not exceed ten percentage-points. But over the 1950's the reverse was the case with only minor exceptions.

The condition and plumbing measure is only shown in Tables 6A and 6B for the close of each period since the measures for the different census years are not comparable. At the close of the 1940's, the overall representation of basically sound housing in early integration tracts was lower than in integration tracts for Boston and Cleveland, and higher for San Francisco. But at the close of the 1950's housing in early integration tracts was of clearly higher quality(as indexed by this measure) than housing in integration tracts.

It is interesting to note that whereas the quality differential in housing between early integration and integration tracts in 1950 was mirrorred in the housing obtained by whites in these areas, nonwhites quite on the contrary obtained better quality housing in early integration tracts than in integration tracts. The difference involved was quite substantial in San Francisco, and somewhat more modest in the other cities. In 1960 the differences were in the same direction for both whites and nonwhites. It is also interesting to note that for all cases except Atlanta, the difference in percentage-points between proportions sound of units occupied by whites and nonwhites was clearly less in early integration tracts than in integration tracts.

On the measure of persons per room, time-trend comparisons by color are only possible for integration tracts, and the overall conclusion there is that nonwhites also benefitted from the general easing of room crowding over time, with two outstanding exceptions: the nonwhite population of integration tracts in Cleveland actually became more crowded over both decades of study, although the change in proportion of units housing more than one person per room was not large in either case. The second exception was in San Francisco where the corresponding proportion rose by six percentage-points between 1940 and 1950.

In Boston and Cleveland, the levels of crowding in white-occupied units were actually higher in early integration tracts than in integration tracts for both decades. As a result, the white-nonwhite differential was less in early integration tracts than in integration tracts, even though the corresponding levels of room crowding among nonwhite-occupied units were not much different from each other. But the differentials involved were not very substantial, and in Boston in 1950 the level of nonwhite crowding in early integration tracts was itself higher than the corresponding level in integration tracts.

For San Francisco in 1950 the overall levels of nonwhite crowding in both early integration and integration tracts were the same, and for Atlanta and San Francisco in 1960, nonwhite-occupied units were less crowded in early integration tracts than in integration tracts.

Concerning family income, time-trend comparisons are only possible for the decade of the 1950's. The income levels of the total populations of both early integration and integration tracts rose substantially between 1950 and 1960. And specifically for integration tracts for which data are available, the income levels of both whites and nonwhites rose in a remarkably similar pattern: the percentage-point increase in proportion of families earning $3,000 or more ranged between thirty and fifty for both whites and nonwhites. In all cases, the percentage-point increase for nonwhites exceeded the comparable figure for whites. The overall result of this trend was that by 1960 the proportion of nonwhite families in integration tracts who earned $3,000 or more was not far exceeded by the corresponding proportion for white families.

At the same time, the corresponding proportions of nonwhite families in early integration tracts were even higher than the proportions in integration tracts. The former, in their case, were high enough that the difference between them and the corresponding proportions for whites had almost disappeared. Indeed, in Atlanta, Boston, and San Francisco in 1960 the proportion of nonwhite families in early integration tracts who earned $3,000 or more exceeded the comparable proportion for whites in integration tracts.

The patterns at the end of the 1940's were less distinctive, but again family income in early integration tracts was generally higher than family income in integration tracts for both whites and nonwhites in Cleveland and San Francisco. In San Francisco the proportion of nonwhite families in early integration tracts having incomes of $3,000 or more also exceeded the comparable proportion of white families in integration tracts. Boston was exceptional in that the respective proportions were less in early integration tracts than in integration tracts for both whites and nonwhites.

In general summary of the main points in this section, it has been observed that across the different stages of racial residential succession in the four cities of study considered together over the total period of study, white and nonwhite levels of homeownership have tended to be highest in segregated white areas, and lowest in areas of segregated black residence. The pattern of change was such that the largest increases in levels occurred in early integration and integration tracts between 1940 and 1950, but the largest increases in the 1950's occurred in growing areas.

Homeownership levels in early integration and integration tracts were also generally high on the whole, noticably so for nonwhites at the close of each period. And specifically among these areas which were undergoing the first stages of racial residential succession(penetration tracts included), the patterns of change over the periods involved articulated with the initial levels of homeownership in such a way that for nonwhites, the highest levels of homeownership at the end of both periods of study were observed in early integration tracts. For whites, however, this was true only for the 1950-60 period - the homeownership levels at the end of the 1940-50 period were actually lowest in these areas.

On further study of the areas undergoing the first stages of racial residential succession (penetration tracts excluded), the basic character of the housing inventory in these areas was found not to have undergone any drastic changes in the two decades of study. It was also found that with only minor exceptions the quality of housing in areas undergoing the very early stage of succession("early integration" tracts) was generally higher than the quality of housing in areas undergoing the next stage("integration" tracts). Also, the racial differential in quality of housing obtained was generally found to be less in the former areas than in the latter.

Finally the socioeconomic status(as indexed by family income) of nonwhites in early integration tracts was found not only to be quite similar in the aggregate to that of whites in the same areas, but also to be generally higher than that of nonwhites in integration tracts and, indeed, in some instances, higher than that of whites in integration tracts.

The discussion to follow below will aim to articulate the ecological analysis of this section with the aggregate-level analysis of the previous section, and then to bring the collated findings to bear on the basic research questions of this paper.

Discussion

All the materials so far arrayed have sought to explicate the role of homeownership in the various processes involved in the urban housing market and the resulting distribution of housing quality by race of occupants. In that attempt, several intermediate research questions have had to be faced. Considered in order, these may be stated as follows: first, what have been the trends in homeownership? Second, what have been the trends in relevant characteristics of the population, and of the general housing inventory? Third, what have been the trends in housing quality obtained? Fourth, in terms of the ecological manifestations of the larger aggregate processes, what have been the area-specific parallels of these processes? Finally, therefore, what general conclusion may be drawn concerning the implications of patterns of homeownership for the trends in racial residential distribution, and the corresponding distribution of housing and neighborhood amenities?

The review of broad trends revealed a general increase in homeownership levels between 1940 and 1960 for the U.S. as a whole, for the four major regions, and for both whites and nonwhites in the four cities of study. The increase was generally greater for whites than for nonwhites, and the levels of homeownership among nonwhites remained lower over the period than the corresponding levels among whites. Concurrent trends in population characteristics were first of continued growth in the total number of households in each city(the representation of nonwhite households increased in three cities but did not change much in the fourth), and secondly of continued increase in family income. The increase in family income seems to have been greater for whites than for nonwhites.

Concurrent trends in general housing characteristics were of varying construction trends, a substantial increase in the representation of one-unit structures in two of the cities(little change in the other two), and a slight decrease in vacancy rates in the first decade of study, followed by a slight increase in the second decade. The final levels of vacancy remained low. Finally, clear and persistent differentials were observed in the overall quality of housing obtained, with blacks being at a disadvantage.

Each of these general conclusions held to varying degrees in the different cities, and for each of the two decades covered by the total period of study. Each general pattern also had manifestations on an areal basis - the trends in homeownership levels in different areas of the cities, for instance, paralleled quite closely the overall trends in homeownership levels.

The review of ecological patterns over the period of study further revealed a clear relationship between homeownership and stage in the process of racial residential succession: the highest levels of homeownership tended to cluster in the areas of segregated white residence, and the lowest levels tended to occur in segregated black areas. Particularly for nonwhites, the areas which were undergoing the early stages of residential succession evidenced a rapid increase in homeownership levels over the periods involved, with the result that the homeownership levels in those areas at the close of each period were generally high - distinctly higher, as a general rule, than the overall level of homeownership for the total nonwhite population of the city involved. And in addition, the level of nonwhite homeownership at the end of each period of study was consistently higher in areas that had undergone early integration during the period than in areas which had undergone the next stage of racial residential succession. The conclusion clearly seems warranted that homeownership has been a significant factor in the process of succession wherein blacks have gained entry into areas of previously all-white residence, and have thereby come to enjoy neighborhood amenities previously unavailable to them.

The analysis of areas undergoing the early stages of succession also established a clear link between succession and the improvement of quality of housing obtained by blacks: the quality of housing obtained in early integration tracts was generally higher than the quality of housing obtained in integration tracts for both whites and nonwhites. Furthermore, the white-nonwhite differential in quality of housing obtained was lower in early integration tracts than in integration tracts, and the quality of housing obtained by nonwhites in early integration tracts was generally higher than the quality of housing obtained by nonwhites in the city as a whole.

With the establishment thus of a linkage between homeownership and succession, and between succession and improvement of quality of housing obtained by nonwhites, a connection is strongly suggested between homeownership and the improvement of quality of nonwhite-occupied housing.

Further evidence bearing on this issue is provided in the direct tabulation in Table 7 of trends in quality of total housing obtained by whites and nonwhites, by tenure, for the four cities of study over the entire period involved. The clear conclusion from this table is that homeowners have enjoyed better quality housing than renters, and this conclusion holds true for both whites and nonwhites. Furthermore, the white-nonwhite differential in quality of housing obtained has been less among homeowners than among renters, and the owner-renter differential has been greater among nonwhites than among whites. The ecological analysis undertaken above essentially details the parallel manifestations and specific explication of these aggregate patterns.

There is still, however, a remaining question to answer. During the review of broad trends above, it was observed that levels of family income had generally risen over the period of study. The ecological analysis also revealed a clear relationship between family income and succession: by the end of each period of study, the family-income profiles of the white and nonwhite populations of each of the areas undergoing the first stages of succession were quite similar. Also, the differences between whites and nonwhites in proportions of all families having incomes of $3,000 or more were less in early integration tracts than in integration tracts, and the nonwhite proportions in early integration tracts were higher than those in integration tracts, and in some cases, than the corresponding proportions for whites in integration tracts. Thus the question arises: what effect does the implied relationship between income and quality of housing obtained(see also Glazer and McEntire 1960) have on the relationship between homeownership and quality of housing?

Table 8 represents an attempt to examine the tenure-quality relationship net of income by showing data for the distribution of housing quality by family income and color of occupants, and by tenure. These data are for 1960, the end of the total period of study, showing thereby the net patterning of these relationships after all the processes reviewed in the analysis above. They also refer to the total SMSA population of the U.S., to give an overall summary of patterns in metropolitan areas. Separate tables for the different cities are omitted for considerations of space, particularly since they would lead to identical conclusions. The major conclusions from Table 8 are clear and definite: at all levels of income, homeownership is associated with better quality housing for both whites and nonwhites. Also, the white-nonwhite differential in quality of housing obtained(excluding age of structure) was less for homeowners than for renters, and the owner-renter differential was greater among nonwhites than among whites, except on the age of structure, and for income classes under $4,000 on the condition and plumbing measure.

On the whole, therefore, it may be concluded that patterns of homeownership have a definite bearing on the differences in quality of housing obtained by whites and nonwhites. In the face of the various distinctive characteristics of the four cities of study over the two decades involved, homeownership has been found to feature quite prominently in the early stages of the process of racial residential succession. Particularly over these early stages, this process has resulted in a definite improvement in the quality of housing obtained by nonwhites. It is also true that particularly in these early stages, this process has involved high levels of socioeconomic selectivity for the nonwhite population, and a question thereby arises as to the extent to which the improvement in quality of housing obtained by nonwhites may be attributed to homeownership independently of the socioeconomic characteristics of the population. But the overall advantage which homeowners enjoy over renters in terms of quality of housing obtained is technically independent of income: the quality differential observed between owners and renters over the total period of study was also found to hold at the end of the period for every level of income, and for both whites and nonwhites.

Thus from the point of view of improving the quality of housing obtained by nonwhites in metropolitan areas of the U.S., homeownership has served as a viable instrument of social change in gradualistic perspective. It may be well to encourage this factor with affirmative national policy, especially inasmuch as nonwhite housing quality is not only improved by homeownership but is also thereby brought closer to corresponding white levels..In addition to the improvement of housing quality on which this study has focused, other reasons may be adduced in support of the desirability of encouraging homeownership among blacks - such as, for instance, the benefit of the accrual of equity to the homeowner(see Kain and Quigley 1970). And if a program of encouraging homeownership is to be undertaken, it should be useful to attempt to delineate the target population toward which such a program may be directed. This is a task for further research.

NOTES

1. Because of the lack of space, the following items have been omitted: (a). a final paragraph dealing with the subsidiary research question of the interchangability of "nonwhite" and "black." (b)Tables 1A through 8. (c). References.

2. The full text and tables are available on request.

FITTING AND ANALYZING POLYMODAL DISTRIBUTIONS OF CONSUMER CHOICES

Purnell H. Benson, Rutgers Graduate School of Business

The Problem of Describing Consumer Choices

This paper reports research experiences and insights from analyzing frequency distributions of consumer choices, culminating with the trimodal problem presented by data published by R. N. Reitter (1969). First, some propositions are offered as premises for work in analyzing distributions of consumer choices.

1. Most dispersions of consumer choices along variations in product characteristics are normal or moderately skewed. This conforms with the psychological fact that most distributions of psychological traits are normal or near-normal, arising as they do from probability principles of genetics or environmental chances. The distribution of educational or economic opportunity is markedly skewed. When it impinges upon human development, dispersions of human traits become skewed like that of taxable incomes, for which the leptokurtic lognormal distribution function can be used.

2. In consumer buying, the pressure of social convention and product availability mold consumer choices in ways which gives distributions stubby tails and sometimes no tails. The penalty of being different crowds consumer populations into platykurtic distributions, flattened with cut-off tails. For this situation, a distribution function frequently used in psychometric work is the arcsine, which is symmetrical without tails. What shall one do when one needs a platykurtic skewed distribution? The Beta distribution is one possibility, but lacks sufficient flexibility.

3. Often it does not matter what distribution pattern we assume consumer choices to have, if the measure of the product characteristic, which is the basis for consumer choice, is converted into normal deviate measurements. That is, sweetness of pudding can be measured with reference to the median as the zero point and the spreadoutness of choices by the consumer population as the scale unit.

By definition, when a product characteristic is measured in this way, a normal distribution will always fit perfectly. The true distribution could be rectangular, and the normal deviate values can still be assigned. What is wrong with this normalization? The difficulty arises when boundaries are drawn between consumer segments. It is usually assumed that these boundaries between product stimuli fall midway between them. If the true distribution is distorted, the midlines fall at different places each time a different method is used for stretching and contracting. This does not matter in market segmentation analysis if the boundary lines for consumer buying choices are very approximate. When the choices which consumers make depend very much upon habituation and advertising, the boundaries become zones, rather than lines.

4. It does matter how data are transformed if the market is highly segmented, as in hard goods or automobiles. Here the accurate representation of buying densities and segment boundaries is crucial. The more precise mathematical solution which suggests itself is to find the functional relationship between the normal deviate value for accumulated choices and the product characteristic physically measured.

One such relationship is a modification in the usual lognormal. Let the relationship between the normal deviate and the logarithm of the product characteristic be rewritten as an exponential relationship between the product characteristic and _e_ raised to the normal deviate as its exponent and then multiply the whole exponential term by the normal deviate. The relationship obtained is skewed and platykurtic. Moreover, by making the exponent in the normal deviate a polynomial, the degree of kurtosis and the skewness are controllable.

In working with this modified lognormal, it can be cast in the form of linear regression if we estimate and subtract the median of the product characteristic Z, then divide the resulting value by the normal deviate X, and then take the logarithm of both sides of the expression (Benson, 1965).

$$\log ((Z - A)/X) = B + CX + DX^2 + \ldots$$

One limitation is the poor behavior of the distribution in the vicinity of the tails. One or both wings ends in an upward projecting lip, instead of a tail, or else one of the tails may be greatly elongated. These are conditions which become pronounced when extreme skewness or platykurtosis bordering on rectangularity are sought. For any ordinary distribution of consumer choices, the tail problem is not serious. In extreme situations, the function becomes discontinuous at intermediate points as well as at the ends.

5. A contrast exists between the fitting of frequency functions by classical methods of statistics and what the consumer researcher is doing. (a) From a practical standpoint, he is not much interested in the tails of distributions, when he is describing distributions of buying choices. The tails mean much in hypothesis testing. But to find out where the best sales prospects are is not hypothesis testing in the usual sense. Hence the consumer researcher does not satisfactorily measure how well an empirical distribution is described simply by means of a chi square test of goodness of fit, which is much affected by the smaller cell frequencies in the tails. In consumer work, the correlation between actual and estimated sales densities is a more meaningful measure of goodness of fit of a distribution function.

(b) With some exceptions, irregular dispersions of buying choices do not arise from any of the familiar probability models. They depend upon product availability, a categorical matter, and social custom, a truncating matter. Exceptions include use of binomial and beta probabilities in analyzing media choice behavior of consumers (Benson, 1971; Greene and Stock, 1967). Also, if time of entrance of the buyer into the market place is a variable of consumer choice, then the exponential and gamma distributions are relevant, the first where a burst of promotion gets K percent of the remaining consumers to buy after each increment of time, and the second where a chain reaction takes place with each consumer telling others by word of mouth buy until all available consumers have bought.

6. Consumer researchers are restricted by lack of adequate data. This is first evident in the fewness of the points at which to measure frequencies of consumer choices. Consumers who try out products will usually try out one or two, and no more, except in the artificial situation of a food-testing laboratory. Moreover, hasty trials of many product stimuli are not typical of the leisurely situation in which consumers reach their buying decisions.

Nor are the data collected of one uniform kind. They may be multiple choice, paired, or monadic. For expensive hard goods, as TV screen size, the influence of advertising upon choice of screen size is much less and market data provide multiple choice data which are reasonably dependable.

More commonly, consumers are asked to try out products. One form of data is paired choices. From the standpoint of consumer judgement, these are incisive data. The problem to which Kuehn and Day (1962) addressed themselves is that two products which are much alike tend to get about a 50/50 split, whether the two products are near the middle of the distribution of consumers or at one side. If the dividing line between consumer choices puts 30% on one side and 70% on the other side, then two products on either side of the line should draw a 30-70 vote. They don't, and what can be done about it? The question has still not been fully answered as an operating thing.

Consumers can be asked whether their ideal buying choice is for a product to one side or the other of an experimental product tried out, that is, do they want more or less of a product characteristic than the experimental formulation given to them to try out? In effect, they are asked to position their choices relative to the product they look at. By taking the accumulated number of consumers who will buy a product up to the one tried, the cumulative frequencies needed for fitting distributions are obtained. Differencing the cumulative frequencies gives class frequencies.

Analysis of Reitter's Trimodal Distribution

R. N. Reitter (1969), following J. O. Eastlack (1964), felt he found an answer to lack of empirical data for multiple choices by asking consumers to mark on a line-interval scale where they feel their ideal choices to be. But whether they do this with precision leaves unanswered questions. As a questionnaire procedure, Reitter's method leaves something to be desired. However, he has published data exhibiting three modes, as shown in Figure 1.

The frequency bars have been smoothed, first by drawing lines which connect first and third quartile points within each class interval with those of the

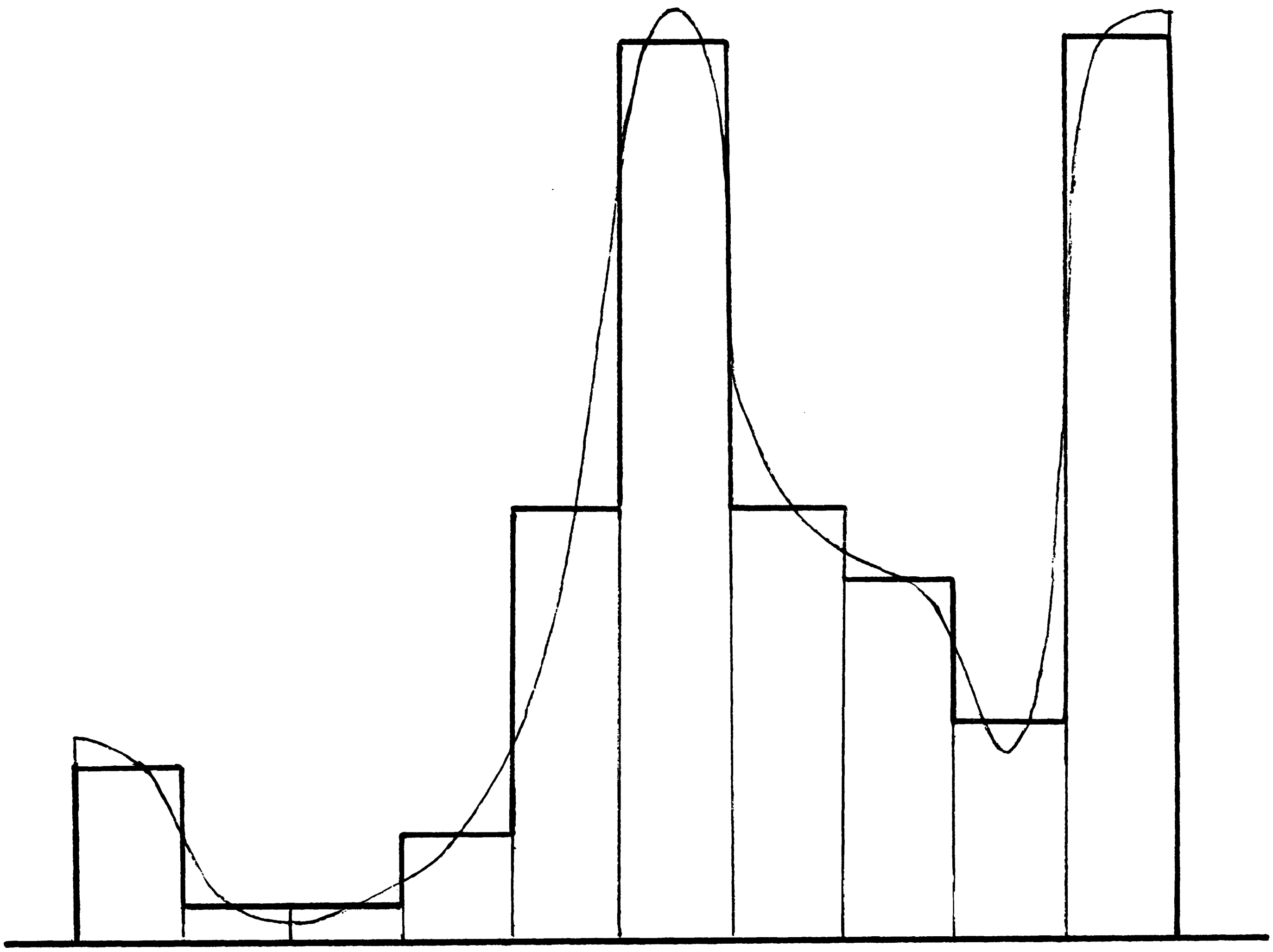


Figure 1. Choices for Strength of Coffee (R. N. Reitter Data)

next class interval, and then drawing a continuous curve which leaves areas between class boundaries unchanged. Coffee drinkers tend to check either the ends of this scale or the middle for how strong they want their coffee. Whether they are lethargic respondents or whether they really feel this way about coffee really does not matter if we are anxious to find data with which to explore the fitting of a polymodal distribution. The marks on the scale are choices made by consumers in some sense.

Before turning to the trimodal problem presented by Reitter, some comments may be made on fitting bimodal distributions. Such distributions are infrequent, and can usually be reduced to distributions of choices for related products, which have been combined. For example, lumping hand portable TV sets with room consoles will yield a bimodal distribution of buying choices. This bimodality can be handled by separating the data by product class to use unimodal analysis.

Otherwise, a workable procedure is to make a normal deviate conversion with the normal deviate a function of the product variable as a quartic polynomial or higher. In order not to waste the end points where the normal deviate is infinity, -3.0 and +3.0 are used as terminal values. Here only .002 of the distribution lie outside. If the data are not too erratic, this will yield a curve which appears decent on the surface. What is a decent curve? The statistician would say one which adequately fits properly collected data. The abundance of data for replicated tests of this kind is lacking. Making the accumulated area a function of the product variable, or vice-versa, is also serviceable to describe the distribution function. The modified lognormal distribution is so stable that it fills in the valley between the two modes. The more flexible linear polynomial is to be preferred.

Returning to the problem of fitting Reitter's data, it may be noted that the distribution has a central mode and J's at either end, and the left hand J reversed. The frequency distribution as it stands has no tails. One might infer that if more categories beyond the present ends had been offered, consumers might have checked them. In this event, the terminal frequencies represent cumulative frequencies up to the inside boundaries of the end categories. Here the class frequencies are taken as Reitter gives them and the distribution is regarded as ending with the outside boundaries of class intervals one and ten.

Eight functional types were tried out, as indicated in Table 1. Seven of them involve polynomial expansions on the right hand side. Terms up to the twelfth power were used, with best fitting terms selected by stepwise regression. This improves the fit for estimating class frequencies, but may do so at the expense of producing absurdities in the density function at intermediate points. Footnotes in Table 1 comment upon the behavior of the density functions.

Table 1

PERCENTAGE FREQUENCIES OF CONSUMER CHOICES ESTIMATED FOR TRIMODAL DATA, USING DIFFERENT TYPES OF FUNCTIONS FOR FITTING THE DISTRIBUTION

Type of Function[a]		Class Interval										Error Variance
	Observed Frequency:[b]	1 5	2 1	3 1	4 3	5 12	6 25	7 12	8 10	9 6	10 25	
$Z = X + X^2 + \ldots$		3[c]	2	5	7	10	13	14	14	11	21[c]	24.6
$X = Z + Z^2 + \ldots$		5[c]	2	-1[d]	5	13	18	18	10	4	26[c]	10.0
$Z = Xe^{X+X^2+\ldots}$		5[c]	1	1	3	16	21	12	9	8	24[c]	3.8
$X = Ze^{Z+Z^2+\ldots}$		5[c]	4	5	8	11	13	13	11	10	20[c]	23.8
$A = Z + Z^2 + \ldots$		5	2	-1[d]	4	14	20	16	9	6	25	5.2
$Z = A + A^2 + \ldots$		5	1	2	2	13[e]	24	12	10	6	25	.4
$F = Z + Z^2 + \ldots$		4	-1	6	0	15	18	7	1	13	37	39.6
$A = \sin^2(X + pi/4) + \ldots$		4	2	1	7	14	17	16	14	8	17	18.6

[a] Z is measurement of product characteristic. X is horizontal deviate. A is accumulated frequency. F is class frequency.

[b] N equals approximately 200.

[c] These frequencies include accumulated proportions beyond class intervals 1 and 10. The density function rises abruptly to infinity outside of these class intervals.

[d] The density function is discontinuous within the class interval. Class frequency is obtained by differencing accumulated frequencies between boundaries of the class interval.

[e] The cumulative function is multivalued for this class interval. The frequency recorded is an average of two differences between class boundaries.

By coincidence, Reitter's data have a pattern which the modified exponential function could effectively represent. This distribution function has a central mode, and can terminate with upward projecting lips at either end. The fit of this function is included, along with simple polynomials which use both the product characteristic or the normal deviate as the dependent variable. Also tried are the accumulated frequencies as a function of the product characteristic, and also class frequencies as a function of the product characteristic. None of these functional formulations really copes satisfactorily with the data, as shown by Table 1. The modified lognormal is a fairly satisfactory fit. But left and right hand lips rise outside of the data.

The best fit, if the density problem is overlooked, is provided by the product characteristic expressed as a function of the accumulated choices. The most serviceable function from an all-round standpoint is the arcsine, listed last in the table.

This last line of solution regards the trimodal distribution as a composite of three separate distributions. Owing to the lack of tails, the use of the arcsine function for the separate modes seems appropriate. Each of the three component distributions has a mean, a standard deviation and a proportion within it. (The third proportion is dependent upon the first two, so there are eight unknowns). The 10 class intervals provide 9 degrees of freedom. There seems no simple solution to such a system other than taking trial values and generating a response surface which represents the residual error term in fitting the frequencies. The final result showed a component projecting to the left of the class with 6 percent in it and 1 percent adjacent. This last fit may imply that Reitter's data ought to include consumer replies before the first class interval and after the tenth class interval.

Still another method, not tested here, abandons the quest for a single mathematical expression to describe the density throughout the range of all of the frequencies. Instead, the frequency data are divided into several overlapping ones. Within each zone a separate distribution function can be fitted and densities calculated by taking the derivative of the cumulative function. This raises interesting questions concerning how well such overlapping functions will blend.

Some may feel that a good deal of time has been used collecting climbing equipment for a mountain climb still not completed. What does one do with a trimodal mountain so difficult to ascend? The problem is a broad one, arising in distributions of events over time, such as time series data, as well as in consumer choice analysis.

REFERENCES

Benson, P. H. Fitting and analyzing distribution curves of consumer choices. Journal of Advertising Research, 1965, 5, 28-34.

Benson, P. H. A solution to define latent segments of media audiences. 1971 Proceedings of the Social Statistics Section---American Statistical Association.

Eastlack, J. O. Consumer preference flavor factors in food product design. Journal of Marketing Research, 1964, 1, 38-42.

Greene, J. and Stock, J. S. Advertising Reach and Frequency in Magazines, New York: Readers Digest Association, 1967.

Kuehn, A. A. and Day, R. L. Strategy of product quality. Harvard Business Review, 1962, 40, 100-110.

Reitter, R. N. Product testing in segmented markets. Journal of Marketing Research, 1969, 6, 179-184.

TOWARD A COMPREHENSIVE VIEW OF THE ATTITUDE-BEHAVIOR RELATIONSHIP: THE USE OF MULTIPLE-SET CANONICAL ANALYSIS

Richard J. Lutz, University of Illinois[1]

John A. Howard, Columbia University

The primary purpose of this paper is to present an application of an advanced multivariate statistical technique-multiple-set canonical analysis- to a problem of current interest in consumer research - the study of the attitude-behavior relationship. M-set canonical analysis appears to offer some advantages over other multivariate techniques in treating some special problems which have arisen in the literature on consumer attitudes. Before introducing the actual problem under consideration here, it will be beneficial to briefly review past research on consumer attitudes which is directly relevant to the current investigation.

Review of the Literature

Current research in consumer psychology has as its goal the explanation of the behavior of the consumer. No longer is the consumer treated as a "black box" which mysteriously treats input from his environment and responds by purchasing some product or service. Rather, the current focus is to examine the decision processes by which the consumer transforms information into action. As might be expected, consumer researchers focusing on such internal processes, have begun to rely more and more heavily on predispositional, rather than overt, states of the organism. Because attitude is seen as a mediating variable intervening between psychological inputs and outputs, it has become a very useful construct to include in a theory of consumer decision-making. Unfortunately, a construct as rich in meaning as attitude can be a double-edged sword, creating almost as many problems as it does provide solutions. This apprehension is borne out when one reviews the psychology literature and finds that attitude as a construct has been a center of controversy for over fifty years (Fishbein, 1967). Nevertheless, the potential explanatory power of the attitude construct is certainly worth the risk of some disagreement among theorists, and attitude has become firmly entrenched as a major variable in the study of consumer behavior, if not by operational consensus (Adler and Crespi, 1968), at least by virtue of the vast amount of research effort it has stimulated (Nicosia, 1966; Sheth, 1967; Howard and Sheth, 1969).

Considerable attitude research in consumer behavior has recently began to build on the theories of Fishbein (1967) and Rosenberg (1956). While these two theories were developed independently and out of different traditions, methodological similarities between the two theories have led to a gradual merging of their applications in the marketing literature. Rather than discuss the similarities and differences between these two theories in their initial formulations, the present discussion will be limited to such applications.

The basic form of the Fishbein-Rosenberg type of model, expressed in functional form as a linear relationship between two variables, is:

$$(1) \quad A_{ij} = f\left(\sum_{i=1}^{n} B_{ijk} I_{ijk}\right)$$

where A_{ij} = consumer i's attitude toward brand j

B_{ijk} = the extent to which consumer i believes that brand j possesses some attribute k which leads to the fulfillment of some desire

I_{ijk} = the importance to consumer i that brand j possesses attribute k (i.e., the importance that brand j fulfills the desire in question)

n = the number of salient attributes

Thus, the consumer's attitude toward a brand is conceptualized within an expectancy-times-value framework,[2] where the brand is evaluated as a goal-object. Beliefs about the brand represent the degree to which the consumer expects the brand to possess attributes which will lead to satisfaction and the importance of each attribute represents how valuable each particular type of satisfaction is to the consumer of that brand. Since the belief and importance components combine multiplicatively, this interaction determines whether each attribute will be of significance in explaining attitude toward a brand. It is not enough that a brand possess a great deal of some attribute; the consumer must also consider it important to derive this type of satisfaction from that brand. Similarly, an attribute which is very important to the consumer will have little effect on attitude if none of the available brands are believed to possess an adequate level of that attribute. This particular approach to the analysis of consumer attitudes should be especially appealing to marketing researchers, who are concerned with actions which will ensure that products conform to consumers' wants and needs. Consider, for example, Kotler's (1967) definition of a product as "a bundle of physical, service, and symbolic particulars expected to yield satisfactions or benefits to the buyer" (Kotler, p. 289). Substituting the word "attributes" for Kotler's "particulars" and assuming that satisfactions and benefits have some value to the consumer completes the correspondence between the marketing product concept and the expectancy-value approach to attitude. Perhaps the large amount of research generated by this model is due in part to the intuitive appeal of the model in a marketing context.

Despite its intuitive appeal, the expectancy-value model has stirred some controversy among marketing researchers since its first application three years ago (Hansen, 1968). A review of the fifteen-odd articles employing this model which have appeared since then reveals three major issues which have confronted marketing researchers regarding the expectancy-value attitude model:

a) Would a disaggregative approach be more satisfactory than the customary summed-score form of the model?

b) What is the relative contribution of each component (i.e., beliefs or importance) in determining the consumer's attitude toward a brand?

c) How strong is the relationship between the expectancy-value attitude measure (B·I) and subsequent measures of market performance; e.g., preference and/or purchase? In other words, what is the market significance of this means of analyzing consumer behavior?

The present discussion will deal with each of these issues in turn. While most of the early applications of the expectancy-value model were strictly extensions of the basic model to a marketing situation (Hansen, 1968, 1969; Bass and Talarzyk, 1969; Bither and Miller, 1969), it was quickly recognized that using a single numerical index to represent the consumer's cognitive structure left something to be desired. The very information which had made the model so intuitively appealing to the marketer, permitting analysis on a set of underlying beliefs, was not being utilized to its fullest extent - it was being collapsed into a single value. This severely hampered the utility of the model for suggesting possible strategies for attitude change. Sheth (1970, p. 8) has expressed the shortcomings of the aggregative model while proposing a disaggregative linear model of the form:[3]

(2) $A_{ij} = f(B_{ij1}I_{ij1} + B_{ij2}I_{ij2} + \text{---} + B_{ijn}I_{ijn})$

There has been virtually no disagreement among marketing researchers that this diaggregative model is superior to the earlier version of the model. As mentioned earlier, it is of primary interest to the marketer to determine which of the attributes of a brand are most significant in contributing to the consumer's attitude and behavior with respect to that brand. The identification of these attributes enables the marketer to utilize his promotional dollar more effectively by stressing those attributes which contribute most to the consumer's preferences or purchase pattern. In some cases, there may be only one dominant belief which, if it can be changed, may lead to a substantial modification of preference or purchase.

The disaggregative approach is well-suited to these purposes, as the relative contribution of each $B_{ijk}I_{ijk}$ element can be determined through statistical estimation procedures such as multiple regression (Sheth, 1969, 1970; Cohen and Houston, 1970) or discriminant analysis (Cohen and Ahtola, 1971). Not surprisingly, the above researchers have reported much higher predictive power, whether the dependent variable is preference or purchase, when the disaggregative model is compared with the aggregative model.

The use of multivariate procedures has revealed many relationships which could not have been identified under the summed-score form of analysis. For example, Sheth (1971), in a large-scale study, used canonical analysis to show that while "taste" was the most significant attribute in determining consumer attitudes toward a convenience food product, two other attributes - "good buy" and "meal substitute" - were more significant in determining the consumer's intentions to buy the brand.

Thus, there is a consensus among marketing researchers that the disaggregative model is the more powerful one for the explanation of consumer behavior. It is at this point, however, that researchers diverge in their opinions regarding the second major issue which has been raised in the use of this attitude model in marketing.

As noted previously, when Sheth (1970) proposed the disaggregative model, he also chose to exclude the value term from the equation, creating a linear model of the form:

(3) $A_{ij} = f(B_{ij1} + B_{ij2} + \text{---} + B_{ijn})$

Several researchers have presented evidence relating to the question of the relative contribution of the belief and importance components in determining attitude under the beliefs-importance model. The results have been conflicting, with some researchers reporting the importance component to be relatively useless (Sheth and Talarzyk, 1970), while others have found it to contribute significantly to explained variance (Hansen, 1969).

Recently, two independent teams of researchers have applied multidimensional scaling to this problem. Moinpour and MacLachlan (1971) created two separate dissimilarity spaces, one using the importance measure to weight the dissimilarities of beliefs across brands of headache remedies, and the other using unweighted dissimilarities. The resultant two-dimensional configurations were practically identical, leading the researchers to conclude that the importance measure is superfluous. Hansen and Bolland (1971), using a similar procedure to compare the beliefs-only and beliefs-importance models, found that while the two measures yielded highly correlated distance scores in the case of student pub-crawling behavior, the beliefs-only model made more correct predictions. However, in the prediction of patronage of self-service car washes, the B·I model was clearly superior. In attempting to explain these conflicting results, Cohen and Houston (1970) and Hughes (1970) have suggested that differences in measurement procedures could have significant effects on the relative power of the models; nevertheless, the issue remains unresolved.

The controversy surrounding the importance component is, however, empirical rather than theoretical. Even the researchers who have presented negative results regarding the inclusion of the importance measure acknowledge that importance is probably implicit in the mind of the consumer (or the respondent in a survey) and thus may already be included in the belief measures (Sheth and Talarzyk, 1970). Cohen and Ahtola (1971) tested this hypothesis and found that generally the correlations between the two components were not high enough to forego the measurement of the importance component.

The final issue of interest here is the market significance of the expectancy-value attitude model. It is in this area that there has been the greatest divergence among researchers in terms of the particular means used to test the model and the statistical methods utilized to assess the relationships between the B·I measure and criterion measures of preference and/or purchase of the brand in question.

Since the summed-score form of the model did not allow the use of powerful multivariate statistical techniques, the earliest evaluations of this form ofthe model used tests of independence (e.g., chi-square and other nonparametric tests) relative to such criteria as brand "appeal" (Bither and Miller, 1969) and actual choice behavior in an experimental situation (Hansen, 1968). Bass and Talarzyk (1969) used the computed beliefs-importance scores to attempt to reproduce the preference ranking of brands within six product classes, adopting a "confusion matrix" of conditional probabilities to present their results. Sheth and Talarzyk (1970) used a simple linear regression model in which brand preference served as the criterion variable and the B·I score was the predictor variable. Results of these studies indicated that the expectancy-value framework was probably a valuable one from which to analyze consumer behavior, but at the same time, the predictions based on the model were not as good as marketing researchers desired. It was at this point that the disaggregative approach was initiated, along with the use of more sophisticated analytical procedures.

Sheth (1969) introduced the use of multiple regression into this research area. A measure of overall liking for the brand (affect) was used as the dependent measure; and separate belief measures were used as predictors, as in Equation (3) above. This analysis yielded better predictive power and also enabled the researcher to identify exactly which attributes were the primary contributors to the explained variance in liking for a brand. Cohen and Houston (1970) used the same approach with similar good results, relating the disaggregative B·I measure to a retrospective report of purchase frequency. Since their analysis included a measure of importance, it represents an application of Equation (2) above. Sheth (1970), in an extension of the basic model to account for situational factors and to better conceptualize the entire decision-making process of the consumer, performed a multiple regression on data gathered in a longitudinal study of consumer attitudes and purchase behavior with respect to three different brands of instant breakfast. The data were gathered at three separate points in time. Sheth used intention to purchase as the criterion variable and belief measures as predictor variables. In eight out of nine cases, a significant R^2 was obtained. Lutz (1971), in a similar analysis performed on a different population and with respect to a service rather than a product, found that the belief measures explained a substantial amount of variance in purchase intentions.

While the multiple regression approach proved to be very useful in predicting affect and behavioral intention from a set of belief and/or importance measures, the prediction of actual purchase behavior seemed more closely related to a discriminant analysis problem, as noted by Cohen and Ahtola (1971). Using a consumer's overall purchase pattern to classify him as loyal to one of three leading brands of toothpaste, they were able to explain 67% of the variance in purchase behavior using the belief-times-importance measures as the predictor variables in a multiple discriminant analysis. The model using only belief measures did only slightly worse, while the aggregative form of the model was able to account for only about 44% of the variance in purchase behavior. While it must be borne in mind that the purchase measure was retrospective in this case, the results of the Cohen and Ahtola study are very encouraging. The discriminant model not only yielded a much better behavioral prediction than did the multiple regression model on the same data, but as in multiple regression, discriminant weights enable the researcher to ascertain exactly which attributes are instrumental in determining differences among consumers' purchase behavior.

Since there are numerous measures of market response which may be of interest to the marketer at a point in time, Sheth (1971) set up another test of the expectancy-value model, using seven belief measures as the predictor set in a canonical analysis. The criterion set consisted of affect, intention, and actual purchase behavior. The first two canonical variates accounted for approximately 60% of the variance in the criterion set. This form of analysis has the particular advantage that it is able to show, within the same framework, which attributes are instrumental in explaining specific measures of market response.

Multiple-Set Canonical Analysis

M-set canonical analysis was proposed as an extension of traditional (2-set) canonical analysis by Horst (1961) as a method for examining the relationships among three or more set of variables. In the 2-set case, the objective of canonical analysis is to derive linear combinations of the two sets of variables such that the correlation between the two sets is maximized; under m-set analysis, the linear combinations of the m sets of variables are derived such that the sum of the intercorrelations among the m sets is maximized. For example, if there were three sets of variables to be considered, m-set canonical analysis would yield a 3x3 matrix of canonical correlations as shown below:

1.00	.78	.63
.78	1.00	.41
.63	.41	1.00

As seen above, the canonical correlations matrix is symmetric and has diagonal values of unity. Just as in 2-set canonical analysis, as many canonical variates can be derived as there are variables in the smallest of the m sets; all of the m possible canonical variates are mutually orthogonal. Also similar to the 2-set case, the solution under m-set analysis provides a weight for each variable in the total set of variables. Thus, m-set canonical analysis performs the same functions and provides the same type of information as does 2-set canonical analysis, but with the distinct advantage of being able to accommodate more than two sets of variables simultaneously.

Horst (1961) suggests several possible applications of m-set canonical analysis:

a) Testing the congruence of factor structures among more than two subpopulations which have responded to the same profile of test batteries.

b) Testing the similarity of response patterns when subjects are exposed to three or more treatment conditions.

c) Testing the similarities among three or more independent test batteries administered to the same population.

Considering only these three categories of problems, many potential marketing applications can be derived. In the area of segmentation analysis, m-set canonical analysis would allow the researcher to split the population into homogeneous groups on the basis of several different demographic and socio-economic variables and then simultaneously compare the groups on the basis of their response patterns to a battery of attitudinal or personality measures.

Another important application of m-set analysis lies in the study of consumer decision processes. There are many classes of variables which have been shown to have significant effects upon the consumer (Sheth, 1967); among these are demographic variables, social class, group influence, interpersonal interaction, marketing communications, attitudes, personality, etc. M-set canonical analysis would allow each of these classes of variables to be included as a separate predictor set for one or more sets of market response variables. Thus, in one global analysis, the researcher could investigate not only the effects of the predictor variables on the criterion set, but also the interrelations among the various classes of predictors. This can be viewed as a type of exploratory procedure at the macro level in that it deals with several different classes of variables simultaneously.

M-set canonical analysis can also be useful at a more molecular level of the study of consumer decision processes for sorting out relationships within a particular class of explanatory variables. It is a problem of this type on which this research focuses.

Definition of the Problem

While more familiar multivariate methods have been used to compare the differential results when one or the other of the two disaggregative forms of the expectancy-value model were employed, m-set canonical analysis can be used to test the effectiveness of both models in the same analysis. Therefore one aspect of the problem to be treated here is to include both forms of the beliefs-importance model as presented in Equations (2) and (3) in an m-set canonical analysis. This will reveal simultaneously the relationship of each model to the criterion variables and also the nature of the relationships between the two alternative models.

The importance of utilizing multiple rather than isolated criterion measures to increase reliability has been articulated by Fishbein (1967). Since previous research in the study of consumer decision processes has focused on two different sets of criterion variables - preferences and purchase behavior - it seems appropriate to include multiple measures of each set of variables in the m-set canonical analysis. This treatment will reveal the relationships between the two sets of criterion variables, as well as identifying their separate relationships with the predictor variables.

Thus the current problem is to investigate simultaneously the relationships among four sets of variables:

1) A set of belief-importance scores
2) A set of beliefs
3) A set of brand preference measures
4) A set of purchase measures

Following are some general hypotheses regarding the relative magnitudes of the canonical correlations among the various sets of variables.

Hypothesis 1: The set of B·I measures will be slightly more related to both criterion sets than will be the set of beliefs-only measures.

Hypothesis 2: Both sets of belief measures will be more closely related to the set of preference measures than to the set of purchase measures.

Hypothesis 3: The set of preference measures will be more closely related to the set of purchase measures than will either of the sets of belief measures.

The first hypothesis is based on previous empirical results which, in general, have shown the beliefs-importance model to be slightly superior to the beliefs-only model. The latter two hypotheses derive from the proposition that consumer preferences intervene between beliefs and actual purchase behavior in the decision process (Lavidge and Steiner, 1961; Howard and Sheth, 1969). The more closely related two sets of variables are in the decision process, the stronger should be the empirical relationship between them (Sheth, 1970). Thus, beliefs should be more closely related to preference than to purchase, and preference should be a better predictor of purchase than beliefs.

The Data[4]

The data used in this analysis were collected as part of the Buyer Behavior Project under the direction of Professor John A. Howard at the Columbia University Graduate School of Business in 1966. A longitudinal panel of housewives was formed through the use of standard probability sampling procedures. Initial contact was made through a mail questionnaire, which was followed at approximately 1-month intervals by three telephone interviews. In addition to responding to a variety of attitudinal and socio-economic questions, each panel member recorded her purchases of a convenience food product over the entire duration of the panel. Since the specific information required in the current study was gathered for only one brand - here called CIB - it is the only brand which will be analyzed. All of the beliefs, importance and preference measures were taken from the mail questionnaire, while the purchase measures were taken from panel diaries.

The beliefs-only measures consisted of the respondent's ratings of CIB on twelve 7-point bipolar scales. The positive ends of these scales are shown in Table 1. A value of 1 represented the most favorable rating and a value of 7 the least favorable rating.

The same twelve beliefs, weighted by importance, were used to form the set of B·I measures. Each belief rating was multiplied with its corresponding importance, which was measured on a 3-point scale from "very important" to not at all important." Thus, each of the B·I measures had a potential range of values from 1 to 21, with a lower score representing a more favorable response. See Table 1 for the actual beliefs included in this set.

The set of preference measures, seen in the table, consisted of three measures of liking for a brand. The measure of affect was obtained through the use of a 7-point bipolar scale, ranging from "In general, I like CIB very much" to, "In general, I do not like CIB at all." The semantic differential measure was derived from four beliefs about CIB which were shown to load on the evaluative factor of a factor analysis (Osgood, Suci, and Tannenbaum, 1957). These four scales include "snack," "low in price," "good buy," and "real flavor." The ratings on these four scales were summed to obtain the semantic differential attitude score, which had a range of possible values from 4 to 28. The Likert scale attitude measure was derived from the respondent's agreement-disagreement ratings on a series of projective-type questions regarding the type of person who would use CIB. After standard item analysis procedures, five scales were selected to represent the Likert score. These scales included "people trying to gain weight," "people who are health conscious," "people who have a health problem," "people who want a quick energy lift," and "people who like snacks." Since the extent of agreement was measured on a 5-point scale, the Likert scale attitude score had a range of possible values from 5 to 25, with lower scores representing more favorable attitudes.[5]

The set of purchase measures, as shown in Table 1, consisted of 3 continuous variables and two dichotomous variables. All of the measures are sound from a conceptual viewpoint as alternative measures of purchase behavior. The only possible drawback is that the inclusion of dichotomous variables in traditional 2-set canonical analysis violates certain assumptions which are necessary to perform tests of significance on the canonical correlations obtained (Green and Tull, 1966). However, in this case, this shortcoming seems to be of minimal importance since traditional 2-set significance tests do not hold for m-set canonical analysis (Horst, 1961). While it is recognized that dichotomous variables are not the most satisfactory measures to use in m-set canonical analysis, these deficiencies are not judged to be critical.

A total of 583 respondents whose data were complete for the scales used in the study constituted the initial sample. One hundred of these respondents were randomly selected to form the subsample which was used to construct the semantic differential and Likert scale on a post hoc basis. This left a total of 483 respondents in the test sample which was used in the main analysis.

Data Analysis[6]

The first step in m-set canonical analysis is to create a supermatrix, G, of correlations from the raw data. Using Horst's notation, this supermatrix appears as:

$$\begin{bmatrix} G_{11} & G_{12} & G_{13} & G_{14} \\ G_{21} & G_{22} & G_{23} & G_{24} \\ G_{31} & G_{32} & G_{33} & G_{34} \\ G_{41} & G_{42} & G_{43} & G_{44} \end{bmatrix}$$

Next, each diagonal submatrix, G_{ii}, is decomposed into the product of two triangular matrices, such that

(4) $t_i t_i' = G_{ii}.$

This procedure orthogenalizes the variables within each set included in the analysis.

(5) $R_{ij} = t_i^{-1} G_{ij} t_j'^{-1}$

Using equation (5), another supermatrix, R, is created representing the correlations among the now orthogonalized variables. As such, the diagonal submatrices of R become identity matrices:

$$\begin{bmatrix} I_1 & R_{12} & R_{13} & R_{14} \\ R_{21} & I_2 & R_{23} & R_{24} \\ R_{31} & R_{32} & I_3 & R_{34} \\ R_{41} & R_{42} & R_{43} & I_4 \end{bmatrix}$$

Subtracting an identity matrix of order 32 (since there are a total of 32 variables in this analysis) yields the supermatrix $_1P$.

$$\begin{bmatrix} 0 & {}_1P_{12} & {}_1P_{13} & {}_1P_{14} \\ {}_1P_{21} & 0 & {}_1P_{23} & {}_1P_{24} \\ {}_1P_{31} & {}_1P_{31} & 0 & {}_1P_{34} \\ {}_1P_{41} & {}_1P_{42} & {}_1P_{43} & 0 \end{bmatrix}$$

The objective of m-set canonical analysis is to derive a linear combination of the variables for each of the four sets such that the function

(6) $\phi_1 = 1' \, {}_1\rho 1 - m$

is maximized. 1 is a unit vector of m elements; ρ is the matrix of canonical correlations; and m is the number of sets in the analysis. The solution for this problem is:

(7) $_1P \, D_{\beta.1} \, 1 = D_{\beta.1} \, {}_1\lambda$

where $D_{\beta.1}$ is a supervector of length 1, representing canonical weights; 1 is a unit vector; and

(8) $_1\lambda = {}_1\rho 1 - 1.$

Thus $_1\lambda$ represents the sum of the elements in each row of the canonical correlations matrix minus the main diagonal element.

It is interesting to note that the notation Horst chose to use in setting up Equation (7) resembles the classic eigenvalue-eigenvector solution so pervasive in traditional multivariate statistics. However, m-set canonical analysis does not rely on the eigenvalue-eigenvector solution. Instead of dealing directly with the variance in the data, Horst chooses to deal with it in an indirect manner, through his maximization of the sum of the elements in the correlation matrix. Horst claims that this procedure yields results analogous to Hotelling's for the 2-set case, and relies on an intuitive proof to extend his solution to m sets of variates. Whether this technique is truly appropriate remains a question for mathematical statisticians.

Equation (7) is used iteratively to reach the solution for the first canonical variate. Using a first approximation to the canonical weights of $1/\sqrt{n_i}$, where n_i is the number of variables included in the ith set, ensures that each $_iD_{\beta.1}$ is of unit length. Thus

(9) $_1P_1 D_{\beta.1} \, 1 = D_{B.1} \, 1$

where $D_{B.1}$ is a supervector of calculated canonical weights. Since

(10) $_1\rho = D_{\beta.1}' \, R D_{\beta.1},$

and

(11) $_2D_{\beta.1} = {}_1D_{B.1} \, {}_1D_\lambda^{-1},$

where

(12) $_1D_{1\lambda}^2 = {}_1D_{B.1}' \, {}_1D_{B.1} \, .$

$_1\lambda$ can be calculated from equation (8) after each iteration. The iterative process continues until $_1\lambda$ stabilizes to some specified degree of decimal accuracy.

To compute the second canonical variate, it is first necessary to derive the supermatrix $_2P$.

(13) $_2P = [I - D_{B.1} D_{B.1}'] \, {}_1P [I - D_{B.1} D_{B.1}'].$

This ensures that the second canonical variate will be orthogonal to the first one. Then the same iterative procedure is followed until $_2\lambda$ stabilizes.

Results

The results of the analysis are shown in Tables 2 through 7. Since the canonical correlations with the purchase measures were of such small magnitude for the second canonical variate (see Table 6), it was decided not to compute the remaining two possible canonical variates.

As can be seen in both Tables 3 and 4, the iterative procedure proposed by Horst did lead to a set of stable λ values. Also, as can be seen from the values of ϕ in the two tables, ϕ is at a maximum when the λ values stabilize.

Tables 5 and 6 show the canonical correlations among the four sets of measures. On the basis of the first set of correlations, only one of the hypotheses is supported - that both sets of beliefs will be more closely related to preference than to purchase. Both sets correlate in excess of .80 with preference, while correlating less than .40 with purchase, for the first canonical variate. The same relationship also appears for the second variate.

Contrary to the first hypothesis, the set of beliefs-only measures correlated more strongly (.92) with preference than did the set of B·I measures (.81), for the first canonical variate, and similarly for the second variate (.47 to .41). In addition, the two sets were approximately equally correlated with the set of behavioral measures. The third hypothesis was similarly rejected, since all three sets of predictor measures (beliefs, B·I scores, and preferences) correlated almost identically with the set of purchase measures on both canonical variates.

The calculated canonical weights are shown in Table 7. For the first variate, three beliefs appear to dominate in the relationship between both sets of beliefs and the criterion measures. "Ease of use," "Meal substitute," and "Delicious" exhibit the three largest weights in both belief sets. Of secondary importance are the beliefs "Snack," "Good buy," and "Real Flavor," again for both sets. This result would tend to suggest that either form of belief measure (B-only or B·I) will yield the same pattern of relationships with criterion variables, even though the magnitude of the relationships may be different.

For the preference set, affect exhibited the highest weight on the first variate, while the semantic differential was weighted highly on both of the variates. Although some of the explanatory power attributed to the semantic differential is artifactual due to the beliefs included in its construction, this artifact does not show up on the first variate. Rather, it seems to dominate the beliefs-preference relationship on the second variate. The three scales which overlap between the semantic differential and the belief measures have the highest weights for both belief sets, while the three weights which had been highest on the first variate have negative weights on the second variate. It is obvious that future research should in some part be directed at obtaining independent measures of preference.

Finally, the negative weights shown by the purchase measures on the first variate were not surprising, since these measures had previously exhibited an inverse relationship with all the other variables in the analysis (see Table 2). Fortunately, this inverse relationship is due only to the direction of the scales used to measure the phenomena. The purchase measure which is most closely related to the other sets of measures is "% of CIB/Total units instant breakfast purchased." This result is interesting in that it suggests that it may not be appropriate to examine a consumer's purchase behavior with respect to any one brand, but rather look at his purchase behavior relative to other brands in the product class. Thus, there may be many influences which determine the absolute amount of any product which will be purchased by a household, but beliefs and attitudes may be instrumental in determining which brand within that product class will be purchased by the consumer. The relatively small correlations between the purchase measures and the other three sets of measures are disappointing, but not unexpected. There are certainly many influences other than beliefs about the brand and liking for the brand which affect a consumer's purchase behavior.

As the results of this analysis suggest, it will be necessary in the future study of consumer behavior to look at the effects of more than one class of variables simultaneously. Multiple-set canonical analysis should be an invaluable tool in that task. The consumer researcher can include not only product-specific attitudes and beliefs as explanatory variables, but also more general influences on the consumer-social class, personality, group interaction, family roles, etc. - and examine the interactions of these variables in their influence on purchase behavior. This, then, will present a more comprehensive view of the attitude-behavior relationship.[7]

Conclusion

An attempt was made to apply multiple-set canonical analysis to the study of the attitude-behavior relationship in consumer psychology. The method is mathematically tractable and yields results which are unobtainable from traditional forms of multivariate analysis. Currently, the primary advantages of m-set analysis seem to be at the exploratory stages of a scientific investigation. In the future, as tests of significance are developed and the precise meaning of canonical weights is determined, m-set canonical analysis should prove to be an even more valuable tool in the continuing study of consumer psychology.

TABLE 1

Variable Set 1: Beliefs only - Equation (3)

CIB is easy to use
CIB is a good substitute for a meal
CIB is low in calories
CIB is delicious tasting
CIB is nutricious
CIB is a good snack
CIB is filling
CIB dissolves easily
CIB is a good energy source
CIB is a good buy for the money
CIB has a "real" (as opposed to artificial) flavor
CIB is a good source of protein

Variable Set 2: B·I (Belief x Importance) scores - Equation (2)

Ease of use
Meal substitute
Low in calories
Delicious
Nutricious
Snack
Filling
Dissolves easily
Energy source
Good Buy
Real flavor
Protein source

Variable Set 3: Preference Measures

Affect
Semantic Differential
Likert Scale

Variable Set 4: Purchase Measures

% CIB purchased of total units of instant breakfast purchased
Purchase - No Purchase (Dichotomous)
Repeat Purchase (Dichotomous)
Number of units of CIB purchased
Number of purchases of CIB

TABLE 3

Iterations for the First Canonical Variate

Iteration	$_1\lambda_1$	$_2\lambda_1$	$_3\lambda_1$	$_4\lambda_1$	ϕ_1
1	1.940	1.666	1.882	0.710	5.498
2	2.114	1.958	2.083	0.917	7.072
3	2.153	2.000	2.091	0.974	7.218
4	2.155	2.007	2.086	1.000	7.248
5	2.155	2.010	2.080	1.011	7.256
6	2.154	2.010	2.076	1.019	7.259
7	2.154	2.011	2.076	1.020	7.261
8	2.154	2.012	2.075	1.021	7.262

TABLE 4

Iterations for the Second Canonical Variate

Iteration	$_1\lambda_1$	$_2\lambda_2$	$_3\lambda_2$	$_4\lambda_2$	ϕ_2
1	1.004	0.995	0.563	0.163	2.725
2	1.156	1.152	0.792	0.228	3.328
3	1.239	1.221	0.913	0.261	3.634
4	1.280	1.254	0.958	0.284	3.776
5	1.298	1.270	0.977	0.295	3.860
6	1.308	1.276	0.982	0.304	3.870
7	1.313	1.278	0.984	0.309	3.884
8	1.316	1.279	0.984	0.313	3.892
9	1.317	1.281	0.983	0.316	3.897
10	1.319	1.280	0.982	0.319	3.900
11	1.320	1.280	0.981	0.321	3.902

TABLE 2

SUPERMATRIX G OF CORRELATIONS

(Decimal points omitted)

100																															
09	100																														
15	12	100																													
41	35	24	100																												
03	35	-05	09	100																											
27	19	14	48	11	100																										
28	33	20	47	20	32	100																									
29	22	07	23	17	20	20	100																								
27	38	18	45	20	35	50	21	100																							
19	40	22	50	17	33	43	15	46	100																						
36	32	20	68	08	44	45	20	46	48	100																					
33	32	20	46	21	36	44	19	60	45	46	100																				
85	13	13	38	04	26	27	27	25	22	31	28	100																			
08	84	11	32	25	20	30	21	33	35	29	25	17	100																		
09	19	67	16	04	06	20	08	21	17	15	22	12	22	100																	
29	32	14	83	09	43	41	17	35	43	55	39	34	33	12	100																
06	34	-04	11	82	17	20	20	23	15	10	26	10	29	12	18	100															
22	14	09	38	06	84	27	19	22	21	34	25	25	18	02	40	13	100														
21	35	15	41	18	26	81	16	42	38	40	34	26	38	21	40	25	27	100													
21	19	00	18	10	20	17	80	19	14	19	19	24	19	05	20	18	24	18	100												
21	33	12	35	18	28	41	13	80	37	36	50	21	33	21	31	28	23	44	21	100											
17	33	15	40	16	30	36	14	40	74	36	40	19	30	16	42	22	23	39	21	39	100										
25	29	09	54	04	41	37	17	34	40	83	38	26	28	11	59	12	39	39	23	34	39	100									
25	31	12	37	18	29	38	16	50	38	40	81	25	31	26	34	36	23	39	21	58	44	42	100								
35	39	22	70	15	46	49	23	50	51	65	45	35	37	16	61	17	37	42	19	38	37	50	38	100							
30	38	28	67	13	69	49	18	51	81	75	52	30	35	21	57	15	53	42	18	40	59	63	43	66	100						
09	20	-04	23	18	34	22	16	21	22	31	24	11	21	03	29	21	38	21	20	28	26	35	27	30	34	100					
-11	-24	-06	-23	-13	-16	-13	-05	-16	-23	-19	-13	-16	-23	-07	-21	-11	-11	-15	-04	-16	-16	-15	-13	-33	-24	-16	100				
-11	-26	-05	-24	-15	-16	-14	-04	-17	-24	-21	-14	-17	-25	-05	-22	-15	-11	-17	-05	-17	-17	-17	-14	-35	-24	-15	96	100			
-06	-27	-03	-19	-15	-10	-07	-07	-09	-18	-15	-11	-14	-26	-05	-20	-16	-07	-12	-07	-09	-12	-14	-11	-29	-16	-12	68	72	100		
-13	-30	-13	-25	-15	-20	-17	-08	-15	-21	-20	-15	-20	-28	-09	-23	-14	-16	-18	-07	-13	-14	-17	-13	-32	-22	-16	59	62	72	100	
-12	-28	-08	-23	-14	-17	-15	-08	-13	-20	-20	-14	-18	-26	-05	-22	-12	-14	-17	-07	-12	-14	-17	-12	-31	-20	-16	60	64	75	94	100

TABLE 5

Canonical Correlations - First Variate

1.000	.879	.924	.351
.879	1.000	.807	.326
.924	.807	1.000	.344
.351	.326	.344	1.000

TABLE 6

Canonical Correlations - Second Variate

1.000	.751	.473	.096
.751	1.000	.406	.123
.473	.406	1.000	.102
.096	.123	.102	1.000

TABLE 7

Canonical Weights for the First Two Canonical Variates

Variable Set	Canonical Weights	
Beliefs Only	$D_{B.1}$	$D_{B.2}$
Ease of Use	0.87	-0.22
Meal Substitute	1.19	-0.51
Low in Calories	0.34	0.33
Delicious	1.13	-0.31
Nutricious	0.15	-0.13
Snack	0.72	0.89
Filling	0.25	0.03
Solubility	0.14	-0.21
Energy Source	0.17	0.03
Good Buy	0.52	0.58
"Real" Flavor	0.37	0.22
Protein Source	-0.05	-0.01
B·I Scores		
Ease of Use	0.95	-0.25
Meal Substitute	1.05	-0.41
Low in Calories	0.19	0.26
Delicious	1.07	-0.27
Nutricious	0.13	-0.18
Snack	0.59	0.86
Filling	0.27	0.01
Solubility	-0.11	-0.19
Energy Source	0.30	0.08
Good Buy	0.44	0.61
"Real" Flavor	0.38	0.27
Protein Source	-0.01	0.04
Preference Measures		
Affect	1.78	-0.50
Semantic Differential	1.05	0.84
Likert Scale	0.23	0.05
Purchase Measures		
% CIB/Total units of Instant Breakfast	-0.85	0.01
Purchase - No Purchase	-0.29	0.15
Repeat Purchase	-0.07	0.24
#Units CIB purchased	-0.46	-0.13
#Purchases CIB	0.08	0.06

FOOTNOTES

[1]The author is both indebted and grateful to Professor Jagdish N. Sheth of the University of Illinois for providing the opportunity and impetus for this paper. The research reported here is one of a series of studies which has been conducted under the guidance of Professor Sheth, and as such, is closely related to much of his work in the attitude-behavior area.

[2]See Cohen and Houston (1970). Fishbein, working from a behavioral orientation, would probably take issue with the nomenclature used here to classify his theory. Cohen and Ahtola (1971) provide a rationale for including Fishbein's model under an expectancy-value framework.

[3]It should be noted that Sheth's formulation did not actually include the importance component since he had concluded from earlier studies that the importance component offered little predictive power (Howard and Sheth, 1969; Sheth, 1969). Nevertheless, the importance component is included here for the sake of comparison with Equation (1).

[4]The author expresses his appreciation to Professors John A. Howard of Columbia University and Jagdish N. Sheth for the use of the Columbia panel data.

[5]For a more detailed treatment of the construction of the semantic differential and Likert scale used in this analysis, the reader is referred to a paper currently in preparation by the author and Professor Jagdish N. Sheth, "A Multimode Investigation of the Attitude-Behavior Relationship."

[6]The author expresses his thanks to Professor Charles Lewis of the Psychology Department of the University of Illinois for his invaluable insights into the analysis proposed by Horst. While Professor Lewis was instrumental in enabling the analysis to be carried out, the author assumes full responsibility for any errors or shortcomings in this research.

[7]The author is grateful to Professors Joel B. Cohen and Paul E. Green for their comments on an earlier version of this paper.

REFERENCES

1. Adler, Lee and Irving Crespi, eds. (1968), Attitude Research on the Rocks (Chicago: American Marketing Association).

2. Bass, Frank M. and W. Wayne Talarzyk (1969), "A Study of Attitude Theory and Brand Preference," in McDonald, Philip R. (ed.), Marketing Involvement in Society and the Economy: 1969 Fall Conference Proceedings (Chicago: American Marketing Association), pp. 272-279.

3. Cohen, Joel B. and Olli T. Ahtola (1971), "An Expectancy x Value Analysis of the Relationship Between Consumer Attitudes and Behavior," paper presented at the Annual Meetings of the Association for Consumer Research, University of Maryland, September.

4. Cohen, Joel B. and Michael J. Houston (1970), "The Structure of Consumer Attitudes: The Use of Attribute Possession and Importance Scores," revised version of paper presented at the Attitude Research and Consumer Behavior Workshop, Urbana-Champaign, Illinois, December.

5. Cooley, William W. and Paul R. Lohnes (1971), Multivariate Data Analysis (New York: John Wiley & Sons, Inc.), pp. 168-193.

6. Fishbein, Martin, ed. (1967), Readings in Attitude Theory and Measurement (New York: John Wiley and Sons, Inc.).

7. Green, Paul E., Michael H. Halpert and Patrick J. Robinson (1966), "Canonical Analysis: An Exposition and Illustrative Application," Journal of Marketing Research, Vol. 3, No. 1, Feb., pp. 32-39.

8. Green, Paul E. and Donald S. Tull (1966), Research for Marketing Decisions (Englewood Cliffs, N.J.: Prentice-Hall, Inc.), pp. 357-64.

9. Hansen, Flemming (1968), "An Experimental Test of a Consumer Choice Model," in King, Robert L. (ed.), Marketing and the New Science of Planning: 1968 Fall Conference (Chicago: American Marketing Association), pp. 393-400.

10. Hansen, Flemming (1969), "Consumer Choice Behavior: An Experimental Approach," Journal of Marketing Research, Vol. 6, No. 4, Nov., pp. 436-43.

11. Hansen, Flemming and Thomas Bolland (1971), "The Relationship Between Cognitive Models of Choice and Non-metric Multidimensional Scaling," paper presented at the Annual Meetings of the Association for Consumer Research, University of Maryland, September.

12. Horst, Paul (1961), "Relations Among m Sets of Measures," Psychometrika, Vol. 26, No. 2, June, pp. 129-49.

13. Howard, John A. and Jagdish N. Sheth (1969), The Theory of Buyer Behavior (New York: John Wiley and Sons, Inc.).

14. Hughes, G. David (1970), "Distinguishing Salience and Valence," paper presented at the Attitude Research and Consumer Behavior Workshop, Urbana-Champaign, Illinois, December.

15. Kotler, Philip (1967), Marketing Management: Analysis, Planning, and Control (Englewood Cliffs, N.J., Prentice-Hall, Inc.).

16. Lavidge, Robert J. and Gary A. Steiner (1961), "A Model for Predictive Measurements of Advertising Effectiveness," Journal of Marketing, Vol. 25, No. 4, Oct., pp. 59-62.

17. Lutz, Richard J. (1971), "A Comparative Analysis of Two Conceptual Models of Behavioral Intention," unpublished paper, University of Illinois, June.

18. Moinpour, Reza and Douglas L. MacLachlan (1971), "The Relations Among Attribute and Importance Components of Rosenberg-Fishbein Type Attitude Model: An Empirical Investigation," paper presented at the Annual Meetings of the Association for Consumer Research, University of Maryland, September.

19. Nicosia, Francesco (1966), Consumer Decision Processes (Englewood Cliffs, N.J.: Prentice-Hall, Inc.)

20. Osgood, Charles, George J. Suci, and Percy H. Tannenbaum (1957), The Measurement of Meaning (Champaign, Illinois: University of Illinois Press).

21. Rosenberg, Milton J. (1956), "Cognitive Structure and Attitudinal Affect," Journal of Abnormal and Social Psychology, Vol. 53, pp. 367-72.

22. Sheth, Jagdish N. (1967), "A Review of Buyer Behavior," Management Science, Vol. 13, No. 12, Aug., pp. 718-56.

23. Sheth, Jagdish N. (1969), "Attitude as a Function of Evaluative Beliefs," paper presented at the American Marketing Association Workshop in Consumer Behavior, Columbus, Ohio August.

24. Sheth, Jagdish N. (1970), "An Investigation of Relationships Among Evaluative Beliefs, Affect, Behavioral Intention and Behavior," working paper, University of Illinois, April.

25. Sheth, Jagdish N. (1971), "Canonical Analysis of Attitude-Behavior Relationship," paper presented at The Institute of Management Science XVIII International Meeting, Washington, D.C., March.

26. Sheth, Jagdish N. and W. Wayne Talarzyk (1970), "Relative Contribution of Perceived Instrumentality and Value Importance Components in Determining Attitudes," paper presented at the American Marketing Association, 1970 Fall Conference, Boston, Mass., August.

A MULTI-BRAND MULTI-SEGMENT MODEL OF CONSUMER BEHAVIOR: ON ESTIMATING THE PARAMETERS

John C. Bieda, Procter & Gamble

Introduction

The purpose of this paper is to report on the findings from an empirical study of the Minimum Chi Square estimation procedure when this estimation procedure is used for estimating the parameters of a heterogeneous linear learning model in the context of consumer purchasing behavior.

In 1965, Massy [1] proposed that the Minimum Chi Square estimation procedure be used to estimate the parameters of the homogeneous two operator linear learning model. Massy was able to show that the expected proportion of families that would purchase a given sequence of brands is a function of the parameters of the two operator model and the raw moments of the initial distribution of the probabilities of purchasing the two brands.

In addition, Massy suggested using data from continuous purchase diary panels to obtain the observed proportion of families that purchased a particular sequence of brands. These are the two pieces of data that are necessary for the Minimum Chi Square estimation procedure as can be seen by formula

$$X^2 = M \sum_{i=1}^{N} \frac{(P_i - \pi_i(\delta))^2}{\pi_i(\delta)} \quad .$$

where M is the sample size for the panel; N is the length of the purchase string; P_i is the observed proportion of families that have purchased string i; π_i is the expected proportion of families that will purchase string i; and δ is the vector of parameters which included the parameters of the linear learning model and the parameters of the initial distribution of purchasing the brands.

Massy used Cramer's proof, that when the above expression is at a minimum the estimated parameters are asymptotically equivalent to maximum likelihood estimates, to justify using this procedure. Massy also pointed out that when the above expression is at the minimum the resulting Chi Square value can be used to test the fit of the model against the null hypothesis that the model does fit the observed data.

In a forthcoming doctoral dissertation, Bieda has expanded the two operator linear learning model to allow for heterogeneity among the population with respect to the model parameters. In a manner similar to that used by Massy, he has been able to show that the expected proportion of families that will purchase a particular sequence of brands is a function of the raw moments of the initial distribution of purchasing the brands, the weights associated with segments of the population that are assumed to be homogeneous with respect to a particular two operator linear learning model and the parameters of the learning models associated with each segment. Thus, Bieda was able to adopt Massy's estimation procedure.

With this background in mind, we now shall take a closer look at the estimation procedure. To minimize the above expression, we would ordinarily take the partial derivatives with respect to each of the parameters, set them equal to zero and solve the resulting set of equations. However, for both the homogeneous and heterogeneous models, the partical derivatives are highly nonlinear and very complicated. Thus, there is a need to resort to some type of non-linear programming algorithm to solve for the minimum.

The algorithm that has been used for the homogeneous two operator in several studies that have been reported in the literature is Pattern Search. Time does not permit a discussion of this nonlinear programming procedure; instead, I will simply refer you to an excellent discussion of it in Wilde's book, Optimum Seeking Methods. [2]

The fact that we have to use a non-linear programming procedure to minimize the Chi Square expression is the basis for the present study. It would seem that unless one can arrive at the true minimum - 1) one has no idea of what statistical properties the estimates have and 2) using the Chi Square value obtained to test the fit of the model would be incorrect since Cramer's proof requires that the Chi Square expression be at the minimum for the observed set of data.

In this study our primary research question then deals with the ability of the non-linear programming, Pattern Search, to arrive at the true minimum.

The Experiment

The design of the experiment for examining this question was factorial in nature. Four formulations of the heterogeneous two operator learning model were examined. These were 1) a Two Beta Equal Lambda Model, 2) a One Beta Equal Lambda Model, 3) a Two Beta, Unequal Lambda Model and 4) a One Beta Unequal Lambda Model.

Here the unequal and equal lambdas refer to whether or not the slopes of the operators are unequal or equal and the one and two betas refer to whether the initial distribution of probabilities is specified as a single beta distribution or a sum of weighted beta distributions.

The procedure used was as follows: 1) For a given formulation a set of parameters was arbitrarily chosen. 2) These parameters were used to obtain a set of exact proportions for purchase strings of length five using the model. 3) The exact proportions were submitted to the non-linear programming algorithm as if they were observed proportions from a consumer panel with another set of parameters used as starting values. 4) The non-linear programming algorithm was allowed to run until there was for all practical purposes no further improvement possible.

For each formulation of the model a set of parameters was chosen for a 1, 2 and 3 segment market. Thus, in all, twelve sets of exact proportions were generated.

Before going into the results, a word needs to be said about the starting values. Pattern Search, like many non-linear algorithms, must be given an initial vector of feasible starting parameters. We used the following procedure to select the starting values. Regardless of how many segments were actually used to generate the exact proportions, the exact proportions were first submitted to a one segment version of the model. The ending parameters were then duplicated, the segment weight cut in half and these values served as the starting values for a two segment version of the same model. If the exact data was generated with a three segment model the ending two segment parameters served as the starting values for the three segment version. Here, however, we took the larger of the two segments duplicated the parameters and split the larger segment weight in half.

In several cases, an alternate procedure was also employed. Here we simply duplicated our initial starting values for the one segment version for say exact data generated for a two segment market, split the segment weight in half and used these values as starting values. We will have more to say about the starting values in the discussion section.

Results

In Table 1, we have presented the Chi Square values that were obtained using the stepwise starting value procedure and those obtained for the selected cases where the second set of starting values were used.

For the equal lambda model the observed Chi Square values were generally quite low when the stepwise starting values were used.

For the two beta formulation the one and two segments observed χ^2's were $.55 \times 10^{-5}$ and $.17 \times 10^{-4}$ respectively. For the one beta formulation the one and three segment cases they were $.49 \times 10^{-7}$ and $.26 \times 10^{-5}$ respectively. The two exceptions to the low observed Chi Square value for the equal lambda formulation were for the two beta 3 segment and one beta 2 segment cases with the values of $.30 \times 10^{-1}$ and $.11 \times 10^{-3}$. However, with the alternative set of starting values both of the observed values dropped to levels comparable with the others $.87 \times 10^{-4}$ for the two beta 3 segment cases and $.29 \times 10^{-4}$ for the one beta 2 segment formulation.

With the unequal lambda model the results are somewhat different. Here we see that, for both the two beta and the one beta formulations, as the number of segments increases, the observed Chi Square steadily decreases. Although not shown in the table using alternative starting values did not improve any of the values.

In Tables 2 through 4, we have presented the estimated parameters for the twelve sets of exact data. Under each of the estimated parameters, in parentheses, is the exact parameter being estimated.

Looking first at Table 2. Here it can be seen that for the Two Beta Equal Lambda Model the estimated parameters are very close to the exact parameters in almost every case. The exception involves the α_{11} and λ for the three segment market. In addition, the beta parameters, while estimating the mean of the initial distribution quite closely, are not estimating the second or higher raw moments very well.

Turning now to Table 3. The estimates for the two brand one beta equal lambda formulation are almost precisely the exact parameters for the one and two segment markets. The beta parameters for the one segment market are also very close, however, for the two segment market they are underestimated. For the three segment market the model parameters are generally close with the exception of the segment weights. The beta parameters, however, are badly overestimated.

In Table 4, it can be seen here that as more segments are added the estimated parameters are closer to the exact parameters. In the three segment case the beta parameters are very close to those used to generate the data, the model parameters in almost every case are also quite close to the exact parameters.

In Table 5, we have a virtual repeat of Table 4. As more segments are added the estimated parameters are closer to the exact parameters. The exception here is in the estimates of the beta parameter. The estimates for the two segment market are closer to the exact beta parameters than the estimates for the three segment market.

Discussion

It will be noted that in these experiments we did not introduce any error into the exact proportion generated from the known set of parameters. The reason for this is quite simple. We are trying to determine if for a given set of observed proportions we can move to the global minimum using Pattern Search and the model that generated the proportions. If we can, then we know from Crámer's proof that the estimated parameters are equivalent to the maximum likelihood estimates.

The results are somewhat mixed in the equal lambda case, the stepwise starting procedure generally provided very low observed values. However, using this procedure a local minima was definitely encountered in the two beta 3 segment case.

For the unequal lambda case a local minima was encountered in both the one and two segment cases. Even when an alternative set of starting values was used other local minima were encountered. Only in the case of the three segment market were the results reasonably close.

In general, the estimates of the beta parameters, for both the equal and unequal formulations, were quite poor.

It should be noted that in no case did the Chi Square value ever reach zero. For these particular experiments, it should have been possible to reach zero since the exact proportions were carried to sixteen decimal places. One explanation besides local minima for why smaller values were not observed lies in round off error in the calculations; another is that the step size in the search routine was not allowed to decrease to a small enough value to obtain the desired accuracy.

Conclusion

Generally speaking, we were able to get quite close to the global minima. However, in several cases the observed Chi Square values were sufficiently far enough away from the known true minima, 0.0, to cast some doubt on the ability of the search routine to reach the global minima in all instances. Particularly disturbing was the fact that when an alternative set of parameters was used for the unequal lambda formation, no improvement was found in the observed Chi Square values.

The general conclusion seems to be that a number of starting values should be used even though one is not guaranteed of reaching the global minima using Pattern Search for estimating the parameters of a multi-brand multi-segment linear learning model.

REFERENCES

(1) Massy, William F., "Estimation of Parameters for the Linear Learning Model," Working Paper #78, 1965.

(2) Wilde, D. J., Optimum Seeking Methods, Prentice Hall, Inc. 1964.

TABLE 1

Observed Chi Square Value for Stepwise and Selected Second Set Starting Values

Date Generated With		Observed X^2 Values Stepwise Starting Values Used	X^2 Values, Second Set of Starting Values Used
2 Beta, Equal	1 Segment	.005528 X 10^{-3}	
	2 Segment	.017580 X 10^{-3}	
	3 Segment	30.035 X 10^{-3}	.08711 X 10^{-3}
1 Beta, Equal	1 Segment	.000049 X 10^{-3}	
	2 Segment	.1076 X 10^{-3}	.02871 X 10^{-3}
	3 Segment	.00261 X 10^{-3}	
2 Beta Unequal	1 Segment	4.397 X 10^{-3}	
	2 Segment	.2880 X 10^{-3}	
	3 Segment	.0159 X 10^{-3}	
1 Beta Unequal	1 Segment	4.225 X 10^{-3}	
	2 Segment	.5668 X 10^{-3}	
	3 Segment	.004082 X 10^{-3}	

TABLE 2

Estimated and Exact Parameters Two Brand Two Beta Equal Lambda Formulation

Exact Parameters Being Estimated in Parentheses

	One Segment Market	Two Segment Market		Three Segment Market		
	Seg. 1	Seg. 1	Seg. 2	Seg. 1	Seg. 2	Seg. 3
α_{11}	0.3145	0.364	0.363	0.30	0.21	0.51
	(0.3145)	(0.3145)	(.382)	(0.382)	(0.3145)	(0.4468)
λ	0.644	0.621	0.621	0.69	0.79	0.47
	(0.644)	(0.644)	(0.618)	(0.618)	(0.644)	(0.54)
α_{12}	0.004	0.19	0.18	0.0007	0.04	0.03
	(0.004)	(0.004)	(0.029)	(0.029)	(0.035)	(0.0282)
Segment Weight		0.41	0.59	0.31	0.12	0.57
		(0.40)	(0.60)	(0.30)	(0.20)	(0.50)
1st Parameter 1st Beta Distribution	4.18	11.87		2.36		
	(22.50)	(22.50)		(22.50)		
2nd Parameter 1st Beta Distribution	1.26	2.16		2.40		
	(6.04)	(6.04)		(6.04)		
Weight 1st Beta Distribution	0.09	0.07		0.27		
	(0.10)	(0.10)		(0.10)		
1st Parameter 2nd Beta Distribution	3.97	4.99		7.80		
	(6.04)	(6.04)		(6.04)		
2nd Parameter 2nd Beta Distribution	14.16	17.13		34.64		
	(22.50)	(22.50)		(22.50)		

TABLE 3

Estimated and Exact Parameters For the
Two Brand One Beta Equal Lambda Formulation

Exact Parameters Being Estimated
In Parentheses

	One Segment Market	Two Segment Market		Three Segment Market		
	Seg. 1	Seg. 1	Seg. 2	Seg. 1	Seg. 2	Seg. 3
α_{11}	0.262 (0.262)	0.24 (0.262)	0.40 (0.41)	0.264 (0.262)	0.38 (0.3135)	0.40 (0.41)
λ	0.683 (0.683)	0.70 (0.683)	0.60 (0.59)	0.622 (0.683)	0.60 (0.557)	0.59 (0.59)
α_{12}	0.0089 (0.0089)	0.01 (0.0089)	0.03 (0.025)	0.0098 (0.0089)	0.024 (0.0165)	0.024 (0.025)
Segment Weight		0.38 (0.40)	0.64 (0.60)	0.44 (0.30)	0.48 (0.30)	0.08 (0.40)
1st Beta Parameter	0.90 (0.899)	4.18 (5.82)		8.42 (13.62)		
2nd Beta Parameter	8.55 (8.50)	8.18 (12.00)		38.50 (62.37)		

TABLE 4

Estimated and Exact Parameters for the Two Brand
Two Beta Unequal Lambda Formulation

Exact Parameters Being Estimated
in Parentheses

	One Segment Market	Two Segment Market		Three Segment Market		
	Seg. 1	Seg. 1	Seg. 2	Seg. 1	Seg. 2	Seg. 3
α_{11}	0.22 (.32)	0.19 (.10)	0.18 (.32)	0.20 (.25)	0.28 (.32)	0.27 (.10)
λ_1	0.78 (.65)	0.79 (.85)	0.79 (.65)	0.61 (.50)	0.70 (.65)	0.70 (.85)
α_{12}	0.03 (.03)	0.05 (.05)	0.04 (.03)	0.02 (.10)	0.03 (.03)	0.03 (.05)
λ_2	0.86 (.80)	0.72 (.70)	0.71 (.80)	0.60 (.60)	0.79 (.80)	0.79 (.70)
Segment Weight		0.53 (.60)	0.47 (.40)	0.50 (.40)	0.30 (.40)	0.20 (.20)
1st Parameter 1st Beta Distribution	6.63 (10.50)	28.03 (10.50)		6.58 (10.50)		
2nd Parameter 1st Beta Distribution	4.40 (4.55)	12.89 (4.55)		3.43 (4.55)		
Weight 1st Beta Distribution	0.31 (.20)	0.21 (.20)		0.22 (.20)		
1st Parameter 2nd Beta Distribution	4.03 (7.35)	11.36 (7.35)		7.40 (7.35)		
2nd Parameter 2nd Beta Distribution	57.15 (59.20)	95.14 (59.20)		62.76 (59.20)		

TABLE 5

Estimated and Exact Parameters Two Brand
One Beta Unequal Lambda Formulation

Exact Parameters Being Estimated in Parentheses

	One Segment Market	Two Segment Market		Three Segment Market		
	Seg. 1	Seg. 1	Seg. 2	Seg. 1	Seg. 2	Seg. 3
α_{11}	0.24 (0.32)	0.19 (0.10)	0.19 (0.32)	0.14 (0.10)	0.25 (0.25)	0.31 (0.32)
λ_1	0.76 (0.65)	0.81 (0.85)	0.81 (0.65)	0.70 (0.85)	0.57 (0.50)	0.68 (0.65)
α_{12}	0.03 (0.03)	0.05 (0.05)	0.05 (0.03)	0.03 (0.05)	0.02 (0.01)	0.03 (0.03)
λ_2	0.87 (0.80)	0.69 (0.70)	0.69 (0.80)	0.79 (0.70)	0.59 (0.60)	0.82 (0.80)
Segment Weight		0.56 (0.60)	0.56 (0.40)	0.18 (0.20)	0.51 (0.40)	0.31 (0.40)
1st Beta Parameter	1.41 (6.04)	5.70 (6.04)		10.27 (6.04)		
2nd Beta Parameter	5.09 (22.05)	21.19 (22.05)		38.28 (22.05)		

TOWARD THE MEASUREMENT OF SOCIAL INDICATORS: CONCEPTUAL AND METHODOLOGICAL IMPLICATIONS*

Ralph M. Brooks
Leslie D. Wilcox
George M. Beal
Gerald E. Klonglan
Iowa State University

INTRODUCTION

The topic of this paper is concerned with the development of a system of measurable social indicators. The widespread interest in social indicators represents a shift in information premises for decision-making in the United States and has come about as a result of a need for more reliable data of transeconomic issues, quality of life, social problems and planned social development. To date, however, the discussion of social indicators has focused more on its potential uses rather than specifying the steps necessary for the development of social indicators. The rapid build-up of interest in social indicator research has produced a rather massive body of literature relative to this topic over the past 5 years (Beal et al., 1971a). As a result, there is no general concensus regarding the nature and definition of social indicators, how social indicators are to be developed and how they are to be used. The objective of this paper is to deal with some of these issues and to attempt to suggest a perspective to provide an adequate definition of social indicators and a strategy for the development of a taxonomy of social indicators for future monitoring of societal conditions.

SOCIAL INDICATOR PERSPECTIVES

The failure to develop common perspectives concerning some of the basic issues to be overcome in the development of social indicators has meant that many of the current discussions surrounding this topic must be viewed as apologies for, or criticisms, of the social indicator movement (Beal et al., 1971b). There has, however, been considerable evidence of maturing of the movement over the past 2 years, with significant efforts being made to cope more systematically with some of these basic issues. Through such efforts, several more clearly defined perspectives and orientations to social indicator research seem to be emerging. The crystallization of these differing perspectives suggest the early stages in the development of "schools" or "persuasions" of social indicators. These perspectives, we believe, reflect the unique interests and needs that underlie individual motivation to obtain better social information.

Since the motivation behind the social indicator movement has generally been the desire to generate usable data, the perspectives and orientation to social indicator research adopted by individuals tend to reflect the unique role each writer visualizes that social indicators will fill in social planning, social development of in the social sciences. These differing perspectives are usually built on quite different definitions of the term, social indicator, and suggest quite different strategies for social indicator development.

Among the various perspectives that seem to be forming within the current social indicator movement, four are especially worthy of brief mention (Wilcox et al., 1971). Perhaps the most common perspective one encounters in current social indicator research is the orientation that regards social indicators as instruments for detecting changes in the "quality of life" of individuals, groups or societies. The strategy of research suggested by this perspective focuses upon the problem of defining "quality of life" and the establishment of quantifiable categories to measure variations in crucial social components of human life conditions. The problems posed by this perspective are, perhaps, the most difficult to quantify and raise issues that cannot be dissociated from normative interests. Current research efforts that reflect this orientation include the work of Becker and de Brigard (1970), Harland (1971), and Jones and Flax (1970).

A second perspective tends to regard social indicators as instruments to monitor progress toward societal goals. This approach has often been suggested as an alternative to the quality-of-life emphasis in an effort to reduce the normative implications inherent in the term quality of life. The problem of establishing generally agreed upon and clearly defined sets of goals, however, has proved highly elusive. One specialized application of the goals approach focuses on program evaluation, in which the goals are largely established by those concerned with the direction of the program. Much of the work of the federal government is reflective of this perspective including the National Goals Research Staff (1970), HEW's work on the preparation of an annual social report, Toward a Social Report (U.S. Department of Health, Education, and Welfare, 1969), and Vestermark's (1968) efforts to develop indicators of social vulnerability.

Another perspective that seems less common but still an integral part of the indicator movement tends to view social indicators primarily as social statistics. The thrust of this type of research focuses on an attempt to assess various aspects of social life by reporting statistical series that reflect change in these social components through time. Rarely does one find any serious effort to show cause, effect and interrelationship between variables or to include such statistics in a larger theoretical system. To gain a clearer picture of this perspective, the reader may find it helpful to examine the work of Tunstall (1970), Agency for International Development (1971), Drewnowski (1966, 1970) and some of the statistical data developed by the United Nations Research Institute (1961, 1966a, 1966b, 1966c, 1969, 1970).

---

*Journal Paper No. J-7037 of the Iowa Agriculture and Home Economics Experiment Station, Ames, Iowa. Project No. 1837.

A fourth perspective, which tends to be more reflective of the work of social scientists, views social indicators as social statistics that measure changes in variables that are components in a social-systems model. Here, the concern is with the monitoring of systems performance and the cause, effect and interrelationship between variables in a social system and how these values change through time. For examples of this perspective, see: Land (1970, 1971), Warren (1970a, 1970b), Wilcox and Brooks (1971a), and Brooks (1971).

This fourth perspective, we believe, offers the most in terms of advancing the development of social indicators. This system, once developed, would show interrelationships between variables and the assessment of causal factors that are necessary in making effective policy decisions. It also minimizes the problem of developing indicators of expressed normative interests of narrow segments of society and refocuses our attention on the monitoring of actual performance of social systems and social groups more objectively. Several general systems models exist in the social sciences; few of them, however, have been explicated to a quantifiable level necessary for the monitoring of social change. Therefore, we believe that the initial step in developing a system of social indicators must focus on the problem of developing a taxonomy of social conditions related to a general model that can provide an explication of quantifiable categories.

TAXONOMIES OF SOCIAL INDICATORS

In the past 3 years, several social-science researchers have addressed themselves to the task of explicating a taxonomy of indicators for such abstract concepts as "quality of life" and the "general good." The attempts were exploratory, but optimistic, as they tried to explicate these higher-level concepts into lower-level indicators that could be eventually quantified.

Rossi (1971), from a social psychological perspective, sought to establish a conceptual scheme to review the component parts of the community. For Rossi (1970:77), social indicators should be based on a model of how social life "works"; they should be small in number and related to potential social policy. The model of "how social life works" will have to be generated because past models are not helpful for the current social indicator needs. That indicators ought to be related to potential social policy is a difficult objective to achieve since current social policy at the community level is not clearly defined.

Becker and de Brigard (1970) attempted to develop a taxonomy of community, based on action planning, with a goal orientation. To these researchers, "quality of life" represents society's overall objective, with the three subcategories of physical, social and economic representing basic societal environments. They suggest that lower-level elements of quality of life are education, housing, health, social services, economic development, public safety, transportation, culture, interpersonal communication, local government and natural resources.

A third attempt of recent years to develop a taxonomy of indicators is presented by the Stanford Research Institute (1969), Toward Master Social Indicators. Master social indicators may be viewed as highly abstract concepts, such as abundance, or intermediate abstractions, such as wealth, utilized in a heuristic model of major societal concern. Their model seeks to demonstrate how low-level concepts can be aggregated into master social indicators of two main elements, one relating to the individual and, the other, to the social system. The elements they chose for aggregating are the components specified in the HEW document of Toward a Social Report.

Each of these three strategies for developing a taxonomy of indicators has started with an optimistic attempt to assess overall quality of life at some macro level. Although two of the studies related their taxonomies to the community, selecting quality of life as the general goal is viewed as macro and presents problems in explication and future analysis. All three strategies have indicated the frustations in attempting to generate a taxonomy to measure the complexity of social life. Yet, all might agree that the current level of social indicator sophistication is at the threshold of what must ultimately be accomplished if useful information is to be provided for future decision-making. The proposed task is difficult and well recognized as such by Hagen (1962:4) who states: As judged by the history of the physical, biological, and social sciences, study in any field is apt to begin with a none-too-ordered description--followed by a cataloging on bases that seem to make sense. As understanding grows, the systems of classification become more closely related to the functioning of interacting elements. Gradually, generalizations about functioning are reached which are useful in predicting future events. As the generalizations gain rigor, they take the form of analytical models of the behavior of the elements being studied. They take the form, that is, of systems.

The three studies discussed thus far have demonstrated the none-too-ordered description of generating taxonomies of social indicators. As yet, the current status of social indicators lacks this rigor and certainly has not acquired the model of the social system described earlier. This will take much concerted effort on the part of social scientists, and continuing to develop taxonomies at perhaps lower levels of abstraction and that are more complete seems a logical step in this larger task. These are lofty goals, and our present abilities to accomplish such tasks are somewhat inadequate. Yet, this challenge may prove to be one of the major contributions to the development of sociology as well as in providing societal guidance in the near future.

BASIC COMMUNITY MODEL

As previously indicated, trying to adapt studies using macro concepts, such as quality of life, to communities is highly complex and thus far has not proved very successful. Quality of life seems to be a relative term and can only be understood after a thorough examination of the empirical referent in question. If one were to delineate the major functions performed in communities and seek to measure that performance, it might be possible to make some statement about that community's level of living or quality of life. What we wish to propose is to focus on the community as the unit of analysis, rather than the state or nation as is commonly selected; also rather than focusing on abstract goals such as the "general good" or "quality of life," we propose that we focus on the basic units and processes of the community system as the phenomena to be explicated and for which social indicators will be developed. To date most studies of community have emphasized economic variables and have not looked at the total community as it relates to the environment.

The definition of social indicator utilized in this paper requires such an indicator to be a component in a social system, collected over time and aggregated or disaggregated according to the specifications of the model (Land, 1970). Furthermore, these indicators must be readily combined measures of indicators from lower levels of abstraction that can be controlled to "show the partial deficits of given subgroups attributable to given causes" (Coleman, 1969:96). To achieve this task will require a broader model than those typically embraced by sociologists and indeed, social scientists. Perhaps, the theoretical model currently in use in sociology that most systematically attempts to relate human behavior and social organization to environment is the ecological model.

In contrast to other models of society, ecology includes more encompassing variables that are judged useful in developing multiple profiles of social and physical aspects of the community. For this reason, we believe that the contributions to the ecological models by Hawley (1950, 1969), Duncan (1961, 1964, 1969), and Duncan and Schnore (1969) might, with some adaptation, help us to achieve a general model of the community ecosystem for understanding and monitoring system performance. If social indicators are to be useful in monitoring the performance of this ecosystem, one obviously must specify the basic components in such a system. Perhaps one of the reasons that present social indicators have not been particularly useful is because there is no general model available capable of allowing a wider range of explanation from which appropriate social indicators can be explicated. A model is needed that is capable of showing the processes that take place and the implications they may have for the conditions of man's social life and the environment in which he lives. It appears that ecological models may come closer to monitoring the community system in this broader sense than do present sociological models of society that focus primarily on the internal social and psychological dynamics of social systems.

COMMUNITY ECOSYSTEM

The community ecosystem is composed of several elements similar to the ecological complex described by Duncan and Schnore (1969). The community ecosystem, however, is conceptualized at a lower level of abstraction than is the ecological complex and will demonstrate slight modifications. This community ecosystem is more than the traditional social-systems approach to the study of social phenomena. It is attempting to include all meaningful activities at the community level that impact individuals in the system. The four elements, which we believe can serve to describe important aspects of life conditions, are environment, population, social organization and culture. These are presented in Figure 1 and will be briefly defined before a partial explication of one of the elements to lower level indicators.

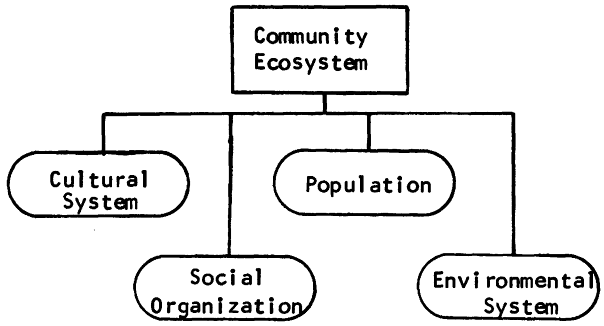


Figure 1. Basic model of the community ecosystem.

Environment

The environment, according to Hawley (1950: 12) "is a generic concept under which are subsumed all external forces and factors to which an organism is actually or potentially responsive." Populations have to exist in some form of natural environment, cope with this environment, and learn to adapt to its ever-changing conditions. In general, the environment sets limits to the size of population it can sustain. Man, with his technology, however, alters it sufficiently to allow for population growth.

Social Organization

Social organization is the social patterning that takes place in the population as individuals compete for limited resources to sustain life. These activities must be regular and systematic, regardless of the size of the social group. An essential component of organization is that smaller units come together to form larger units or wholes. According to Gould and Kolb (1965: 661): Social organization is a relatively stable set of functioning interrelations among component parts (persons or groups) which are not possible, by themselves in the components. Social organizations evolve as structures of such relations in such a way as to fulfill functions in a manner more efficient and durable than could be achieved by unorganized persons.

Population

In statistics, a population is defined as an aggregate of objects about which information is desired, but for which only a sample is selected for investigation. For social sciences, population generally refers to the number of inhabitants of a given territoriality and frequently is concerned with the characteristics of individuals. Population will therefore be concerned with more than demographic characteristics. Our major interest will be with developing multidimensional profiles of those individuals and subgroups within the community and not the personality system. This system of social and physical characteristics of individuals will be explicated, in part, into a taxonomy of lower-level indicators.

Cultural system

The cultural system consists of patterns of behavior transmitted by symbols, the traditional ideas and attached values that are considered interdependent within the given territoriality and systems of knowledge including technology. The cultural system is considered to be a very important component of the community ecosystem and is noted as a component in the model. Technology may be considered as one important subsystem of the cultural system that must be monitored because of the impact it will have on areas of social life.

Interrelationship of community ecosystem elements

These four elements then, make up or compose what we have termed the community ecosystem. The elements in the community ecosystem interact and are interrelated in much the same manner as are the elements in the ecological complex. Figure 2 includes the four elements, with hypothesized interrelationships. A basic assumption is that the ecosystem's purpose is to benefit the humans in that system.

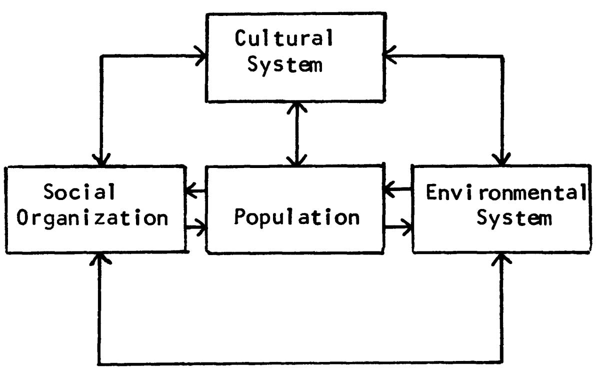


Figure 2. Interrelations between the elements of the community ecosystem.

The environment is taken as a given in the ecosystem. By itself, unaffected by humans, it experiences little change. By placing a population within the environment, however, the ecosystem begins to experience loss of resources and basic alterations as man begins to adapt to his surroundings. As man competes for resources, he soon learns that, by organization, he can more effectively utilize both human and physical resources in a process of adaptation to new situations. As social organization takes place, the environment becomes increasingly artificial, resulting in new social organization. This interaction between social organization, population and the environment takes place within the cultural system.

Social organization is also considered to be the mobilization of both human and physical resources for the delivery of services to the population within the community ecosystem. Therefore, one major interest might be in the impact of these services on that population. Vital questions might be: What services are available? Who has access to them? How are they being utilized? What are the effects of a changing environment on the population? To assess these questions will require a multidimensional profile of the individuals within the system. In other words, delineating a taxonomy of these four major elements might allow us to begin to make inferences regarding the various dimensions of quality of life and, at the same time, to develop measures for those dimensions.

The discussion of social organization as the mobilizing of resources could also be viewed as the input to the community, with the impact on the individuals within the system as the output. In other words, it may be possible to assess the net costs and benefits of the services and current social conditions to the individuals within the system.

Each of the four elements of the community ecosystem in Figure 2 could be explicated to lower-level indicators, which would allow the assessment of current social conditions within the community. In Figure 3, the four elements are again presented; however, each is explicated initially to one lower level of subindicators.

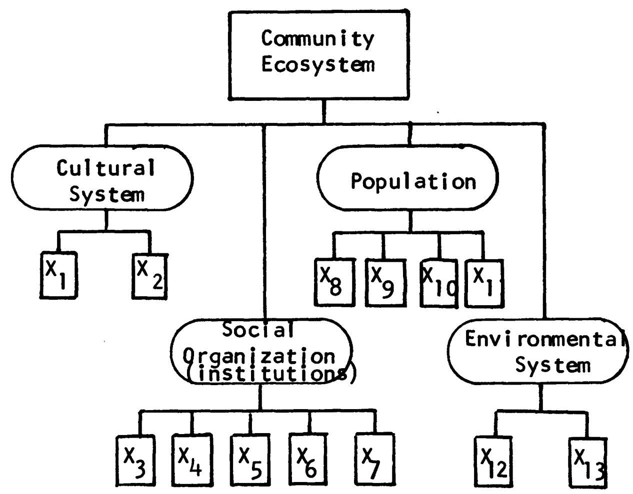


Figure 3. General taxonomy of the community ecosystem.

X_1= Values, X_2= Knowledge, X_3= Religion, X_4= Polity, X_5= Family, X_6= Economy, X_7= Education, X_8= Institutional and Social Patterns, X_9=

Physical Environmental Characteristics, X_{10}= Organic Characteristics, X_{11}= Cultural Esthetics, X_{12}= Social, X_{13}= Physical.

These are only some very general categories and are not necessarily comprehensive of all the subelements that may need to be included. In Figure 3 the cultural system contains the total symbolic system, of which two important subsystems of values, beliefs and ideologies included in the X_1 category and knowledge of which technology would be an important part included in the X_2 category. Social organization emphasizes an institutional approach to society and contains, at a minimum, the subelements polity, family, economic, religion and education. The element of population is explicated to four subelements of institutional and social patterns, physical environmental characteristics, organic characteristics and cultural esthetics. The environmental system is explicated to two subelements of social and physical and also is viewed as a major influence on other community-ecosystem components.

A complete explication of the subelements included in Figure 3 would indeed be a major task. This is not the objective of this paper, nor will it be claimed that the subelements that are explicated will, in fact, be complete. We have, however, attempted to continue this basic explication and present, in the Appendix, a more extensive discussion of one part of the model--population, along with supporting figures to demonstrate possible initial lower-level explications.

METHODOLOGICAL NEXT STEPS

Our objectives in this paper have been to suggest a perspective and definition of indicators, as well as a strategy for their development. The perspective thus far views social indicators as components in an ecological system, and we have dealt primarily with a general discussion of the community ecosystem. Before indicators can be developed, however, considerable investment must be made in research to determine how well this general model will allow the explication of social indicators that reflect the actual life conditions of persons living in a community. For this to be realized will require considerable efforts by all social scientists. To outline a more complete strategy of social indicators, it is necessary to consider additional steps to be utilized in the development of this general model.

What we are proposing is a 4- to 5-year plan of study designed to utilize this taxonomy in the process of inductive model building. The first year would be primarily devoted to a continuation of the explication of the taxonomy. The various components of the ecosystem model will be explicated to a quantifiable level, with the needed epistemic links between the various levels of the taxonomy. Before this taxonomy can be effective in measuring the life conditions of individuals in the community or the performance of that community, it will be necessary to obtain a complete enumeration of the important properties of that system at the empirical level. We do not believe that focusing on current quality of life or social problems in the development of social indicators can provide the information system needed for effective policy decisions because what is important to us today may not be of crucial concern in the future. Because these are potentially invisible problems, we believe the ecosystem approach has merit for it allows us to explicate a wider range of conditions related to the society, individual, culture and environment than would be possible in research efforts focusing on immediate normative concerns. We also believe that any meaningful measure of life conditions should reflect, in part, the perception of people living in a community; therefore, part of this first year's activity will include a field reconnaissance in which we will engage in extensive interviews with influentials, leaders and members of the community to gain an understanding of their perceptions of the community.

The second year of our plan of study will be engaged in the refinement of our taxonomy and the operationalization and development of measures of the low-level concepts. The refinement of the taxonomy will be done largely on the basis of our field reconnaissance wherein we will attempt to include the perception of the members of the community that we study.

To develop measures for our lower-level concepts, we propose to utilize existing techniques as much as possible, to make revisions in these measures where necessary and to develop new measures where none exist now. By focusing our study on existing measurement techniques, we believe that, in many instances, there will be existing data sources and data-collection procedures that can be utilized in this type of monitoring system. Our objective will be to suggest refinement in existing data-collection procedures and to suggest new procedures only where necessary.

At the end of the second year and the beginning of the third year, our plan is to attempt a field survey, primarily aimed at testing the validity and reliability of our measures, and to collect pilot data that can be utilized in the initial attempts to build inductively a systems model. The remainder of the third year will be devoted to a refinement of the taxonomy and measurement techniques where necessary and beginning the data analysis.

The data analysis during the third, fourth and fifth years will be largely aimed at an attempt to develop time series through replication studies, to utilize existing statistical techniques for combining lower-level indicators to provide higher-level indicators of greater theoretical value, and to develop controlled indicators wherever possible. And through the use of computer simulation we will attempt to establish interrelationships between a wide range of variables that will allow the development of models to assess social change.

Quite obviously this is an approach that will require the expenditure of considerable investments of time and energy before an information system can be developed that will allow better assessment of quality of life and current life conditions. We recognize this is a very ambitious undertaking, but also believe that, if social indicators are to be useful for policy decisions, we must make this investment and approach the task in a scientific manner.

BIBLIOGRAPHICAL ENTRIES

Agency for International Development. 1971. Summary Economic and Social Indicators 18 Latin American Countries: 1960-1970. Office of Development Programs, Bureau for Latin America: Agency for International Development. Washington, D.C.

Beal, George M., Ralph M. Brooks, Leslie D. Wilcox and Gerald E. Klonglan. 1971a. Social indicators: Bibliography I. Department of Sociology and Anthropology. Sociology Report No. 92 (January). Ames, Iowa: Iowa State University.

Beal, George M., Gerald E. Klonglan, Leslie D. Wilcox and Ralph M. Brooks. 1971b. "Social indicators and public policy: Toward an alternative approach." A paper presented at the Annual Meetings of the American Sociological Association (August 30-September 3) Denver, Colorado.

Becker, Harold S. and Raul De Brigard. 1970. Considerations on a Framework for Community Action Planning. Middletown, Connecticut: Institute for the Future, Working Paper WP-9.

Brooks, Ralph M. 1971. Social Indicators for Community Development: Theoretical and Methodological Considerations. Unpublished Ph.D. Dissertation. Iowa State University.

Coleman, James S. 1969. "The methods of sociology." Pp. 86-114 in Robert Bierstedt (ed.), A Design for Sociology: Scope, Objectives, and Methods. Monograph 9, The American Academy of Political and Social Sciences (April).

Drewnowski, Jan. 1966. Social and Economic Factors in Development - Introductory Considerations on Their Meaning, Measurement and Interdependence. Report No. 3. February. Geneva, Switzerland: United Nations Research Institute for Social Development.

Drewnowski, Jan. 1970. Studies in the Measurement of Levels of Living and Welfare. Report No. 70.3. Geneva, Switzerland: United Nations Research Institute for Social Development.

Duncan, Otis Dudley. 1961. "From social system to ecosystem." Sociological Inquiry (Spring): 140-149.

Duncan, Otis Dudley. 1964. "Social organization and the ecosystem." Pp. 36-82 in Robert E. L. Faris (ed.), Handbook of Modern Sociology. Chicago: Rand McNally and Company.

Duncan, Otis Dudley. 1969. "Human ecology and population studies." Pp. 678-716 in Philip M. Hauser and Otis Dudley Duncan (eds.), The Study of Population: An Inventory and Appraisal. Chicago: University of Chicago Press.

Duncan, Otis Dudley and Leo F. Schnore. 1969. "Cultural, behavioral, and ecological perspectives in the study of social organization." Pp. 70-85 in Walter L. Wallace (ed.), Sociological Theory. Chicago: Aldine Publishing Company.

Gould, Julius and William L. Kolb (eds.) 1965. A Dictionary of the Social Sciences. New York: The Free Press.

Hagen, Everett E. 1962. On the Theory of Social Change. Homewood, Illinois: The Dorsey Press.

Harland, Douglas G. 1971. Social Indicators: A Framework for Measuring Regional Social Disparities. Ottawa, Canada: Department of Regional Economic Expansion. (July 1, Second Draft).

Hawley, Amos H. 1950. Human Ecology. New York: The Ronald Press Company.

Hawley, Amos H. 1969. "Human ecology." Pp. 64-70 in Walter L. Wallace (ed.), Sociological Theory. Chicago: Aldine Publishing Company.

Jones, Martin V. and Michael J. Flax. 1970. The Quality of Life in Metropolitan Washington (D.C.). Washington, D.C.: The Urban Institute.

Land, Kenneth C. 1970. "Social Indicators." Pp. 121-173 in Robert B. Smith (ed.), Social Science Methods. New York: The Free Press.

Land, Kenneth C. 1971. "On the definition of social indicators." A paper presented at the Annual Meeting of the Population Association of America (April) Washington, D.C.

National Goals Research Staff. 1970. Toward Balanced Growth: Quantity with Quality. U.S. Government Printing Office (July 4) Washington, D.C.: Superintendent of Documents.

Rossi, Peter H. 1970. Community Social Indicators. The Johns Hopkins University Center for the Study of Social Organization of Schools Report No. 85.

Stanford Research Institute. 1969. Toward Master Social Indicators. Menlo Park, California: Educational Policy Research Center, Stanford Research Institute.

Tunstall, Daniel B. 1970. "Developing a social statistics publication." U.S. Office of Management and Budget. A paper presented at the Annual Meetings of the American Statistical Association (December 27-30) Detroit, Michigan.

United Nations Research Institute for Social Development. 1961. International Definition and Measurement of Levels of Living, An Interim Guide. Sales No. 61.IV.7.

United Nations Research Institute for Social Development. 1966a. Social and Economic Factors in Development, Report No. 3.

United Nations Research Institute for Social Development. 1966b. The Level of Living Index, Report No. 4.

United Nations Research Institute for Social Development. 1966c. Cost-Benefit Analysis of Social Projects. Report No. 7.

United Nations Research Institute for Social Development. 1969. Levels of Living and Economic Growth: A Comparative Study of Six Countries

1950-1965. Geneva, Switzerland.

United Nations Research Institute for Social Development. 1970. Contents and Measurement of Socio-Economic Development. Report No. 70.10. Geneva, Switzerland.

U.S. Department of Health, Education, and Welfare. 1969. Toward a Social Report. Washington, D.C.: U.S. Government Printing Office.

Vestermark, S. D., Jr. 1968. Indicators of Social Vulnerability. McLean, Virginia: Human Science Research, Inc. 285 pages and summary.

Warren, Roland. 1970a. "The good community--what would it be?" Journal of Community Development Society 1 (Spring):14-24.

Warren, Roland. 1970b. "Toward a non-utopian normative model of the community." American Sociological Review 35 (April): 219-228.

Wilcox, Leslie D. and Ralph M. Brooks. 1971a. "Social Indicators: An alternative approach for future research." A paper prepared for the Rural Sociological Society Annual Meeting.

Wilcox, Leslie D. and Ralph M. Brooks. 1971b. "Toward the development of social indicators for policy planning." A paper presented at the Ohio Valley Sociological Society.

Wilcox, Leslie D., Ralph M. Brooks, George M. Beal and Gerald E. Klonglan. 1971. "Social indicators: Recent trends and selected bibliography." Sociological Inquiry (In press).

APPENDIX

The community ecosystem is made up of 4 basic elements, population, environment, social organization and culture. This appendix focuses on an initial explication of one of these elements of the community ecosystem--population. This procedure is presented to demonstrate how the community ecosystem can be used in explicating lower level social indicators for future monitoring of social conditions.

Population

The term population, as used in this paper, is not concerned with human personality. Furthermore, population is not individual data. The concern with population for this research strategy is to develop social indicators to provide a quantitative profile of the social and physical characteristics of the total population of the community derived from aggregated individual data. These indicators will attempt to measure the existing social and physical conditions of that population and monitor the changes in these conditions through time. The interest in population includes the delivery of services that might be derived from other elements in this complex as well as the basic population characteristics that operate independent of those elements. The needed data must contain the total characteristics of the population and how it is altered and impacted by other elements in the complex, especially the social organizations' ability to deliver services to the individuals. Social indicators are to monitor existing conditions through time as experienced by individuals within the territoriality. Satisfaction or statements pertaining to the quality of life of the residents in a given territoriality are assumed to be derived by inferences from the data.

Population is aggregated individual data and is expected to play a vital role in understanding how effectively community services are being delivered to the individuals in the system. Aggregated data allows generalizing to other population groups, however, to assess the performance of the community will necessitate focusing on the question of disaggregation.

By disaggregating to subgroups in the community it would appear that the monitoring and awareness of community conditions would be more complete. Again seeking to monitor individual satisfaction and quality of life entangles one in monitoring normative type statements. All that indicators can be expected to do is monitor what the conditions are. Individual satisfaction and statements about the current quality of life must come from inferences based on disaggregation. For this reason, it is important to consider Coleman's category of combined conditions discussed earlier in the paper. But, before conditions can be combined for the purpose of inference, it will be necessary to know what the current conditions are. Indicators in the population element of the community ecosystem are measures of the social and physical characteristics that are generalized from an aggregate and are therefore aggregated data. It is recognized, however, that aggregating can tend to blur the impact of the system elements in terms of the individuals in the system. To overcome this "blurring," social indicators must be disaggregated to lower levels. Thus far, in the initial stage of this research strategy, it would seem imperative that the population within the community be disaggregated on the basis of age, sex, ethnicity (religion, national origin and race), place of residence in terms of geographical location, territoriality, and socio-economic conditions based on one of the common indexes of education, occupation and income. These still are basically concerned with aggregates, and it is quite possible that the operational measures developed for the subindicators in the taxonomy would reflect a more extensive disaggregation as the attempt is made to monitor change through time. It is hoped that this type of effort will allow assessment of the costs and benefits accrued to the individuals in the community system. An assessment of the population component of the community ecosystem is necessary and needed in order to understand the impact of the other components in the basic system.

<u>Population system indicators</u>

To understand what is meant by population, Figure 4 is presented with four major indicators of the population system. Each of these four will in turn be briefly defined to demonstrate how they are in fact different. This taxonomy is exploratory. To our knowledge such a task has never been attempted and although it is not complete, never-the-less, it will be illustrative of the next steps in this particular effort to monitor societal conditions.

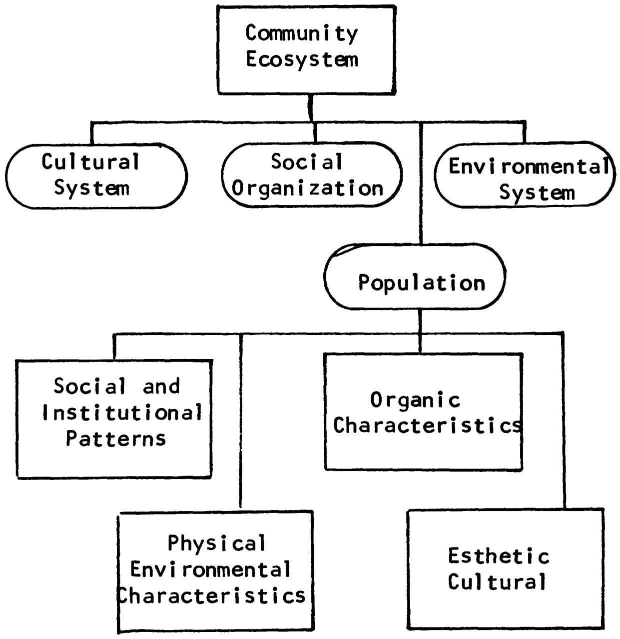


Figure 4. Initial taxonomy of the population component in the community ecosystem.

Social and institutional patterns

This indicator is defined as the variable patterns of individual involvement in and utilization of the processes and services of the institutional organization and facilities of the community. It is therefore concerned with the degree to which those services are in fact delivered rather than establishing their existence--the latter would be the task of the social organization operationalization.

Physical environmental characteristics

This indicator of the population system is concerned with the physical and environmental conditions in which the population lives and how these conditions change through time. These, like all other characteristics, will be impacted and have costs and benefits accrued to individuals through the delivery of services. The interest is in the current state of the individual's conditions resulting from the environment in which he lives.

Organic characteristics

This indicator is defined as the variable patterns of individual processes and services utilized to maintain the physical organic conditions of individuals in the community. Two important organic conditions are health and nutrition.

Esthetic/cultural system

The interest in this system is not in the usual scientific sense of culture. Rather, this indicator of the population system is concerned with the esthetic cultural conditions of the population. Of interest in this system might be the impact of fine arts, leisure and recreation and areas of entertainment on the individuals in the community system that contribute to a more complete understanding of the individual's "well-being" in this area. It is therefore defined as the variable patterns of individual involvement in, and utilization of the cultural and esthetic processes of the community.

The development of a taxonomy of these four subindicators of the element population is indeed a laborious task. Only the next lower level of indicators will be presented for these four subindicators of population. A complete taxonomy of social indicators would require the enumeration, not only of these components, but also the explication of the elements of social organization, culture and environment.

Social and Institutional Patterns

Figure 5 presents the initial explication of this subindicator. There are probably other subindicators of this category that are not included in Figure 5; however, these five are, at least in part, assumed to be the minimum to be considered in further explications. Each of the five can be logically explicated into at least four to six additional sublevels and probably more before the indicators are at a low enough level of abstraction to develop measurements.

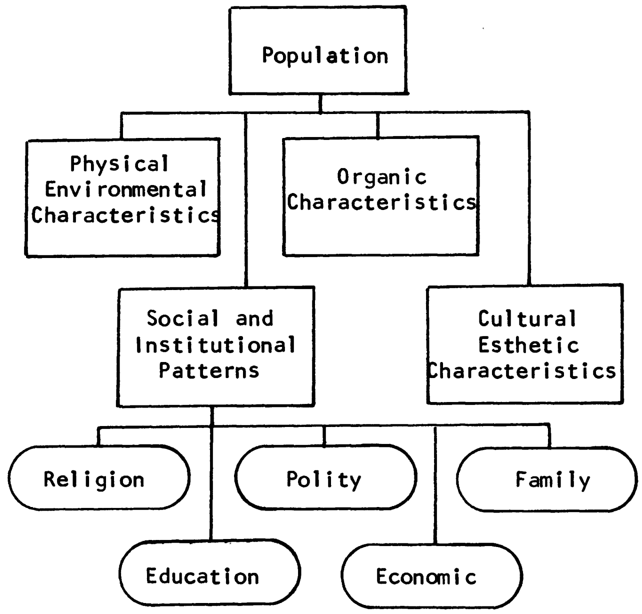


Figure 5. Explicating the element of social and institutional patterns.

One of the basic problems encountered in developing a taxonomy is the decision as to which subconcept belongs in which category. Ideally, one should use as mutually exclusive categories as possible, but, it is difficult to attain this level of expertise in a discipline that has multidimensional concepts and extensive mutual causality among variables.

Polity

Polity is the subindicator of the "social and institutional patterns" selected for further explication and is broadly defined as the services one would assume to be delivered by the community and what benefits they are for the individuals. The major interest is in the costs and benefits to individuals in reference to these services, are they available and do all members of the community participate in them on an equal basis?

It is possible to demonstrate how this component could be partly explicated to lower level indicators. Figure 6 is one possible delineation of this indicator. The five subcategories are social order, public maintenance, social welfare, political participation and political socialization. Social order is defined as the maintenance of safety or securing the community residents from threat of danger, harm or loss. Further explication might include public safety and public justice. Public maintenance is defined as those activities carried out by the government to maintain or improve the physical well-being of the community. Social welfare is defined as the organized efforts by a community for the social betterment or general improvement in the welfare of its members. Measures of social welfare should reflect the manner in which various subgroups have access to and utilize the social welfare services. Political participation is defined as those voluntary activities by which the members of a society share in the selection of officials and, directly or indirectly in the formation of public policy. The concern might be with voting behavior which would include who is registered to vote and who actually votes. Political socialization is often defined as a process whereby individuals incorporate into their own attitudinal structure politically relevant behavior patterns of their respective social groups and society. A next step for the development of a taxonomy of polity would suggest developing lower level indicators for the five elements in Figure 6.

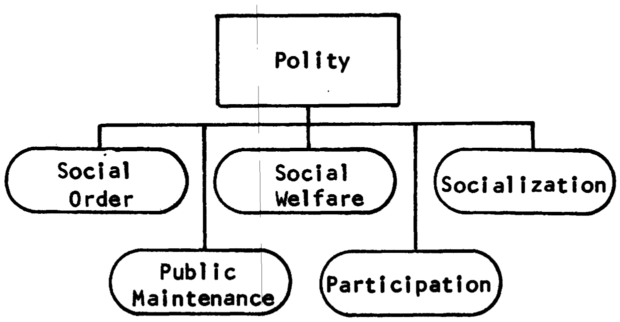


Figure 6. Initial taxonomy of polity.

What may exist in one community may not exist in another. There may be deprivation in a community because a particular service is not provided by the polity and the individuals therefore must seek a desired benefit from another community. It is also assumed that some services are delivered unequally. Therefore, a major part, not only of the polity, but also of the entire explication must be considered in terms of intercommunity and intracommunity comparison of subaggregates of the population.

Physical Environmental Characteristics

This element of the population system has, as a basic concern, the present state of the individual's well-being in reference to his physical environment. Figure 7 is the initial explication of this indicator, which is composed of three subindicators. The individual is the unit of basic concern in this explication. His physical environment is, however, enhanced or detracted depending on the adequacy of his immediate surroundings, the neighborhood in which he lives and the community. The subindicator of individual is explicated to include the physical (man-made) environment and the natural environment. One subindicator of the physical environment could be housing. Transportation could also be a subelement of this subindicator. For natural environment, the concern is with the current state of the air, water and land.

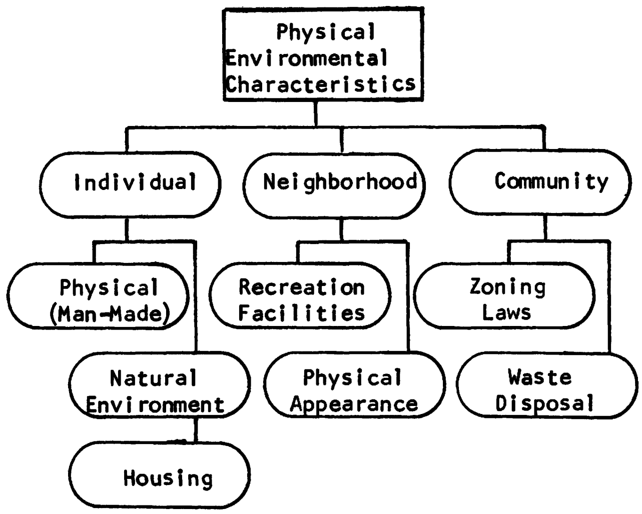


Figure 7. Initial taxonomy of the physical environmental characteristics.

In the neighborhood subindicator are included recreation facilities and the physical appearance. In recreation the concern is with the access to and use of facilities such as pools, bike trails, parks and school grounds. There are other concerns in this area; however, it is believed that these four give an indication of the type of services and resources that were mobilized in the social organization system for delivery in this system of social and institutional patterns.

The last of the three subindicators of "physical environmental characteristics" is community. It could be further explicated to include zoning laws and waste disposal which are considered important in enhancing the physical environment. Important questions might be: Do the individuals in the community have access to a public dump? Do they have city pickup of solids and trash, or must they rely on some other means of disposal? What are the zoning laws and how can they contribute to enhancing the physical environment should provide direction in explicating the category of "zoning laws" to lower-level indicators for the purposes of assessing current social conditions in the community.

Organic Characteristics

The third indicator of the population system is the category of "organic characteristics" of the individuals in the community system. Figure 8 presents this indicator with three possible subindicators. Health may be considered a resource to maintain the organic well-being of the individual in the community system. Subindicators of this indicator would be concerned with access to medical services, frequency of visits to these medical facilities, types of diseases cured during past years, type of insurance carried by the individuals and assessments of the current state of mental health.

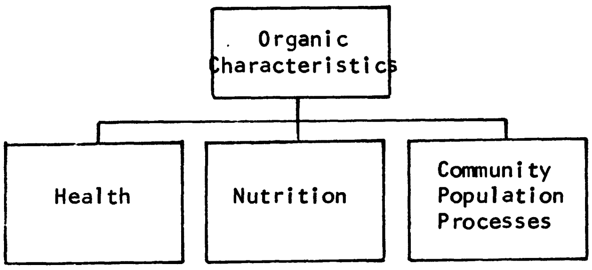


Figure 8. Initial taxonomy of the organic characteristics.

Nutrition is considered a resource utilization, and it is assumed that calorie intake, percentage of net income spent for food, regularity of meals, and type of diet may be possible measures of the nutritional state of well-being of community residents.

The third subindicator is community population processes. In general discussions of population characteristics at least five different variables are likely to be mentioned. These variables, furthermore, are often referred to as the major population processes. Figure 9 presents these five variables with initial taxonomies for fertility, marriage and mortality. The other two processes are mobility and migration. The partial explication of community population processes is presented to demonstrate the types of data and statistics that are needed and how they are related to higher level indicators in the community ecosystem.

This terminates the partial taxonomy of social indicators. At these lower levels are where the social indicators become closer to the empirical level and more easily lend themselves to future quantification. Again, this procedure is definitely none too ordered, but does suggest a strategy for delineating components and indicators of polity in the community system.

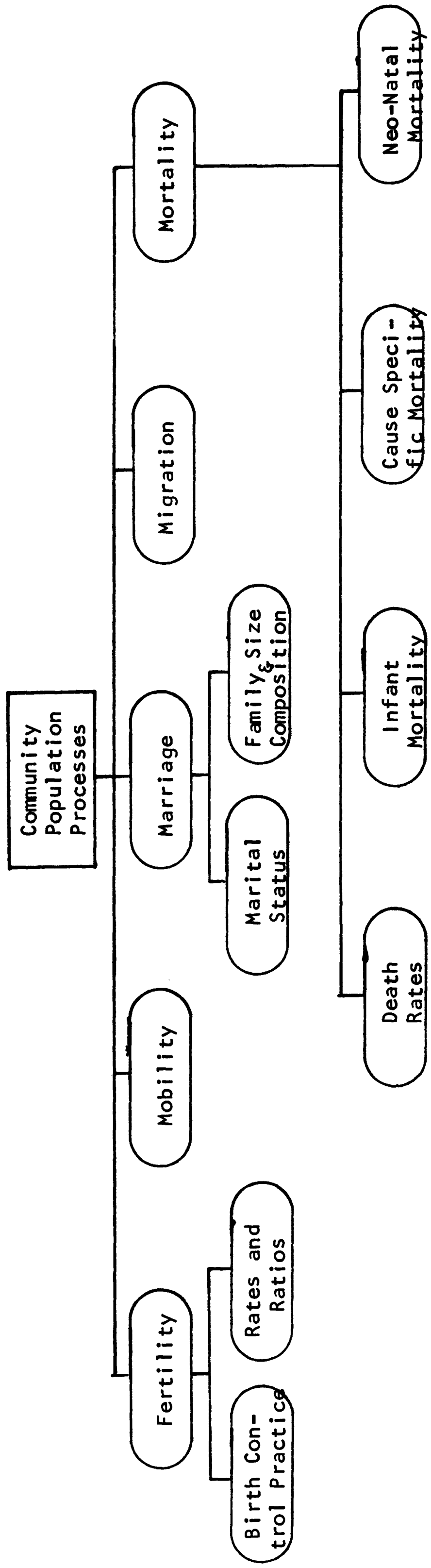


Figure 9. Initial taxonomy of the community population processes.

CLUSTER ANALYSIS: AN APPLICATION TO TYPOLOGY OF URBAN NEIGHBORHOODS

Herbert Bixhorn, District of Columbia Government

1. Introduction

Cluster analysis is the name given to a body of methods for partitioning a heterogeneous collection of objects into groups or clusters in which the objects tend to be similar. In this paper a particular type of cluster analysis is introduced and applied to the problem of classifying geographic subareas of a city into a meaningful typology. The objects to be classified here are census tracts of a city, each tract having a set of variables associated with it. Tracts are considered to be similar or to belong to the same cluster if their values on these variables are similar according to some criterion. The description of the criterion function used in clustering will be more conceptual than rigidly mathematical. The reader who is acquainted with matrix algebra will find a complete discussion of the subject in Rubin and Friedman [1]. The main purpose of this paper is to show some of the advantages of one type of cluster analysis over methods now in common use. To aid in this, both artificial examples and results of analyses performed on tracts of Washington, D. C. will be given.

Before discussing the method of clustering used in this paper, we will review two commonly used methods of classification: summed-ranks and principal components. This will give some indication of the problems encountered in classification.

2. Summed-ranks

The method of summed-ranks will be introduced by first discussing the method of ranking on one variable.

EXAMPLE: Suppose we wish to partition a set of 12 census tracts on the basis of median family income. (See Table 2a at end of text.) The tracts are ranked from lowest to highest on income as shown in Table 2b. If the tracts were to be divided into 2 groups, all tracts with ranks 1-6 would be in one group and all those with rank 7-12 would be in the other. Similarly if 3 groups were to be formed, the first group would contain tracts with ranks 1-4; the second group, ranks 5-8; and the third group, ranks 9-12.

Let us now plot the income of the 12 tracts and denote the partition into 2 and 3 groups by the parentheses around the x's representing income (Figures 2a and 2b). Two difficulties become clear here:

a. There is no indication what the optimum number of groups is.
b. Even if we assume that either 2 or 3 is the correct nunber of groups, the groups themselves do not appear to be "natural."

As an example of b., notice that in the partition into 2 groups, the tract with income of $11,000 appears to be distant from others members of its group. This difficulty is caused by the distortion of distances in the ranking process. A grouping that might appeal to our intuition is given in Figure 2c. (Notice that we intuitively pick the "correct" number of groups while at the same we determine group composition.) This grouping seems reasonable because distances between groups appear large with respect to distances between points in the same group. These distinctions disappear in ranking. The differences in income between tract 7 and 10 is $500 while that between tract 10 and 1 is $5,000. The difference in ranks, however, is 1 in each case. (Table 2b).

The method of summed-ranks is a simple extention of the method of ranking on one variable. Let p variables be measured on each census tract. The tracts are ranked on each variable separately, the p ranks are summed for each tract, and this sum is finally ranked.

EXAMPLE: Let each of 8 tracts have a median family income and median education of household head associated with it (Table 2c). Each tract can be plotted as a point in 2-dimensional space as shown in Figure 2d. Tracts 3 and 7 exhibit quite different behavior and are therefore distant from each other on the graph. A glance at Table 2c, however, reveals that they have the same rank. The difficulty here is that a 2-dimensional problem is being forced into 1 dimension. Although it was reasonable to order the tracts on each variable separately, there was no justification for ordering the tracts on both variables simultaneously. Only in the case where two variables are highly correlated is it valid to represent their ordering by the summed-rank.

The difficulties shown in this example occur in higher dimensions and are compounded with the problem of distortion of distances, illustrated above in 1-dimension.

A reverse type of problem can also occur. Assume three variables are measured on each tract and that two are highly correlated. These two variables may be different names for the same phenomenon and yet they are treated as being independent. They are therefore given more weight than they are due in the method of summed-ranks.

3. Principal Components

The method of principal components often allows us to replace an initial set of variables with one index number. The method is demonstrated graphically for the 2-dimensional case. Let two variables x_1, x_2 be measured on each tract and plotted as in Figure 3a. An axis is drawn through the origin such that the sum of squares of the perpendicular distances of the points to the axis is minimized. This axis is called the principal component. The tract is now represented by one number: the distance from the origin of its projection on the principal component axis. This number takes the form $y = c_1x_1 + c_2x_2$, where the c's are known constants.

Representing each tract by its principal component value is justified only if the dispersion of points is primarily along the direction of the principal component axis. If this is the case, the tracts can be ranked and grouped on the basis of this value. This procedure gives rise to many of the same problems encountered in summed-ranks:

a. One index number can usefully replace the original variables only if the variation is primarily in the direction of the principal component axis. For this to happen the variables must be highly correlated.

b. If the principal component values are ranked to form groups, the problem of distortion of distance and the number of groups to consider again arises.

4. Cluster Analysis

The particular clustering technique applied here explores the structure of multivariate data in search of "clusters" by means of a certain criterion function. Each object has p variables associated with it and therefore can be represented by a point in p-dimensional space. The criterion function measures the ratio of the total dispersion of all points to the pooled dispersion of points within clusters. The goal is to find a grouping or clustering of points which maximizes the criterion function.

One-dimensional case

Consider the configuration of points $x_1, x_2, \text{---}, x_6$ with the two possigroupings shown in Figures 4a and 4b. The groups (or clusters) in 4a appear more compact, i.e. the dispersion or scatter of points within each group appears small with respect to the total scatter of all points.

Total scatter T is expressed mathematically as follows:

$$T = \sum_{i=1}^{6} (x_i - \bar{x}_T)^2 \text{ where } \bar{x}_T = 1/6 \sum_{i=1}^{6} x_i$$

The pooled-within groups scatter W is given by

$$W = W_1 + W_2$$

For Figure 4a,

$$W_1 = \sum_{i=1}^{4} (x_i - \bar{x}_1)^2 \text{ where } \bar{x}_1 = 1/4 \sum_{i=1}^{4} x_i$$

$$W_2 = \sum_{i=5}^{6} (x_i - \bar{x}_2)^2 \text{ where } \bar{x}_2 = 1/2 \sum_{i=5}^{6} x_i$$

For Figure 4b,

$$W_1 = \sum_{i=1}^{3} (x_i - \bar{x}_1)^2 \text{ where } \bar{x}_1 = 1/3 \sum_{i=1}^{3} x_i$$

$$W_2 = \sum_{i=4}^{6} (x_i - \bar{x}_2)^2 \text{ where } \bar{x}_2 = 1/3 \sum_{i=4}^{6} x_i$$

The criterion function is defined as the ratio T/W. Notice that T is constant under both groupings. Therefore maximizing T/W is equivalent to minimizing W. If the grouping in 4a is actually better than that in 4b then its value for W should be smaller. To find the optimum grouping into two clusters, all possible assignments of the points into two groups should be attempted until T/W is maximized.

In general the criterion function for the 1-dimensional case is defined as follows. Let x be a variable measured over each of n objects (here tracts).

Suppose the tracts are partitioned into g groups with the first group containing n_1 tracts with respective values $x_{11}, x_{12}, \text{---}, x_{1n_1}$;

the second group containing n_2 tracts with respective values

$$x_{21}, x_{22}, ---, x_{2n_2};$$

the g-th group containing n_g tracts with respective values

$$x_{g1}, x_{g2}, ---, x_{gn_g}.$$

Then total scatter is given by

$$(1) \quad T = \sum_{i=1}^{g} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_T)^2 = n\,\sigma_T^2, \quad \sigma_T^2$$

is the variance of the entire collection of points;

$$\bar{x}_T = \frac{1}{n} \sum_{i=1}^{g} \sum_{j=1}^{n_i} x_{ij}$$

Pooled-within group scatter,

$W = W_1 + W_2 + --- + W_g$ where

$$(2) \quad W_i = \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2 = n_i\,\sigma_i^2, \quad \sigma_i^2$$

is the variance of points in the i-th group;

$$\bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij}$$

Criterion function = T/W. All possible assignments of n points into g groups are attempted. The grouping which maximizes T/W is considered optimum.

<u>Two-dimensional case</u>

Let two variables x, y be measured over each tract and the tracts be partitioned into g groups as in the preceding paragraph. (The notation for the subscripts of the y's will be the same as that for the x's.) Then total scatter is given by the 2x2 determinant $|T|$

$$\text{where } T = \begin{bmatrix} n(\sigma_x)_T^2 & n[cov(x,y)]_T \\ n[cov(x,y)]_T & n(\sigma_y)_T^2 \end{bmatrix}$$

$$n[cov(x,y)]_T = \sum_{i=1}^{g} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_T)(y_{ij} - \bar{y}_T)$$

and $n(\sigma_x)_T^2$ is given by equation (1). The expression for $n(\sigma_y)_T^2$ is completely analogous.

Pooled-within group scatter is given by $|W|$ where
$W = W_1 + W_2 + --- + W_g$

$$\text{where } W_i = \begin{bmatrix} n_i(\sigma_x)_i^2 & n_i[cov(x,y)]_i \\ n_i[cov(x,y)]_i & n_i(\sigma_y)_i^2 \end{bmatrix}$$

$n_i(\sigma_x)_i^2$ is given by equation (2),

$n_i(\sigma_y)_i^2$ is calculated in the same way, and

$$n_i[cov(x,y)]_i = \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)(y_{ij} - \bar{y}_i)$$

The criterion function is $|T| / |W|$.

It should be noted that the total scatter determinant

$$|T| = n^2 \left\{ (\sigma_x)_T^2 (\sigma_y)_T^2 - [cov(x,y)]_T^2 \right\}$$

may be thought of as n^2[(length of scatter) x (width of scatter) - (overlap due to correlation)], i.e. the total area of scatter. Similarly $|W|$ can be considered the pooled-within group area of scatter. As in the 1-dimensional case, all possible assignments of the n points into g groups are attempted. The grouping which maximizes $|T| / |W|$ is considered the optimum.

The concepts presented above can be extended to any dimension p. In multivariate statistical theory, pxp determinants such as $|T|$ and $|W|$ (excluding the factor of n^2) are known as generalized variances and are often interpreted as representing volumes of dispersion.

EXAMPLE: Calculation of criterion function for 2-dimensional case. Consider the partitioning into two groups of the following points:
(0,6), (2,12), (10,2), (12,4), (12-2), (14,4).

The criterion function will be calculated for two possible clusterings into two groups (see Figure 4c).

Clustering A: First group contains (0,6), (2,12).

$$\bar{x}_1 = 1, \ \bar{y}_1 = 9$$

$$2(\sigma_x)_1^2 = (0-1)^2 + (2-1)^2 = 2;$$

$$2(\sigma_y)_1^2 = (6-9)^2 + (12-9)^2 = 18$$

$$2[cov(x,y)]_1 =$$
$$(0-1)(6-9) + (2-1)(12-9) = 3+3 = 6$$

$$W_1 = \begin{bmatrix} 2 & 6 \\ 6 & 18 \end{bmatrix}$$

Second group contains (10,2), (12,4), (12-2), (14,4).

$$\bar{x}_2 = 12, \ \bar{y}_2 = 2$$

$$4(\sigma_x)_2^2 = 8; \ 4(\sigma_y)_2^2 = 24;$$

$$4[\text{cov}(x,y)]_2 = 4$$

$$W_2 = \begin{bmatrix} 8 & 4 \\ 4 & 24 \end{bmatrix}$$

$$W = W_1 + W_2 = \begin{bmatrix} 10 & 10 \\ 10 & 42 \end{bmatrix} \ ;$$

$$W = (10)(42) - 10^2 = 420 - 100 = 320$$

Clustering B: First group contains (0,6), (2,12) (10,2), (12,4)

$$\bar{x}_1 = 6, \ \bar{y}_1 = 6$$

$$4(\sigma_x)_1^2 = 104; \ 4(\sigma_y)_1^2 = 56;$$

$$4[\text{cov}(x,y)]_1 = -52$$

$$W_1 = \begin{bmatrix} 104 & -52 \\ -52 & 56 \end{bmatrix}$$

Second group contains (12,-2), (14,4).

$$\bar{x}_2 = 13, \ \bar{y}_2 = 1$$

$$2(\sigma_x)_2^2 = 2;$$

$$2(\sigma_y)_2^2 = 18, \ 2[\text{cov}(x,y)]_2 = 6$$

$$W_2 = \begin{bmatrix} 2 & 6 \\ 6 & 18 \end{bmatrix}$$

$$W = W_1 + W_2 = \begin{bmatrix} 106 & -46 \\ -46 & 74 \end{bmatrix} \ ;$$

$$|W| = (106)(74) - (46)^2 = 5728$$

For both clusterings:

$$\bar{x}_T = 8.33 \quad \bar{y}_T = 4.33$$

$$6(\sigma_x)_T^2 = 171.33; \ 6(\sigma_y)_T^2 = 107.33;$$

$$6[\text{cov}(x,y)]_T = -92.67$$

$$|T| = \begin{vmatrix} 171.33 & -92.67 \\ -92.67 & 107.33 \end{vmatrix} = 9801.12$$

Clustering A: $|T| \ / \ |W| = 30.63$
Clustering B: $|T| \ / \ |W| = 1.71$

It was obvious from Figure 4a that A is a much better clustering than B. This has now been verified by the larger value of A's criterion function.

It is clear from the definition of the criterion function that the difficulties observed in the methods of summed-ranks and principal components have been eliminated. Distance is preserved by use of variances to measure dispersion, correlation between variables is accounted for by the covariance term in the scatter matrices, and the use of generalized areas or volumes rids us of the notion of strict ordering of objects.

The question of how many groups to take remains. Regardless of the number of groups taken, the total scatter remains the same. The pooled-within group scatter for the optimum grouping decreases, however, as the number of groups is increased. This, of course, causes an increase in $|T| \ / \ |W|$. Experience indicates that log $|T|/|W|$ tends to reach a plateau at a certain point, and an increase in the number of groups gives diminishing returns. The point at which the plateau begins is taken as the optimum number of groups.

5. A Clustering Computer Program

An IBM computer program employing the methods of section 4 has been written in G or H level FORTRAN and in 360 assembler language. (See Rubin and Friedman [2]. Some of the material in this reference is identical to that in [1]. The remaining material concerns other methods of clustering and instructions for utilizing the programs.) In addition to performing the computations directly related to the determination of clusters, the program produces auxiliary output which is necessary for a complete understanding of the clustering process. Two examples of this are the plot of tracts in eigenvector space and the calculation of discriminant weights.

In any classification problem it is not unusual to have objects which do not clearly belong to any group. In analysis of the census tracts of Washington, D. C., there were often tracts with values of certain variables which placed them far from the mean of any group. A plot of the tracts in a certain eigenvector space (see [1]) enables us to identify such outliers.

The discriminant weights indicate which variables play the greatest role in distinguishing one cluster from another. A complete discussion of discriminant analysis can be found in Anderson [3] and Morrison [4].

Examples of eigenvector plots and discriminant weights are given in the next section.

6. Clustering Census Tracts of Washington, D. C.

Cluster analysis was applied separately to three different sets of variables measured on each census tract of Washington, D. C. The tracts composing each cluster were listed in the output of the program mentioned in Section 5. The mean value of every variable over each cluster was also computed. It is this set of mean values which characterizes the cluster.

Data for two of the sets of variables, "conditions surrounding birth" (1969) and welfare (1967), came from agencies in the District of Columbia government. The tracts in this case were based on 1960 census tract boundaries and were 122 in number.

The third set of variables was meant to serve as a general socio-economic indicator. The data and tract boundaries were taken from the 1970 Census. The 1970 tracts were 147 in number.

Table 6a gives the mean values of the clusters formed on the basis of five "conditions surrounding birth" variables for the year 1969. There is a clear ordering from BEST to WORST groups simultaneously on all variables. It is worth noting the unequal number of tracts in each group. This would not have occurred in the method of summed-ranks.

In practice the rule for determining the optimum number of groups is often quite vague. Exactly where the point of diminishing returns occurs in the criterion function is not always obvious. For this case, however, the choice seemed clear. Table 6b indicates that it is reasonable to take three groups.

The discriminant weights in Table 6c show that the percent of mothers under 20 years of age had a primary role although prematal care was also important in distinguishing between BEST and MEDIUM groups. Age of mothers was again dominant, although to a less extent, and both prenatal care and illegitimacy had secondary roles in distinguishing MEDIUM from WORST groups.

The next set of variables considered was the caseload, expressed as a percent of the population at risk, in each of four welfare categories (Table 6d). Here there is a high degree of skewness with the great majority of tracts belonging to the BEST group. Although the group in the second column is labelled MEDIUM, it is the worst in AFDC. This is a clear case where the tracts cannot be ordered on all variables simultaneously. Table 6e demonstrates the difficulty in choosing the optimum number of groups. Either 3 or 4 seemed appropriate.

Figure 6a is a plot of the tracts in eigenvector space as explained in [1]. The tracts in the BEST group are plotted as B's, etc. The BEST group appears more compact than the other groups. By means of other output from the computer program, it is possible to identify the outlying tract represented by the encircled "M" in the MEDIUM group and to determine which variables caused it to be so distant from its group mean.

The clusters formed in an analysis of four variables chosen as a general socio-economic indicator illustrate an interesting phenomenon (Table 6f). Although there is generally clear ordering of all variables from BEST to WORST, the distinction sometimes disappears, as in comparing the matriarchy and overcrowding indices between the POOR and WORST groups. The incomplete plumbing index seems to be the dominant variable in distinguishing between these groups. The discriminant weights in Table 6g verify this.

7. Additional Remarks

The method of clustering discussed in this paper is one of several which may be appropriate for classifying census tracts of a city. [2], for example, presents other methods and also considers various options to be used with the method applied here. In future studies, we plan to use a hybrid clustering model which will employ judgments from subject matter experts as well as mathematical techniques. One significant result of this will be the subjective weighting of variables before they are entered into the clustering process. At present all variables are assumed to be of equal importance.

REFERENCES

[1] Rubin, J. and Friedman, H. P., "On Some Invariant Criteria for Grouping Data," Journal of the American Statistical Association, December 1967.

[2] Rubin, J. and Friedman, H. P., "A Cluster Analysis and Taxonomy System for Grouping and Classifying Data," IBM Program Library, August 1967.

[3] Anderson, T. W., An Introduction to Multivariate Statistical Analysis, John Wiley and Sons, Inc., 1958.

[4] Morrison, D. F., Multivariate Statistical Methods, McGraw Hill, Inc., 1967.

TABLE 2-a
Distribution of Income by Census Tract
(in $1,000 units)

Census Tract No.	1	2	3	4	5	6	7	8	9	10	11	12
Median Family Income	11.0	4.5	12.0	13.0	4.0	14.0	5.5	5.0	18.0	6.0	20.0	19.0

TABLE 2-b
Census Tracts Ranked by Income

Rank	1	2	3	4	5	6	7	8	9	10	11	12
Income	4.0	4.5	5.0	5.5	6.0	11.0	12.0	13.0	14.0	18.0	19.0	20.0
Tract No.	5	2	8	7	10	1	3	4	6	9	12	11

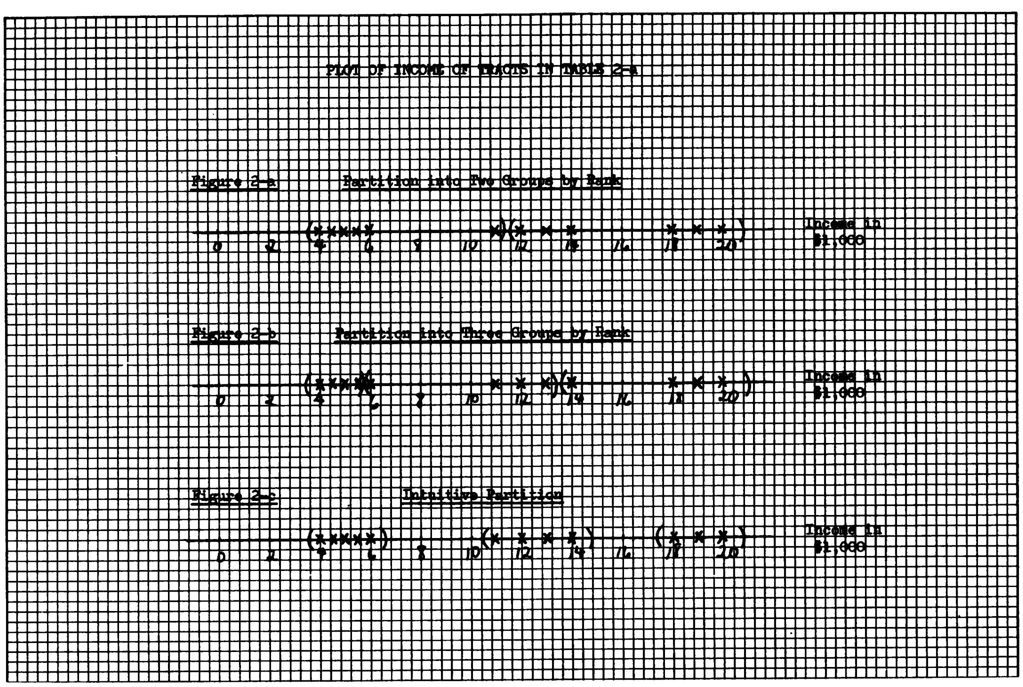
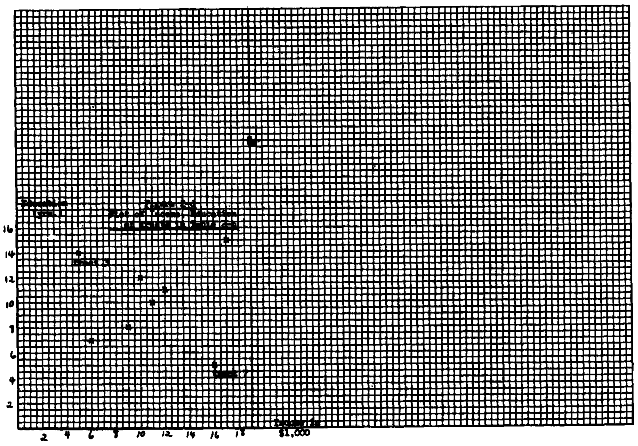
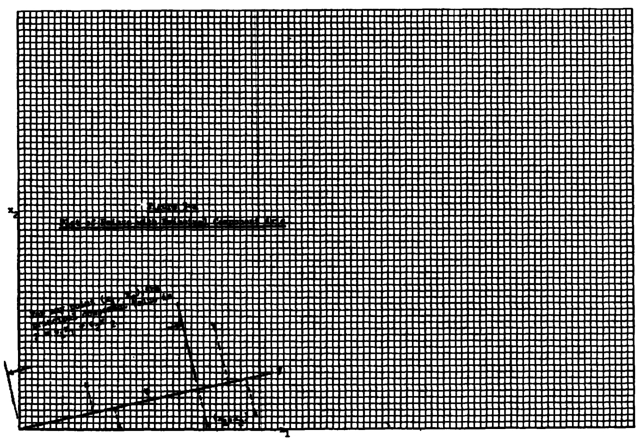


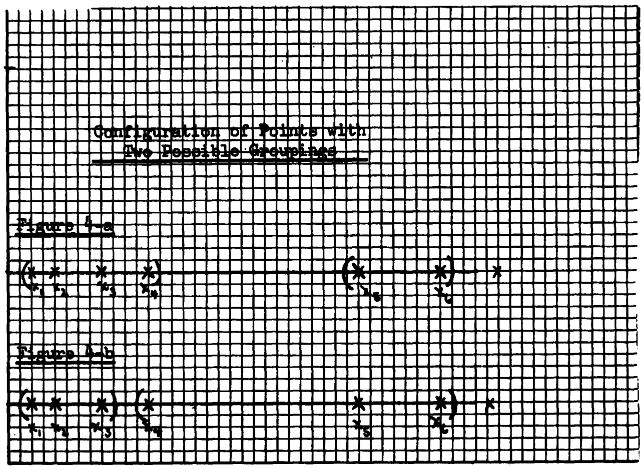
Table 2-c

APPLICATION OF SUMMED RANKS TO DISTRIBUTION OF INCOME AND EDUCATION OVER EIGHT TRACTS

Census Tract No.	1	2	3	4	5	6	7	8
Income ($1,000 units)	11	12	5	10	9	6	16	17
Education (yrs.)	10	11	14	12	8	7	5	15
Rank on Income	5	6	1	4	3	2	7	8
Rank on Education	4	5	7	6	3	2	1	8
Summed-rank	9	11	8	10	6	4	8	16
Final Rank	5	7	3½	6	2	1	3½	8







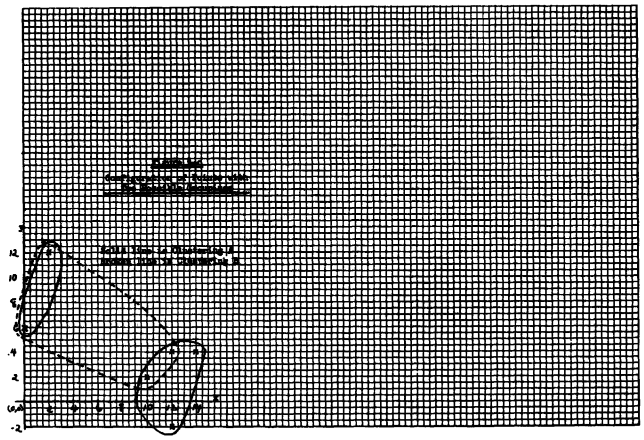


Table 6-a

CONDITIONS SURROUNDING BIRTH - 1969

GROUP MEANS IN PERCENT*			
	BEST	MEDIUM	WORST
Mothers Under Age 20	5.2	23.2	38.8
No or Inadequate Prenatal Care	8.2	25.0	35.4
Birth Weight Under 5½ Lbs.	7.0	13.4	15.2
Illegitimate Births	12.9	27.4	51.0
Infant Mortality	2.1	3.2	3.2
Number of Tracts	30	51	41

* Base population is the total number of births.

Table 6-b

Maximum Value of Criterion Function by Number of Groups

NUMBER OF GROUPS	CRITERION FUNCTION	INCREMENT
2	1.25	
		1.21
3	2.46	
		.73
4	3.19	
		.45
5	3.64	
		.43
6	4.07	

Table 6-c

Discriminant Weights Between Groups

VARIABLES	DISCRIMINANT WEIGHTS	
	Best-Medium	Medium-Worst
Mothers Under Age 20	.78	.51
No or Inadequate Prenatal Care	.45	.29
Birth Weight Under 5-1/2 Lbs.	-.19	-.08
Illegitimate Births	-.002	.27
Infant Mortality	-.04	-.08

Table 6-d

WELFARE - 1967

PUBLIC ASSISTANCE CATEGORIES

GROUP MEANS IN PERCENT			
	BEST	MEDIUM	WORST
Old Age Assistance	2.7	7.7	9.0
Aid to Families with Dependent Children	3.4	23.2	12.0
Aid to Permanently and Totally Disabled	.6	1.5	3.1
General Public Assistance	.2	.3	.6
Number of Tracts	94	10	18

Table 6-e

WELFARE - 1967

MAXIMUM VALUE OF CRITERION FUNCTION BY NUMBER OF GROUPS

Number of Groups	Criterion Function	Increment
2	1.22	
		.86
3	2.08	
		.86
4	2.94	
		.76
5	3.70	
		.76
6	4.46	

Figure 6-a

PLOT OF CENSUS TRACTS IN EIGENVECTOR SPACE OF WELFARE VARIABLES

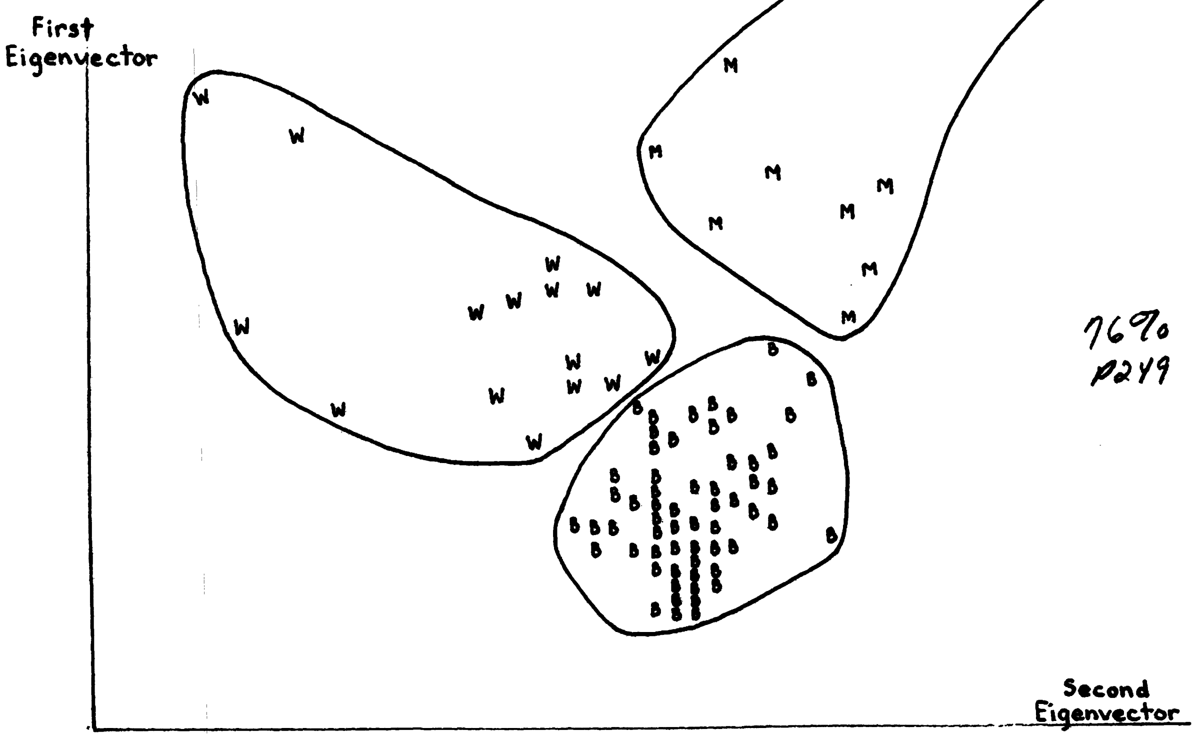


Table 6-f

SELECTED SOCIO-ECONOMIC VARIABLES

GROUP MEANS - 1970				
	BEST	MEDIUM	POOR	WORST
Median Family Income*	$17,000	$8,600	$6,700	$4,800
Matriarchy Index	15.5%	26.4%	38.4%	39.1%
Overcrowding Index	1.9%	8.8%	22.4%	21.4%
Incomplete Plumbing	.8%	2.2%	2.1%	16.8%
Number of Tracts	22	62	58	5

* This is not yet available from the 1970 Census. The values were estimated for each tract from a regression model developed by Westat Research, Inc., Rockville, Maryland.

Table 6-g

SELECTED SOCIO-ECONOMIC VARIABLES

VARIABLES	DISCRIMINANT WEIGHTS POOR - WORST
Median Family Income	- .03
Matriarchy Index	- .21
Overcrowding Index	.05
Incomplete Plumbing Index	1.10

DISCUSSION

Theodore Suranyi-Unger, Jr., The George Washington University

The quest for improved and more scholarly methods of studying, describing, and analyzing the quality of life has, for a long time engaged the energy of social scientists. Intuitively, we all have similar thoughts in everyday conversations on the quality of life, and we meet with fair success in conveying the idea, when we say that certain people or peoples live "better" or "worse" than others. But, as Brooks and Wilcox have driven it home this morning, the accurate description and comparative analysis of the quality of life with the sophistication and scholarly rigor befitting a social scientist, presents a formidable, and as yet unsurmounted problem.

To paraphrase Brooks and Wilcox, each student of the quality of life has his own perspectives and orientation and each tends to visualize a unique role for indicators used in the description of the quality of life and their relevance to social planning, social development or in the social sciences.

Some two hundred years ago, Jeremy Bentham introduced utilitarianism into the body of social thought of his day. This concept was rapidly picked up by the then budding discipline of economics, and to this day, appropriately or inappropriately, economists employ a calculus of utility vs. disutility as a gauge of the quality of life. With this, I am afraid that I have revealed myself as an economist, but I hope that the rest of my discussion will not be too slanted in favor of my discipline.

For reasons that are well known to many in this group, utility vs. disutility is no answer to the types of questions raised by Brooks and Wilcox. Yet, these are the questions that society is also raising today not only because it is the current fad, but also because social scientists and students of society in the past, have not been successful in devising relevant methods of analysis. Hence, the present quest for social indicators.

Social indicators address, or perhaps I should say will, at some future date, address what Brooks and Wilcox call "transeconomic" issues. I would like to argue with that designation a little bit, because in my view those issues are every bit as "economic" as issues surrounding say, productivity or employment. After all, if we define economic topics as those having to do with the allocation of scarce resources among competing ends, very few topics in the social sciences would be left out of that designation. The relevant point here is that traditional economic theory for the largest part is not well suited for the study of the economic issues that Brooks and Wilcox prefer to lable "transeconomic." But if and when viable social indicators as discussed by Brooks and Wilcox are developed, they will also serve as novel and highly useful economic indicators. I will come back to this shortly; right now, let me point out what I found particularly significant in the papers of Brooks and Wilcox and of Bixhorn, in the present context.

Throughout their paper, Brooks and Wilcox show an awareness of the importance of reality, i.e., of a de facto set of norms, customs and practices of a community, in the construction of social indicators. This is brought out in the discussion of the systems model, and it is brought out in such phrases as, "... any meaningful measure of life conditions should reflect, in part, the preconception of people living in a community." I take this to mean that any achievements and accomplishments should be evaluated in reference to a going set of quests. These quests are defined in terms of de facto reality rather than on the basis of a preconceived set of normative conditions. What we are saying here is that it makes little sense to judge the performance of one community in terms of the value and goal structure of another community. If this is what Brooks and Wilcox had in mind, I couldn't agree more. If not, perhaps they will consider the point in their future work.

Without the presence of a relevant and realistic set of standards, the level of performance, although accurately observed and described, holds limited meaning. This is my chief concern about the otherwise fine work being done at the UN in Geneva, as quoted by Brooks and Wilcox. What social or economic significance can be attached to the level of newspaper distribution or caloric intake or nutritional balance, if the members of a given community or society don't care to read newspapers, don't wish to maximize (or minimize) caloric intake, and patently ignore norms of nutritional balance? I admit, I have not seen the latest work done at the UN, but a couple of years ago, my observations held true.

There is a small world of economists in which I claim membership, and in which primary interest is focused on the quality of life in reference to living standards and living levels. In that small world, there has been a longstanding debate concerning the proper meaning and definition of living standards vs. living levels. The substance of the debate is immaterial right now; suffice it to say that standards are usually thought of as something aspired, levels are something achieved.

This distinction sounds reasonable, but the measurement-oriented social analyst soon discovers that the distinction is largely anecdotal and will not readily lend itself to rigorous analytical treatment, either in inductive or deductive terms. My present research in the field is designed to sharpen the analytical rigor and widen the empirical content of the concepts of living standards and living levels. This, too, is social indicators research.

In case I have created the impression that this is a brand new approach to social indicators research, let me hasten to point out that there exists a wealth of analytical and empirical material in the field, much of it contributed by our chairman, Helen Lamale and by the Bureau of Labor Statistics. My own concept of living standards vs. living levels differs somewhat from that of Lamale and the BLS, inasmuch as I conceive of standards in terms of the de facto within a group--this can be an income group or a community--and of levels as the attainment of individual consuming and home producing units in reaching those standards.

The standards, of course, can pertain to any type of undertaking that may be of interest to the sociologist, economist, political scientist, and of course, the statistician. The levels are measured strictly in relation to the given standard.

The pertinence of the living standards and living levels concept to the work of Brooks and Wilcox should be obvious, although I will return to it in a moment. Right now, I would like to observe that the method of cluster analysis, as proposed by Bixhorn, if I understand it correctly, appears to be an eminently well suited tool in the identification of living standard classes as well as in the study of living levels within a standard class. And if the levels of living within a standard class show sufficient clustering in certain patterns of consumption, home production, or other forms of social behavior, this may well warrant a reconsideration and redefinition of the living standard class using, of course, cluster analysis. As society changes, so does the structure and composition of living standard classes; hence the need for continuing surveillance and redefinition.

In conclusion and summary, the present day search for social indicators is hamstrung by the insistence on generally applicable and objective standards of evaluation. This approach is replete with arbitrary value judgments and constantly seeks to apply the standards of one society in evaluating the performance of another. Some of these generally applied standards or norms lay a claim to scientific objectivity; this may be justified but of what real interest is that to the social scientist studying the quality of life? I am most pleased to see that Brooks and Wilcox do not subscribe to this approach, although, from time to time, they do appear to throw wistful glances in the direction of universally applicable social indicators. I do not mean to be totally cynical on this topic. Objective norms have their proper place, but before attempting to develop universally applicable, scientific, and objective social indicators, I believe we would do well in paying a great deal more attention to relevant standards observed by a community and the relevant levels of attainment.

This is an inductive approach, and as such, it is in absolute need of workable statistical tools, such as the Bixhorn-type cluster analysis. And even if this method of approaching the problem of social indicators does not offer instant normative appraisal of the quality of life, it will give us a useful illustration of the goal structure and attainment structure of a given standard group or community.

DISCUSSION

Ethel E. Vatter, Cornell University

There are two possible ways of reviewing a paper such as the one presented by Dr. Brooks. One would be to give it empathetic review from the viewpoint of its own methodology; the other would be to critically evaluate it from the viewpoint of methodology in economics. The latter may seem unfair, but I gather not any more so than if the evaluation were done by someone in cultural or behavioral sociology, as distinguished from ecological sociology.[1]

First, one cannot help but express admiration for a project that will (and I quote):

> "...develop a taxonomy of social conditions related to a general model that can provide an explication of quantifiable categories...
>
> ...collect pilot data that can be utilized in initial attempts to build inductively a systems model...
>
> ...develop time series through replication studies...
>
> ...establish interrelationships between a wide range of variables that will allow the development of models to assess social change (my italics)."

Such development of a universal system of social accounting is far more ambitious in scope than was our national income and product accounts. The latter are related to a half-dozen theoretical constructs developed by Keynes and his followers in the 1930's and 1940's: e.g.--aggregate demand and supply functions; propensities to consume and invest; multipliers and accelerators. The micro and macro sides of economic theory were never well integrated, but the research and policies that emerged from the Keynesian approach were enough to keep economists going; that is, until now.

Today economists' interest in the quality of life (and in social indicators) are related to problems of environmental pollution, discrimination (in access to housing, health, education, and employment) poverty, crime, etc. Economists, somewhat belatedly, admit the inadequacies of theory that excludes social costs from private pricing, and are working to extend existing theory to incorporate these costs. They are also concerned with production functions for various categories of public expenditures related to the problems mentioned above. And they are beginning to find a systems approach helpful, perhaps I should say, necessary in evaluating the side effects (i.e., feedback) from specified private and public behaviors.

What concerns me is the nature of the grand theorizing that implicitly directs the model explicated on page 16. Here the community ecosystem is classified first into (1) Cultural System, (2) Social Organization, (3) Population, and (4) Environment, and then sub-classified into thirteen sub-categories: values and knowledge under (1); religion, polity, family, economy, and education under (2); institutional and social patterns, physical environmental characteristics, organic characteristics and cultural esthetics under (3); and finally, social and physical under (4). Are these sub-categories of equal levels of abstraction? Do they not themselves require structure as a director to data gathering?

I believe this confusion is reflected in the title of the paper: Toward the Measurement of Social Indicators... Am I wrong in thinking that Social Indicators are themselves "Measures", and that the authors are proposing the creation of a system of community social indicators, i.e. measures, that can be aggregated and disaggregated (as can the national income and product accounts) as required, and used to understand, monitor and hopefully, predict social change under alternative actions? I like the ecological framework but need assurance that meaningful generalizations can be made from the Iowa setting to more industrialized areas of the U.S. I know that rural-urban populations are increasingly similar in their life styles, but would community organization and decision-making in Des Moines be similar to that in other major metropolitan areas? Is a small town in Iowa like a small town in New York State? What does existing sociological knowledge tell us about the rural-urban continuum? Also, where is there provision for system linkage?

Then there is the assumption on page 12 that the ecosystem's purpose is "to benefit humans." Unfortunately, humans, like communities, are not homogeneous. They cannot all be benefited by the same social policy. There will be resistance to change. When I lived in Iowa, there were serious discussions between economists and sociologists as to whether or not certain communities--those not especially amenable to economic development efforts--should be allowed to "wither on the vine". In Philadelphia, where I spent my sabbatical leave, there was resistance to low-cost housing because it meant racial integration, resistance to urban development because it meant displacing persons from their accustomed homes, and everybody was against raising taxes for anything. When one group wins at the polls or through court decisions or by picketing, it is frequently at the expense of other groups. Conflict, not consensus, is the "name of the game".

I like the discussion on pages 12 to 14 on the "interrelationships of community ecosystem elements", but am puzzled by the statement that "social organization is also considered to be the mobilization of both human and physical resources for the delivery of services to the population within the community ecosystem." It is true that "no man is an island unto himself", but also that under our system, man sells his services (is this mobilization?)--in a private or public market--in exchange for money that in turn is used to buy the goods and services of others. True, more and more services are being provided through the public market via taxation, but the exchange mechanism still remains and affects other elements in the subsystem, Social Organization, as family polity, and religion. Again, not everything is equal, but one must make some tentative decisions regarding the relative importance of dependent and independent variables that recognizes the conflict involved in these changing forms of exchange.

Finally, I must reveal my own bias. Although I think that research based on the ecological approach may contribute more to the understanding of pressing human problems than research based on culture or individual behavior, I personally am skeptical that general knowledge of interrelationships between a wide range of variables will result in the development of models useful for assessing social change. In the first place, I assume sociologists already know a lot about these interrelations, but like economists not enough about discontinuity and novelty. These are not easily recognized nor uniformly interpreted, even in a systems approach. My own inclination would be to settle for a more modest focus--say on the planning and evaluation of social programs, since that is where the action is these days. For this we need models that are dynamic in character and that trace out the consequence of major new government programs or investments over time, and in considerable social, economic, and geographic detail. To my way of thinking, social indicators would thus emerge from social programs and avoid the limitations that exist in traditional theorizing in both sociology and economics. New theory would evolve from the new social actions that the indicators are supposed to evaluate, and not the other way around, as this proposal suggests.

1/ See "Cultural, Behavioral, and Ecological Perspectives in the Study of Social Organization", by Otis Dudley Duncan and Leo F. Schnore, American Journal of Sociology, Vol. LXV, Sept. 1959, pp. 132-146.

PUBLIC SECTOR INFLUENCE ON INTERSTATE MIGRATION OF THE ELDERLY*

Steve L. Barsby and Dennis R. Cox,
University of Arizona

In his well-known 1956 article, Charles Tiebout argued that geographic mobility could serve as a substitute for a market for public goods.[1] Families would move to those localities which provided the most desirable mix of public goods and modes of finance. Communities, in order to reach their optimum sizes, would attempt to structure their public sectors in such a way that the optimum number of residents would be attracted. These two forces -- family mobility and community competition for residents -- are analogous to normal market forces -- purchaser mobility among sellers and seller competition for buyers. Under certain mobility and information conditions, this "market" process would result in the optimum production of public goods and the optimum distribution of these goods among citizens. This argument does not apply of course to "national public goods," such as national defense. It could apply, however, to local public goods, such as police protection, fire protection, local court systems, education, medical facilities, and possibly others. Perhaps the most important implication of the Tiebout model is the suggestion that local decision making about the mix of public services and modes of finance need not be "responsive" to the wishes of the local citizenry to assure optimal production and distribution of local public goods. Instead the citizens, by their responses to local public sector variations, achieve these optimal results. Democratic and undemocratic decision making procedures will be equally effective, provided families are sufficiently mobile and communities are sufficiently numerous.

In view of the importance of this implication, it is surprising that the fundamental premises of this theory have not received more attention -- especially empirical attention. Tiebout's first and perhaps most crucial assumption is that families are mobile and responsive to variations in local public sectors. Tiebout urged that this assumption ". . . should be checked against reality [1,p.423]." We are not aware of any efforts to perform this check explicitly.[1] In this paper we present the results of a simple test of the mobility assumption -- results which offer little in support of it.

This paper is extracted from a larger study of interstate mobility of elderly persons (aged 65 and older). The elderly comprise a particularly appropriate group for this test for at least two reasons. First, it is plausible that the public sector preferences of elderly persons are more homogeneous within the group than are the preferences of the population as a whole, so that by separating this group out we may more clearly discern the relationship between public sector variation and migration. For example, the elderly might be expected to be less concerned with expenditures on health care facilities than the general population; or, since many older persons have little current income (financing consumption out of wealth) they might be expected to be less repulsed by income taxes than by property taxes, relative to the general population.

The second reason that this group is particularly appropriate is that older persons have less attachment to the private sector of the economy, and so can give more weight to the public sector in making locational choices. Younger families, still relying principally on labor income, must compromise their public sector preferences with private sector opportunities. In contrast, social security benefits, annuity income, dividends, interest and rental income can be received in any location. Families which are more reliant on these sources of income, or simply on depletion of wealth, to finance consumption are freer to choose their residential locations on the basis of public sector preferences.

While the age-group focus of our study has certain advantages in this application, its concentration on interstate, rather than intercommunity, mobility is certainly a disadvantage. The obvious reason is that within states there can be considerable variation among communities in the mix of public services and the modes of finance. Descriptions of the public sectors of states are only averages and cover possibly large local variability. In addition, intrastate mobility escapes our attention. We hope in the near future to be able to expand our investigation to intercommunity mobility.

Our statistical model is

$$M_{i,j} = E_{i,j}\,\alpha_j \prod_{k=1}^{9} X_{i,k}^{B_{k,j}} \prod_{h=1}^{55} Y_{i,j}^{B_{h,j}} \quad ,$$

where

$M_{i,j}$ signifies the j^{th} migration rate for the i^{th} state,

$X_{i,k}$ signifies the value of the k^{th} public sector variable for the i^{th} state,

$Y_{i,h}$ signifies the value of the h^{th} non-public sector variable for the i^{th} state,

α_j is a constant scale adjustment for the j^{th} migration rate,

$E_{i,j}$ is the multiplicitive error term for the i^{th} state, and the j^{th} migration rate, and

$B_{k,j}$ and $B_{h,j}$ are elasticities of the explana-

*This investigation was supported in part by grant number 10-P-56029/9-01 from the Social and Rehabilitation Service and Social Security Administration, U.S. Department of Health, Education and Welfare, Washington, D.C.

The authors also acknowledge the use of the facilities of the University of Arizona Computer Center.

tory variables X and Y with respect to the j^{th} migration rate. Elasticities are estimated by converting all the values of the M's, X's and Y's to logarithms (base e) and performing conventional least squares fitting to the transformed variables.

The migration rates of each of the 48 contiguous states employed as dependent variables in this paper are computed from the 1960 Census of Population twenty-five percent sample. They are the number of in (out) migrants into (out of) a state between 1955 and 1960 aged 65 and older (65-69) divided by the population of the state aged 65 and older (65-69) in 1960. These rates are computed for males, females, and all sexes. Thus we obtain twelve migration rates for each state.

The public sector variables are obtained from various sources. They are

- X_1: Minimum number of years of residence required for old age assistance eligibility, 1950-1960 average;
- X_2: Maximum monthly old age assistance payment for one person, 1950-1960 average;
- X_3: Dollar amount of special old age state income tax exemption, 1963;
- X_4: Dollar amount of special old age state income tax credit, 1963;
- X_5: Per capita state and local property tax receipts, 1957-1962 average;
- X_6: Per capita state and local income tax receipts, 1950-1960 average;
- X_7: Per capita state and local expenditures on education, 1950-1960 average;
- X_8: Per capita state and local expenditures on health and hospitals, 1950-1960 average;
- X_9: Per capita state and local expenditures on public welfare, 1950-1960 average.

The remaining explanatory variables (Y) developed for our larger study will not be described in detail here. Most of them could be described as private sector variables. They consist of previous (1949-1950) migration rates, wage rates in various industries, unemployment rates, labor force participation rates, turnover rates, industrial structure variables, average education level, housing occupancy rate, monthly rental rate, and a geography variable (north vs. south).

With respect to most of the explanatory variables, we entertained certain a priori notions of the direction of the effect of the variables on the migration rates. These will be described for the public sector variables for in-migration. We expected opposite effects for out-migration. The first two variables (X_1 and X_2) describe the relative generosity of state old age assistance programs. Given the generally low incomes of elderly persons, many of them should be concerned with the availability and level of old age assistance payments. It seems reasonable to suggest, then, that less strict residency requirements and higher maximum benefits levels will be associated with greater in-migration rates. We therefore expected a negative elasticity for X_1 and a positive elasticity for X_2.

Variables three and four reflect special tax advantages for elderly persons. We expected positive elasticities for both of them. It should be noted that the special exemptions and credits are not very large in the states which have them, so that it is perhaps too much to expect them to influence locational choices. We expected negative elasticities on both of the average tax level variables (X_5 and X_6), with perhaps property taxes having the larger (absolute value) elasticity for reasons described earlier.

The last three public sector variables partially describe the mix of public services provided, and represent potential "offsets" to income and property tax payments. We predicted that state and local government expenditures for education (variable X_7) would have a negative elasticity on the grounds that most persons in our age group do not have school-age children, and as a result such expenditures act simply to increase taxes with no offsetting benefits. On the other hand, persons over age 65 are relatively important beneficiaties of health and hospital services, and of public welfare expenditures. Consequently we expected variables X_8 and X_9 to have positive elasticities.

Our resluts are contained in Tables 1 and 2.[2] Table 1 reports the simple correlation coefficients (r) between each of the public sector variables and each of the migration rates. Significant coefficients are underlined and coefficients with signs conforming to our expectations are indicated by asterisks. These coefficients for variables X_4, X_6, X_8, and X_9 are not significant (at the .05 confidence level) for any of the migration rates. The coefficient for X_3 (amount of special old age state income tax exemptions) is significant for only one of the twelve migration rates, and it has the "wrong" sign. The remaining four variables display fairly consistent patterns. They are generally significant for in- or for out-migration rates, but not for both. However, signs of the significant coefficients for variables X_1, X_2, and X_7 are opposite to our expectations. This leaves per capita state and local property tax receipts (X_5). For this variable the correlation coefficients for out-migration rates are significant at the .01 confidence level and have the expected sign, indicating that high property tax levels repulse elderly persons.

Table 2 contains estimated elasticities for the public sector variables. These elasticities come from regression equations containing non-public sector variables as well, but elasticities for these variables are reported here. The coefficients of determination (R^2) in Table 2 apply to the entire equations, including non-public sector variables. Most of the elasticities in the table are zero. The reason for this is that we did not include a variable in a regression equation unless its regression coefficient would be significant at the .05 level of confidence. Elasticities with the expected signs are indicated by asterisks.

The most noteworthy thing about the results in Table 2 is that the elasticity for the single independent variable giving significant, predicted results in Table 1 (X_5) is significant in only one of the twelve equations. While it still

Table 1

Simple Correlation Coefficients[a] Between the Public Sector Variables and Migration Rates

Migration Rates	Public Sector Variables[b]								
	1	2	3	4	5	6	7	8	9
Gross In-Migration Males Aged 65+	.311	.106*	.157*	-.115	.001	-.011*	.300	-.046	-.146
Gross In-Migration Females Aged 65+	.315	.189*	.223*	-.115	.135	-.018*	.364	.059*	-.185
Gross In-Migration Total Aged 65+	.315	.150*	.193*	-.118	.072	-.016*	.331	.011*	-.170
Gross Out-Migration Males Aged 65+	.073*	.276	.241	-.113*	.479*	-.152	.237*	.215	-.115*
Gross Out-Migration Females Aged 65+	.086*	.354	.308	-.087*	.438*	-.072	.231*	.121	-.155*
Gross Out-Migration Total Aged 65+	.040*	.257	.258	-.123*	.448*	-.152	.202*	.102	-.161*
Gross In-Migration Males Aged 65-69	.325	.067*	.126*	-.091	-.019	.007	.330	-.050	-.141
Gross In-Migration Females Aged 65-69	.338	.135*	.175*	-.099	.096	-.011*	.341	.045*	-.175
Gross In-Migration Total Aged 65-69	.332	.102*	.151*	-.097	.040	-.004*	.335	-.001	-.161
Gross Out-Migration Males Aged 65-69	.087*	.319	.253	-.102*	.500*	-.149	.286*	.218	-.057*
Gross Out-Migration Females Aged 65-69	.074*	.266	.231	-.081*	.437*	-.128	.213*	.161	-.127*
Gross Out-Migration Total Aged 65-69	.080*	.295	.245	-.093*	.473*	-.141	.250*	.191	-.096*

*Expected sign.

[a]Single underlining indicates significance at .05 confidence level; double underlining indicates significance at .01 level.

[b]Defined in text.

Table 2

Migration-Rate Elasticities[a] of Public Sector Variables

Migration Rates	Public Sector Variables[b]									R^{2}[c]
	1	2	3	4	5	6	7	8	9	
Gross In-Migration Males Aged 65+					-.26*			-.73		.97
Gross In-Migration Females Aged 65+									-.26	.97
Gross In-Migration Total Aged 65+						.03		-.51	.18*	.98
Gross Out-Migration Males Aged 65+						-.04				.81
Gross Out-Migration Females Aged 65+			.02						-.17*	.82
Gross Out-Migration Total Aged 65+										.78
Gross In-Migration Males Aged 65-69						.05		-1.07		.97
Gross In-Migration Females Aged 65-69						.04		-.62	-.31	.99
Gross In-Migration Total Aged 65-69								-.71	-.17	.96
Gross Out-Migration Males Aged 65-69				-.08*						.87
Gross Out-Migration Females Aged 65-69						-.04				.92
Gross Out-Migration Total Aged 65-69						-.05				.90

*Expected sign.

[a]All elasticities are significant at .05 confidence level.

[b]Defined in text.

[c]Coefficient of determination applies to entire equation, including some variables not shown here.

has the predicted sign, it appears in an equation for in-, rather than out-migration as was the case in the previous table. X_5 does not appear more often because it is highly correlated with a binary variable for geography (states below the 37th parallel have the value two while states above the parallel have the value one for this variable). When the geography variable enters a regression equation, the partial correlation of X_5 with the out-migration rates becomes insignificant.

The bulk of the entries in Table 2 are for variables which had no significant simple correlations with the migration rates -- X_6, X_8 and X_9. Additionally, elasticities for variables X_3, X_5, X_6 and X_9 are low. The largest of these is .31. On the other hand, migration elasticities for X_8 (expenditures on health and hospitals) are relatively large, and suggest that high expenditures for health care facilities strongly discourage in-migration of elderly persons.

CONCLUSION

Data on migration of elderly persons offer scant support at best for Tiebout's crucial assumption about mobility in relation to the public sector. Most simple correlations between the public sector variables and migration rates were either non-significant or had "wrong" signs. Public sector variables rarely were significant in our complete migration equations. Three did not appear at all. Three more appeared in only one equation. Of the three remaining independent variables, only one contained elasticities of the predicted sign.

As usual, in studies of this kind, our results contain many caveats and may not be regarded as conclusive. We hope to pursue this matter with better data and, possible, greater perception in the future.

FOOTNOTES

[1]The largest body of relevant literature deals with location of industry, rather than location of population. Undoubtedly, factors which affect business location choices ultimately also affect family location choices. It could happen, though, that the public sector affects business location through its affect on family location and therefore on wage rates. Most research on the effect of taxes on business location has failed to reveal strong, consistent relationships. For a survey of much of this literature, see John F. Due, "Studies of State-Local Tax Influences on Location of Industry," National Tax Journal, June, 1961, pp. 163-173. A more recent paper, which deals approximately with mobility patterns of middle-sized cities is Raymond J. Struyck, "An Analysis of Tax Structure, Public Service Levels, and Regional Economic Growth," Journal of Regional Science, 1967, No. 2, pp. 175-183. A clarifying note on this article appeared in Journal of Regional Science, 1969, No. 2. Struyck finds strong inverse relationships between rates of population growth and levels and rates of changes of taxes in fifty cities.

[2]It should be noted that there are some purely statistical factors affecting the correlations between our dependent and some of the independent variables. Since all the migration rates are defined as ratios where the denominators are population figures, and several of the independent variables also are ratios where the denominator is population (all the "per capita" variables), positive spurious correlation is introduced. On the other hand, since in-migration, ceteris paribus, raises population and thereby lowers all the per capita independent variables, negative correlation is introduced. For out-migration this effect produces positive correlation. We should, therefore, be more than usually suspicious of positive correlation for out-migration rates. We should also, perhaps, be somewhat more lenient than usual in judging the significance of negative correlation coefficients.

REFERENCES

1. Tiebout, Charles M., "A Pure Theory of Local Expenditures, Journal of Political Economy, Vol. 64, No. 3, October, 1956, pp. 418-424.

A SOLUTION TO DEFINE LATENT SEGMENTS OF MEDIA AUDIENCES

Purnell H. Benson, Rutgers Graduate School of Business

Procedure

In media analysis, data are supplied for the proportions of the population reached by different vehicles of communication within a limited number of exposure opportunities. From these data, projections are made of the exposures which result with combinations of vehicles whose opportunities for exposure extend over substantial periods of time. These projections are based upon various probabilistic models as to how the frequencies of exposure are numerically related from vehicle to vehicle and exposure to exposure. The Beta function (Greene and Stock, 1967; Hyett, 1958) and the bivariate normal distribution (Benson, 1969) are two such approaches.

Let us assume that the population which has exposure opportunities is made up of segments whose sizes are $X_1, \ldots X_n$ with independent probabilities $P_1, \ldots P_n$ of these segments being exposed to an advertising vehicle in a single exposure opportunity. We then consider what takes place when opportunities for exposure to pairs of vehicles (or for two successive exposures of the same vehicle) exist. The four possibilities are for exposure to both vehicles, the first but not the second, the second but not the first, and exposure to both vehicle opportunities.

Since we assume that probabilities of exposure within each segment are independent from opportunity to opportunity, each of the four combinations of exposure and non-exposure has a frequency of occurrence which is equal to the sum of the binomially combined probabilities within segments. A system of equations relates each of the four cell frequencies for each pair of vehicles to corresponding algebraic terms for the X's and the P's. From the system of equations, the corresponding X's and P's can be calculated, if sufficient combinations of vehicles and exposure opportunities are included.

If refinement of the solution is sought through using a large number of segments, the iterative solution for both X's and P's is not rapid. Alternatively, a large number of segments of equal size can be assumed. Then the system of equations is of second degree only. For M segments, the equations take the following form:

$$MC_{00.ij} = (1 - P_{k.i})(1 - P_{k.j}) + \ldots + (1 - P_{m.i})(1 - P_{m.j}),$$

$$MC_{10.ij} = P_{k.i}(1 - P_{k.j}) + \ldots + P_{m.i}(1 - P_{m.j}),$$

$$MC_{01.ij} = (1 - P_{k.i})P_{k.j} + \ldots + (1 - P_{m.i})P_{m.j},$$

$$MC_{11.ij} = P_{k.i}P_{k.j} + \ldots + P_{m.i}P_{m.j}$$

C_{00}, C_{10}, C_{01} and C_{11} refer to cell frequencies for non-exposure to either vehicle I or J, exposure to I but not J, exposure to J but not I, and exposure to both I and J. Subscripts i and j refer to the pair of vehicles, and subscripts _k_ and _m_ refer to audience segments.

The solution uses a full set of trial values of variables to solve for the same variables as linear multipliers. The matrix of equations is repeated a second time in the system for solution in order to permit solving for the same variables as linear multipliers in one-half of their involvement in the equations. Approximation proceeds by averaging the trial values with the corresponding solution values. Faster convergence occurs if the average is obtained by taking the square root of the product of the trial and solution values.

Application to Media Exposure Data

Through the courtesy of W. R. Simmons, syndicated data for exposure of adults to six magazines were made available for analysis. These data give the proportions of consumers exposed to each magazine in a single opportunity, and the proportions exposed to one or both of the exposure opportunities for pairs of magazines, including a magazine paired with itself for two successive exposure opportunities. From this information frequencies in each of the 4 cells in the two-exposure tabulations are determined.

Six vehicles paired in all combinations provide 57 independent observations, 3 for each of the 15 pairs of vehicles and 2 for each of the 6 vehicles paired with itself. Before using these cell frequencies as entries for the dependent variable, residual constants on the right hand side of the equal sign were transposed and combined with the entry for the dependent variable. For instance, cell C_{00} which is set equal to the products of ones minus probabilities has combined with it those terms which do not provide an entry for the multipliers of the independent variables.

If we use 8 segments for which probabilities of exposure for each of 6 vehicles are to be found, the number of unknowns is then 48, within the 57 degrees of freedom. In addition, there are segments of the population who do not see any of the 6 magazines in a single exposure. This residual group comprises approximately 40% of the population. These may be represented by an additional 5 segments for whom the exposure probabilities are set equal to zero throughout the solution. The probability entries for these are also transposed and combined with entries for the dependent variable.

To assemble the input for multiple regression analysis to define the probabilities of exposure of each segment to each vehicle, a machine program was prepared. The output from this program is the matrix of entries for multiple regression analysis. Values are bounded by 0.01 before averaging and by 1.0 after averaging.

The matrix of trial probabilities for exposure of the 8 active segments and the 5 inactive segments is given in Table 1. Inspection of Table 1 indicates which segments of the population share common exposure to different magazines.

A practical application of the definition of latent segments is to use the associated exposure probabilities to make projections for exposures achieved by media schedules with multiple vehicles and multiple insertions. Existing methods based upon the Beta distribution are inadequate when exposure patterns for the audiences of different vehicles are heterogeneous. The use of probabilities for latent segments may prove to be a useful and inexpensive way out.

As a check upon the efficiency with which exposure projections can be made, the frequency pattern for 8 exposure opportunities was calculated for each of the 6 magazines. The results are compared with the usual Beta projections in Table 2. Since Beta analysis is

Table 1

INITIAL TRIAL VALUES AND PROBABILITY VALUES AFTER FIVE CYCLES OF ITERATION

		Segment												
Vehicle	Cycle	1	2	3	4	5	6	7	8	9*	10*	11*	12*	13*
1	1	.90	.20	.20	.20	.20	.20	.20	.20	.00	.00	.00	.00	.00
	5	.33	.00	.96	.63	.62	.30	.00	.00	.00	.00	.00	.00	.00
2	1	.20	1.00	.20	.20	.20	.20	.80	.90	.00	.00	.00	.00	.00
	5	.22	.22	.46	.35	.66	.24	1.00	.48	.00	.00	.00	.00	.00
3	1	.20	.20	1.00	.20	.20	.20	.90	.80	.00	.00	.00	.00	.00
	5	.15	.01	.57	.22	.19	.22	.96	.72	.00	.00	.00	.00	.00
4	1	.20	.20	.20	.90	.20	.20	.20	.20	.00	.00	.00	.00	.00
	5	.38	.02	.24	.90	.00	.44	.02	.16	.00	.00	.00	.00	.00
5	1	.30	.30	.30	.30	1.00	.50	.30	.90	.00	.00	.00	.00	.00
	5	.47	.10	.24	.72	.90	.62	.04	.99	.00	.00	.00	.00	.00
6	1	.20	.20	.20	.20	.50	.90	.40	.40	.00	.00	.00	.00	.00
	5	.00	.02	.25	.30	.61	.82	.45	.02	.00	.00	.00	.00	.00

* These values are set as fixed during the iterations.

Table 2

ESTIMATION OF PERCENTAGES REACHED BY EACH OF SIX ADVERTISING VEHICLES AFTER EIGHT EXPOSURE OPPORTUNITIES, USING INITIAL TRIAL PROBABILITIES AND EXPOSURE PROBABILITIES ESTIMATED AFTER FIVE CYCLES OF ITERATION

		Initial Values		Values after 5 Cycles	
Vehicle	Beta Calculation	Estimate	Error	Estimate	Error
1	34	53	+19	38	+4
2	63	55	- 8	58	-5
3	51	55	+ 4	49	-2
4	32	53	+21	38	+6
5	55	59	+ 4	52	-3
6	32	56	+24	39	+7

considered acceptable for making projections for a single magazine, this provides a test of latent segment projection. The results look encouraging for what may be accomplished with larger numbers of segments and vehicles.

Relation to Lazarsfeld's Work

The historical debt to and the inspiration from the work of Lazarsfeld for the analysis reported here is large. His creation of latent structure analysis two decades ago is one of methodological landmarks of our century. (Stouffer et al, 1950). I have returned to this type of work with somewhat different reasons and emphasis than his, being first preoccupied with media scheduling.

Some of the procedural differences between Lazarsfeld's development of latent structure analysis and the lines I pursue may be kept in mind.

1. The latent segment framework introduced here for analysis of media exposure assumes a multiplicity of latent segments of equal size, some of which may be regarded as duplicates. The degree of the system of equations is correspondingly reduced.

2. A paired matrix of probability cross-products of no higher than order two is used, leading to a quadratic system of simultaneous equations. The solution to these is relatively straightforward, involving a progressive averaging of trial and obtained values. The multiplicity of solutions appears to be equivalent to factorial combinations of segments, all of which yield sets of roots which are identical except for factor designation.

3. All four cells of the 2-by-2 cross tabulation of probabilities P_i and P_j for two classes of response and non-response are introduced, including $(1 - P_i)$ $(1 - P_j)$, $(1 - P_i)\ P_j$, $P_i\ (1 - P_j)$, and P_iP_j. Thus far the fit of probabilities is limited to dichotomous response categories.

4. The model is fitted by multiple regression of the observed cell frequencies, instead of an exact solution with the number of unknowns equal to the number of degrees of freedom in the system. This may help to cope with problems of fallible data.

5. No conclusions are yet drawn concerning the latent structure of factors to account for the segment probabilities of media exposure. This is work for the future. Doubtless an endless variety of multiple factor structures can be investigated in relation to the segment probabilities which are calculated. All types of behavioral response, including brand behavior, seem amenable to latent structure analysis.

Conclusion

The present inquiry was undertaken to establish the probabilities of exposure to magazines for population segments.

The segments are latent in that their exposure probabilities are computationally inferred, not actually observed. These segments are useful as a means of representing the population of magazine readers or any population exhibiting dichotomous responses of various kinds.

The conceptual representation of latent segments in terms of probabilities of response is a useful starting point in analyzing the relationship of segment response to other population variables.

One form of such analysis is to infer the structure of factors which accounts for the segment exposure probabilities. Another line of analysis is to relate exposure probabilities to observed personality and population variables. One practical application is to use the exposure probabilities for segments to project reach and frequency of exposure for media schedules with multiple insertions over a period of time. This is also a test of the validity of latent structure analysis. Another application is to establish more precise trends in brand buying behavior for evaluating market sales. The machine program is available to others on a shared cost basis. There are many opportunities to be explored.

REFERENCES

Benson, P. H. Bivariate normal distribution to calculate media exposure. Journal of Advertising Research, 1969, 9, 41-47.

Greene, J. and Stock, J. S. Advertising Reach and Frequency in Magazines. New York: Readers Digest Association, 1967.

Hyett, G. P. The measurement of Readership. Paper read to the Statistics Seminar, London School of Economics, February, 1958.

Stouffer, S. A., Guttman, L., Suchman, E. A., Lazarsfeld, P. F., Star, S. A., and Clausen, J. A. Studies in Social Psychology in World War II, vol. 4, Measurement and Prediction. Princeton: Princeton University Press, 1950; Lazarsfeld, P. F. and Henry, N. W. Latent Structure Analysis. New York: Houghton-Mifflin Co., 1968.

Response Bias in Reports of Father's Education

Michael E. Borus, Michigan State University
Gilbert Nestel, Ohio State University*

1. INTRODUCTION

Father's educational attainment is an important explanatory variable in studies by sociologists, educationists, and economists. It has been shown to be a significant predictor of youngsters' intergenerational occupation mobility [1] educational attainment [4,5] and success in school [7]. Yet the predictive power of this variable may be related to the inaccurate responses individuals (usually youth) give to questions on their fathers' educational attainment. Systematic response biases have been found in a number of other common variables secured from surveys [e.g., 2,6,9].

The evidence on response errors in estimates of father's education is very limited. Nonresponse rates appear to vary with age, with the lowest rates of refusal among teenagers and young adults.[1] In the only study we could find where the extent of the response error in reporting this variable was measured, Blau and Duncan concluded that there was ". . . no general tendency for the OCG respondents to exaggerate the attainment of their fathers, except that considerable numbers of OCG respondents appear to have classified their fathers as high school graduates when they should have been reported as completing only one to three years of high school" [1, p. 15]. This finding must be considered very tentative, however, since it was based on a rather circuitous and complicated estimation procedure.[2] Furthermore, Blau and Duncan presented only aggregate data and did not search for systematic biases which might be offsetting such as have been found in earnings data (2).

In this paper we make use of a unique sample of households in which the father and son were interviewed separately at different times within the same year and asked to report the number of years of schooling completed by the father. After a brief description of this data we turn to a summary of the bivariate relationships between each of three different measures comparing the two responses and selected characteristics of the son, his family, and his community. We then highlight some of the findings from two different multivariate techniques used to quantify the extent of these relationships and to assess the relative importance of the different explanatory factors. The final section provides an overall summary of these findings and some suggestions for further research.

2. THE DATA

The data for this study were obtained from two of the four National Longitudinal Surveys (LGS) of the civilian noninstitutional population in the United States. Each of the samples was selected by the Bureau of the Census under a contract with the Department of Labor with the initial interviews of the two relevant age-sex cohorts for this study--approximately 5,000 young men between the ages of 14 and 24 years and a comparable number of men between 45 and 59 years of age--completed in 1966. In each of the four surveys nonwhites were oversampled and comprised about 30 percent of the total sample.[3]

In an attempt to contain the costs of administering these surveys over a five-year period the Bureau of the Census allowed certain households to be represented in more than one cohort group. In the initial survey of the older men, one out of three households had at least one additional member represented in one of the four age-sex cohort samples, while three out of four households in the young boys' sample were multiple-respondent households. In total, there were 931 households, consisting of 936 men and 1167 boys, where at least one man and one boy in the same household were interviewed. Among the possible man-boy combinations, 1,013 were found to be father-son relations.[4] The sample was then reduced by 44 cases because the father failed to report his educational attainment. In another 56 cases the son did not report the years of schooling completed by his father and these observations also were eliminated. The remaining 913 cases contained father-son responses to similarly worded questions on the number of years of schooling completed by the father. Each of the respondents was also asked to report on selected demographic, economic, and geographical characteristics. Thus, it was possible to study not only the extent of the response discrepancy to the educational attainment variable but also to identify some of its correlates.

3. BIVARIATE ANALYSIS

The extent of the discrepancy between the son's reporting of his father's educational attainment and the father's response is measured in three ways: by the arithmetic difference in the two reported responses; by the absolute value of the response difference; and by an indicator variable which takes the value of "1" if the son and father report identically and a value of "0" if the responses differ. We began our analysis by examining the gross relationship between each of these dependent variables and a variety of possible explanatory variables (these findings are presented in columns 3, 5, and 7 of Table 1.)

Even though we do not provide in this paper a formal theory of the incidence and magnitude of the response variation from which a set of contributory variables can be identified we nevertheless expect to find that the discrepancy between what the son reports and his father's response is related to the various characteristics listed in the previous section. Among the demographic factors considered are the age of the son, his father's age, the differences between the two ages, the size of the household, and the son's color. Also included in this category are educational status of the son, the number of years of schooling he has completed, and his estimate of the educational attainment of his father. The geographic place of residence of the household and whether it is located in a central city of an SMSA are considered. The list of economic factors includes the labor force status of the son, the occupational category of the job his father held for the longest period in the 12 months prior to the son's interview, and the son's estimate of his family's income.

The strong color difference in responses among the respondents is perhaps the single most significant finding from this preliminary analysis. Whereas in the total sample 61 percent of the 913 father-son responses are identical, among the 686 whites it is 68 percent, while among the 211 blacks it falls to 37 percent. The greater likelihood of a response discrepancy among the blacks also suggests that the distribution of the latter responses will be more variable and thus on the average contain the larger error.[7] This is what we find in Table 1; the mean absolute discrepancy for this color group is about one year, which is double the magnitude found among the whites.

The likelihood of a discrepancy in response is substantially reduced at what are traditionally called terminal attainment points. The likelihood of a father-son match is about three of four cases for sons who report 8, 12, or 16 or more years of schooling completed by their fathers. The corresponding probability of a match for other reported attainments of the father is considerably smaller and varies between four or five of every ten comparisions. The lowest likelihood of a match is found among sons who report their fathers' educational attainment as less than eight years of schooling (Table 1).

We also found the difference in responses to be associated with several other characteristics. Whereas the mean arithmetic difference is less than one-tenth of one year when the age difference between the father and his son is 21 to 25 years it increases to almost four-tenths of a year when the age difference is between 41 and 45 years. Households with family sizes of at least 10 members (including father and son) are considerably more likely to have fathers and sons report differently and to have larger mean discrepancies in responses than smaller households. Sons who report their fathers employed in white-collar occupations are much more likely to agree with their fathers' responses on educational attainment, while those who say their fathers are in service occupations are most likely to report differently. Finally, boys who were enrolled in school were more likely than those not enrolled to match their fathers' responses.

The bivariate analyses, however informative and suggestive, fail to control for the interrelationships among the set of explanatory variables. For example, since color, household size, and educational attainment are intercorrelated, it is difficult to distinguish which of these variables, if any, is significant. To redress this limitation, we also analyze the data using multivariate techniques.

4. MULTIVARIATE ANALYSIS

Two multivariate techniques were used--Multiple Classification Analysis (MCA) and the AID program developed by Morgan and Sonquist [8]. The MCA program assumes an additive relationship among the variables with the parameters estimated by ordinary least squares procedures.[9] The AID analysis is designed to uncover nonadditive or interaction effects among the factors and is a sequential application of a one-way two-factor analysis of variance test.[10]

4.1 Additive Specification--MCA Findings

The results of the MCA analysis do not drastically change the findings deduced from the bivariate relationships (See columns 4, 6, and 8 of Table 1). The singular importance of the color variable still prevails even after statistically controlling for the other specified factors. The mean absolute discrepancy among blacks is now a little less than one year but still double the mean discrepancy reported by the whites. The adjusted likelihood of a father-son match remains at about two in three for the whites but increases to almost one in two for the blacks.

The likelihood of a mismatch continues to be above the sample mean where the son reported his father's educational attainment as less than 8 years, 9 to 11 years, or 13 to 15 years and in the case of a response in the 0-7 year category it is 22 percentage points below the mean. Similarly, the higher-than-average likelihood of a match if the son reported his father's educational attainment level at a transition point is also evident; in almost three of every four cases the father reported the identical number of years of schooling completed. Finally, the mean absolute discrepancy in response, while expected to be larger at nontransition points, also continues to be largest among sons who reported their fathers' educational attainment as less than eight years. For these respondents the adjusted mean absolute discrepancy is about one year, which is about one and one-half times the sample average.

The less-than-average likelihood of a father-son match for households with 10 or more members and the above-average mean discrepancy in response is not substantially altered by the multivariate findings. Whereas in the entire sample about three of five father-son responses are identical, among the largest-sized households only two of five responses coincide. Similarly, whereas in the entire sample the mean absolute discrepancy is about three-fifths of a year among these households it is slightly more than one year.

The larger-than-average likelihood of a father-son match that was observed in the bivariate relation when the son reported his father as employed in a white-collar occupation is no longer evident with the introduction and control of other factors. In contrast, the low likelihood of a match when the father is reported in a service occupation, while increased as a result of the multivariate analysis, is still about 10 percentage points below the average for the entire sample.

There continues to be some evidence that boys who are enrolled in school at the time of interview are more likely than those not enrolled to report in the same way as their fathers do. The adjusted likelihood of a matching response for boys out of school is about nine percentage points below the likelihood for those in school. Once again the higher likelihood of a match is accompanied by a smaller mean absolute difference.

The coefficient of multiple determinantion adjusted for the number of explanatory variables is a summary measure frequently used to evaluate the goodness of fit of a statistical relation. Since each specification attempts to answer a different question,[11] it is not too surprising to find that the regressor variables explain only 4.5 percent of the total variance in the arithmetic difference formulation, approximately 10.3 percent of the variation when the dependent variable is the absolute difference in response, and 19.3 percent of the total variance in the linear probability model.

4.2 Nonadditive Specification--AID Findings

The results of the AID analysis further highlight the importance of color and the son's estimate of the educational attainment of his father and help to identify other variables which interact with these factors to explain the degree of variation in the father-son responses to the number of years of schooling completed by the father. In both the absolute difference and the likelihood of a matching response specifications the color of the son is the basis for the initial split of the sample. With the latter dependent variable the two color subgroups are then further dichotomized by the educational attainment of the father as reported by his son. As expected, the overall likelihood of a match among the blacks (0.374) is substantially reduced when the sons reported their fathers' attainments as less than eight years (0.313) and it is increased above the group average if they report attainments of eight or more years (0.544). No other specified variable has sufficient explanatory power to split these latter two subgroups further.

The pervasive importance of the son's estimate of his father's educational attainment is also evident in the subsample of nonblacks (whites and other-than-blacks). The sample is split repeatedly by this variable and the terminal groups highlight the greater likelihood of a matching response when the son reports one of the terminal attainment points. The likelihood of a match for this color group varies from about two in very five cases when the son reports his father's education as between nine and eleven years to about four of five cases if he reports an attainment of eight years.

The likelihood of a match among the nonblacks where the son reports his father has completed at least a high school education is also affected by the age of the father and the occupation group he is reported employed in by his son. The likelihood of a match is inversely related to the age of the father and is significantly reduced below the group average if the father is reported employed in the white-collar occupation.

There is additional evidence that place of residence also affects the likelihood of a match for the sample of nonblacks where the sons report the lowest educational attainments for their fathers. The likelihood of a father-son match is only one in five for nonblack households living in central cities of SMSA but greater than six in ten for those who live outside the central city or do not live in an SMSA.

As we indicated earlier the difference in color is also the basis for the initial split of the sample when the dependent variable is defined as the absolute difference in response. In this specification, however, the geographic locational characteristics of the blacks and nonblacks affect the magnitude of the absolute response difference. This magnitude is also influenced by the size of the family income of the blacks and the age of the father and the labor force status of the son among the nonblacks. For example, the mean absolute error is less than one-third of a year for nonblack households not located in the largest populated urban areas and where the son was not employed at the time of interview, while it is almost two years for blacks who live in a noncentral city of an SMSA.

5. CONCLUSIONS

The findings from the multivariate analysis highlight the greater than average likelihood of a response difference and a larger mean discrepancy in reporting among blacks than nonblacks and among boys who did not report their fathers' educational attainments as one of the terminal points. We also found that boys from very large households, who reported their fathers employed in service occupations, or who were not enrolled in school, were also more likely than others to respond differently from their fathers. There is also evidence of an interaction effect between color and locational characteristics of place of residence on the size of the mean absolute discrepancy.

These findings indicate the need for caution when using a son's report of his father's educational attainment. Moreover, our estimates may be understated since we have limited the comparisons to fathers and sons living in the same households. In addition, since this analysis is restricted to specific father-son age categories the findings may not generalize to other age group comparisons. Thus there is a need to replicate this study for other universes and also to determine a set of "correction factors" which may be used to adjust the responses of the sons to improve the predictive power of this variable.

FOOTNOTES

*The authors are Associate Professor of Labor and Industrial Relations, Michigan State University and Research Associate, Center for Human Resource Research, The Ohio State University, respectively. This paper was written while Dr. Borus was Research Associate, Center for Human Resource Research and Visiting Associate Professor of Economics, The Ohio State University.

[1]Bowles and Levin report that 50 percent of first graders, 40 percent of third graders, 41 percent of sixth graders, 21 percent of ninth graders and 11 percent of twelfth graders did not respond when asked to report their fathers' educational attainment [3, p. 7]. Blau and Duncan found that for adult males the percentage of nonreporters increased with age, rising from 5.4 percent for those 20 to 24 years of age to 18.2 percent for men aged 55 to 64 years [1, p. 472].

[2]Blau and Duncan arrayed their sample by year of birth, used vital statistics records to estimate the distribution of father's year of birth, and compared the educational attainment reported in the 1940, 1950, and 1960 Censuses for each age cohort with the distribution

as reported by the respondents. For a complete description of this process see Blau and Duncan [1, p. 463-66].

[3] A more complete description of these is found in Parnes, et al., [10, 11].

[4] In order to qualify as a potential match the man had to report that he had at least one son in the household and the boy had to list his father as a member of his household. The son's age, as reported in the man's household record, also had to agree with the age reported by the boy when he was interviewed. In addition, the father's age reported in his son's record had to coincide with the age reported by the man. If these criteria were satisfied then a father-son relation was established.

The surveys of the men and boys were not conducted simultaneously, however. The boys were interviewed in October and November of 1966 and the men in June of the same year. Thus the age of the father when reported by his son could exceed by one year the age that the father reported. Similarly, the father could understate his son's age by one year because of the different dating of the two interviews. Our matching procedure allowed for these possibilities.

In households where more than one boy (man) was interviewed each man-boy combination was treated separately. Thus to the extent that there were multiple father-son relations in the same household they were considered as individual observations in the analysis.

[5] The overall nonresponse rate--5.8 percent--was expected to be small because of the age distribution of the boys and the fact that they were living in the same households as their fathers. Higher than average rates were found among boys whose fathers reported between one and four years of schooling completed, among the nonwhites, among boys with six to eight years of schooling, and among boys whose fathers were 55 years and older.

[6] The independence in reporting was made possible by the four-month interval between the father-son interviews. There is the possibility that during this period the father completed an additional year of formal schooling. We believe, however, that this likelihood is very small, particularly for the age group of men and the short time interval under consideration, and we discount, therefore, this possibility.

[7] We have been deliberately careful in this discussion to avoid any inference that the father's response is necessarily accurate or even that it is likely to be more accurate than what his son reports. To the extent that the father reports his educational attainment inaccurately it would not be too surprising to find that he selects one of the terminal educational points as a response (if you include 104 respondents in the 16 years or more category, better than one-half of the fathers reported their educational attainments as 8, 12, or 16 years. Alternatively, we are suggesting the possibility that both respondents report inaccurately. Unfortunately, our data do not provide a means to test for the accuracy of the father's response.

[8] It is also of interest to observe that the conditional distribution of the educational attainment of the father as reported by his son is skewed to the right. The son's response is necessarily bounded from below by zero for very low attainments reported by his father. Nevertheless, even when the alternatives are more symmetrical (educational attainments of the father between 8 and 10 years) the son's response is more likely to exceed than understate that of his father. It is not too surprising to find, therefore, that the mean discrepancy in the sample is positive.

[9] It needs to be mentioned that the traditional "t" and "F" tests of maintained hypotheses involving linear combinations of the unknown parameters are somewhat suspect in this study in two of the three specifications. The frequency distribution of the absolute discrepancy measure is clearly asymmetrical, positively skewed, and bounded from below by zero. Since the least squares estimators in this case are weighted sums of nonsymmetrically distributed variables (the weights depend on the data matrix) one has to defer to the central limit theorem to argue that in repeated sampling these estimators will be normally distributed. The speed by which this convergence takes place, particularly when the universe is finite and the observations not independent, is not known, however.

In the case of the linear probability model (or discriminant function) the dependent variable is binomially distributed and the estimators are known to be inefficient unless a weighted least squares estimation procedure is used (however, the estimated probabilities may exceed one or fall below zero in which case further complications are introduced). If one also adds to these statistical complications the fact that our sample was selected by a multistage probability design, and therefore the traditional standard error formulas based on simple random sampling may be in error, and that the universe sampled is finite, there is some justification on our part to de-emphasize throughout this paper all tests of hypotheses.

[10] At each stage in the analysis the algorithm searches the data to identify a binary split of the codes of one of the variables which among the class of all binary partitions for all specified variables reduces the error sums of squares of the dependent variable by the largest amount. The search procedure continues until there are no eligible groups remaining which have sufficient sample cases or where the total sums of squares in each of the eligible groups do not exceed some earlier assigned constant. The program also terminates if the between sums of squares of the maximum partition of a variable fails to exceed another preassigned constant. In this study no group is eligible to be "split" unless it contains at least 10 sample cases and the total sums of squares in the group is at least one-tenth of 1 percent of the total sums of squares in the sample. No binary partition of a variable is allowed unless the between sums-of-squares associated with this partition exceeds six-tenths of 1 percent of the total sums of squares in the sample.

[11] It should be recalled in this context that we have standardized the set of regressor variables in each of the specifications and varied only the definition of the dependent variable. The specification involving the arithmetic difference in father-son response involves both a sign and magnitude consideration whereas the absolute difference abstracts from the direction of the discrepancy while still retaining the metric of the difference. In contrast, the linear probability formulation asks only whether or not the father and son respond identically. The later specification thereby abstracts from both the sign and magnitude of the discrepancy.

REFERENCES

[1] Blau, Peter M., and Duncan, Otis Dudley, The American Occupational Structure, New York: John Wiley & Sons, Inc., 1967.

[2] Borus, Michael E., "Response Error in Survey Reports of Earnings Information," Journal of the American Statistical Association, 61 (September 1966), 729-38.

[3] Bowles, Samuel, and Levin, Henry M., "The Determinants of Scholastic Achievement--An Appraisal of Some Recent Evidence," The Journal of Human Resources, 3 (Winter 1968), 3-24.

[4] Coleman, James S., et al., Equality of Educational Opportunity, Washington, D.C.: U.S. Office of Education, 1966.

[5] Conlisk, John, "Determinants of School Enrollment and School Performance," The Journal of Human Resources, 4 (Spring 1969), 140-57.

[6] Ferber, Robert, The Reliability of Consumer Reports of Financial Assets and Debts, Studies in Consumer Savings Number 6, Urbana, Illinois: University of Illinois Bureau of Economic and Business Research, 1966.

[7] Masters, Stanley H., "The Effect of Family Income on Children's Education: Some Findings on Inequality of Opportunity," The Journal of Human Resources, 4 (Spring 1969), 158-75.

[8] Morgan, James, and Sonquist, John A., The Detection of Interaction Effects, Ann Arbor: Survey Research Center, University of Michigan, 1964.

[9] Neter, John, and Waksberg, Joseph, Response Errors in Collection of Expenditures Data by Household Interviews: An Experimental Study, Technical Paper No. 11, U.S. Bureau of the Census, Washington, D.C.: U.S. Government Printing Office, 1965.

[10] Parnes, Herbert S., Miljus, Robert C., Spitz, Ruth S., and Associates, Career Thresholds, Vol. I, Manpower Research Monograph No. 16, U.S. Department of Labor, Washington, D.C.: U.S. Government Printing Office, 1970.

[11] Parnes, Herbert S., Fleisher, Belton M., Miljus, Robert C., Spitz, Ruth S., and Associates, The Pre-Retirement Years, Vol. I, Manpower Research Monograph No. 15, U.S. Department of Labor, Washington, D.C.: U.S. Government Printing Office, 1970.

Table 1 Gross and Net Mean Difference in the Education Attainment of the Father as Reported by Father and Son, by Selected Characteristics and Type of Discrepancy, 1966

Type of discrepancy

Characteristics	Sample cases		Mean arithmetic difference		Mean absolute difference		Likelihood of identical responses	
	N (1)	% (2)	Gross relationship (3)	Net relationship (4)	Gross relationship (5)	Net relationship (6)	Gross relationship (7)	Net relationship (8)
Total	913	100.0	0.239	0.239	0.631	0.631	0.611	0.611
Age difference (years)								
21-25	22	2.4	0.091	-0.373	0.545	0.453	0.545	0.545
26-30	240	26.3	0.217	0.079	0.500	0.521	0.679	0.627
31-35	409	44.8	0.311	0.290	0.682	0.697	0.604	0.598
36-40	203	22.2	0.108	0.285	0.670	0.634	0.576	0.631
41-45	39	4.3	0.384	0.794	0.744	0.691	0.487	0.584
Color (son)								
White	686	75.1	0.172	0.175	0.504	0.542	0.682	0.654
Black	211	23.1	0.450	0.448	1.056	0.925	0.374	0.471
Other	16	1.8	0.313	0.222	0.438	0.554	0.688	0.644
Household size								
2	8	0.9	0.250	0.343	0.500	0.255	0.500	0.622
3	140	15.3	0.271	0.285	0.614	0.684	0.611	0.564
4-6	558	61.1	0.142	0.140	0.543	0.573	0.658	0.636
7-9	142	15.6	0.331	0.355	0.697	0.610	0.570	0.636
10-26	65	7.1	0.800	0.720	1.292	1.107	0.308	0.443
Father's educational attainment[b]								
0-7	205	22.5	0.132	-0.083	1.010	0.939	0.370	0.392
8	153	16.8	-0.020	-0.035	0.386	0.387	0.732	0.731
9-11	140	15.3	0.407	0.384	0.821	0.792	0.436	0.440
12	240	26.3	0.342	0.311	0.458	0.491	0.783	0.775
13-15	61	6.7	0.262	0.454	0.623	0.668	0.508	0.509
16+	114	12.5	0.342	0.528	0.412	0.482	0.789	0.764
Father's occupation group[b]								
White-collar	292	32.0	0.188	0.079	0.476	0.592	0.702	0.605
Blue-collar	418	45.8	0.225	0.236	0.689	0.689	0.572	0.609
Service	48	5.3	0.542	0.512	1.000	0.819	0.417	0.495
Farm worker	104	11.4	0.231	0.447	0.673	0.727	0.577	0.613
Armed forces	1	0.1	a	a	a	a	a	a
NA	50	5.5	0.380	0.516	0.620	0.489	0.660	0.771

Table 1 (continued)

Characteristics	Sample cases		Type of discrepancy: Mean arithmetic difference		Mean absolute difference		Likelihood of identical responses	
	N (1)	% (2)	Gross relationship (3)	Net relationship (4)	Gross relationship (5)	Net relationship (6)	Gross relationship (7)	Net relationship (8)
Educational status (son)								
Enrolled	697	76.3	0.195	0.220	0.585	0.603	0.641	0.632
Not enrolled	216	23.7	0.380	0.301	0.778	0.721	0.514	0.545
Educational attainment (son)								
0-7	51	5.6	0.373	0.302	0.765	0.411	0.549	0.774
8	94	10.3	0.213	0.262	0.766	0.685	0.511	0.571
9-11	426	46.7	0.221	0.269	0.662	0.654	0.598	0.594
12	172	18.8	0.279	0.218	0.523	0.519	0.662	0.670
13-15	145	15.9	0.206	0.153	0.552	0.726	0.676	0.574
16+	25	2.7	0.280	0.154	0.520	0.702	0.600	0.540
Residence								
Urban: one million or more	249	27.3	0.313	0.290	0.707	0.732	0.594	0.590
Urban: 250,000-999,999	120	13.1	0.417	0.337	0.750	0.762	0.542	0.542
Urban: less than 250,000	76	8.3	0.395	0.369	0.789	0.775	0.618	0.673
Urban: outside urban	135	14.8	0.178	0.226	0.504	0.593	0.644	0.574
Rural	332	36.4	0.108	0.141	0.548	0.491	0.633	0.652
NA	1	0.1	a	a	a	a	a	a
SMSA								
SMSA central city	258	28.3	0.333	0.164	0.736	0.527	0.554	0.612
SMSA noncentral city	300	32.9	0.337	0.372	0.670	0.758	0.630	0.579
Not SMSA	355	38.9	0.087	0.181	0.521	0.599	0.637	0.638
Age (son) years								
14-15	257	28.1	0.206	0.092	0.634	0.604	0.588	0.571
16-17	281	30.8	0.206	0.199	0.676	0.703	0.612	0.596
18-19	207	22.7	0.227	0.268	0.546	0.606	0.647	0.619
20-21	84	9.2	0.381	0.478	0.619	0.555	0.679	0.747
22-23	65	7.1	0.308	0.500	0.708	0.640	0.492	0.592
24	19	2.1	0.421	0.538	0.632	0.507	0.632	0.763
Family income[b]								
Under $1,000	18	2.0	-0.444	-0.329	0.444	0.366	0.667	0.731
1,000-2,999	71	7.8	0.451	0.382	0.958	0.576	0.366	0.569
3,000-4,999	91	10.0	0.264	0.233	0.725	0.561	0.549	0.6337
5,000-7,499	175	19.2	0.074	0.149	0.657	0.647	0.634	0.660
7,500-14,999	370	40.5	0.254	0.240	0.514	0.576	0.665	0.623
15,000 or more	135	14.8	0.400	0.362	0.681	0.840	0.615	0.519
NA	53	5.8	0.170	0.229	0.698	0.713	0.566	0.580
Labor force status (son)								
Employed	489	53.6	0.303	0.284	0.716	0.680	0.558	0.577
Unemployed	51	5.6	0.235	0.222	0.667	0.613	0.647	0.686
Out of labor force	243	26.6	0.107	0.128	0.428	0.498	0.704	0.657
Never worked	130	14.2	0.246	0.282	0.677	0.701	0.623	0.627
Age of father (years)								
45-47	272	29.8	0.199	0.339	0.522	0.617	0.665	0.630
48-50	252	27.6	0.258	0.243	0.663	0.614	0.619	0.642
51-53	212	23.2	0.307	0.305	0.618	0.602	0.604	0.609
54-56	108	11.8	0.250	0.137	0.898	0.823	0.481	0.510
57-59	69	7.6	0.101	-0.213	0.565	0.539	0.594	0.590

a Means are not presented when number of sample cases is less than 5.

b Reported by son.

FACTORS WHICH INFLUENCE THE RELIABILITY OF HUMAN JUDGMENTS OF STRESS

D. R. Brogan, University of North Carolina at Chapel Hill

Purpose of Investigation

It was desired to estimate the reliability with which persons could make judgments regarding the stress produced in them by various newspaper headlines. One reason for such estimation was that persons were to be selected from a North Carolina rural community to act as judges of stress for certain segments of the community. Naturally, persons selected to act as judges should be able to make reliable judgments regarding situations which produce stress. In addition, it was desired to investigate the factors which contribute to a person's having a high or low reliability in these judgments of stress. The factors reported on here are demographic characteristics of the persons.

Measuring Instrument

Two questionnaires were administered by personal interview to each subject in the sample in spring, 1970. There was a one month time span between the two interviews. The first questionnaire contained one hundred newspaper headlines. Some examples are "Eight per cent interest rate approved,""Impeach President Nixon," "A nuclear war threat is seen," "Franklinton mayor is returned to jail," "North Carolina schools plan various fee hikes." There was an equal representation of local, state, and national headlines in these 100 headlines.

Subjects were asked to rate each newspaper headline on a five point scale ranging from stressful (1) to reassuring (5), with the midpoint (3) used for a neutral reaction, i.e. neither stressful nor reassuring. In addition to the subject's responses to the 100 newspaper headlines, demographic data were collected on each subject.

The second questionnaire, administered one month later, contained fifty newspaper headlines from the original set of 100 on the first questionnaire. Keeping a fairly equal representation of national, state, and local headlines, the least ambiguous headlines in the first questionnaire were selected for the second questionnaire. The second questionnaire also contained questions about the stress produced by various life events such as a new birth in the family, a change in job, etc. The analysis of this life event data is reported elsewhere.[1]

Subjects

A previous study of the county had drawn a random sample of households, and then a random sample of one adult within each household.[2] The sample thus obtained was judged to be representative of the county, based on 1960 census data. The sample for this investigation was drawn by selecting a household adjacent to the household selected in the previous study, and then selecting at random one adult within each household. By this method, 111 adults were selected for this investigation.

In the second interview, 96 persons out of the original 111 were followed up, and these 96 persons constitute the sample for this investigation. The following are the demographic characteristics of this sample: 88% are rural residents, while the other 12% are residents of a small town. 63% are white; 51% are female. 73% have lived in the community for over 20 years, reflecting a very stable population. 10% are in the age range 20-29, 25% in the range 30-39, 27% in the range 40-49, 21% in the range 50-59, and 17% are 60 and above. 79% are married, 12% are widowed, 5% are single, and 4% are divorced or separated. 13% have some college education, 33% have a high school education, 38% have a junior high school education, and 16% have less than 6 years of formal schooling. 5% are employed in professional occupations, 10% in white collar occupations, 10% in skilled occupations, and 19% in semi-skilled occupations. 20% are farm laborers, and 32% are housewives. 23% have an annual income less than $3000, 33% have an annual income of $3000 to $6000, 28% have an annual income of $6000 to $9000, and the remaining 16% have an annual income of $9000 or over. 63% of the subjects are heads of households.

Measures of Reliability

The most commonly used measure of reliability is a correlation coefficient. Hence, a correlation coefficient between the responses on interview 1 and interview 2 was obtained for each of the 96 subjects by adding over the responses to the fifty newspaper headlines common to each interview. Letting X_{ijk} denote the response of person i (i=1,...,96) to newspaper headline j (j=1,...,50) at time k (k=1,2), then the measure of reliability COR for person i is given by

$$COR(i) = \frac{\sum_{j=1}^{50} (X_{ij1}-\bar{X}_{i.1})(X_{ij2}-\bar{X}_{i.2})}{\left[\sum_{j=1}^{50} (X_{ij1}-\bar{X}_{i.1})^2 \sum_{j=1}^{50} (X_{ij2}-\bar{X}_{i.2})^2\right]^{1/2}}$$

where

$$\bar{X}_{i.k} = \frac{1}{50}\sum_{j=1}^{50} X_{ijk}, \quad \text{for } k=1,2.$$

Note that this is not the structure of a correlation coefficient in the usual sense because, for a given person, the responses to the 50 newspaper headlines are probably not independent of one another. However, it is used here for a reliability measure in much the same way that it is used in cluster analysis as a measure of similarity between two units.[3] In the sample of 96 subjects, COR ranged from a low of .12 to a high of .99. Table 1 gives the frequency distribution of the values of COR. For this measure, a low score means low reliability and a high score means high reliability.

The second measure of reliability is the Fisher-z transformation of the correlation coefficient discussed above, i.e. FISH. For a given value of COR,

$$FISH = 1/2 \ln \frac{1+COR}{1-COR} .$$

The Fisher-z transformation was considered as a potential reliability measure because it is approximately normally distributed and, perhaps then, the distribution of the 96 values of FISH would be approximately normally distributed. In the sample of 96 subjects, the value of FISH ranged from .12 to 2.63. Table 1 gives the frequency distribution of the values of FISH. For this measure, a low score means low reliability and a high score means high reliability.

Correlation coefficients, although used extensively as reliability coefficients, aren't always the best indication of reliability since they only measure the linear relationship between two variables. For example, in this study, a subject could have answered each of the 50 newspaper headlines on the first interview with a judgment of 1 (stressful), and all 50 headlines on the second interview with a judgment of 5 (reassuring). The reliability, as measured by COR, would be 1.0. This doesn't sound intuitively reasonable, since this subject is making very different judgments on the same headline at the two interview times. One way to measure reliability without this drawback is to use a squared distance measure, also common to the field of cluster analysis.[3] The sum of squared deviations for each person i is given by

$$SSD(i) = \sum_{j=1}^{50} (X_{ij1}-X_{ij2})^2.$$

In the sample of 96 persons, this value ranged from a low of 3 to a high of 155. In order to use this measure so that a low value means low reliability, an adjusted squared deviation measure (ASD) is defined by

$$ASD = 1 - SSD/155.$$

This measure ranged from a low of 0.0 to a high of .98 in the sample of 96 persons. Table 1 gives the frequency distribution of ASD.

Table 2 gives the correlation coefficients between these three measures of reliability. In general, they are highly correlated with one another. Hence, it probably doesn't make too much difference which measure is used to select persons from the sample who would make "good," i.e., "reliable," judges. ASD is recommended, however, since it correlates highly with the other two and, in addition, has more of an intuitive appeal for this particular investigation.

Independent Variables

The independent variables which were used in an attempt to predict the degree of reliability are:

1) Residence (1=rural, 2=small town)

2) Race (1=white, 2=nonwhite)

3) Sex (1=male, 2=female)

4) Marital Status (1=not married, 2=married)

5) Head of Household (1=yes,2=no)

6) Occupation
   1=professional
   2=white collar
   3=skilled
   4=semi-skilled
   5=unskilled

7) Years Lived in Community
   $1= \le 1$ year
   $2= 1 <$ years ≤ 3

3=3 < years ≤ 6
4=6 < years ≤ 10
5=10 < years ≤ 20
6=years > 20

8) Age (1=20-29, 2=30-39, 3=40-49, 4=50-59, 5=60 and above)

9) Education
1=some college or college graduate
2=high school graduate
3=less than high school, but more than junior high
4=less than 6 years or no schooling

10) Annual Family Income
1=income < 3000
2=3000 ≤ income < 6000
3=6000 ≤ income < 9000
4=9000 ≤ income < 15,000
5=income ≥ 15,000

Note that variables 1 thru 5 are dichotomous, nominal variables. Variables 6 thru 10 are categorized, ordinal variables.

Method of Analysis

The three dependent variables COR, ASD, and FISH were considered separately. For each dependent variable, the 10 independent variables were used in a stepwise multiple regression analysis in an effort to determine which variables influenced the measures of reliability.

Variables 7 thru 10 could be used as indicated above, or else the median value of the category could be assigned to each person within the category. Both methods were used here, and they gave similar results.

Results

Only the results obtained by using the independent variables as indicated previously will be discussed, since the results obtained by using the midpoint of the categories for variables 7 thru 10 were very similar.

1. Regression Analysis of COR

The correlations between COR and each of the independent variables ranged, in absolute value, from .001 to .18. COR was most highly correlated with race (r=.18), head of household (r=-.13), and income (r=-.13). All of these correlations are, of course, very small. These three variables enter the stepwise regression equation in the above order and give a multiple R of .22 (multiple R^2=.05). If all independent variables are allowed into the equation, multiple R increases only to .24 (multiple R^2=.06). Hence, none of the independent variables really explain a significant amount of the variability in COR. (None of the F tests for a regression effect were significant at α=.05.)

2. Regression analysis of FISH

The correlation of FISH with the 10 independent variables ranged, in absolute value, from .06 to .23. The highest correlations were with race (r=.23), income (r=-.16), sex (r=-.11), and head of household (r=-.11). The first two variables entering the stepwise regression equation were race and sex. This yielded a multiple R of .26 (multiple R^2=.07). The F ratio, with (2,93) df, was 3.421, which is statistically significant at α=.05. However, the multiple R^2 is still comparatively small.

3. Regression analysis of ASD

The correlation of ASD with the 10 independent variables ranged, in absolute value, from .001 to .135. The highest correlations were with head of household (r=-.135) and race (r=.131). Using these two variables, multiple R was .18(multiple R^2=.03), and it was nonsignificant at α=.05.

Discussion

Only the reliability measure FISH showed a statistically significant relationship with any of the independent variables, although the multiple R was still quite small. Of interest are the correlation coefficients of highest value. For example, race had the highest correlation with the dependent variables FISH and COR and the second highest correlation with ASD. In all three cases, the correlation coefficient indicated that blacks have higher reliability coefficients than whites, although the magnitude of this relationship was statistically significant only for the reliability coefficient FISH. Head of household was another variable showing some of the highest correlations with the dependent variables. All three correlation coefficients indicated that persons who are head of households have higher reliability coefficients, although, again, the particular correlation coefficients are not statistically significant at α=.05.

In general, one can conclude that none of the independent variables are useful in predicting the degree of reliability of the judges.

FOOTNOTES

[1]Miller, F. T., W. K. Bentz, and D. R. Brogan (1971). "Perception of Life Crisis Events: A Comparative Study of Rural and Urban Americans," submitted to Cross-Cultural Psychology.

[2]Edgerton, J. Wilbert, Willard Bentz, and William Hollister (1970). "Demographic Factors and Responses to Stress Among Rural People," Amer. J. Public Health, 60, 1065-1071.

[3]Sokal, R. R. and P. H. Sneath (1963). The Principles of Numerical Taxonomy. W. H. Freeman, San Francisco, California.

Table 1

Frequency Distribution of COR, ASD, and FISH

COR		ASD		FISH	
Range	Frequency	Range	Frequency	Range	Frequency
.0-.10	0	.00-.10	4	.00-.20	1
.11-.20	1	.11-.20	8	.21-.40	2
.21-.30	1	.21-.30	5	.41-.60	8
.31-.40	1	.31-.40	18	.61-.80	23
.41-.50	6	.41-.50	18	.81-1.00	36
.51-.60	11	.51-.60	12	1.01-1.20	8
.61-.70	27	.61-.70	13	1.21-1.40	9
.71-.80	27	.71-.80	8	1.41-1.60	3
.81-.90	16	.81-.90	5	1.61-1.80	0
.91-1.00	6	.91-1.00	5	1.81-2.00	3
				2.01-2.20	0
Minimum .12		Minimum 0.0		2.21-2.40	1
Maximum .99		Maximum .98		2.41-2.60	1
Median .71		Median .46		2.61-2.80	1
				Minimum .12	
				Maximum 2.63	
				Median .88	

Table 2

Correlation Coefficients Between COR, ASD, and FISH

	COR	ASD
ASD	.859	-
FISH	.886	.840

ERRORS OF MEASUREMENT AND 'ECOLOGICAL' CORRELATION

John J. Chai, Syracuse University

1. INTRODUCTION

The term "ecological correlation" refers to the correlation coefficient based on ecological data, which means the variables describing properties of groups (e.g., averages or proportions for city blocks, enumeration districts, census tracts, etc.) [12,21].

Robinson [21] demonstrated ecological correlations cannot be used to represent individual correlations. Since then, L.A. Goodman [12] generalized Robinson's work and discussed some special cases in which ecological correlations may be used to represent individual correlations.

Despite the problems associated with ecological data, they are still used by many sociologists and scientists. The reasons for this practice may vary from case to case. Ecological data may be used because the main interest may be in studying group characteristics or the relationship between group and individual variables, or simply because the only data available are ecological data.

In most cases, the variables we observe or measure are subject to errors of measurement. Recently there has been a considerable amount of work in the study of errors of measurement [e.g., 2,3, 7,13,14,16,20,22,23].

Quite recently, Chai [4], Cochran [8], Horvitz and Koch [15], Koch [17], and Mandansky [18] have contributed to the development of the theory and application of errors of measurement in surveys beyond the univariate case.

The purpose of this paper is to study the combined effect of errors of measurement and ecological (grouped) data on estimation of the ordinary pearsonian product-moment correlation coefficient when the estimator used is based on a sample of ecological data. We present the mathematical model for the component bias factors of the estimator first and a discussion of the estimates of the component bias factors next.

2. MODEL

For the sake of simplicity, we consider a simple random sample of very large size n taken from a finite population of size N. This sample is then "interpenetrated" into M subgroups, each subgroup containing $\bar{n} = \frac{n}{N}$ elements. We assume that each of M interviewers is assigned to a subgroup and that the collection and processing of data are designed in such a way that there is no correlation between the response errors of any two units in different subgroups. This is to say that correlated errors are expected only within subgroups. We further assume that the survey is repeatable under a constant survey condition and that the finite multipliers for every subgroup $(1-\frac{\bar{n}}{\bar{N}})$ and for the entire sample $(1-\frac{n}{N})$ can be ignored.

The model shown below, under the above simple assumptions, may be an over-simplification of the real world; but the modification and/or extension of the model for more realistic survey conditions can easily be made [e.g., 4,6,13].

Let x_{ijt} and y_{ijt}, respectively, be the sample responses for the j-th individual unit of the i-th subgroup recorded at the t-th measurement. And let $\bar{x}_{it}$ and $\bar{y}_{it}$, respectively, be the sample average responses for the i-th subgroup (group) recorded at the t-th measurement.

Following Hansen et.al., [13], we write:

$$x_{ijt} = X_{ij} - d_{ijt} \qquad (1)$$

$$y_{ijt} = Y_{ij} - e_{ijt} \qquad (2)$$

where X_{ij} and Y_{ij} are the conditional expected values, i.e.,

$$X_{ij} = \underset{t}{E}(x_{ijt} \mid i,j) \qquad (3)$$

$$Y_{ij} = \underset{t}{E}(y_{ijt} \mid i,j) \qquad (4)$$

and d_{ijt} and e_{ijt} are the "response deviations" of x_{ijt} and y_{ijt}.

Suppose that we are interested in estimating the correlation coefficient of expected values for individual units (X_{ij} and Y_{ij}), ρ $(\rho=\sigma_{XY}/\sigma_X\sigma_Y)$ from a sample of grouped data.

Let the estimator of ρ be the Pearsonian product-moment formula based on the sample grouped data, i.e.,

$$r_{At} = \frac{s_{\bar{x}\bar{y}(t)}}{s_{\bar{x}(t)}s_{\bar{y}(t)}} \qquad (5)$$

where $s_{\bar{x}\bar{y}(t)}$ is the between-area sample covariance observed for the t-th trial and $s_{\bar{x}(t)}$ and $s_{\bar{y}(t)}$ respectively are the between-area sample standard deviations observed for the t-th trial.

Now, let

$$\rho_A^* = \frac{\underset{st}{EE}\, s_{\bar{x}\bar{y}(t)}}{(\underset{st}{EE}\, s^2_{\bar{x}(t)}\ \underset{st}{EE}\, s^2_{\bar{y}(t)})^{1/2}} \qquad (6)$$

Then, it is shown [5][1/] that, under the survey conditions assumed in this paper,

$$\rho^*_A = (\rho) \cdot (\varepsilon_1) \cdot (\varepsilon_2) \qquad (7)$$

where ε_1, the component bias factor due to errors of measurement only, is defined by

$$\varepsilon_1 \doteq \frac{1 + \sigma_{de}/\sigma_{XY}}{[(1 + \sigma^2_d/\sigma^2_X)(1 + \sigma^2_e/\sigma^2_Y)]^{1/2}}; \qquad (8)$$

and ε_2, the component factor due to grouping and interaction between errors of measurement and grouping, is defined by

$$\varepsilon_2 \doteq (\rho_A/\rho) \cdot (1/\varepsilon_1) \cdot \frac{(1+\sigma_{de(B)}/\sigma_{XY(B)})}{[(1+\sigma^2_{d(B)}/\sigma^2_{X(B)})(1+\sigma^2_{e(B)}/\sigma^2_{Y(B)})]^{1/2}} \qquad (9)$$

where σ^2_d, σ^2_e and σ_{de} respectively are "simple response variance and covariance" [13]; σ^2_X, σ^2_Y, and σ_{XY} are the variance and covariance of expected values for ungrouped data; $\sigma^2_{d(B)}$, $\sigma^2_{e(B)}$, and $\sigma_{de(B)}$ are the variance and covariance of the response deviations for grouped data [2/]; $\sigma^2_{X(B)}$, $\sigma^2_{Y(B)}$, and $\sigma_{XY(B)}$ are the variance and covariance of expected values for grouped data: and $\rho_A=\sigma_{XY(B)}/\sigma_{X(B)}\sigma_{Y(B)}$ is the ecological correlation coefficient for expected values (see reference [5] for further details).

Let, the third term of Equation (9) above be denoted by ε_3. We call ε_3 the bias component factor due to grouping. Furthermore, we define the interaction term, I by the ratio of ε_3 to ε_1, i.e.,

$$I = \varepsilon_3/\varepsilon_1 \qquad (10)$$

Then, from Equation (9) we have

$$\varepsilon_2 = (\rho_A/\rho)\cdot(I) \qquad (11)$$

and from Equation (7) we have

$$\rho_A^* = (\rho_A)\cdot(\varepsilon_3) \qquad (12)$$

If I = 1 (no effect due to interaction), then we have $\rho_A^* = (\rho_A)\cdot(\varepsilon_1)$ (13)

We may express Equation (7) above using the definition

$$\rho^* = \underset{st}{EE}\, s_{xy(t)}/[(\underset{st}{EE}\, s^2_{x(t)})\cdot(\underset{st}{EE}\, s^2_{y(t)})]^{1/2} = (\rho)\cdot(\varepsilon_1) \qquad (14)$$

where $s^2_{x(t)}$, $s^2_{y(t)}$, and $s_{xy(t)}$ respectively are the sample variance and covariance of observed values for ungrouped data. Hence,

$$\rho_A^* = (\rho^*)\cdot(\varepsilon_2) \qquad (15)$$

In summary, we have for grouping effect only:

$$\rho_A/\rho = (\varepsilon_1)\cdot(\varepsilon_2)/(\varepsilon_3) \qquad (16)$$

combined effect of errors of measurement and grouping:

$$\rho_A^*/\rho = (\rho_A/\rho)\cdot(\varepsilon_3) \text{ if } I \neq 1 \qquad (17.1)$$

$$\rho_A^*/\rho = (\rho_A/\rho)\cdot(\varepsilon_1) \text{ if } I = 1 \qquad (17.2)$$

3. ESTIMATION OF COMPONENT BIAS FACTORS

A detailed discussion of the estimation procedures used to estimate ε_1, ε_2 and ε_3 are given in reference [5] and a brief summary of the estimators used is given in the Appendix of this paper.

The sample estimates of the bias factors are calculated for some housing variables and are summarized in Table 1. These estimates are obtained from two different sources--(1) the 1960 Census of Population and Housing as the original data and a probability sample of 5000 housing units located in approximately 2500 area segments of the United States in October 1960 (six months after the 1960 Census) for reinterview purposes [22] and (2) the six-city sample data used for the purposes of evaluating the quality of housing units at the Bureau of the Census in 1964-65 [23].

The first set of sample data is used primarily to estimate the simple response variance components $(\sigma^2_d/\sigma^2_X, \sigma^2_e/\sigma^2_Y)$ and the covariance component $(\sigma_{de}/\sigma_{XY})$, and the second sample data was used exclusively to estimate the averages of the correlated component of response variances $(\bar{\Delta}_d, \bar{\Delta}_e)$

and covariance $(\bar{A}_{de})$[3/]. The ecological data used are for city blocks, enumeration districts (ED), and census tracts.

4. DISCUSSION

First we discuss the grouping effect only (ρ_A/ρ) (see column 3 of the table) and secondly we study the combined effect of errors of measurement and grouping (ρ_A^*/ρ) (see columns 6 and 7 of the table).

Grouping Effect

The estimates of ρ_A^*/ρ^* estimated by r_{At}/r_t (see Appendix) in Column 2 of the table reflect the estimates of the ratio of the ecological correlation to individual correlation based on observed values, which, of course, are subject to errors of measurement; whereas the estimates of ρ_A/ρ show the grouping effect only (no errors of measurement are included).

The estimates of ρ_A^*/ρ^* show results that are quite similar to the ones given by the earlier experimental works (Gehlke and Biehl [10], Robinson [21], Duncan and Davis [9], Abel and Waugh [1] and Pritzker and Selove [23]). In other words, the earlier works showed that (1) the estimates of ρ_A^* are greater than ρ^* (i.e., $\rho_A^*/\rho^*>1$) and that (2) the estimates of ρ_A^* are greater for a large group than for a small group.

However, the estimates of ρ_A/ρ do not necessarily follow the same patterns as the ones given by the estimates of ρ_A^*/ρ^*. We note first that the estimates of ρ_A/ρ are smaller than the estimates of ρ_A^*/ρ^* for most of the cases given in this study.

This is evident, since

$$\frac{\rho_A^*}{\rho^*} = \left(\frac{\rho_A}{\rho}\right)\cdot\left(\frac{\varepsilon_3}{\varepsilon_1}\right) = \left(\frac{\rho_A}{\rho}\right)\cdot(I)$$

and the estimates of the interaction term I are significantly greater than one for most cases (Column 5). In fact, a comparison of the estimates of ρ_A/ρ with those of "I" reveals the interaction effect to be stronger than the grouping effect.

Furthermore, we note that, unlike the estimates of ρ_A^*/ρ^*, half of the estimates of ρ_A/ρ given in this paper does not increase as the area size increases.

This seems to imply that the ecological correlation in the absence of errors of measurement has an attenuating effect on the ordinary estimator of $\rho(r_{At})$ rather than an inflating effect as was indicated by the earlier works.

Also, the estimates of "I" given in this paper definitely cast some doubts on the possibility that "random errors" (Yule and Kendall [24] may cancel out when individual units are grouped and when the size of the group increases.

Combined Effects [See Equations (7), (17.1), and (17.2)

Columns 6 and 7 show the estimates for the combined effect $(\varepsilon_1.\varepsilon_2)$. The estimates of $(\varepsilon_1.\varepsilon_2)$ are greater than one for all cases except one, meaning that the estimator r_{At} over-estimates ρ, on the average, for most of the cases studied.

Furthermore, the estimates of $(\varepsilon_1.\varepsilon_2)$ for larger-size groups are greater than those for small-sized groups. This indicates that the bias due to errors of measurement and grouping are increasing as the size of ecological groups increases.

It is interesting to compare the estimates of $\varepsilon_1=\rho^*/\rho$, component bias factor due to errors of measurement when no grouping is made (Column 1), with the estimates of $\varepsilon_2=(\rho_A^*/\rho).(I)$ (Column 2); for the estimates of ε_2 are greater than the ones for ε_1 in practically all of the cases considered. This, of course, suggests that the grouping (ρ_A/ρ) and interaction (I) effects are greater than the effect due to errors of measurement alone.

The estimates of the component bias factors presented above simply illustrate that the estimation of the individual correlation ρ using the Pearsonian product-moment estimator based on ecological data (r_{At}) is affected not only by grouping error but also by errors of measurement and by the interaction of the two. Although more study based on more variables are needed, this study clearly demonstrates the possible bias due to errors of measurement and to the use of the estimator of the ecological correlation coefficient for estimation of the individual (ungrouped) correlation coefficient.

Appendix
ESTIMATORS USED

Detailed account of estimation procedures is given by reference [5]. Only a brief summary of the estimators used is given below.

1. Estimator of ε_1,

To estimate ε_1, the factors

σ_{de}/σ_{XY}, σ_d^2/σ_X^2 (or σ_e^2/σ_Y^2) must be estimated. The estimator used for σ_d^2/σ_X^2 (or σ_e^2/σ_Y^2) is [4]:

$$\left(\frac{g}{2s^2_{x(T)}}\right)\cdot\left(1-\frac{g}{2s^2_{x(T)}}\right)^{-1} \quad \text{(A-1)}$$

where

$$g = \frac{1}{n}\sum_i^M\sum_j^{\bar{n}} (x_{ijt} - x_{ijt'})^2 \quad \text{(A-2)}$$

is the "gross difference rate" [14] and is the estimator of σ_d^2: and

$$s^2_{x(T)} = \frac{1}{2}(s^2_{xt} + s^2_{xt'}) \quad \text{(A-3)}$$

is the estimator of $\sigma^2_{x(T)} \doteq \sigma^2_X + \sigma^2_d$

Hence, $\frac{g}{2s^2_{x(T)}}$ is the estimator of

$\sigma_d^2/\sigma^2_{x(T)}$, "index of inconsistency" [14]. The estimator of σ_{de}/σ_{XY} is [4]:

$$\left(\frac{h}{2s_{xy(T)}}\right)\cdot\left(1-\frac{h}{2s_{xy(T)}}\right)^{-1} \quad \text{(A-4)}$$

where

$$h = \frac{1}{n}\sum_i^M\sum_j^{\bar{n}} (x_{ijt} - x_{ijt'})(y_{ijt} - y_{ijt'}) \quad \text{(A-5)}$$

is the estimator of σ_{de} (see reference [20]) and $s_{xy(T)}$ [see (A-3) above] is the estimator of $\sigma_{xy(T)} \doteq \sigma_{XY} + \sigma_{de}$.

2. Estimator of ε_2

Noting that

$$\varepsilon_2 = \rho^*_A/\rho^*$$

we use

$$\frac{r_{At}}{r_t} \quad \text{(A-6)}$$

as the estimator of ε_2. Where

$$r_t = \frac{s_{xy(t)}}{s_{x(t)}s_{y(t)}} \quad \text{(A-7)}$$

3. Estimator of ε_3

To estimate ε_3, the factors

$\sigma_{de(B)}/\sigma_{XY(B)}$ and $\sigma^2_{d(B)}/\sigma^2_{X(B)}$ (or $\sigma^2_{e(B)}/\sigma^2_{Y(B)}$) must be estimated,

where $\sigma_{de(B)}$, $\sigma^2_{d(B)}$, and $\sigma^2_{e(B)}$ are given by $\bar{\Delta}_{de}$, $\bar{\Delta}_d$, $\bar{\Delta}_e$ (see footnote 2).

The estimator of $\bar{\Delta}_d/\sigma^2_{X(B)}$ (or $\bar{\Delta}_e/\sigma^2_{Y(B)}$) is:

$$\left(\frac{\hat{\bar{\Delta}}_d}{s^2_{\bar{x}(t)}}\right)\cdot\left(1-\frac{\hat{\bar{\Delta}}_d}{s^2_{\bar{x}(t)}}\right)^{-1} \quad \text{(A-8)}$$

where

$$\hat{\bar{\Delta}}_d = \frac{\bar{n}\,(s^2_{\bar{x}(t)} - s^2_{x(t)}/\bar{n})}{\bar{n}-1} \quad \text{(A-9)}$$

is the estimator of $\hat{\bar{\Delta}}_d$ and $s^2_{\bar{x}(t)}$ is the estimator of $\sigma^2_{x(B)(T)} \doteq \sigma^2_{X(B)} + \bar{\Delta}_d$

The estimator of $\bar{\Delta}_{de}/\sigma_{XY(B)}$ is:

$$\left(\frac{\hat{\bar{\Delta}}_{de}}{s_{\bar{x}\bar{y}(t)}}\right)\cdot\left(1-\frac{\hat{\bar{\Delta}}_{de}}{s_{\bar{x}\bar{y}(t)}}\right)^{-1} \quad \text{(A-10)}$$

where

$$\hat{\bar{\Delta}}_{de} = \frac{\bar{n}\,(s_{\bar{x}\bar{y}(t)} - s_{xy(t)}/\bar{n})}{\bar{n}-1}$$

is the estimator of $\bar{\Delta}_{de}$ and $s_{\bar{x}\bar{y}(t)}$ is the estimator of $\sigma_{xy(B)(T)} \doteq \sigma_{XY(B)} + \bar{\Delta}_{de}$

Acknowledgements: The author is grateful for Professor John Neter of the University of Minnesota and Dr. Tepping of the Census Bureau for their valuable suggestions. Views expressed in this paper

are those of the author and not necessarily those of the Bureau of the Census.

1/ Reference [5] shows that:

$0[{}^{EE}_{st}\ (r_{At} - \rho^*_A)] \doteq 0(\frac{1}{M})$

2/ For the survey conditions defined here,

$\sigma^2_{d(B)} \doteq \bar{\Delta}_d$, $\sigma^2_{e(B)} \doteq \bar{\Delta}_e$, and $\sigma_{de(B)} \doteq \bar{\Delta}_{de}$, where $\bar{\Delta}_d$, $\bar{\Delta}_e$, and $\bar{\Delta}_{de}$ respectively are the average of the correlated component of the response variance and covariance (the correlated component of the response variance and covariance are based on the intraclass correlation coefficient of response deviation for subgroups).

3/ Although the sample data used to estimate $\bar{\Delta}_d$, $\bar{\Delta}_e$, and $\bar{\Delta}_{de}$ are different from the sample estimating the simple response variance and covariance components, the estimates of $\bar{\Delta}_d$ and $\bar{\Delta}_e$ obtained from the six-city data seem to show the order of magnitudes and the patterns of variation for the different ecological groups similar to the ones estimated at the Census Bureau for other variables based on a much larger scale survey [2,5].

REFERENCES

[1] Abel, Martin. and Frederick V. Waugh, "Relationships Between Group Averages and Individual Observations," Agricultural Economics Research Vol. XVIII, No. 4, October 1966, pp. 105-115.

[2] Bailer, B.A., "The Effect of Varying the Size of the Interviewer's Assignment Area on ρ_d, the Interclass Correlation Between Response Deviations of Different Units Assigned to the Same Interviewer." Unpublished article, Center for Research in Measurement Methods, U.S. Bureau of the Census, September 28, 1967.

[3] Bailer, B.A., "Recent Research in Reinterview Procedures," Journal of the American Statistical Association, 63, March 1968.

[4] Chai, John J., "Correlated Measurement Errors and the Least Squares Estimator of the Regression Coefficient," Journal of American Statistical Association, To be published in September Issue 1971.

[5] Chai, John J., A Study of Effects of Correlated Measurement Errors and Grouping on the Ordinary Least Squares Estimator for the Regression Coefficient and Ordinary Estimator of the Product-Moment Correlation Coefficient of a Finite Bivariate Population, Ph.D. Thesis, University of Minnesota, 1968.

[6] Cochran, W.G., Sampling Techniques, second ed., New York: John Wiley & Sons, Inc., 1963.

[7] Cochran, W.G., "Errors of Measurement in Statistics," Technometrics, Vol. 10, No. 4, 1968.

[8] Cochran, W.G., "Some Effects of Errors of Measurement on Multiple Correlation," Journal of the American Statistical Association, 65 (1970) No. 329.

[9] Duncan, O.D. and Davis, B., "Alternative to Ecological Correlation," American Sociological Review, Vol. 18 (1953), pp. 665-666.

[10] Gehlke, C.E. and Biehl, K., "Certain Effects of Grouping Upon the Size of the Correlation Coefficient in Census Tract Material," Journal of American Statistical Association, 29 (1934 Suppl.)

[11] Goodman, L.A. "Ecological Regressions and Behavior of Individuals," American Sociological Review, 18 (1953), pp. 663-664.

[12] Goodman, L.A., "Some Alternatives to Ecological Correlation," American Journal of Sociology, (1959) pp.610-625.

[13] Hansen, M.H., Hurtwitz, W.N., and Bershad, M.A., "Measurement Errors in Censuses and Surveys," Bulletin of the International Statistical Institute, 38 (1961), pp. 359-374.

[14] Hansen, M.H., Hurwitz, W.N., and Pritzker, Leon, "The Estimation and Interpretation of Gross Differences and the Simple Response Variance," Contributions to Statistics Presented to Professor P.S. Mahalanobis on the Occasion of His 70th Birthday, Calcutta, India: Pergamon Press, 1964

[15] Horvitz, D.G., and Koch, Gary G., "The Effect of Response Errors on Measures of Association," New Development in Sampling (edited by N.L. Johnson and H. Smith, Jr.) pp. 247-282 John Wiley & Sons, Inc., New York, 1969.

[16] Kish, L., "Studies of Interviewer Variance for Attitudinal Variables," Journal of American Statistical Association, 57 (1962) No. 297.

[17] Koch, Gary G., "The Effect on Non-Sampling Errors on Measures of Association in 2X2 Contingency Tables" Journal of American Statistical Association, 64 (1969) pp.851-64.

[18] Mandansky, A., "The Fitting of Straight Lines When Both Variables Are Subject to Error," Journal of the Royal Statistical Society, 54 (1959), pp. 173-205.

[19] Menzel, H., "Comments on Robinson's Ecological Correlations and the Behavior of Individuals, "American Sociological Review, XV (1950), p. 674.

[20] Pritzker, Leon, "Response Research Branch Report No. 66," April 27, 1966, U.S. Bureau of the Census.

[21] Robinson, W.S., "Ecological Correlations and the Behavior of Individuals," American Sociological Review, 15 (1950), pp. 115-120.

[22] U.S. Bureau of the Census, Evaluation and Research Program of the U.S. Census of Population and Housing, 1960 Accuracy of Data on Housing Characteristics, Series ER 60, No. 3, Department of Commerce, Bureau of the Census Washington, D.C., 1964.

[23] U.S. Bureau of the Census, Measuring the Quality of Housing, An Appraisal of Census Statistics and Methods. Working Paper No. 25, Department of Commerce, Bureau of the Census, Washington, D.C.. 1967.

[24] Yule, G.V., and Kendall, M.G., An Introduction to the Theory of Statistics, New York: Hafner Publishing Co., 1950.

TABLE 1. ESTIMATES OF COMPONENT BIAS FACTORS

City and Variables	Estimate of: $\varepsilon_1=\rho^*/\rho$	$\varepsilon_2=\rho^*_A/\rho^*$			$\rho_A/\rho=\varepsilon_1\cdot\varepsilon_2/\varepsilon_3$			$\varepsilon_3=\rho^*_A/\rho_A$			$I=\varepsilon_3/\varepsilon_1$			$\varepsilon_1\cdot\varepsilon_2=\rho^*_A/\rho$ when $I\neq 1$			ρ^*_A/ρ when $I=1$		
		Block	ED	Tract	Block	ED	Tract	Block	ED	Tract	Block	ED	Tract	Block	ED	Tract	Block	ED	Tract
Units deteriorating vs. Owner occupied units:																			
Camden, New Jersey	1.05	1.01	1.30	3.67	.28	.87	1.14	3.73	1.57	3.37	3.55	1.50	3.21	1.06	1.36	3.85	.29	.91	1.20
Cleveland, Ohio	1.05	2.05	2.34	3.54	.40	.51	.42	5.31	4.87	8.85	5.06	4.64	8.43	2.15	2.46	3.72	.42	.54	.44
Fort Wayne, Indiana	1.05	2.00	2.88	3.88	.58	1.73	.80	3.64	1.75	5.06	3.47	1.67	4.82	2.10	3.02	4.07	.61	1.82	.84
Louisville, Kentucky	1.05	1.46	1.88	2.84	.71	.64	.67	2.14	3.07	4.42	2.04	2.92	4.21	1.53	1.97	2.98	.75	.67	.70
Shreveport, Louisiana	1.05	1.58	2.00	2.58	.47	.43	.45	3.53	4.86	6.02	3.36	4.63	5.73	1.66	2.10	2.71	.49	.45	.47
South Bend, Indiana	1.05	1.74	2.36	3.67	.52	.66	1.60	3.49	3.74	2.41	3.32	3.56	2.30	1.83	2.48	3.85	.55	.69	1.68
AVERAGE	1.05	1.64	2.13	3.36	.49	.81	.85	3.64	3.31	5.02	3.47	3.15	4.78	1.72	2.24	3.53	.51	.85	.89
Units deteriorating vs. Units with both for exclusive use:																			
Camden, New Jersey	.66	1.72	2.33	3.55	.29	.30	**	3.90	5.04	**	5.90	7.64	**	1.14	1.54	2.34	.19	.20	**
Cleveland, Ohio	.66	1.79	2.11	2.93	.34	.23	.59	3.56	6.09	3.33	5.39	9.23	5.05	1.18	1.39	1.93	.22	.20	.39
Fort Wayne, Indiana	.66	1.19	2.08	3.00	.52	.64	1.20	1.33	2.12	1.64	2.02	3.21	2.48	.79	1.37	1.98	.34	.42	.79
Louisville, Kentucky	.66	1.95	2.50	3.03	.25	.32	.88	5.03	4.89	2.30	7.62	7.41	3.48	1.29	1.65	2.00	.16	.21	.58
Shreveport, Louisiana	.66	1.65	2.12	2.77	.18	.25	.55	6.18	5.71	3.32	9.36	8.65	5.03	1.09	1.40	1.83	.12	.16	.36
South Bend, Indiana	.66	1.78	2.71	4.04	.18	.24	**	6.50	7.60	**	9.85	11.51	**	1.17	1.79	2.67	.12	.16	**
AVERAGE	.66	1.68	2.31	3.22	.29	.33	.80	4.42	5.24	2.65	6.69	7.94	4.01	1.11	1.52	2.12	.19	.22	.53

** Undefined, since the estimates of $\bar{\Delta}_{de}/\sigma_{xy(B)} \doteq 1$

Source: References [4 and 5]

AN ECONOMETRIC MODEL OF TAIWAN WITH POPULATION POLICY VARIABLES

Kong Chu, Terry Seaks, John Wall, Denis Karnosky, Owen Irvine & Thomas Naylor, Carolina Population Center

Introduction

In most previous attempts to construct econometric models of developing countries, population is treated as an exogenous variable which affects the economy in question but is not effected by it. That is, causality is usually assumed to operate in a single direction. This paper takes a somewhat different point of departure, namely that the economy of a country and the size of its population are jointly determined, and if one wants to understand the process of economic development he must attempt to delineate these relationships of mutual causality.

In this paper we describe a 24-equation econometric model of the economy of Taiwan which was estimated using two-stage least-squares on the basis of annual data for the period 1953-1968. Sixteen of the equations of the model are behavioral equations. The major sectors of the model include consumption, investment, foreign trade, production, monetary, labor force, and population. Unique to this model is the inclusion of equations explaining the birth rate and death rate for the economy of Taiwan. Of particular interest is the fact that the birth rate equation includes population policy variables. (Four alternative specifications of the birth rate equation are presented.) Computer simulation experiments for purposes of validation of the model are also included. Using the mean absolute percent error as a measure of goodness-of-fit for the simulations, the model is shown to yield extremely promising results over the data base period 1953-1968.

The variables and equations for the model are defined and specified in the following section.

TAIWAN MODEL: 1953-1968

Definition of Variables

Endogenous Variables

Birth = Live births per 1000

C = Consumption expenditure in millions of 1963 NT dollars

Curr = Currency in millions of NT dollars

DD = Demand deposits in millions of NT dollars

Death = Deaths per 1000

Exp = Exports in millions of 1963 NT dollars

I = Total investment in millions of 1963 NT dollars

I_{agr} = Investment in agriculture in millions of 1963 NT dollars

I_{ind} = Investment in industry in millions of 1963 NT dollars

I_{ser} = Investment in services in millions of 1963 NT dollars

Imp = Imports in millions of 1963 NT dollars

K_{agr} = Capital stock in agriculture in millions of 1963 NT dollars

K_{ind} = Capital stock in industry in millions of 1963 NT dollars

K_{ser} = Capital stock in services in millions of 1963 NT dollars

L_{agr} = Labor employed in agriculture in thousands

L_{ind} = Labor employed in industry in thousands

L_{ser} = Labor employed in services in thousands

$NNP_{\$}$ = Net national product in millions of NT dollars

NNP_{real} = Net national product in millions of 1963 NT dollars

P = Implicit price level, 1963=100

Pop = Population in thousands

Y_{agr} = Agricultural output in millions of 1963 NT dollars

Y_{ind} = Industrial output in millions of 1963 NT dollars

Y_{ser} = Services output in millions of 1963 NT dollars

Exogenous Variables

Inv = Changes in inventories in millions of 1963 NT dollars

IWT = Index of World Trade, 1963=100

Over 60 = Percent of population over age 60

Time = Time with 1953=1,...,1968=16

WPI/PTE = Ratio of World Price Index to price of Taiwan exports, 1963=100

Policy Variables

CMR = Crude marriage rate per 1000

*DVFP = Dummy variable for family planning program, equals zero from 1953 to 1963, one from 1964 to 1968

Policy Variables (continued)

Ex	= Yearly expenditure of family planning per thousand population in constant 1963 NT dollars (1953-1963, Ex=0)
I_{govt}	= Government investment in millions of 1963 NT dollars
Lit	= Literacy rate in percent
*Loops	= Yearly acceptances of interuterine devices per thousand population (1953-1963, Loops=0)
*Loops + Pills	= Yearly acceptance of interuterine devices plus pill users
Q	= Total money supply in millions of NT dollars
r	= Interest rate, December of each year
Y_{govt}	= Government output in millions of 1963 NT dollars

---

*These can be policy variables if alternate Birth equations 15a, 15b, and 15c are used. See "Population: Alternative Birth Equations."

Behavioral Equations

Consumption

$$C = \underset{(48.90)}{.6965}\ NNP_{real} + \underset{(8.27)}{8325.2376} \qquad (1)$$

$\bar{R}^2$ = .99, D.W. = 1.14

Investment

$$I_{agr} = \underset{(4.91)}{.0217}\ NNP_{real} + \underset{(2.65)}{.0455}\ Y_{agr_{t-1}} \qquad (2)$$

$\bar{R}^2$ = .94, D.W. = 1.82

$$I_{ind} = \underset{(14.39)}{.1288}\ NNP_{real_{-1}} - \underset{(-4.17)}{745.4407}\ r \qquad (3)$$

$\bar{R}^2$ = .89, D.W. = .45

$$I_{ser} = \underset{(14.66)}{.3009}\ Y_{ser} - \underset{(-4.6302)}{594.8094}\ r \qquad (4)$$

$\bar{R}^2$ = .92, D.W. = 1.00

Note: In all equations, the values placed in parentheses below coefficients are t-statistics. $\bar{R}^2$ is the multiple correlation coefficient adjusted for degrees of freedom. D.W. is the Durbin-Watson statistic. Ln denotes the natural logarithm.

Foreign Trade

$$LnImp = \underset{(18.54)}{.9060}\ LnI + \underset{(2.35)}{.1373}\ LnInv \qquad (5)$$

$\bar{R}^2$ = .99, D.W. = 2.36

$$LnExp = \underset{(200.00)}{2.0476}\ LnIWT + \underset{(2.96)}{.8804}\ Ln\frac{WPI}{PTE} \qquad (6)$$

$\bar{R}^2$ = .96, D.W. = 1.66

Production

$$LnY_{agr} - LnK_{agr} = \underset{(19.78)}{.7662}(LnL_{agr} - LnK_{agr}) + \underset{(21.33)}{1.7738} \qquad (7)$$

$\bar{R}^2$ = .96, D.W. = .90

$$LnY_{ind} - LnK_{ind} = \underset{(20.44)}{.6558}(LnL_{ind} - LnK_{ind}) + \underset{(16.71)}{2.2550} \qquad (8)$$

$\bar{R}^2$ = .97, D.W. = .57

$$LnY_{ser} - LnK_{ser} = \underset{(19.47)}{.7692}(LnL_{ser} - LnK_{ser}) + \underset{(18.55)}{2.4410} \qquad (9)$$

$\bar{R}^2$ = .96, D.W. = .55

Monetary

$$DD = \underset{(26.67)}{.1150}\ NNP_{\$} - \underset{(-5.62)}{468.4520}\ r \qquad (10)$$

$\bar{R}^2$ = .96, D.W. = .44

$$Curr = \underset{(40.02)}{.0646}\ NNP_{\$} \qquad (11)$$

$\bar{R}^2$ = .97, D.W. = .38

Labor Force

$$L_{agr} = \underset{(131.12)}{.2158}\ Pop - \underset{(-28.38)}{53.7030}\ Time \qquad (12)$$

$\bar{R}^2$ = .95, D.W. = .98

$$L_{ind} = \underset{(19.67)}{.0290}\ Pop + \underset{(5.12)}{8.6671}\ Time \qquad (13)$$

$\bar{R}^2$ = .94, D.W. = .69

Labor Force (continued)

$$L_{ser} = \underset{(88.26)}{.0813}\ Pop \qquad (14)$$

$\bar{R}^2$	D.W.
.93	.95

Population

$$Birth = \underset{(-4.111)}{-2.2659}\ Ex + \underset{(12.109)}{3.6305}\ CMR + \underset{(3.458)}{1.0866}\ Death_{-2} \qquad (15)$$

$\bar{R}^2$	D.W.
.95	1.20

$$Death = \underset{(-9.83)}{-.0013}\ NNP_{real}/POP - \underset{(-4.80)}{.1173}\ Lit + \underset{(15.51)}{547.1673}\ Over\ 60 \qquad (16)$$

$\bar{R}^2$	D.W.
.91	2.18

Population: Alternative Birth Equations

$$Birth = \underset{(2.71)}{.7702}\ Death_{-2} - \underset{(-5.74)}{4.0014}\ DVFP + \underset{(5.90)}{4.0220}\ CMR - .5734 \qquad (15a)$$

$\bar{R}^2$	D.W.
.97	1.75

$$Birth = \underset{(-4.985)}{-4.4717}\ Loops + \underset{(13.582)}{3.7818}\ CMR + \underset{(3.246)}{.94111}\ Death_{-2} \qquad (15b)$$

$\bar{R}^2$	D.W.
.96	1.48

$$Birth = \underset{(-4.957)}{-3.8519}\ (Loops+Pills) + \underset{(13.590)}{3.7601}\ CMR + \underset{(3.328)}{.96062}\ Death_{-2} \qquad (15c)$$

$\bar{R}^2$	D.W.
.96	1.32

Identities

$$I \equiv I_{agr}+I_{ind}+I_{ser}+I_{govt} \qquad (17)$$

$$K_{agr} \equiv K_{agr_{-1}} + I_{agr} \qquad (18)$$

$$K_{ind} \equiv K_{ind_{-1}} + I_{ind} \qquad (19)$$

$$K_{ser} \equiv K_{ser_{-1}} + I_{ser} \qquad (20)$$

$$NNP_{real} \equiv Y_{agr}+Y_{ind}+Y_{ser}+Y_{govt} \qquad (21)$$

$$P \equiv NNP_{\$}/NNP_{real} \qquad (22)$$

$$Pop \equiv Pop_{-1}(1+Birth/1000-Death/1000) \qquad (23)$$

$$Q \equiv DD + Curr \qquad (24)$$

Description of the Model

Consumption

The consumption equation (1) relates total consumption to the real net national product. Real net national product was used rather than disposable income because of the unavailability of a suitable data series for disposable income or the component series necessary to compute disposable income from net national product. The decision to use net national product was based on the assumption that it was probably a more appropriate proxy for disposable income than any other variable for which data were available.

The annual marginal propensity to consume was found to be approximately .70 which is in the expected range for Taiwan.

Investment

Investment was disaggregated into the agricultural, industrial, and service sectors. Agricultural investment (2) was found to be related to current net national product and the net national product of the agricultural sector in the preceding year. Earlier regressions indicated that agricultural investment was not significantly related to the interest rate. This specification appears to be tenable, especially considering the lack of modern credit facilities serving the agricultural sector in Taiwan.

Investment in both the service sector (3) and the industrial sector (4) was found to be related to the current output of that sector and to the supply price of capital--that is, the long-term interest rate. The aggregation of such capital intensive industries as the electric industry and the transportation industry under the service sector represents an argument for the observed relationship between service investment and the long-term interest rate.

Foreign Trade

Both the export and import equations were estimated in their log-linear forms, as is frequently the practice. In the export equation (6), the Index of World Trade is used as a proxy for the rest of the world's demand for Taiwan's produce. Exports were also found to be related to the prices of Taiwan's export commodities relative to the prices of goods traded in world markets. We expect a positive sign on the ratio of the World Price Index to an index of the price of Taiwan's exports; as world prices rise relative to Taiwan's prices, Taiwan's exports become relatively more attractive. The coefficient of WPI/PTE, .88, is an elasticity--the elasticity of demand for Taiwan's exports with respect to the price ratio (WPI/PTE). The obtained value is in the expected range. Imports (5) were found to be related to two measures of investment demand, fixed capital formation (I) and additions to inventory (Inv). This is a reflection of the fact that a very large and growing portion of Taiwan's imports are used for investment. (The situation is slightly changed after 1970; imports of consumer goods have been increasing steadily as the government eases up on the permits for consumer goods imports.)

Production

The three equations describing sector outputs--(7), (8), and (9)--all assume the familiar form of linearly homogeneous Cobb-Douglas functions. They possess the theoretically attractive properties of constant returns to scale, positive first partials, and negative second partials. Other specifications such as quadratic functions or homogeneity of some degree other than zero were tried, but in all cases they yielded significantly worse results.

A note of explanation is necessary for the capital series that we employed for each of the three sectors. The only available breakdown by sectors gave annual current price investment figures. We converted these to a real series by applying appropriate sector deflators to agriculture and industry and applying the Consumer Price Index to the service sector. The real investment figures were then summed from 1951 to time t, yielding sector capital for t=1953 to t=1968. Some bias will clearly result for early years where sector capital stock is underestimated, but it is probably quickly dominated by rapidly rising investment expenditures that began to occur in the mid-fifties as Taiwan's rate of development increased. Because of the fact that investment rose rapidly and out of fear of introducing further bias, no attempts were made to extrapolate the investment figures backwards or correct the early years capital figures.

Monetary Sector

Equations (10) and (11) and identity (24) comprise the monetary sector of the model. They provide the three endogenous variables--demand deposits, currency, and nominal NNP. These variables are influenced by the exogenous interest rate and the total money supply.

In its present specification we have essentially a simple quantity theory that is reminiscent of the "classical" macroeconomics. Increases in the money supply affect prices via equations (24), (11), and (22). However, the real sector of the model is unchanged by such increases or decreases.

Equation (10) estimates the quantity of demand deposits as a function of nominal NNP and the interest rate. A higher nominal NNP requires a higher level of deposits, but since the interest rate acts as a "price" for demand deposits, we find the expected negative sign on r. Changes in the interest rate will affect not only demand deposits, but also investment in the industrial and service sectors.

Labor Force

The three "supply of labor" equations (12), (13), and (14) relate total population to the labor force in the three sectors. At first we attempted some conventional labor supply equations based on wage rates and wage shares. In many cases the signs of the coefficients were wrong, and often significantly so. Wages were then abandoned in favor of the simple population and time trend arguments that appear in the equations. This produced satisfactory prediction in all cases and even yielded significant evidence of a time trend movement of workers out of agricultural employment and into the industrial sector. This is logically consistent with the country's attempts to develop industrially. (Further, during the period 1953-1968 there was no significant time trend in the labor equation for the service sector.)

Demographic Sector

The demographic sector consists of two behavioral equations--one estimating the crude birth rate (15) and the other the crude death rate (16)--and an identity (23) for calculating total population. Crude birth rates and death rates were used rather than age-specific or standardized rates to avoid unnecessary complications in the specification of the equations.

The birth rate equation (15) relates the crude birth rate to the death rate lagged two periods and the crude marriage rate. Perhaps the most interesting aspect of the birth rate equation is the inclusion of a population policy variable, Ex, which reflects annual expenditures for family planning per thousand population for the years 1964 through 1968. This variable takes on a value of zero over the period 1953 through 1963 since the Taiwan family planning program was not inaugurated until 1964. The three alternative specifications of the birth rate equation are also included--equations (15a), (15b), and (15c). Each of these equations includes the death rate lagged two periods and the crude marriage rate as explanatory variables. In equation (15a) the effects of the national family planning program are reflected through the use of a dummy variable which takes on the value of zero for the years 1953 through 1963, when Taiwan had no family planning program, and the value 1 for the period 1964 through 1968--the period in which Taiwan did have a family planning program. Equation (15b) follows a slightly different specification in terms of the population policy variable. In this case the annual acceptance rate of inter-uterine devices per thousand population was used as the policy variable. Finally, in equation (15c) the yearly acceptance rate of inter-uterine devices plus pill users was the policy variable. In each specification of the birth rate equation, the population policy variable was clearly significant at the .01 level, indicating that Taiwan's family planning program had indeed led to a reduction in the birth rate in Taiwan.

The specification of the birth rate equation draws heavily on the recent work of T. Paul Schultz, who has successfully explained the changes in fertility in a number of developing countries [2]. Schultz has formulated an econometric model which attempts to explain the frequency of births in a population in terms of three groups of factors that influence parents' desires for births: (1) the family size goal or number of surviving children that parents want; (2) the incidence of death, mainly among offspring, which necessitates an adjustment in birth rates to achieve any given family size goal; (3) the effect of uncertainty in the family formation process where births, deaths, and remarriage are unpredictable.

The death rate equation (16) contains environmental variables--average income and the literacy rate--and a third variable whose effect is simply that of compensating for the effect of a changing age structure on the crude death rate. Several attempts to identify one or more variables representing measures of the availability of medical personnel or health care delivery services were found to have insignificant effects on the death rate. Neither registered medical personnel per thousand population nor health delivery facilities per thousand population were found to have coefficients which were significantly different from zero.

On the other hand, average income may be assumed to represent a type of measure of social well-being, and the literacy rate may represent a measure of the accessability to modern health information.

The third variable--percent of total population over sixty--represents an attempt to come up with a crude proxy for the age structure of the population of Taiwan. As the mortality rate in younger age groups declined rapidly, with the decline of the birth rate lagging behind, Taiwan's population became much younger. Therefore, part of the decline in the crude birth rate is attributable to the lowering of the average age of the population along with the lower age-specific mortality rates in younger age groups. This may be a transitory phenomenon; as both the birth rate and death rate continue to fall, the population will again have an older age structure which will cause an increase in the death rate, ceteris paribus.

Computer Simulations

A severe test of the validity of a large-scale econometric model is how well does the model predict the behavior of the observed values of the system when the model is treated as a closed-loop simulation. To validate a model of Taiwan, we have repeatedly solved the model for the endogenous variables of the system in terms of the exogenous variables, policy variables, and lagged endogenous variables. The values of the lagged endogenous variables are those generated by the model in the preceding period, thus making the simulation a completely closed-loop procedure.

Since the model contains several non-linear equations, it was necessary to use the Gauss-Seidel method to solve the system of simultaneous non-linear difference equations. Since space limitations do not permit giving complete simulation results, a table of the mean absolute percentage error for eleven of the more important variables is given below.

In future experiments with the model we expect to experiment with the effects of alternative population, monetary, and fiscal policies on the behavior of the Taiwan economy.

Table 1

Simulation Results

Variable	Mean Absolute Percentage Error
Birth Rate	2.3
Death Rate	8.3
Population	1.1
NNP (real)	7.3
Consumption	5.2
Imports	17.4
Investment	17.8
Currency	8.7
Demand Deposits	9.2
NNP (nominal)	7.1
Price Level	13.1

Data Sources

Bank of Taiwan
Central Bank of China
Government Bureau of Audit and Statistics
Ministry of Communication
Ministry of Economic Affairs
Ministry of Finance
Population Registration Statistics of Taiwan (monthly bulletin)
Provincial Department of Commerce
Report of Taiwan Labour Statistics
Taiwan Customs
Taiwan Demographic Fact Book (annual publication)
Taiwan Power Company

Bibliography

United Nations Economic Commission for Asia and the Far East. Sectoral Aspects of Long-Term Economic Projections with Special Reference to Asia and the Far East, Development Programming Techniques Series Number 6. Bangkok, Thailand, 1967.

Schultz, T. Paul. "An Economic Model of Family Planning and Fertility," Journal of Political Economy, LXXVII (March-April 1969).

Schultz, T. Paul. "The Effectiveness of Family Planning in Taiwan: A Proposal for a New Evaluation Methodology." P-4069, the RAND Corporation, Santa Monica, California, April 1969.

LIFETIME INCOME VARIABILITY AND INCOME PROFILES

Martin David, Institute for Social Research
University of Michigan, and University of Wisconsin

1. Introduction

Two types of dynamics provide the macroeconomic context that generates individual incomes. 1) Individual skills, energy, and human capital imply systematic variation in earnings as individuals age. 2) Changes in the price level, productivity, and factor endowments imply global changes in the yield of human, physical, and financial assets held by individuals.

In this study the aggregative effects of both types of dynamics are taken as given, and the proper subject for another investigation. The questions raised by our analysis pertain to the nature and variability of individual income experiences within this global context:

1) Does an individual establish and maintain a relative income position within the birth cohort and population group to which he belongs? Or does income variation reflect random movement relative to the mean experience of the group?

2) Does aging produce increases in the variance of income?

3) Are changes in relative income position randomly distributed in time, or is some simple stochastic process involved?

4) To what extent do the age-income dynamics vary significantly among identifiable population groups?

Considerable interest attaches to these questions. Policy makers would like to know to what extent poverty is a transitory phenomenon. Forecasters would like to know the stability of income as an important factor influencing consumption choices. Tax experts are interested in variability as it affects the need for averaging under a progressive tax law.

In this paper we describe the pattern of income received by individuals filing at least five tax returns in Wisconsin from 1947 to 1959, with observations in at least four pairs of adjacent years. We fit a trend to the pattern observed for each individual. The results may be displayed simply as a distribution of such individual parameters (David and Miller [1970]) or they may be interpreted as observations in a random coefficients model. The latter approach will be developed in the following section. Tests of alternative models are presented in section 3. The paper concludes with a discussion of the relationship between occupational mobility and the model's parameters.

2. Specification of the Model

To normalize income observed for persons of different ages, income of each individual in each year was measured relative to the mean for his birth cohort in that year. Thus the dependent variable in this study is defined

$$y_{it} \equiv \frac{Y_{it}}{C_t(B_i)} . \qquad [1]$$

Y_{it} is the adjusted gross income received by $\underline{i}$ in year $\underline{t}$. $C_t(B_i)$ is the mean income for all individuals born within an interval of years that includes the birth year B_i of the individual in question.[1]

The variable y_{it} is invariant to inflation that systematically raises all sources of adjusted gross income. It is also invariant to any change in real return to human, physical, or financial capital that is uniform over the cohort. For example, an improved technology that provides increased return to both capital and labor does not affect y_{it}.

For each birth cohort the expected value of y_{it} is the ratio of expected adjusted gross income for the taxpayer universe to the estimated mean Census income received by the cohort in that year.

Nothing about the choice of y_{it} as a dependent variable imples that a particular age-income profile must apply for any particular group. The global age-income profile for a birth cohort $\{C_t(B_i)\}$ is an average of disparate experiences. Indeed, we may suppose that the population is partitioned into K groups that are homogeneous with respect to human and other capital. It is then reasonable to suppose a common response of incomes within the group to market forces, and a common development of income with aging. We designate that individual i is in the k^{th} group by a prescript; ${}_ky_{it}$ thus refers to the relative income position of the i^{th} individual within the k^{th} group.

Two models are assumed for the generation of ${}_ky_{it}$. The first assumes a systematic trend in relative income position:

$${}_ky_{it} = {}_k\alpha + {}_k\beta(t-1959) + {}_ku_i + {}_kv_{it} \qquad [2]$$

where

$${}_ku_i \sim N(0, {}_k\sigma_u^2) \qquad {}_kv_{it} \sim N(0, {}_k\sigma_v^2) \qquad [3]$$

$$\mathrm{Cov}({}_ku_i, {}_kv_{it}) = 0 \quad i \in k \text{ and } 1948 \leq t \leq 1959$$

$$\mathrm{Cov}({}_kv_{it}, {}_kv_{it'}) = 0 \quad i \neq i' \in k, \; t \neq t' \text{ and } 1948 \leq t, t' \leq 1959$$

We refer to this model as the simple trend model.

The parameters ${}_k\alpha$ and ${}_k\beta$ establish a sytematic trend for the adjusted gross income of the k^{th} group relative to the mean income estimated for the cohort. The parameter ${}_k\beta < 0$ indicates that the group loses relative income position during the period observed; ${}_k\beta > 0$ indicates the converse. The parameter ${}_k\alpha$ indicates the expected relative income position for the group in 1959. In addition, individual characteristics determine the distribution of individuals about the group trend according to ${}_ku_i$. ${}_ku_i < 0$ implies that on the average an individual occupies a position below the trend line for the group. Finally, in any given year the relative income position of an individual is determined by a random drawing from the error process that determines ${}_kv_{it}$.

These concepts are illustrated in Figure 1a.

${}_k\alpha$ is given by the ratio ao/bo. A value of ${}_k\beta>0$ is reflected in the larger slope of aa'relative to bb'. Finally cc' reflects $u_i > 0$, where the i^{th} individual has an average relative income position greater than that of the group to which he belongs. The collection of all individual experiences in the group determines the ${}_k\sigma_u$ shown in Figure 1b. The random motion about the trend for the individual is measured relative to the displaced trend observed for the individual shown in the Figure by ${}_k\sigma_v$.

The second model is identical to the first, except that ${}_kw_{it}$ replaces ${}_kv_{it}$ in equation [2]. ${}_kw_{it}$ is defined by the autoregressive relation:

$$ {}_kw_{it} = \rho_i \, {}_kw_{i,t-1} + {}_kv_{it} \qquad -1 < \rho_i < 1 \qquad [4] $$

We refer to this model as the autoregressive model. It will be convenient to distinguish parameters of the autoregressive model by using primes; ${}_k\sigma_u$ refers to the trend model, ${}_k\sigma'_u$ to the autoregressive model. To assure that the variance of ${}_kw_{it}$ is finite, ρ_i must have an absolute value less than unity. If $\rho_i = 1$ the process is a random walk, and estimates of ${}_k\alpha'$, ${}_k\beta'$, ${}_k\sigma'_u$, and ${}_k\sigma'_v$ cannot be obtained from the usual autoregressive transform (Malinvaud [1966] 379-82). We refer to the estimators of the parameters by using corresponding Roman letters ${}_kr_i$, ${}_ka'$, ${}_kb'$, ${}_ks'_u$, ${}_ks'_v$.

Choice between the autoregressive and the simple trend versions of the stochastic model can be made on the basis of the variance explained by the models. We test the hypothesis that estimation of r_i fails to contribute significantly to the explained variance, using the conventional F-test.

Interpretation of the Models

As $\{C_t(B_i)\}$ is the time series of expected income for a cohort, the model provides that the expected income of the k^{th} group is equal to

$$ C_t(B_i)({}_k\alpha + {}_k\beta[t - 1959]. $$

If the experience of a subpopulation is exactly the same as the cohort taken as a whole ${}_k\alpha = 1$ and ${}_k\beta = 0$.

The model most nearly parallel to our autoregressive model is that of Fase [1970, 10-11]. He develops a stochastic model for the distribution of earnings for individuals of different ages. Upon entry into the labor force at age s earnings e_s are lognormally distributed according to $\Lambda(e_s; \mu_s, \sigma_s^2)$. Subsequent earnings are autoregressively developed from the drawing obtained at age s:

$$ \ln e_t = \ln e_{t-1} + \xi(\tau - t) + u_t \qquad t = s+1, s+2, \dots T \qquad [5] $$

where u_t is normally distributed and

i) $E(u_t) = 0$

ii) $\text{var}(u_t) = \sigma^2$

iii) $\text{cov}(\ln e_{t-i}, u_t) = 0 \quad i = 1,2,\dots t-s$

iv) $\text{cov}(u_t, u_{t+r}) = 0 \quad$ for $t > \sigma$ and $r \neq 0$

For comparison with our autoregressive model, equation [5] can be written as

$$ {}_ke_{it} = {}_ke_{i,t-1}\, \eta \exp[-{}_k\xi(t-1959) + {}_ku_{it}] $$

where η is the appropriate constant resulting from the substitution of 1959 for τ.

Fase's model is purley multiplicative and depends on the individual's earnings in the last period. Our autoregressive model is additive in the error terms and depends on cohort income changes as well as individual income changes. Rearranging terms in Appendix equation [A1] gives

$$ \rho \; \frac{Y_{i,t-1}}{C_{t-1}(B_i)} + u_i + v_{it}] $$

$$ \begin{aligned} Y_{it} = {} & \rho_i Y_{i,t-1} \left(\frac{C_t(B_i)}{C_{t-1}(B_i)}\right) \\ & + (1-\rho_i)\, C_t(B_i)({}_k\alpha + {}_k\beta[t-1958] \\ & + C_t(\beta_i)({}_k\beta\rho_i + [1-\rho_i]u_i + v_{it}) \end{aligned} \qquad [7] $$

Aside from the difference between logarithmic and linear formulations, stochastic process [7] is more general than [6] as the formulation permits a weighting of past income experience of the individual (1^{st} term on ther.h.s. of [7]) with the global experiences of the cohort (2^{nd} term on the r.h.s. of [7]).

Both models permit estimation of an initial variance in income positions for the group and a variance associated with the subsequent development of incomes.

3. Empirical Results

Table 1 presents a summary of our two models as fit to individual time series on adjusted gross income for males in Wisconsin for the years 1948-59. Parameters are estimated as appropriately weighted means of individuals' estimates. The justification for this procedure and the formulas used are shown in Balestra and Nerlove [1966, pp. 606-8, equations (43-44)]. The mean intercept estimated exceeds unity for both models. This result is to be expected from the fact that $C_t(B_i)$ was estimated for all income receivers, while the time series data for y_{it} are only available for individuals filing two consecutive tax returns at least four times during the peiod under observation. Persons who file tax returns intermittently because of low incomes will be systematically undersampled in our data. Highly mobile persons who are intermittently out of the state will also be undersampled.

The mean trend is slightly negative. Again this result was to be expected from the discrepancy between the taxpayer universe and the population of income receivers. Young earners, whose incomes are too small to tax, enter the labor force and raise the average cohort income by more than the average annual growth of an individual's income. (See David and Miller [1970], 83-84). A negative trend results for young cohorts.

The autoregressive model is only valid when $-1 < r_i < 1$. Because the models were estimated for individuals and then pooled to obtain ${}_ka', {}_kb'$, some of the r_i were inadmissible.[2] Cases with inadmissible values were excluded from consideration. Individuals for whom $.95 < r_i \leq 1.00$ were also excluded from the tabulations as the estimators for α_i', and β_i' become increasingly unstable as r_i approaches unity. Thus there are 218 individuals for whom we were not able to extract a valid value of the autoregressive coefficient on the basis of an individual's time series. Persons whose time series were excluded reported a shorter time series of data than the remaining population. This is shown in column (2) of Table 1 where the average number of degrees of freedom available for estimating the two models is shown. The autoregressive model entails one more parameter than the simple trend model, and an average of 6.14 degrees of freedom per individual time series are available for parameter estimation. However, individuals with inadmissible r_i were associated with 3.06 degrees of freedom.

A brief consideration of the difference between the simple trend and the autoregressive model suggests the appropriate test to evaluate the contribution of ρ_i. For each individual the autoregressive specification results in one additional regressor. A test of the contribution to the explained variance due to ρ_i, using a standard F-test would be appropriate. For a group of individuals, we can obtain an estimate of contribution to explain variance by pooling and can compare it to the estimate of pooled error variance. The latter can be obtained from line B of table 1.[3] Over the entire population variance explained by r_i fails to exceed what might be expected on the basis of chance ($F=0.945$; $F_{.01}=1.00$). Thus it appears that autoregression, if at all important, is restricted to small population subgroups.

In contrast to this global test for significance of r_i, a t-test applied to the individual time series yields far more significant cases than would be expected by chance (711 cases compared to roughly 70 that could be expected from chance alone). Further investigation of population characteristics leading to autoregression appeared desirable. I feel certain some constraints on the estimation of r_i would be desirable.

Interestingly, the cases in which significant r_i are estimated for individuals involve longer than average time series of information (line C, table 1). This finding might have been anticipated from the fact that the parameter estimates are more reliable when a longer series of information is available. What was not anticipated is that the average value for significant autoregressive coefficients was substantially negative. The autoregressive model was formulated on the hypothesis that a persistence of income position from year to year would lead to positive autocorrelation in the residuals from the trend line; the opposite was observed.

A simple explanation may be offered for the self-employed. Persons in self-employed occupations may make some arbitrary allocations of cost to accounting periods. Thus it is possible that a year of unusually high outlay resulting in exceptionally low income would be followed by another year of unusually low outlay and correspondingly higher profit. For non-self-employed occupations no obvious explanation for negative values of r_i presents itself. Because substantially more cases of significant r_i occur than would be expected on the basis of chance, further investigation of this phenomenon is required.

The principal conclusion to be drawn from Table 1 is that the autoregressive model fails to provide a superior explanation of the relative income variation of individuals than the simple trend model. Moreover ${}_ks_u'$ is larger than ${}_ks_u$ estimated for the simple trend model.[4] The latter finding implies that less unexplained differences among individuals occur when the simple trend model is used.

Study of birth groups did not reveal a subpopulation for which the autoregressive model was significant.

Classification of the population by principal occupation revealed two groups for whom the autoregressive parameters made a significant contribution to explained variance - self-employed businessmen and the relatively large group that includes semi-skilled and unskilled laborers. In both cases the mean r_i is negative; this supports our hypothesis for the self-employed, but we cannot offer an explanation for the laboring group. A summary of the model parameters appears in the lower portion of Table 1.

It is the case that roughly one-fifth of all cases contain a significant autoregressive component, and this component is substantially in excess of the proportion that could be expected on the basis of chance. Until some positive theory for the predominantly negative autoregression can be formulated and tested, it appeared wise to restrict attention to the simple trend model. That model gives a more satisfactory fit over the whole population; the autoregressive model cannot be directly applied in those cases where r_i is close to unity or inadmissible; the autoregressive model also produces a worse fit for the vast majority of individuals for whom no significant autoregressive component in the error term can be isolated.

4. Lifetime Income Patterns and Individual Income Variance

The estimated model of relative income position can be interpreted as a picture of the lifetime income experiences of different individuals. To generalize from the period of observation, 1948-1959, one must assume that relative income positions are determined by a typical pattern for a lifetime career, and are not influenced by the peculiarities of the labor market of the 1950's. In what follows we rely on that assumption and assemble lifetime estimates of ${}_k\alpha$, ${}_k\beta$, ${}_k\sigma_u$ and ${}_k\sigma_v$ for different occupational groups. In these charts the occupation last reported on a tax return determines the classification. The results suggest some typical career patterns and

and give a feeling for the importance of systematic stratification of the distribution as a factor producing increased variance of income among older persons.

Using the estimates to represent a lifetime pattern generalizes the usual cross-section hypothesis, namely that differences between individuals of different ages reflect a dynamic picture for a given individual with the passage of time. What we assume here is that the differentials obtained by studying a 5- to 12-year history for a single individual can be assembled for many individuals to give an accurage picture of the dynamics of income over a lifetime.

Since the data at hand reflect income reported for income tax purposes, neither young nor old persons are adequately represented. All of the 1930-34 cohort entered the labor force during our period of observation and to a large extent began filing tax returns at the time of entry. Those who received college training entered the labor force later than the bulk of their birth cohort. Thus many of the 1925-29 cohort and some of the 1915-24 dohort entered the labor force. A few of the members of the 1895-1904 cohort leave the labor force because of sickness or involuntary unemployment; most of the cohorts born prior to 1895 are retired. Filing of tax returns for the individuals born prior to 1895 is likely to indicate continued labor force attachments on more than a casual basis since few individuals have sufficient assets to require the filing of a tax return following retirement. (Mean earnings observed for this age cohort, 1947-1959, equalled $3131, about three-fourths of the value for the sample as a whole). Care must be used in interpreting results for the youngest and oldest cohorts.

Figure 2 shows the expected relative income position of persons in different occupation groups as they grow older. Each point reflects the income position of workers in an occupation relative to all the members of their birth cohort. Professionals and managerial workers show substantial improvements in relative income position as they age. (A part of this trend may reflect the great shortage of professional workers during the 1950's rather than a "typical" career development.) A slight upward trend in the relative position of sales workers appears with increasing age, while other occupations exhibit more or less random variation about a fixed relative income position. The picture of Figure 2 indicates the manner in which the age-income profile observed in a cross-section should be modified for various occupational groups.

In Figure 3 values of ${}_ks_v$ are graphed for individuals according to their birth cohort and occupation in the last year observed. After entry into the labor force is completed, the values of ${}_ks_v$ show systematic increased for professional, managerial workers, and self-employed businessmen. Clerical, service, and semi- and unskilled workers exhibit no increases in standard error of estimate as older cohorts are observed. For many groups the variance of income declines again as they enter years in which members of their cohort retire. (The retired are excluded from the chart, even though they may have had employment during the period.)

We conclude that professional and managerial groups show a combination of improved relative income position and increased income variation over their lifetime. The increased income variation for the self-employed businessman is not associated with a systematic improvement in relative income position, while a large group of clerical, service, and semi- and unskilled workers show neither a trend in relative income position nor a change in within individual income variance relative to the mean for their dohort.

To complete the picture of careers we present estimates of ${}_k\sigma_u$ as a measure of the heterogeneity of individual experiences. Interpersonal variation association with the intercept shows the same stability for clerical, service, and semi- and unskilled workers as ${}_ks_v$. Whatever stratification, or layering, of the income distribution exists at the time of entry into these occupations persists throughout the lifetimes of these individuals. The professional and managerial workers show a radically different pattern. Interpersonal variation in the intercept increases substantially with age. (The extent of variation appears proportional to the value of ${}_ks_v$. As a consequence we conclude that the trends of individuals' income positions results in increasing layering or stratification of income with age.

This conclusion requires a re-examination of the model in [2]. The model must permit random variation in both trend and intercept over individuals. This is the random coefficients model (Swami [1970]). The extension of the model can be written as

$$ {}_ky_{it} = ({}_k\alpha + u_i) + ({}_k\beta + q_i)(t-1959) + {}_kv_{it} $$

where u_i, q_i are drawn from a bivariate normal distribution with a variance-covariance matrix

$$ {}_k\Sigma = \begin{pmatrix} {}_k\sigma^2_u & {}_k\mu \\ {}_k\mu & {}_k\sigma^2_q \end{pmatrix} \qquad [8] $$

The least squares technique used to estimate [2] provides unbiased estimators of ${}_k\alpha$, ${}_k\beta$. Estimates of ${}_k\Sigma$ are consistent, but not efficient (Swami [1970]).

In the context of [8] it is quite plausible that interpersonal variance about the intercept increases with age. That finding implies ${}_k\mu > 0$; that is, persons with relatively greater than average relative income positions within the k^{th} group also exhibit larger trends than the mean for the group.[5] ${}_k\mu > 0$ results in a "fanning" of income experiences with time, displayed in Figure 5. The interpersonal variance between the two individuals whose relative income experiences are shown in cc' and dd', will be greater at time 2 than at time 1. The positive deviation of u is associated with q > 0 so that the cov(u,q) > 0.

The increasing interpersonal variance ${}_ks_u$ for managers and professionals in different birth cohorts is corroborated by a weak, but positive correlation between a_i' and b_i' within these occupations.[6] The sample as a whole, shows no correlation between these parameters. To summarize in another way, the heterogeneity of trends in relative income position represents random movement of individuals with respect to their group intercept in the case of clerical, service, and semi- and unskilled workers. In the case of professional and managerial workers the trends persist systematically for particular individuals so that interpersonal variation rises as those individuals age.

5. Occupational Mobility and Income Variation

Study of occupational mobility gives still another insight into the variability of earnings and income experience of individuals. Table 3 presents estimates of the simple trend model for a variety of groups defined by observed changes in occupational status. The major occupation groups are the groups

- professional
- semi-professional
- managers
- self-employed businessmen
- farmers
- clerical workers
- sales
- service
- skilled workers
- semi- and unskilled workers

Several of these groups were further classified into detailed occupations; however, lack of precision in the self-reporting of blue-collar occupations prevented any detailed classification of those large occupational groupings. As a consequence the detailed occupations reflect fine distinctions within professional and managerial groups and crude differentiations (if any) among occupations included in the various categories.

In table 2, entry and departure from the labor force is treated as a change in major occupational group. In part A individuals who reported no change in major occupational group clearly show the most favorable trend in relative income position and the largest interpersonal variation by comparison with those individuals who reported some kind of change in occupational status. Individuals who reported more than one change in major occupational grouping clearly had the least favorable trend in relative income position and showed the least interpersonal variation. (Such persons must report at least three of the ten occupations listed above, or two such occupations and a change in labor force participation.) Among persons who reported one change in major occupation those who were in the new occupation a relatively short period of time reported more favorable income experiences. The meaning of this finding is confused by the fact that movement into or out of the labor force is counted as a shift in occupational group.

Table 2 employed clarifies the latter problem (also Table 10, Schroeder and David [1970]). Only persons who were employed in the labor force during the entire period for which they reported income for tax purposes are included. In this population the most favorable relative income trends are experienced by persons who have experienced a change in their detailed occupation, while the least favorable experience is again reported by persons with more than two occupational affiliations. Experiences of persons who remain employed and have two occupational affiliations are more favorable and show less interpersonal variation than those who enter or leave the labor force. (Compare row 2 all, with row C, employed).

Some insight into the meaning of these associations between relative income trends and occupation change can be gleaned from further classification of the groups in Table 2 by birth year cohorts. Our a priori hypotheses would be that occupation mobility would be likely to produce the most favorable change in income early in a career. Mobility between ages 55-64 is also likely to produce favorable impacts on income, given the manner in which the table is generated. Persons in the older age group who attempt a change in occupation and do not succeed will retire from the labor force altogether and will be excluded from the tabulation.

Unfortunately we cannot determine whether occupation change causes an improvement in relative income position as the model of the individual time series includes observations from years prior to the change and years following the change. Lack of an association in the expected direction would disprove our hypotheses, so that the data presented in Table 3 will indicate whether we should reject the null hypothesis. In fact, Table 3 indicates that persons born in 1925-34 who make a major shift in occupation have a superior income position in 1959 to those who do not. Interpersonal variation in those groups, as indicated by ${}_k s_u$ is also smaller. Finally, there does not appear to be a radical difference in the values of ${}_k s_v$ for mobile and non-mobile persons (considering the relatively small number of persons observed in some of the groups shown).

6. Conclusions

This descriptive lifetime history of income offers challenging subjects for analytical study.

a) The negative autoregression observed does not correspond to usual conceptions of the persistence of income positions.

b) The increased interpersonal variance (Figure 4) of the incomes of older professionals, managerial workers and the self-employed needs to be associated with a causal model related to skills and motivation.

c) The homogeneity of experiences of mobile workders needs to be related to the labor markets that they search, so that a mechanism causing the findings in Table 3 can be identified.

d) Lastly macroeconomic modeling should be undertaken to determine whether the relative income movements of professionals and managers during the 1950's is a peculiarity of the period investigated, or whether the lifetime income pattern is in some way associated with the evolution of their skills.

APPENDIX

Estimation of Parameters

The simple trend model was estimated by ordinary single equation least squares for each individual. However only those observations were used that met the information requirements of the autoregressive model. It might seem that by not using the additional values available for the simple trend model we have unnecessarily restricted ourselves to less efficient estimates of their parameters than what could easily be obtained. However since of major concern were the comparisons among the models that could be made we did not wish to confound these comparisons by utilizing the additional data.

The autoregressive model was estimated by the following transform (the subscript i has been suppressed from u_i, α_i, β_i, ρ_i and the data to simplify the notation):

$$y_t - \rho y_{t-1} = \alpha - \rho\alpha + \beta(t-1959) - \rho\beta(t-1-1959) + w_t - \rho w_{t-1} + (1-\rho)u \qquad [A1]$$

Rearranging terms and substituting v_t for $w_t - \rho w_{t-1}$ allows this to be put in the form

$$y_t = a_o + a_1 y_{t-1} + a_2(t-1959) \qquad [A2]$$

[A2] was estimated directly by ordinary least squares. a_i are defined by

$$a_1 = \rho$$
$$a_2 = (1-\rho)\beta \text{ and}$$
$$a_o = \rho\beta + (1-\rho)(\alpha + u) \qquad [A3]$$

so that we can solve for the original parameters by

$$\rho = a_1$$
$$\beta = a_2(1-a_1)^{-1} \text{ and}$$
$$\alpha + u = (1-a_1)^{-1}[a_o - a_1 a_2 (1-a_1)^{-1}]. \qquad [A4]$$

Similarly the estimates of the a's in [A2] were inserted into [A4] to yield estimates of the parameters: $\hat{\alpha} + \hat{u}$, $\hat{\beta}$ and $\hat{\rho}$ in terms of $\hat{a}_o$, $\hat{a}_1$ and $\hat{a}_2$. The variances of the parameters were approximated using the variance-covariance matrix of the $\hat{a}$'s and the partial derivatives of the parameters with respect to the a's from [A4] in the usual Taylor series expansion (derivatives in terms of a's have been converted to their parameter equivalents):

$$\text{Var}(\hat{\rho}) = \text{Var}(\hat{a}_1),$$
$$\text{Var}(\hat{\beta}) = (1-\hat{\rho})^{-2}[\hat{\beta}^2\text{Var}(\hat{a}_1) + \text{Var}(\hat{a}_2) + 2\hat{\beta}\text{Cov}(\hat{a}_1\hat{a}_2)],$$
$$\begin{aligned}\text{Var}(\hat{\alpha}+\hat{u}) = {} & (1-\hat{\rho})^{-2}\text{Var}(\hat{a}_o) + [\hat{\alpha} + \hat{u} - \hat{\beta} - \hat{\rho}\hat{\beta}(1-\hat{\rho})^{-1}]^2\text{Var}(\hat{a}_1) \\ & + \hat{\rho}^2(1-\hat{\rho})^{-2}\text{Var}(\hat{a}_2) + 2(1-\hat{\rho})^{-1}[\hat{\alpha}+\hat{u}-\hat{\beta} - \hat{\rho}\hat{\beta}(1-\hat{\rho})^{-1}]^2\text{Cov}(\hat{a}_o\hat{a}_1) \\ & - 2\hat{\rho}(1-\hat{\rho})^{-1}[\text{Cov}(\hat{a}_o\hat{a}_2) + [\hat{\alpha} + \hat{u} - \hat{\beta} - \hat{\rho}\hat{\beta}(1-\hat{\rho})^{-1}]\text{Cov}(a_1 a_2)].\end{aligned}$$

These estimates were than aggregated according to Balestra and Nerlove [1966, 606-08] to obtain pooled parameters for the desired population.

FOOTNOTES

1. The birth year intervals used were: 1860-74, 1875-84, 1885-94, 1895-1904, 1905-14, 1915-24, 1925-29, 1930-34, 1935 and over. Adjusted gross income is defined by Wisconsin tax law and parallels the Federal definition. $C_t(B_i)$ was estimated from Census data on all income sources, so that the expected value of y_{it} is not necessarily unity.

2. An alternative model could be studied in which $r_i = {}_kr$ and a single autoregressive parameter is estimated for the k^{th} group. Inadmissible cases are unlikely with this alternative.

3. To test the significance of the autoregressive specification for the entire sample, we included the time series with inadmissible r_i. Thus identical populations are compared. As an alternative, the inadmissible series could be excluded from the pooled estimates for the simple trend model. That procedure subsamples the data on a rather arbitrary basis.

4. The value of ${}_ks_u = 1.329$ corresponds to the ${}_ks'_u = 1.558$ estimated in line C, Table 1, for the population with admissible r_i.

5. ${}_k\mu < 0$ may ultimately result in greater interpersonal variance, but there will be an intervening period where ${}_k\sigma_u$ is smaller than its initial value.

6. Actually the data are available for principle occupation, not last occupation reported.

REFERENCES

1. Balestra, P. and Nerlove, M. [1966] "Pooling Cross Section and Time Series Data in the Estimation of a Dynamic Model: The Demand for Natural Gas" Econometrica, 34 (July, 1966) 585-612.

2. Bauman, R., David, M., and Miller, R.F. "The Wisconsin Assets and Incomes Studies Archive", Social Science Information (December, 1967) 49-70.

3. David, M. [1969] "Time-Series Versus Cross-Section Lifetime Earnings Patterns in Different Occupational Groups", Proceedings of the Business and Economics Section of the American Statistical Association (August 21, 1969).

4. David, M. [1971] "Legislation, Enforcement and Filing of Tax Returns", National Tax Journal, 25 (September, 1971).

5. David, M. and Miller, R.F. [1970] "A Naive History of Individual Incomes in Wisconsin, 1947-1959", Journal of Income and Wealth, March, 1970.

6. Fase, M. [1970] An Econometric Model of Age-Income Profiles, Rotterdam: Rotterdam University Press.

7. Malinvaud, E. [1966] Statistical Methods of Econometrics, Chicago: Rand McNally and Co., 379-382.

8. Schroeder, L. and David, M., "Long- vs. Short-Term Impacts of Mobility on Earnings", Proceedings of the Social Statistics Section American Statistical Association, December 1970.

9. Swami, P.A.V.B., "Efficient Inference in a Random Coefficients Model", Econometrica, 38 (March 1970) 311-324.

TABLE 1

Occupation, model	Number of Individuals	Degrees of Freedom Per Individual	Intercept Mean a	Intercept Std.Dev. $k^s u$	Mean Trend b	Error Std. Error of Estimate $k^s v$	Error Mean r
All							
A. Simple trend	3740	7.14	1.055	1.365	-0.0138	0.518	--
B. Autoregressive[a/]	3522	6.34	1.079	1.558	-0.0177	0.525	0.0154
C. Significant [b/]	711	7.44	1.180	1.255	-0.0129	0.435	-0.1274
D. Not significant	2811	6.06	1.054	1.625	-0.0190	0.551	0.0517
Self-employed businessmen							
A. Simple trend	281	7.55	1.068	1.844	-0.0265	0.804	--
B. Autoregressive[a/]	269	6.77	--	--	--	0.779	-0.015
C. Significant[b/]	47	n.a.	1.224	1.482	-0.0975	n.a.	-0.172
D. Not significant	222	n.a.	0.979	1.293	-0.0132	n.a.	0.019
Semi-skilled and unskilled							
A. Simple trend	1248	6.86	0.856	0.477	-0.0216	0.209	--
B. Autoregressive[a/]	1168	6.06	--	--	--	0.204	-0.018
C. Significant[b/]	237	n.a.	0.889	0.546	-0.0293	n.a.	-0.140
D. Not significant	931	n.a.	0.885	1.295	-0.0226	n.a.	0.050

a/ $-1 < r_i \leq 0.95$. The inadmissible cases have 3.06 degrees of freedom per individual.

b/ Probability of observing r_i when $\rho_i = 0$ is less than 0.02.

TABLE 2

Category	Number	Intercept Value k^a	Intercept Std.Error $k^s u$	Trend k^b	Standard Error $k^s v$
All					
Unique major occupation	1920	1.134	1.562	-.0096	.589
Changed major occupation group	1820	0.970	1.115	-.0184	.428
More than once	562	0.874	0.645	-.0243	.370
Only once					
Less than 21% observations in last occupation	617	1.046	1.089	-.0143	.424
21-99% of observations in last occupation	641	0.989	1.419	-.0171	.479
Employed					
A. Unique detail occupation	1821	1.121	1.558	-.0108	.599
B. Unique major occupation, change in detailed occupation	71	1.157	0.752	.0014	.238
C. Two major occupations	801	1.050	0.821	-.0058	.371
D. Three or more occupations	204	0.918	0.486	-.0165	.381
E. All employed	(2897)	1.09	1.321	-.0095	.525

TABLE 3

Occupation observed 1947-1959		Birth Year 1860-1865	1885-1894	1895-1904	1905-1914	1915-1924	1925-1929	1930-1934
A.	k^a	1.01	1.24	1.27	1.09	0.99	0.89	0.89
B.		*	*	1.40	1.15	1.01	*	*
C.		*	1.33	1.09	1.01	0.98	0.95	1.08
D.		*	*	0.86	0.83	0.96	0.99	0.81
E.		1.09	1.27	1.22	1.05	0.99	0.92	0.96
A.	$k^s u$	1.38	2.20	1.84	1.40	0.79	0.50	0.76
B.		*	*	0.98	0.71	0.54	*	*
C.		*	1.79	1.10	0.70	0.65	0.49	0.40
D.		*	*	0.45	0.38	0.52	0.58	0.39
E.		1.30	2.09	1.66	1.20	0.72	0.51	0.61
A.	$k^s v$	0.75	0.68	0.57	0.35	0.25	0.24	0.31
B.		*	*	0.11	0.32	0.11	*	*
C.		*	0.37	0.63	0.32	0.25	0.24	0.31
D.		*	*	0.11	0.29	0.25	0.23	0.39
E.		0.72	0.24	0.57	0.33	0.25	0.22	0.32

*less than 10 observations.

A.- E. refer to the row stubs in Table 3.

Figure 1a

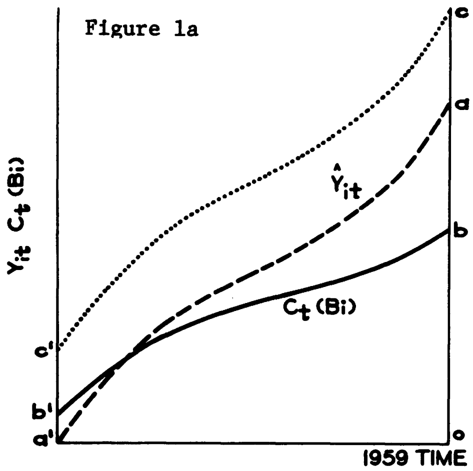


Figure 1b

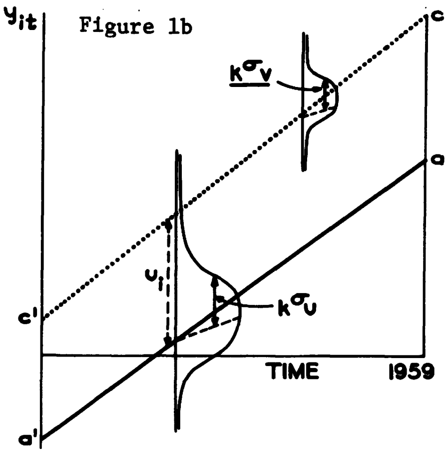


Figure 5

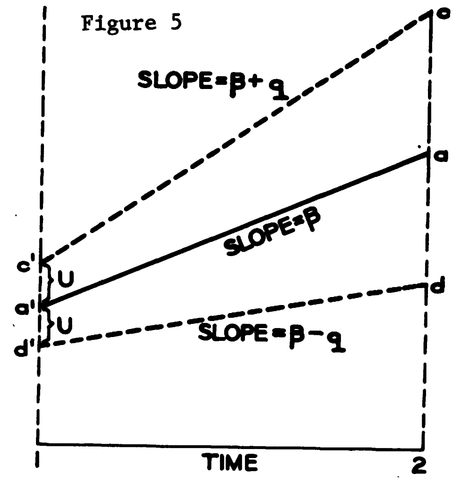


Figure 2

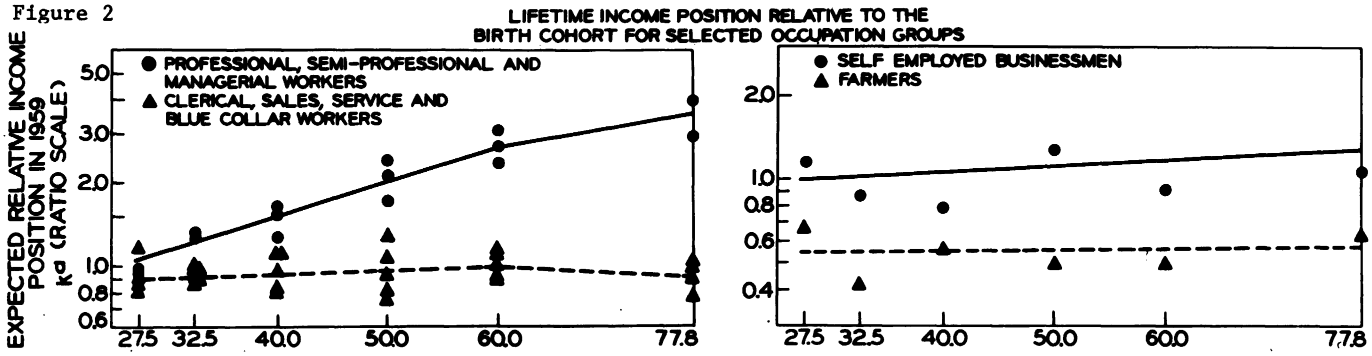


Figure 3

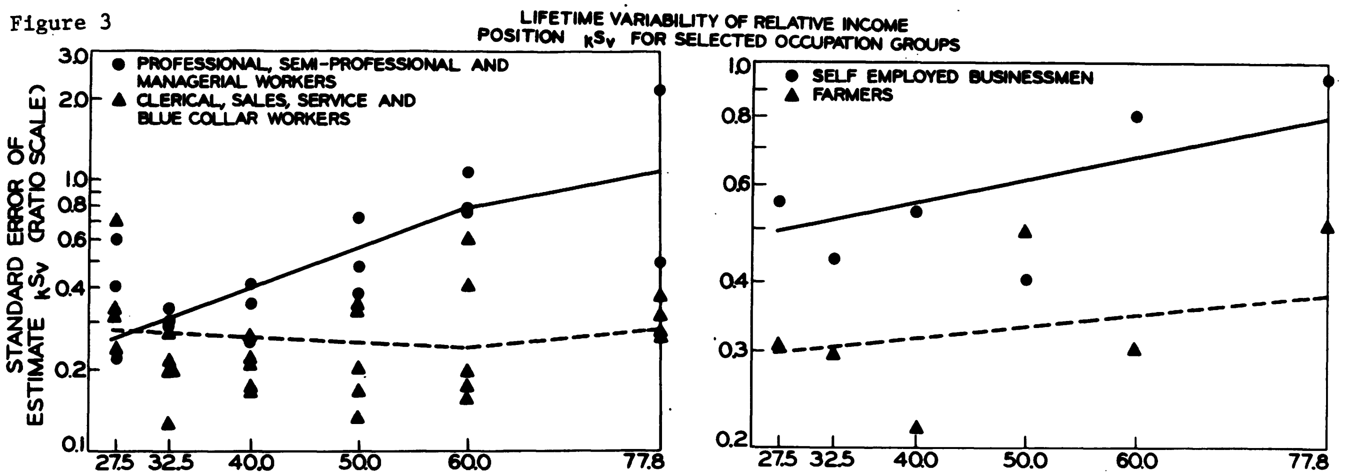
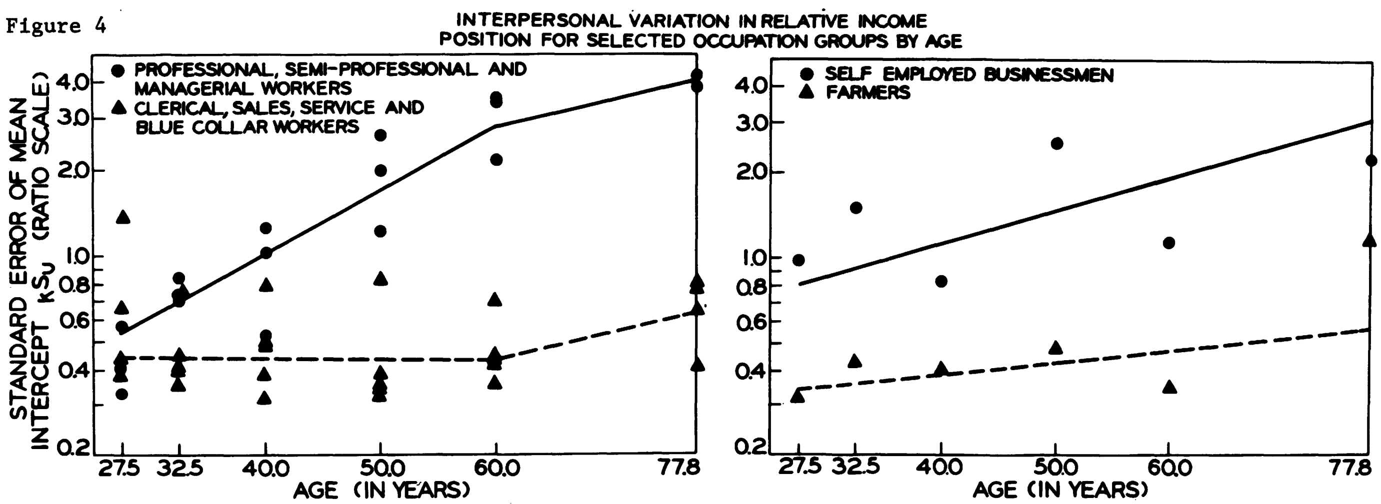


Figure 4



SOCIO-ECONOMIC DETERMINANTS OF INFANT MORTALITY: A PRELIMINARY INVESTIGATION

Paul F. Dickens III and Timothy D. Hogan, Arizona State University

I. Introduction

In recent years infant mortality has been the subject of increasing concern, perhaps for two related reasons. First, the overall decline of infant mortality in the United States has slackened, even though regional differences in rates persist.[1] Secondly, our infant mortality experience, and even that of areas with the lowest rates compares unfavorably with the experience of other developed nations (6,8,15); for, as Table I demonstrates, several European nations have rates lower than Utah, the state with the lowest infant mortality rates in the United States.

Table I

Infant Deaths Per 1,000 Live Births, 1967[2]

United States	22.4
Utah	16.6
Denmark	15.8
Finland	14.8
Netherlands	13.4
Sweden	12.9
Norway	12.8

For planning purposes and policy decisions, information relating to the characteristics closely associated with the infant mortality experience for areas is needed, i.e., what characteristics do areas with high infant mortality have? Such information on areas would seem to have more value for policy decisions and resource allocation than identifying such characteristics for individuals.[3]

As a step in this direction, this paper presents some preliminary analysis of infant mortality rates that exist among areas in one region of the United States. We plan to expand the study to the country as a whole using 1970, as the data becomes available. But for this preliminary study we restricted our analysis to the southeastern portion of the country, because it is one of the areas with the highest overall rates and because in this region great variation exists among localities with respect to infant mortality, income levels, and other variables.

II. Methodology and Data

The study employs multivariate regression analysis of cross-section data to investigate which characteristics of local areas are most closely related to the area's level of infant mortality.

The sample utilized in our analysis consisted of 1960 data for 551 counties in eight southeastern states.[4] This represents approximately 75 percent of the 734 counties in this region. The remainder were excluded either because complete data series could not be obtained or because their population was less than 10,000 people. This minimum population figure was required in order to prevent rather small absolute changes in the number of infant deaths from causing major fluctuations in mortality rates.

In addition to the infant mortality measures, nine explanatory measures were included in the analysis. The specific variables utilized are listed in Table II.

Table II

PBHOS = Infants born in hospitals/total infants born

IMR = Infants mortality/total infants born

NNMR = Neonatal mortality/total infants born

PNW = Percent non-white

PPOOR = Percent of Families with income below $3,000

FR = Children under five per 100 women 15 to 49 years of age

POPHSD = Population per household

BEDR = Short term general hospital beds/population

PCPR = Physicians who's primary duty is patient care/population

PNMR = Post Neonatal mortality/total infants born

PURB = Percent of population in urban areas

Data relating to the various infant mortality measures was collected from the 1960 edition of <u>Vital Statistics</u> (19). The population and socio-economic variables were compiled from the 1960 census (18).

In some discussion of the problem of infant mortality, the supply of doctors and of hospital facilities are mentioned as one important determinant of the level of infant mortality (3,6,8,15). In order to test this hypothesis our study explicitly includes variables relating to the supply of health resources (in the form of the non-federal physician rate and the hospital bed rate).[5] The data on non-federal short-term general hospitals was compiled by us from American Hospital Association sources (1). The physician data was collected from American Medical Association sources (2) and unlike the remainder of the data series it refers to 1963, since this was the earliest year for which the data was available.

The investigation of whether such supply variables are closely related to infant mortality is also important because infant mortality is often used as a summary index of the quality of medical care received by inhabitants of an area (4, p.559).

Since the causes of infant mortality have generally been found to vary systematically with age (11,14,15), this study has disaggregated the infant mortality statistics into two subclasses, neonatal mortality (deaths occurring during the first month) and post-neonatal mortality (those deaths which occur during the remainder of the first year of life) to study whether the strength of relationships between the two segments of infant mortality and the explanatory variables are substantially different.

III. Results and Interpretation

Initially, several interesting points can be made of the simple correlations presented in Table III. For example, the supply variables, bed rate and physician rate are not highly correlated with the percent of infants born in hospitals, while the percent nonwhite and the percent poor are, in fact, the explanatory variables which are most closely correlated (in both cases a negative relation was indicated) with use of hospitals. Several possible explanations might be advanced for these results. First poor, and particularly nonwhite poor, individuals may be excluded from hospital facilities, due to either racial discrimination or to lack of wealth. Alternately, the explanation may be cultural, i.e., these groups may prefer to have their children at home. It is also possible that transportation costs and/or differential automobile ownership contribute to this pattern.

The socio-economic variables are found to be more highly correlated with post-neonatal mortality than with neonatal deaths. These findings tend to support the evidence from previous empirical studies, which utilized death records.[6] In a like manner, the correlation between the supply variables and neonatal mortality was found to be much weaker than the relation between these supply measures and post-neonatal mortality. Such results can most probably be attributed to the previously mentioned tendency for different causes of death at different ages. Neonatal deaths are usually due to immaturity or congenital malformations, while infectious diseases (which are more amenable to medical care) cause a high proportion of post-neonatal deaths (11,14,15).

Table III

Simple Correlation Matrix

IMR	1.000000										
PNW	.4865077	1.000000									
PPOOR	.3227890	.3593604	1.000000								
PBHOS	.4528181	-.7504172	-.5954418	1.000000							
BRDR	-.06011388	.04585699	-.1344326	.05504764	1.000000						
PURB	-.09191584	.02671853	-.6230910	.2231697	.2451806	1.000000					
POPHSD	.2802555	.4517182	.3745008	-.4888915	-.1643905	-.3340864	1.000000				
POPR	-.1210380	-.08241280	-.4435359	.2614871	.3634300	.5362321	-.3431448	1.000000			
FR	.3437361	.6395716	.3029297	-.5876336	-.1166915	-.1342311	.5943598	-.2737308	1.000000		
NNMR	.7959253	.2256201	.1511545	-.2196922	.004495067	-.02931197	.1067010	.01463263	.1670217	1.000000	
PNNR	.7604760	.5433187	.3589164	-.4952951	-.1018580	-.1169306	.3379406	-.1796855	.3757098	.2121593	1.000000

The regression results are presented in Table IV. The estimated t-values for the regression coefficients were utilized to test the significance of the relations between the dependent and the explanatory variables. The figures in parentheses directly beneath each estimated regression coefficient is its estimated standard error. Those cases in which the relation was indicated to be statistically significant are designated by asterisks, with one, two, or three asterisks signifying respectively the 0.10, 0.05, and 0.01 level of significance.

Table IV

Regression Results

	R^2	d.f.
PBHOS = 1.43053 - .00439139PNW*** - .005123589 PPOOR - .002257438 FR*** - .02892547 POPHSD* - .0003337689 PURB (.0003155885) (.0034848296) (.0007366395) (.01658236) (.0002917778) +12.82410 PCPR + .5455547BEDR (14.95495) (1.623369)	.6986443	543
PBHOS = 1.419589 - .004798745 PNW*** - .004912236 PPOOR*** - .002367216 FR*** - .02704132 POPHSD* (0002856565) (.0003638259) (.0007222757) (.01592381)	.6976261	546
IMR = .02514097 + .0001990799 PNW*** + .0001227387 PPOOR** - .005737891 PBHOS - .2296512 BEDR + .00001153091 PURB (.00003612912) (.00005122902) (.004129646) (.1562339) (.00002811173) + .0006227159 POPHSD + .4097812 PCPR + .000009102862 FR (.001600193) (1.440096) (.00007149755)	.2686925	542
IMR = .0210015 4 + .0002382151 PNW*** +.0001462724 PPOOR*** (.00002202075) (.00003386011)	.2618274	548
NNMR = .02000289 + .00004571283 PNW* - .003189529 PBHOS + .00004791284 PPOOR + 1.137347 PCPR + .00004232896 FR (.00002719337) (.003108269) (.00003855864) (1.083920) (.00005381420) - .0005818479 POPHSD - .01613419 BEDR + .000001984518 PURB (.001204420) (.1175928) (.00002115891)	.06172415	542
NNMR = .0241644 + .00005174459 PNW** - .004729550 PBHOS* (.00002335400) (.002574220)	.05671490	548
PNNR = .005138086 + .0001533671 PNW*** + .00007482593 PPOOR** - .2135170 BEDR** + .001204563 POPHSD - .002548363 PBHOS (.00002125062) (.00003013216) (.09189448) (.0009412104) (.002428997) - .00003322610 FR - .7275664 PCPR + .000009546398 PURB (.00004205382) (.8470435) (.00001653491)	.3407515	542
PNNR = .004516408 + .0001688288 PNW*** + .00009087789 PPOOR*** - .2458189 BEDR*** (.00001301855) (.00002018004) (.08579566)	.3359214	547

Since the percent of infants born in hospitals varied substantially among localities and is commonly used as a measure of medical services in an area, we treated this variable as a dependent as well as an independent variable in the regression analysis. The only variables which were found to be statistically significant at the 0.10 level or better were socio-economic measures. The hypothesis that the supply of either medical services or hospital facilities affect the percent of infants born in hospitals must be rejected on the basis of these results, since the regression coefficients of both the physician rate and the bed rate were smaller than their standard error. The coefficient of determination, R^2, was approximately .70, which may be interpreted as the percentage of the total variance in the percent born in hospitals which was explained by the variables included in the regression equation.

In the regressions relating to total infant mortality, none of the supply variables proved to be significant. Both the percent nonwhite and the percent poor were found to be significant at the 0.01 level.

The separate analysis of neonatal and post-neonatal mortality did produce substantially different results. The regression equations for neonatal mortality demonstrated that differences in socio-economic and/or health supply factors could explain little of the variance in neonatal mortality rates ($R^2 = 0.06$). However, the relation between neonatal mortality and two variables, the percent nonwhite and the percent of infants born in hospitals, was found to be statistically significant. This tends to support the previous findings (7,16,22) of little correlation between socio-economic status and neonatal mortality and at the same time the existence of racial differentials in neonatal mortality (11, 14,15,22). The failure of either socioeconomic or supply variables to adequately explain the pattern of neonatal mortality is not surprising in view of the predominance of such deaths due to congenital malformations and immaturity, neither of which are directly measured by our variables.

In the post-neonatal mortality equations on the other hand, the regression coefficients for the percent nonwhite, the percent of families with income of less than $3,000, and the bed rate were found to be statistically significant at the 0.01 level. Furthermore, the coefficient of determination indicated that these socioeconomic and supply variables explained approximately 34 percent of the total variance in post-neonatal mortality--quite a difference in explanatory power when compared with the low R^2 of the neonatal equations.[7]

While the availability of hospital services, as measured by the bed rate, was indicated to be significantly related to post-neonatal mortality, the regression results did not indicate a significant correlation between the physician supply and the level of post-neonatal mortality.

IV. Conclusion

This preliminary study has emphasized some interesting facets of the infant mortality experience as well as the need for further research. The difference between the ability to explain post-neonatal and neonatal mortality by means of socio-economic and health supply variables was striking. Our preliminary results also lend little support to the argument that high infant mortality rates reflect a "shortage" of medical and hospital resources.

Further extension of this study will allow us to test additional hypotheses, concerning the impact of maternal health programs, and the disaggregation of physician data by race will permit examination of the effect of physician race on racial mortality differentials. Another possible extension, if unpublished data is made available, will be to test the significance of variations in average birth-weight by county, since low birthweight is generally considered to be highly correlated with the probability of infant death. With the expansion of the study to national dimensions, we also hope to study the regions of the country separately in order to investigate whether substantial differences between regions exist.

FOOTNOTES

[1] For a discussion of the change in the trend see Moriyama (13). For a summary of the differences among regions in infant mortality see Moriyama (12) and Hunt (10).

[2] The infant mortality figures for the United States and for European nations were collected from the 1967 edition of Vital Statistics of the U. S. (20, pp. 2-6). U. N. Statistical Yearbook (17, p. 100).

[3] This need for planning and for implementation is emphasized by Hunt (10, p. 11) in a previous study in which those counties throughout the United States with excess infant mortality were identified. Further, using census tract data for the Chicago and Washington, D.C. metropolitan areas, this study also analysed the relationship between poverty areas and the level of infant mortality and found substantial poverty-nonpoverty differentials. Most other studies of the correlates of infant mortality, such as (5, 7,9,16,21) have utilized individual death records, rather than studying the relation between area characteristics and the area's infant mortality experience.

[4] Alabama, Georgia, Kentucky, Louisiana, Mississippi, North Carolina, South Carolina, Tennessee.

[5] At an earlier stage in the analysis, we tried to employ a binary variable in order to indicate whether or not a hospital was present in the county as an alternative measure of the supply of hospital facilities, but it was not found to be significant in any equation.

[6] In a Providence, R. I. study, Stockwell (16) found no relationship between neonatal mortality and a measure of socioeconomic status, while post-neonatal mortality and socioeconomic status were indicated to be significantly correlated. In a similar Boston study, Donabedian, et.al., (7,p. 1089) also concluded that neonatal mortality is much less sensitive to socioeconomic differentials than post-neonatal mortality.

[7] As mentioned previously in the discussion of the correlation matrix results, the evidence from the regression analysis supports previous findings that post-neonatal mortality is more sensitive than neonatal mortality to differences in socioeconomic status and other institutional factors (7,11,14,15,16).

BIBLIOGRAPHY

1. American Hospital Association, Hospitals, Journal of the American Hospital Assoc., Guide Issue, August 1961, Part II.

2. American Medical Association, Distribution of Physicians in the U. S., 1963, Vol. I Regional, State, County, Chicago: The American Medical Association, 1963.

3. Armstrong, A., "Infant Mortality - Some Possible Determinants," Canadian Journal of Public Health, 57 (June,66), 263-8.

4. Bogue, D. J., Principles of Demography, New York: John Wiley and Sons, 1969, p. 559.

5. Chase, H., Father's Occupation, Parental Age, and Infant's Birth Rank: The Relationship of Certain Biologic and Socioeconomic Factors to Fetal, Infant, and Early Childhood Mortality, Washington: Children's Bureau, 1964.

6. Chase, H., "International Comparison of Perinatal and Infant Mortality," Vital and Health Statistics, Series 3, No. 6, March 1967.

7. Donabedian, A., Rosenfeld, L., Southern,E., "Infant Mortality and Socioeconomic Status in a Metropolitan Community," Public Health Reports, 80 (Dec. 65), 1083-94.

8. Falkner, F., ed., Key Issues in Infant Mortality, Report of a Conference on Reducing Infant Mortality, Bethesda, Md.: National Institute of Child Health and Human Development, 1969.

9. Hartman, E., Sayles, E., "Some Reflections on Births and Infant Deaths Among Low Socioeconomic Groups," Minnesota Medicine, 48 (Dec. 1965), 1711-18.

10. Hunt, E.P., "Infant Mortality and Poverty Areas," Welfare in Review, 5 (Aug. 67), 1-12.

11. McCarthy, M.A., "Infant, Fetal, and Maternal Mortality, U.S., 1963," Vital and Health Statistics, Series 20, No. 3, Sept. 1966.

12. Moriyama, I.M., "Infant Mortality Trends, U.S. and Each State, 1930-64," Vital and Health Statistics, Series 20, No. 1, Nov. 1965.

13. Moriyama, I.M., "Present Status of the Infant Mortality Problem in the United States," American Journal of Public Health, 56 (April 66), 623-25.

14. Shapiro, S., Schlesinger, E., Nesbitt,R.E.Jr. "Infant and Perinatal Mortality in the U.S.," Vital and Health Statistics, Series 3, No. 4, 1965.

15. Shapiro, S., Schlesinger, E.,Nesbitt,R.E.Jr., Infant Mortality, Perinatal, Maternal and Childhood Mortality in the United States, Cambridge: Harvard University Press, 1968.

16. Stockwell, E.G., "Infant Mortality and Socioeconomic Status: A Changing Relationship," Milbank Memorial Fund Quarterly, 40 (Jan. 62), 101-11.

17. United Nations Statistical Office, Statistical Yearbook, 1970, New York, 1971.

18. U. S. Bureau of Census, U. S. Census of Population, Vol. I,Characteristics of the Population, Parts 2, 12, 19, 20,26 35, 42, 44. Washington, D. C., 1963.

19. U. S. Public Health Service National Vital Statistics Division, Vital Statistics of the United States 1960, Vol. II, Mortality, Part B, Washington, D. C., 1963

20. U. S. Public Health Service, Division of Vital Statistics, Vital Statistics of the United States, 1967, Volume II Mortality, Part A, Washington, D. C., 1969.

21. U. S. Public Health Service, National Vital Statistics Division, Vital Statistics of the United States, Volume I--Section 3 Natality - Local Area Statistics -Tables, Washington D. C. 1963.

22. Willie, C.V., Rothney, W.B., "Racial, Ethnic, and Income Factors in the Epidemiology of Neonatal Mortality," American Sociological Review, 27 (Aug. 62), 522-26.

MEASURES OF ASSOCIATION BETWEEN A DICHOTOMOUS AND A CONTINUOUS VARIABLE

Janet Dixon Elashoff, Stanford University

Introduction How should the degree of association between a dichotomous variable and a continuous variable be measured? The usual answer is to use the point biserial correlation coefficient. This coefficient, however, is specially designed for the case in which the conditional distribution of y, the continuous variable, given the value of x, the dichotomous variable, is normal, and the mean of the conditional distribution of y depends on x but the variance does not.

Goodman and Kruskal (1954) have argued persuasively that a measure of association for cross classifications should be chosen with a particular underlying model and a purpose in mind. Many different models could be proposed to describe a relationship between a dichotomous and a continuous variable; two general models will be discussed here. A measure of association might be examined with many different purposes in mind. In this paper, some measures of association are suggested which are appropriate for the purpose of screening y variables for use in predicting x. That is, we propose measures of association appropriate for the classification problem.

The basic model to be discussed is one in which the x variable takes on the values 1 and 2 with probabilities (1-p) and p respectively. The distribution of y given x is $F_x(y)$. The problem is to decide how useful the y variable would be in assigning new individuals to x categories, given n_1 observations with x = 1 and n_2 observations with x = 2, with $N = n_1 + n_2$ and observations y_{ij}, i = 1,2 and j = 1,...n_i. We discuss two situations: (1) $F_1(y)$ and $F_2(y)$ differ only in the mean; (2) $F_1(y)$ and $F_2(y)$ may have different variances as well as different means.

Model 1 $F_1(y)$ and $F_2(y)$ differ only in the mean. If normality is assumed, the point biserial correlation coefficient, ρ, is appropriate. The probability of misclassification using y is a function of Δ, the distance between $F_1(y)$ and $F_2(y)$, where

(1) $$\Delta = \frac{\mu_2 - \mu_1}{\sigma}$$

and ρ is a function of Δ,

(2) $$\rho = \Delta\sqrt{\frac{p(1-p)}{1 + p(1-p)\Delta^2}} \ .$$

The maximum likelihood estimator of ρ is

(3) $$r_{pb} = \hat{\Delta}\sqrt{\frac{\hat{p}(1-\hat{p})}{1 + \hat{p}(1-\hat{p})\hat{\Delta}^2}}$$

where

(4) $$\hat{p} = n_2/N$$

(5) $$\hat{\Delta} = \frac{N(\bar{y}_2 - \bar{y}_1)}{\sum_{j=1}^{n_1}(y_{1j}-\bar{y}_1)^2 + \sum_{j=1}^{n_2}(y_{2j}-\bar{y}_2)^2} \ .$$

Conditional on n_1 and n_2, a test of ρ = 0 can be based on the usual t test;

(6) $$t = \frac{r_{pb}\sqrt{N-2}}{\sqrt{1-r_{pb}^2}}$$

has a t distribution with N-2 degrees of freedom when ρ = 0.

When Model 1 is true, but we are unwilling to assume normality, it is natural to consider nonparametric classification procedures based on ranks. Das Gupta (1964) suggests a classification procedure based on the sample cumulative distribution function. Define

(7) $$\hat{F}_i(a) = c_i/n_i$$

where c_i is the number of observations $y_{ij} \leq a$. Let y' be an observation to be classified. Then assign an individual to category x = 1 if

(8) $$\left|F_1(y') - \tfrac{1}{2}\right| < \left|F_2(y') - \tfrac{1}{2}\right| \ .$$

Using this classification procedure, the probability of misclassification is a function of π, where

(9) $$\pi = P(\ y_2 > y_1\ )$$

is the probability that a y observation from x = 2 is larger than a y observation from x = 1. For a dichotomous and a continuous variable, Goodman and Kruskal's measure of association γ reduces to

(10) $$\gamma = 2\pi - 1 \ .$$

The Mann-Whitney U statistic provides an estimator of π and

(11) $$\hat{\gamma} = 2U/n_1 n_2 - 1 \ .$$

Conditional on n_1 and n_2, a test of γ = 0 can be made using tables for the U statistic.

Another method of classification using ranks was developed by Stoller (1954) for the situation where $F_x(y)$ is absolutely continuous and the optimal discrimination rule consists of classifying an individual into category 1 if $y \leq a^*$ and into category 2 otherwise. The probability of a correct classification using any cutoff point a is

(12) $$Q(\ a\ ) = (1-p)F_1(a) + p(1 - F_2(a))$$

and a natural measure of association is

(13) $$\lambda_1 = 1 - \frac{P(\ \text{misclassification}\,|\,\text{y known}\ )}{P(\ \text{misclassification}\,|\,\text{y unknown}\ )} = \frac{Q(\ a^*\ ) - m}{1 - m}$$

where $m = \max(p,(1-p))$.

A distribution-free estimate of Q(a) for any a is obtained by substituting $\hat{p} = n_2/N$ and

$\hat{F}_1(a)$ in (12) to obtain

(14) $$\hat{Q}(a) = \frac{1}{N}(n_2 + c_1 - c_2) \; .$$

The point a^* is estimated using the point a for which $\hat{Q}(a)$ is maximized. Thus letting

(15) $$d^+ = \max_a(c_1 - c_2)$$

(16) $$\hat{Q}(a^*) = \frac{1}{N}(n_2 + d^+) \; .$$

If it is not known a priori whether $\mu_2 > \mu_1$ or $\mu_2 < \mu_1$, the rule can be extended by letting

(17) $$d^- = \max_a(c_2 - c_1)$$
$$d = \max(d^+, d^-)$$

and defining

(18) $$\hat{Q}(a^*) = \begin{cases} \frac{n_2 + d}{N} & d^+ > d^- \\ \frac{n_1 + d}{N} & d^+ < d^- \end{cases}$$

and

(19) $$\hat{\lambda}_1 = \frac{\hat{Q}(a^*) - \hat{m}}{1 - \hat{m}} \; .$$

This derivation assumes that we want to estimate p from the sample at hand. However, if $n_1 = n_2$, or if we can assume p = .5, the formulation is simplified and the distribution theory is known. Define

$$D^+ = \max_a(\hat{F}_1(a) - \hat{F}_2(a))$$

(20) $$D^- = \max_a(\hat{F}_2(a) - \hat{F}_1(a))$$

$$D = \max(D^+, D^-) \; ;$$

these are the well-known Kolmogorov-Smirnov statistics. Then

(21) $$\hat{\lambda}_1 = D$$

and a test of $\lambda_1 = 0$ conditional on n_1 and n_2 can be based on tables of the D statistic.

Some general properties of these three measures of association are obvious. The measures ρ and γ range from -1.0 to +1.0 while λ_1 must lie between 0 and +1.0. For fixed $F_1(y)$ and $F_2(y)$, γ is unaffected by the value of p, but ρ and λ_1 decrease as $|p - .5|$ increases. The estimator $\hat{\lambda}_1$ can be expected to have a positive bias. The measure $\hat{\rho}$ can be expected to be much more strongly affected by the presence of outliers.

The behavior of these association measures is illustrated in two examples. Example 1 is calculated on the data shown in Table 1 which is generated by a normal shift model with $\Delta = 1$. The estimates are $r_{pb} = .43$, $\hat{\gamma} = .55$, and $\hat{\lambda}_1 = .67$; all are significantly different from zero at the 5% level; note that r_{pb} has the smallest and $\hat{\lambda}_1$ the largest value. For example 2, the data from Table 1 was used again, except that the largest observation in category 1 was changed from 7.6 to 27.6. For example 2, the estimates are $r_{pb} = .34$, $\hat{\gamma} = .46$, and $\hat{\lambda}_1 = .60$; both $\hat{\gamma}$ and $\hat{\lambda}_1$ are still significant at the 5% level. Note that the effect of the outlier on $\hat{\lambda}_1$ was considerably smaller than on r_{pb} and $\hat{\gamma}$.

TABLE 1. DATA GENERATED FROM A NORMAL DISTRIBUTION WITH $\sigma^2 = 100$

$\mu_1 = 0$	$\mu_2 = 10$
-16.9	-19.0
-12.2	-9.9
-6.7	-5.8
-3.4	0.7
-2.1	1.6
-1.0	8.1
-0.9	10.4
-0.8	10.7
0.7	10.8
1.5	11.0
1.8	11.1
2.9	12.8
3.7	21.5
3.9	22.1
7.6	26.0

These three measures of association provide reasonable and interpretable measures of association for the classification problem where only a difference in means is of interest. But what about the situation in which the variances may differ also?

<u>Model 2</u> The conditional distributions $F_1(y)$ and $F_2(y)$ may differ in variance as well as in mean. If normality is assumed, the probability of misclassification using a quadratic classification rule is a rather messy function of the means and variances which suggests no simple overall measure. As an ad hoc two-stage procedure, one could examine r_{pb} first and if it were not found to be significant, take a look at the F statistic.

A one-stage procedure can be obtained by extending the Stoller classification procedure to a rule in which an observation is assigned to category 1 if $y \le a_1^*$, or $y > a_2^*$. Again, the cutoff points a_1^* and a_2^* are estimated by maximizing the estimated probability of a correct classification and the measure λ is used. When p is estimated from the data,

(22) $$\hat{\lambda}_2 = 1 - \frac{(a - d^+ - d^-)}{N(1 - m)}$$

where

(23) $$a = \begin{cases} n_2 & \hat{a}_1^* < \hat{a}_2^* \\ n_1 & \text{otherwise} \end{cases} \; .$$

For p = .5

(24) $\hat{\lambda}_2 = D^+ + D^-$

where D^+ and D^- are given in (20). Using definition (24), the distribution of $\hat{\lambda}_2$ conditional on n_1 and n_2 is given by Gnedenko (1954).

In examples 1 and 2, where $\hat{\lambda}_2 = .73$, $\hat{\lambda}_2 = .67$ respectively, $\hat{\lambda}_2$ is only slightly larger than $\hat{\lambda}_1$. It is larger than $\hat{\lambda}_1$ because of the -19.0 observation in category 2 which is smaller than all the observations in category 1. In small samples like these, $\hat{\lambda}_2$ will be overly sensitive to one observation.

Example 3 has been calculated on the data shown in Table 2 which was generated by a normal model with $\mu = 0$ and $\sigma_1 = 10$, $\sigma_2 = 40$. The estimates of the association measures are $r_{pb} = .19$, $\hat{\gamma} = .08$, $\hat{\lambda}_1 = .40$, $\hat{\lambda}_2 = .73$. The estimates of ρ and γ are small. Although not significant at the 5% level, $\hat{\lambda}_1$ is fairly sizeable by the standards one is used to with measures of association. Of course, if the variances are quite different, one could often expect to do better even with a one-sided classification rule than would be obvious from examination of a difference in means. The estimated $\hat{\lambda}_2$ is significant at the 5% level.

TABLE 2. DATA GENERATED FROM A NORMAL DISTRIBUTION WITH $\mu = 0$

$\sigma_1 = 10$	$\sigma_2 = 40$
-25.1	-57.0
-13.0	-50.6
-10.9	-23.6
-6.6	-17.7
-6.1	-15.4
-4.3	-14.0
-1.5	-10.6
-1.3	6.6
0.1	7.6
3.2	28.0
3.8	42.8
4.0	56.4
9.0	59.0
10.3	61.2
13.5	69.2

The measure $\hat{\lambda}_2$ would seem to be a useful measure of association for a dichotomous and a continuous variable for the classification problem in which means, variances, or both may differ.

Additional research to determine large and small sample properties of these sample measures of association for specific choices of $F_1(y)$ and $F_2(y)$ is underway. Investigation of other models for the relationship between dichotomous x and continuous y and other problems requiring a measure of association should lead to alternative measures.

REFERENCES

Das Gupta, Somesh (1964). "Non-parametric classification rules." Sankhyā, Series A, 26, 25-30.

Gnedenko, B. V. (1954). "Kriterien für die Unveränderlichkeit der Wahrscheinlichkeitsverteilung von zwei unabhängigen Stichprobenreihen." Mathematisches Nachrichten, 12, 29-66.

Goodman, L. A. and Kruskal, W. H. (1954). "Measures of association for cross classifications." Journal of the American Statistical Association, 49, 732-764.

Stoller, David S. (1954). "Univariate two-population distribution-free discrimination." Journal of the American Statistical Association, 49, 770-777.

THE INFLUENCE OF SEASON OF CONCEPTION ON OBSTETRIC PROBLEMS AND CASUALTIES

Carl L. Erhardt, Jean Pakter, and Frieda G. Nelson, Department of Health, City of New York

An earlier paper before this Section demonstrated empirically that infant mortality could increase as much as 40 percent among infants conceived during the same one-month interval without detection of the increase as a significant rise in infant mortality in any single month of delivery. (1) The failure to detect the increase is due to the varying times at which the infants subjected to the increase are delivered. In that paper we described seasonal variation for fetal losses over a three year period of conceptions, demonstrating the substantial excess loss among fetuses conceived during the rubella outbreak of 1963-64.

The present report is based on redistribution by month of the last menstrual period (LMP) of all live births and fetal deaths reported in New York City from 1960 to 1968. It covers 1,397,465 reported conceptions from January 1, 1960, to December 31, 1967, an eight year period. Of the total, 829,901 of the mothers were classified as white, 356,701 as nonwhite, and 211,863 as Puerto Rican, meaning that they were born on that island. These three ethnic divisions are used in New York City because obstetric experience differs among the three groups and each is reasonably large.

Several studies (2,3,4) have indicated that the LMP date is sufficiently accurately reported for the purposes of this research. It is estimated that 1.2 percent of the live births and 5.0 percent of the fetal deaths (2.2 percent of the total pregnancies reported) were lost to the study because of lack of information about the LMP date. In order better to detect and compare seasonal variations, since the level of the rates differs by ethnic group, all data were converted to a seasonal index by dividing the monthly rates for the aggregated eight-period by the total rate for this octennium. To reduce random monthly variability and yet emphasize as far as possible monthly fluctuations, a two-month moving average, centered, was then calculated.

It is our purpose in this report to give a summary view of our findings rather than the details about any single aspect.

Seasonal Patterns of Conceptions

Although there are differences among the ethnic groups in the depth of the trough of conceptions during the early part of the year, they all follow a pattern of marked rise in frequency of conceptions toward the end of the year, usually about November. This pattern holds remarkably, regardless of hospital service (private or ward), maternal age, ethnic group or pregnancy order. A similar pattern is found for illegitimates of all three ethnic groups. The general pattern can be seen in Chart 1, where the combined data for all groups are illustrated to show that the shape of the annual curves (dash line) closely follows that for the aggregated eight-year experience (solid line). (5)

On the other hand, no evidence of seasonal variation could be found for the relative frequency of multiple births nor for sex ratios. (5)

Birthweight and Duration of Gestation

Mean birthweight is higher among summer conceptions (July to September) and the lowest mean birthweight occurs among fall and late winter (first quarter) conceptions. This pattern tends to be confirmed when we review the percent of births under 2,501 grams by month of conception. Yet, when this proportion for July to September is tested against the proportion for the rest of the year, the result is significant (at less than the .05 level) during the third quarter only for four of the 12 groups tested (ethnic group, sex, service). Nevertheless, although the differences were usually slight, the proportions of infants of low birthweight still were lower among these summer conceptions than among conceptions at other times of the year for seven of the eight remaining groups.

Altogether, these findings suggest a hypothesis that if the crucial first trimester of pregnancy coincides with the season when the risks of common viral and bacterial infections are least, the opportunities for optimal fetal growth and development or even survival will be enhanced. Of course, this is also the period of the year with maximum sunlight in the New York area and the time when fresh vegetables become available.

There is no doubt that birthweight and duration of gestation are associated. A peak in the mean duration of gestation was found also in the July-September period, but this was not the only peak and the crest varied among the ethnic groups. Moreover, the mean completed weeks of gestation appears to be too crude a measure for precise determination of the facts. When the proportion of deliveries at less than 36 weeks was considered, the higher proportions were found in the early part of the year and after midyear. Here again, the findings tended to be consistent with those for birthweight but with some shift. However, when one examines Chart II, showing both mean birthweights (by sex) and the percent of deliveries under 36 weeks of gestation (all live births), the inverse relationship between the two variables is evident. On the other hand, when the percent under 2,501 grams in birthweight is charted against the percent less than 36 weeks of gestation, a positive relationship by month of conception appears.

In summary, both birthweight and duration of gestation have a seasonal pattern, but it is not of marked degree, especially for birthweight. The evidence suggests that infants conceived during the summer months have a somewhat longer gestational interval and weigh more than those conceived in most other months. The variations in the means do not appear of much practical importance, but the proportions of low weight babies and those of short gestational interval do seem to vary enough seasonally to have clinical import.

Perinatal Losses

In general, and specifically for each ethnic group, maternal age and pregnancy order, perinatal mortality rates are seasonally high among winter conceptions and decline to a low about October. Neither component of the perinatal loss rate (late fetal and early infant deaths) departs from this pattern. Chart III is presented to show this pattern as indicated by the fetal death component, which is depicted in the lowest bank. Inclusion of this chart has the merit of demonstrating that a quite similar pattern generally exists for fetal losses at earlier gestational intervals.

These losses were unusually high, as was observed in the earlier study (1), during the latter part of 1963 and early 1964 during the course of the rubella epidemic in New York City. However, the rates were also unusually high from December 1964, through September, 1965. Thus far, we have been unable to explain the high rates for this interval.

Complications of Pregnancy

As used in this report, complications of pregnancy are those conditions listed as a check-off item on the birth certificate form as "conditions present during pregnancy". They include such conditions as eclampsia, preeclampsia, hypertensive disease, heart disease, German measles (and trimester), tuberculosis, neoplasms and syphilis. As might be expected, the only one of the specific conditions that shows unequivocal and marked seasonal variation is rubella, an infectious disease with recognized seasonal variation in incidence. Chart IV illustrates the findings for several of these conditions and the marked seasonal pattern for rubella occurring in the first trimester of pregnancy is obvious. Tuberculosis appears peculiarly low among pregnancies starting late in the year, but findings seem negative for the other conditions shown on this chart. Chart V is included merely to demonstrate that combining rubella reported as occurring in different trimesters of pregnancy can conceal much information. The seasonal pattern is seriously dampened, as indicated by the upper panel. The lower panel makes it clear that women conceiving at different times of the year encounter the seasonal impact of rubella at different stages of pregnancy, and that the aggregation of these three curves produces the apparently mild seasonal pattern of the upper panel.

For preeclampsia, eclampsia and hypertensive disease (not shown here), there is some evidence that they are most likely to occur (or be exacerbated) among pregnancies starting during the middle of the year, with perhaps extra hazard for whites from preeclampsia during most of the first half of the year.

Congenital Anomalies

Congenital anomalies were coded in detail following to major extent the rubrics of the eighth revision of the International Classification of Diseases. Only those pregnancies reaching at least 17 weeks of gestation were included, since it was considered unlikely that anomalies would be identified or reported when pregnancy terminated earlier. However, to maximize the ascertainment from the existing records, malformations reported as causes of infant deaths were included whenever such malformations had not been reported on the corresponding birth certificates of the infants. Moreover, as many as three separate anomalies were coded in each case. The figures utilized here represent, therefore, the frequency of each anomaly and not counts of infants with anomalies.

Despite aggregation of data over an eight year interval, small numbers were found for most anomalies. Although suggestive

CHART I

ACTUAL AND EIGHT-YEAR AVERAGE RATIOS* OF OBSERVED TO EXPECTED CONCEPTIONS, BY MONTH OF CONCEPTION : NEW YORK CITY 1960-1967

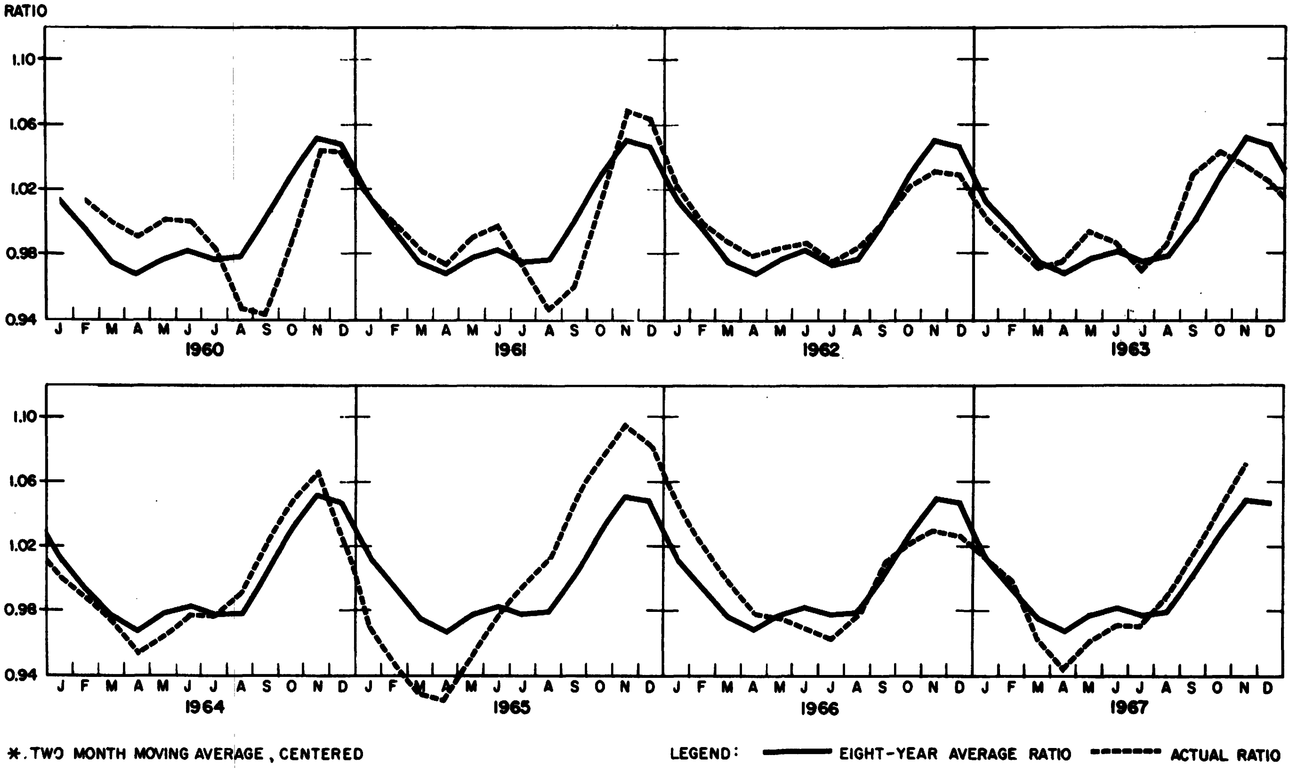


CHART II

MEAN BIRTHWEIGHTS BY SEX, AND PERCENT UNDER 36 WEEKS GESTATION, OF SINGLE LIVE BIRTHS* EXPRESSED AS MULTIPLES OF THE OCTENNIAL AVERAGES** BY ETHNIC GROUP, SERVICE AND MONTH OF CONCEPTION: NEW YORK CITY, 1960-1967

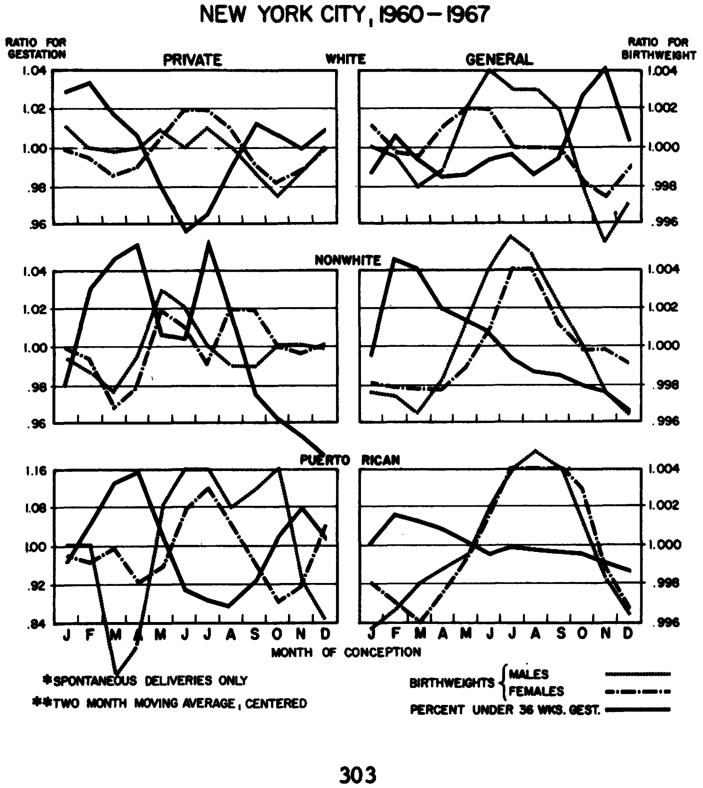
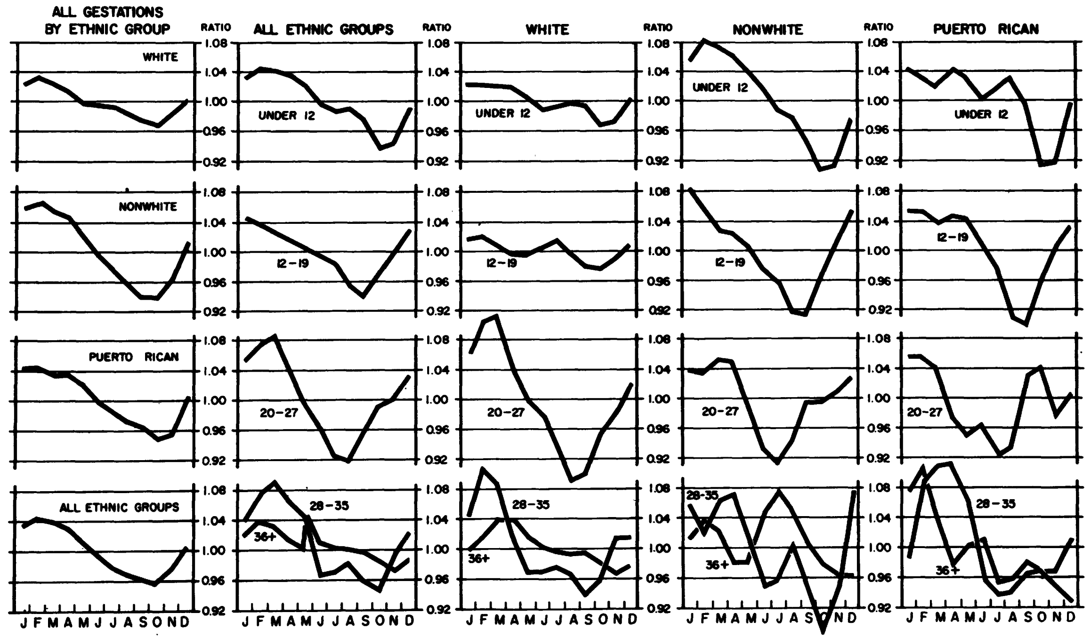


CHART III

REPORTED FETAL DEATHS* PER 1,000 CONCEPTIONS EXPRESSED AS MULTIPLES OF THE OCTENNIAL RATE** BY MONTH OF CONCEPTION, ETHNIC GROUP AND DURATION OF GESTATION (WEEKS): NEW YORK CITY, 1960-1967

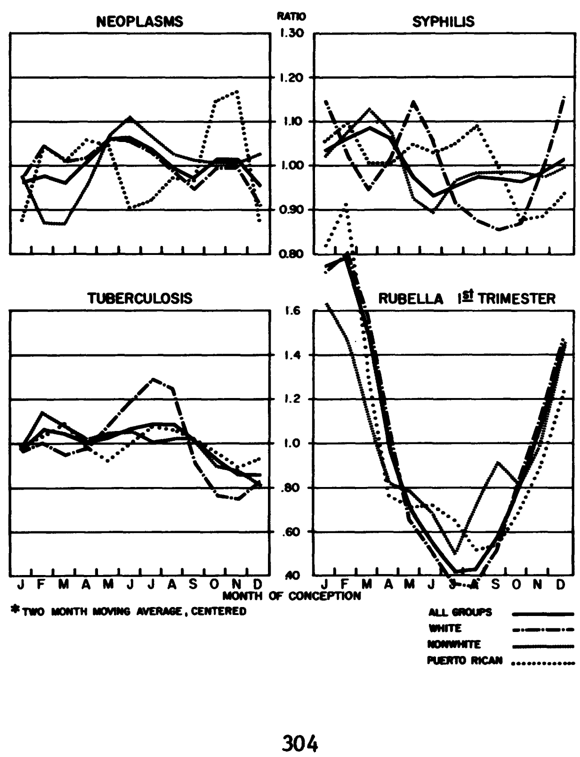


*EXCLUDES THERAPEUTIC ABORTIONS
**TWO MONTH MOVING AVERAGE, CENTERED

CHART IV

COMPLICATIONS OF PREGNANCY

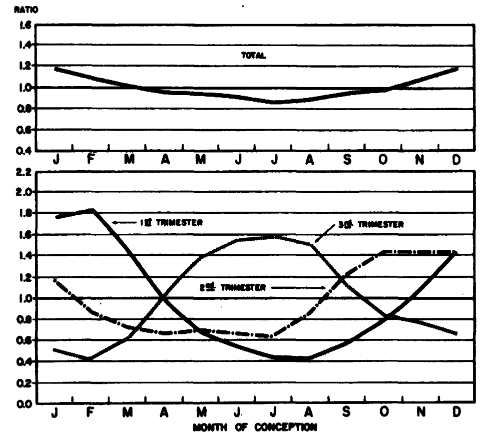
SELECTED COMPLICATIONS OF PREGNANCY PER 100,000 CONCEPTIONS REACHING AT LEAST 17 WEEKS GESTATION. AVERAGE MONTHLY RATES EXPRESSED AS MULTIPLES OF THE AVERAGE OCTENNIAL RATE*, BY ETHNIC GROUP AND MONTH OF CONCEPTION: NEW YORK CITY, 1960-1967



*TWO MONTH MOVING AVERAGE, CENTERED

CHART X

REPORTED FREQUENCY OF RUBELLA AS COMPLICATION OF PREGNANCY
PER 100,000 CONCEPTIONS REACHING AT LEAST 17 WEEKS GESTATION.
AVERAGE MONTHLY RATES EXPRESSED AS MULTIPLES
OF THE AVERAGE OCTENNIAL RATE*
BY MONTH OF CONCEPTION AND TRIMESTER OF OCCURRENCE:
NEW YORK CITY, 1960-1967



*TWO MONTH MOVING AVERAGE, CENTERED.

indications of variation by season of conception were sometimes noted for one or two of the ethnic groups or for either sex, or for either maternal age group into which the data were subdivided, we rejected a conclusion that such seasonal variation actually exists in most instances and wavered in others. We have looked for common patterns between the indicated population subgroups and have been disinclined usually to assert that seasonality exists when such a common pattern could not be reasonably identified. In so doing, we may have overlooked a true seasonal pattern that for some reason actually applies to only one ethnic group or one sex or one maternal age group.

We have concluded that there exists a possibility of higher risk of hemolytic disease among infants conceived about midyear, especially among nonwhites and Puerto Ricans. It also appears that in the New York City vicinity the frequency of anencephaly may be relatively high among spring conceptions. Clubfoot appears to be a higher risk among second and third quarter conceptions and reduction deformities of the limbs among those during the first and last quarters. There is a weak indication that other limb deformities may occur most frequently among fall conceptions. For both Mongolism and polydactyly, the influences are believed to be largely genetic, but an impression is obtained that a superimposed environmental influence may exist.

Summary

Conception is most likely to occur during the last quarter of the year in New York City, regardless of other factors. But infants conceived during the third quarter tend to be slightly heavier and of somewhat longer gestation than those conceived during the rest of the year. Perinatal and early fetal losses are relatively high among pregnancies starting in winter and become progressively lower until about October. Rubella in the first trimester is most frequent, as expected, among winter conceptions and conditions related to eclampsia appear to rise in midyear. Variation by season of conception for congenital malformations was not clear for most conditions, but anencephaly seems relatively high among pregnancies starting in spring and clubfoot among those starting throughout the middle of the year.

References

1. Erhardt, Carl L., Nelson, Frieda G. and Pakter, J., "Season of Conception: A Neglected Analytic Variable". Proc. Soc. Stat. Sec., ASA, 1966

2. Treloar, Alan E., et al., "Analysis of Gestational Interval". Am. J. Obs. and Gyn. 99:34, Sept. 1, 1967

3. Hammes, Laurel M. and Treloar, Alan E., "Gestational Interval from Vital Records". Am. J. Public Health 60:1496, Aug., 1970

4. Erhardt, Carl L., "Accuracy of Reported LMP Data on Birth Certificates". Dept. of Health, City of New York, 1968 (mimeo)

5. Erhardt, Carl L., Nelson, Frieda G. and Pakter, J., "Seasonal Patterns of Conception in New York City". Am. J. Public Health (to be published)

Acknowledgments

This research was supported in part by Grant H-106, Maternal and Child Health Service, Department of Health, Education and Welfare. Acknowledgment is gratefully recorded to Mrs. Esther Bigus for her many contributions to the completion of the project.

THE USE OF MUTUALLY INTERDEPENDENT VS. MUTUALLY INDEPENDENT SCHOOL SYSTEM OUTPUTS IN ESTIMATING EDUCATION PRODUCTION FUNCTIONS*

Thomas G. Fox, The Pennsylvania State University

1. INTRODUCTION

During the past few years, school system "outcomes" (or outputs) such as "academic achievement" and "dropout prevention or 'holding power'" have been studied in relation to the school system resources available to "produce" these outputs. The functions relating school system inputs to outputs are known as education production functions in the literature of the economics of education. Though it is unequivocally accepted that very complex schooling processes "turn-out" more than one type of schooling or education output, previous empirical work in this area has statistically emphasized the mutual independence of education outputs while most authors verbally acknowledged output interdependence. This study builds on previous empirical work by estimating an interdependent output high school production function for a large-city high school system using simultaneous euqations and comparing results obtained with ordinary-least-squares estimates. The sensitivity of the statistical significance of coefficients obtained to the weighting method used is then considered.

The analysis is developed in three sections. Section 2 contains an overview of education production functions with interdependent outputs and the accompanying structural equations together with a discussion of alternative methods of weighting variables. Two-stage-least-squares (2SLS) regression estimates of the structural equations in the high school production function are presented in Section 3. In Section 4, ordinary-least-squares (OLS) regression analysis is applied to the structural equations with one output variable deleted to obtain naive estimates of reduced-form coefficients. Alternative estimates of the reduced-form coefficients are derived from 2SLS estimates of the coefficients in the structural equations. Estimated coefficients obtained are then comparatively analyzed. Conclusions are restated in the last section.

2. OVERVIEW

In a very general form, an education production function with two interdependent outputs can be written as

$$F(SI, PC, GE, AA, HP) = 0 \qquad (1)$$

where SI (school system resource inputs), PC(relevant personal pupil characteristics), and GE(general socioeconomic environment) are--in the short-run--given, predetermined, and therefore exogenous. The output variables AA (academic achievement) and HP (school holding power; i.e., 1 - dropout rate) are jointly determined as endogenous variables within the school system in interaction with the exogenous variables and with each other. Thus the education putput variables are hypothesized to be mutually interdependent.

Equation (1) contains two production functions, one for the production of academic achievement

$$AA = F_1(SI, PC, GE, HP) \qquad (2)$$

and one for the production of holding power --

$$HP = F_2(SI, PC, GE, AA). \qquad (3)$$

These are the structural equations for the education production function and are to be solved simultaneously.

Authors of the earlier single-equation studies [eq., 2,3,4] have recognized that since the size of each school affects the volume of school system inputs, input variables tend to be highly intercorrelated if measured as, say, total teacher man-years, when school enrollments vary from very small to very large. To reduce multicollinearity among the exogenous variables, it has been common practice to express all variables as averages per student, then re-insert school size by including enrollment or attendance as a separate variable. Thus variables usually include average achievement per student, median family income, average class size, among others.

Such a method of deflating "total school inputs" and "total school outputs" may or may not facilitate statistical estimation of underlying relationships among variables. In this study, each equation will be estimated in "average" form and in "total" form so that the statistical results can be compared.

Space does not permit a review of the numerous difficulties inherent in defining and measuring variables for education production functions. These problems have been carefully considered elsewhere [1,2,3]. The reader is forewarned, however, that data limitations and inadequate knowledge of "schooling technology" severely constrain the applicability of the education production function concept to all of the empirical studies that the author is aware of, including the present study. Measurement is imperfect, variables for pupil characteristics are usually unavailable, census data or attendance area or district-wide averages must often be used as "proxies" for pupil characteristics and for the general environment, (and hence empirically the PC's have been dropped from equations 1 and 2).

In addition to problems in specifying the proper variables, there are significant issues involved in specifying the mathematical form of the production functions. For the empirical work involved in this study, a log-linear education production function is assumed, as this type of function has proved highly useful in other production function studies. The structural equations assumed to represent the education production function are

$$AA' = N' \; X_3^{\alpha_3'} \; X_4^{\alpha_4'} \; X_5^{\alpha_5'} \; X_7^{\alpha_7'} \; X_8^{\alpha_8'} \; X_9^{\alpha_9'} \; X_{10}^{\alpha_{10}'} \; HP^{\lambda'} + b_{12}'X_{12} + b_{13}'X_{13}$$

*The author gratefully acknowledges computational services provided by the Computation Center, The Pennsylvania State University.

$$HP'' = N'' X_2^{\alpha_2''} \; X_4^{\alpha_4''} X_5^{\alpha_5''} X_6^{\alpha_6''} X_7^{7''} X_8^{8''} X_9^{9''} X_{10}^{10''} AA^{2''} + b_{12}'' X_{12} + b_{13}'' X_{13} \quad (5)$$

where

- X_2 = student employment, a proxy for ability of students to reduce the opportunity cost of a high school education in terms of foregone earnings
- X_3 = students planning on college attendance, a proxy for student preferences
- X_4 = high school attendance area income, a proxy for income of students' families
- X_5 = student class hours in vocational courses to adjust for curricula-mix differences among schools
- X_6 = years teaching experience of teachers, a proxy for the quality of man-years of teacher time
- X_7 = teacher man-years
- X_8 = auxiliary-service man-years (primarily teachers not in class-rooms such as librarians)
- X_9 = text and library book expenditures, a proxy for materials, supplies, other non-building educational capital inputs
- X_{10}= coded building age--weighted by attendance, a proxy for capital plant facilities of different ages and sizes
- X_{12}= dummy variable for racial composition of high school attendance area = 1 if less than 11% NW, otherwise 0, a proxy for racial mix of students
- X_{13}= dummy variable for racial composition of high school attendance area = 1 if 11% - 45% NW, otherwise 0, a proxy for racial mix of students
- AA = 11th grade reading achievement, a proxy for 9th-12th grade academic achievement
- HP = holding power of schools (1 minus drop-out rate for "average" weight equations, multiplied by number of students for "total" weight equations)
- N', N" = constant terms in equations

Date are for the 39 Chicago public high schools and are discussed at length in Burkhead-Fox-Holland [2].

The mathematical form assumed is log-linear in the parameters α_i, requiring all variables with α_i coefficients to have non-zero values before any production is forthcoming and has the well-known desirable economic features associated with Cobb-Douglas-type production functions. In simultaneous equation form each exogenous variable has both a direct affect <u>and</u> an indirect affect on <u>each</u> endogenous output variable: teacher man-years, for example, are expected to directly effect student achievement, and have an indirect affect operating through the impact of holding power on achievement as holding power is simultaneously affected by changes in teacher man-years. The total affect of each exogenous variable upon each endogenous variable, consists of the sum of its direct and indirect effects. The total impact of exogenous variables is found from reduced-form estimates of the structural equations. Hence any statistical estimating techniques which capture only the direct impact of exogenous input variables upon endogenous output variables can yield erroneous conclusions if estimated coefficients are interpreted as estimates of total effects.

3. 2SLS ESTIMATES

Estimates of the coefficients for the structural equations of the high school production function obtained by application of 2SLS techniques are displayed in Table 1, with standard errors reported in parentheses. The first pair of coefficient columns are for equation (4), the second pair pertain to equation (5). Headings titled "average units" are equations expressed in averages per student, per school, form; the headings "total units" are for totals by school.

Looking first at the structural equations for achievement, we observe that the weighting method affects the number of statistically significant coefficients, if a two-tailed t-test is applied and we require coefficients to be at least twice the size of their respective standard errors (significant coefficients are marked with asterisks). Two coefficients, one socioeconomic (X_4, income) and one for a school system input (X_8, auxiliary service man-years) are significant with "average" weights. Seven coefficients--three socioeconomic (X_4, income, and X_{12} and X_{13} for race), one for student preferences (X_3, students planning on college attendance), two school inputs (X_8, auxiliary service man-years and X_{10}, building age-attendance weighted) and the endogenous holding power variable (HP)--are significant in the "total form" structural equation for AA. The coefficients on the exogenous variables in the structural equations record only their direct affects upon achievement; all indirect affects operate through the other endogenous variable, holding power.

Interestingly, where coefficients are significant in <u>both</u> "average" and "total" form (X_4, X_8) they differ from each other by less than one standard error of either coefficient and hence are not statistically different.

The equation in "total" form explains almost 98 percent of the variance in total AA, whereas in "average" form, the same set of variables accounts for 93 percent of the variance in average AA. The endogenous variable HP is significant only in the "total" weighted structural equation for the production of achievement. Apparently, expressing all variables as "averages per students" to reduce multicollinearity among exogenous variables leads to a weaker explanatory equation with fewer individually significant coefficients.

The second pair of columns in Table 1 pertain to equation (5). The structural equation for holding power in "total" form explains almost 99 percent of the variance in total holding power; when estimated in "average" form, the same set of variables explain only 60 per cent of the variance in holding power. Excluding constant terms, the "total" form yields 6 coefficients whose values are more than twice the size of their standard errors, the "average" weighted equation has 4 significant coefficients. In the two instances where both "average" and "total" coefficients are significant (AA, X_7) they differ in magnitude by more than two standard errors for teacher-man-years but by less than two standard errors for achievement (when only the standard error for the "total" weight coefficient is used), and their signs are all positive.

TABLE 1
TWO-STAGE-LEAST-SQUARES ESTIMATES OF STRUCTURAL EQUATIONS
FOR A HIGH SCHOOL SYSTEM PRODUCTION FUNCTION

		Achievement		Holding Power	
		Average Units	Total Units	Average Units	Total Units
AA	Achievement-Stanines (Endogenous)			.444* (.010)	.381* (.146)
HP	Holding Power (Endogenous)	4.022 (3.544)	.171* (.072)		
X_{12}	Race=1 if < 11% NW, Otherwise 0	.001 (.031)	.081* (.021)	- .003 (.007)	- .038* (.011)
X_{13}	Race=1 if 11-45% NW, Otherwise 0	- .024 (.825)	.052* (.026)	.008 (.009)	- .015 (.013)
X_2	Student Employment			- .024 (.034)	.112* (.054)
X_3	Students Planning on College Attendance	.082 (.171)	.227* (.107)		
X_4	Area Income ($100)	.526* (.086)	.489* (.172)	- .228* (.061)	- .015 (.097)
X_5	Vocational Class Student Hours	- .067 (.080)	- .019 (.092)	.018 (.023)	.099* (.027)
X_6	Years Teaching Experience			- .001 (.005)	- .012 (.012)
X_7	Teacher Man-years per 100 Students	- .114 (.061)	.025 (.272)	.053* (.026)	.326* (.075)
X_8	Auxiliary Service Man-years per 100 Students	.223* (.068)	.159* (.073)	- .087* (.027)	- .046 (.044)
X_9	Text and Library Book Expenditures ($)	.003 (.010)	- .031 (.067)	.007 (.018)	.065* (.025)
X_{10}	Building Age-Weighted by Attendance	.014 (.025)	.042* (.021)	- .007 (.011)	- .001 (.026)
	Constant	-7.099 (7.008)	.056 (1.037)	1.605* (.117)	1.359* (.144)
	R^2	.927	.978	.598	.987

Though it would be interesting to study and analyze the indivdual variables at greater length, deducing certain plausible educational implications, we continue to concentrate on comparisons. As previous education production functions have used single equation estimating models rather than simultaneous-equation models, we now turn to the single-equation comparisons.

4. REDUCED-FORM ESTIMATES

In this section, alternative techniques for assessing the total statistical impact of exogenous variables upon endogenous output variables are considered using achievement, as the endogenous variable.

Suppose first, that some researcher specifies his education production functions exactly as we have specified the structural equations (4) and (5). Suppose, further, that for arbitrary or a priori reasons he concludes that the education outcomes are mutually independent and hence not jointly determined. He therefore deletes $HP^{\lambda'}$ from (4) and $AA^{\lambda''}$ from (5): the resulting single-equation estimating models contain only exogenous variables hypothesized to affect only the single dependent variable in each equation. This version of statistical estimating models, deliberately misspecified (from the context of the previous section) is labled as "simple OLS" in Table 2.

If the analyst recognizes that exogenous variables have both direct and indirect effects on endogenous variables, he will no doubt be interested in estimating their total effect on each endogenous variable. This is found by eliminating all but one endogenous variable from the production function in order to obtain the reduced-form equation of the structural education production function. Two different statistical procedures can be employed to estimate these reduced-form coefficients, OLS and 2SLS regression techniques. In estimating reduced-form coefficients, it is widely recognized in the econometrics literature that OLS yields biased and inconsistent results so 2SLS is preferred [5, p. 189]. In many economic empirical studies, actual estimates of coefficients by either OLS or 2SLS techniques frequently yield similar results, though 2SLS is theoretically preferrable to OLS. Both techniques are utilized herein. Previous empirical studies of education production functions where various education output variables, one at a time, have been regressed against a common set of independent variables

TABLE 2
ACHIEVEMENT: COMPARISONS OF ALTERNATIVE ESTIMATES OF SINGLE EQUATION HIGH SCHOOL SYSTEM PRODUCTION FUNCTIONS

		Average Unit Weights			Total Unit Weights		
		Simple[a] OLS	Reduced-Form Coefficients OLS	Reduced-Form Coefficients 2 SLS	Simple OLS	Reduced-Form Coefficients OLS	Reduced-Form Coefficients 2 SLS
X_{12}	Race=1 if < 11% NW, Otherwise 0	.032 (.019)	- .010 (.014)	.014	.083* (.020)	.009 (.013)	.214
X_{13}	Race=1 if 11-45% NW, Otherwise 0	.001 (.059)	.001* (.000)	- .010	.055* (.022)	.001 (.000)	.134
X_2	Student Employment		.048 (.779)	.123		.101 (.055)	.054
X_3	Students Planning on College Attendance	- .112* (.029)	- .065* (.025)	- .104	.250* (.051)	.083* (.032)	.649
X_4	Area Income	.553* (.096)	.467* (.073)	.498	.526* (.075)	.492* (.061)	1.389
X_5	Vocational Class Student Hours	.002 (.053)	- .023 (.047)	- .006	.001 (.071)	- .041 (.030)	- .006
X_6	Years Teaching Experience		.006 (.016)	.005		.007 (.015)	.006
X_7	Teacher Man-years	- .141* (.066)	- .332* (.062)	- .126	.083 (.112)	- .020 (.090)	.231
X_8	Auxiliary Service Man-years	.202* (.076)	.068 (.066)	.162	.168* (.066)	.046 (.055)	.431
X_9	Text and Library Book Expenditures ($)	- .013 (.047)	- .027 (.716)	- .032	- .017 (.035)	- .020 (.030)	- .057
X_{10}	Building Age-Weighted by Attendance	.013 (.028)	- .006 (.023)	.018	.043 (.025)	.015 (.020)	.120
	Constant	.858* (.224)	1.122* (.183)	.819	.303 (.179)	1.608* (.241)	.823
	R^2	.901	.940		.976	.986	

Notes: [a]OLS estimates of structural equations for achievement after deleting high school holding power as a variable.

using OLS (such as Burkhead-Fox-Holland) can be properly interpreted as (naive) estimates of the reduced-form of structural education production functions even though this was not the original intent of the authors.

Turning to Table 2, note that the deliberately misspecified simple OLS equation yields as many significant coefficients for exogenous variables (four) as the OLS model, using average unit weights, and simple OLS yields twice as many, when total unit weights are utilized, yet the OLS always explains more of the variance in academic achievement. The number of statistically significant coefficients is quite sensitive to the specification of variables in the estimating equations. (As existing tests of significance for 2 SLS estimated reduced-form coefficients, especially from small samples, are questionable, these are not presented: 2 SLS reduced-form coefficients will be compared with significant simple OLS and OLS coefficients.) The signs of the coefficients are sensitive to the model used in only one case: where average unit weights are used, the race variable X_{13} has a positive sign for the simple OLS and OLS variants, a negative sign for 2 SLS variant. Values for some coefficients are similar (e.g., average weight, X_4 coefficients are within 2 standard errors of each other); yet differ dramatically when the weighting system changes (e.g., X_4). Similar conclusions emerge from analysis of reduced-form equations for holding power, not presented herein to conserve space.

5. CONCLUSIONS

We find the empirical results obtained for the high school production function to be highly sensitive to the estimation techniques employed and variable weights utilized. If these characteristics are also embodied in other education production function studies, and they probably are, then potential operating policy conclusions should not be based on the empirical work, unless a careful sensitivity analysis has been developed which clearly delimits the applicability of the statistical models.

In this study, the 2 SLS "total" variant appears to provide the best statistical results.

In the estimation of the structural education production functions at least with this set of data, variables expressed in "total" weights yield equations with higher coefficients of multiple determination and a larger number of statistically significant coefficients than obtained when structural equations use "average" weighted variables. Further, the fact that three of the four coefficients on endogenous variables are statistically significant supports the hypothesis that the educational outputs of academic achievement and holding power are mutually interdependent.

Turning to the single equation models, we know that, theoretically, the simple OLS variant is improperly specified, given the production function hypothesized, yet it often yields more significant coefficients for exogenous variables than does OLS. Hence empirical results are frequently sensitive to specifications of variables in the production functions, when single equation estimating techniques are employed. Also, it is well-known that OLS gives a biased and inconsistent estimate of reduced-form coefficients: when compared with reduced-form coefficients derived from 2 SLS estimates, OLS estimated reduced-form coefficients are frequently quite different. The total statistical impact of several of the exogenous variables is sensitive to both the weighting method used and the statistical technique used. The 2 SLS method has the advantage of explicitly showing interdependence among the endogenous variables. Much more work on education production functions is needed before users of these studies can be sure that empirical results are not simple statistical artifacts.

SELECTED REFERENCES

[1] Bowles, Samuel, "Towards an Education Production Function," Education, Income and Human Capital, W. Lee Hansen, ed., Studies in Income and Wealth, Vol. 35, National Bureau of Economic Research, (New York: Columbia University Press, 1970), pp. 11-61.

[2] Burkhead, Jesse, Thomas Fox, and John Holland, Input and Output in Large-City High Schools, Syracuse, N.Y.: Syracuse University Press, 1967.

[3] Kiesling, Herbert J., "Measuring a Local Government Service: A Study of School Districts in New York State," Review of Economics and Statistics (1967), 356-67.

[4] ________________, "The Relationship of School Inputs to Public School Performance in New York State," The RAND Corporation, Santa Monica, California (1969).

[5] Walters, A.A., An Introduction to Econometrics, New York: W.W. Norton & Company, Inc., 1970.

THE VALIDITY OF INCOME AND WELFARE INFORMATION REPORTED BY A SAMPLE OF WELFARE FAMILIES

Teh-wei Hu, The Pennsylvania State University

I. Introduction

Survey information is often subject to error [1 and 2]. When the errors have systematic bias, analyses based on this survey information become questionable. This paper will examine two kinds of bias in a survey of welfare families. First is the difference of income between the amount reported from the household interview and the amount recorded at the county welfare agency. Second is the number of families interviewed who denied that they received public assistance, in the form of either cash assistance or medical assistance. A further analysis of sociodemographic factors associated with the differences is also presented. The results of these analyses may contribute to an understanding of the nature of possible bias in earnings and public assistance information from a survey of welfare families.

Although the original intention was to examine the difference between the amount of public assistance reported from the household interview and the amount recorded at the county welfare agency, the data we have collected is not sufficient to study this problem.

II. Data

In a study by this author and his colleagues [4], a household sample of 650 families in a coal mining county of Pennsylvania was obtained. The main reason for choosing this county as the sample area was that 40 per cent of the households in the county had been or were under welfare programs during 1962-68. The household information was obtained through questionnaire interviews performed by school nurses, since all households in the sample had children attending elementary schools at the time the study was conducted. The household interviews provided the income information and also indicated whether the families had been or were still receiving welfare assistance. The public assistance reported by the households dealt only with medical assistance for children. On the other hand, the county welfare agency had welfare assistance and income information for these households. The 650 households contained about 240 welfare families. Of these 240 families, there were only 89 for which matching information on incomes--from household interview questionnaires and the county welfare agency--was available. Of the 240 welfare families, there were 186 families that had information for the analysis of factors that affect their admission or denial of receiving welfare assistance from the government.

Table 1 summarizes the means and standard deviations of socioeconomic characteristics of various groups contained in the study sample. Ethnic origin of the head of the household is not included because the sample area contains only 3 per cent nonwhite population. Therefore, it was not possible to make inferences about reported income difference between white and nonwhite households.

It can be seen from column (1) that the household disposable income reported during interviews

Table 1
Household and Family Characteristics of Various Sample Groups

Variables	Sample to be Examined on Income Difference (1)	Sample to be Examined Concerning Denial of Participation Under The Welfare Program	
		Admitted (2)	Not Admitted (3)
Age of Household head	39 (10)[a]	37 (7)	38 (7)
Percentage Employed, Household Heads	--	74% (44%)	100%
Percentage of Household Heads Living with their Spouses	--	79% (41%)	100%
1968 Monthly Income Reported from Interview[b]	$345 (180)	$333 (273)	$525 (257)
1968 Monthly Income Recorded at Agency	$99 (116)	--	--
Years of Education of Household Head	12.9 (13.2)	11.9 (10.8)	11.7 (2.0)
Sample Size	89	136	50

Notes: [a]Values in the parentheses are the standard deviations of the variables

[b]These incomes are take-home pay (disposable income from all sources).

was about 250 per cent more than the amount recorded at the county welfare agency. Comparing columns (2) and (3), it can be seen that about 27 per cent of the 186 households denied having received either cash assistance or medical assistance. This percentage is much higher than the estimate of less than 10 per cent provided by David [2]. The denial group claimed $192 more (monthly income) or about 57 per cent more than the admitted group. Although the age and education levels of the heads of households are similar between the two groups, the marital and employment status are different. The household heads of the denial group had 100 per cent employed, and all were living with their spouses. The admitted group on the other hand, had 74 per cent employed and 79 per cent living with their spouses. These are sample means of sociodemographic factors between the two groups. The questions are: what are the sociodemographic variables that can explain the difference of income between the reported interviews and the amount recorded at the county welfare agency? And what are the sociodemographic variables that can explain why the interviewed families denied that they received public assistance? The next section will apply the regression technique to answer these two questions.

III. Factors Affecting the Reporting Differences

There is no way of knowing whether the income information obtained from the household interview or the county agency is a correct one. However, it is at least possible to study the income differences from these sources. It is conceivable that welfare families tended to underreport their income to county welfare agencies so that they can qualify to obtain or to maximize their welfare assistance from the government. On the other hand, considering that the household interviews for this study were conducted by the school nurse, the families may have tended to overreport their incomes to make a "good impression" or to "save face" in front of the interviewer. The income difference (D--i.e., income reported from interview minus the income recorded at county welfare agency during 1968, in dollars), was employed as a dependent variable. The age of household head (A), the income information obtained from the household interview (Y), and educational level of household head (E) served as explanatory variables in the regression equation. The estimated regression equation is as follows:

$$D = \underset{(488)}{2410} + \underset{(423)}{724A_1} + \underset{(410)}{1017A_2} + \underset{(0.06)}{1.09Y} + \underset{(11)}{9E}$$

$$R^2 = .77 \qquad N = 89$$

where $A_1 = 1$ for age less than or equal to 34, $A_1 = 0$ otherwise; $A_2 = 1$ for age between 35 and 44, $A_2 = 0$ otherwise. The classification of the ages above 44 are omitted and entered into the intercept. The values in parentheses are standard errors of coefficients. N is the sample size. It can be seen that, except for the education variable, each coefficient is statistically significant at the 5 per cent level, one-tailed test. Although the education variable was specified in dummy variable form, the results were not statistically significant. Therefore, a continuous form of education variable is presented in the model.

The results suggest that the higher the level of household-interviewed income the greater is the difference of income between the reported amount from interview and its recorded amount at the agency. According to the "beta coefficient," household income is the most important factor among these independent variables in explaining the difference between two sources of income. The coefficients of the age variables indicate that the age group between 35 and 44 shows the largest difference between the two kinds of income.

The second question to be examined relates to the factors associated with the families denying participation under welfare programs. A dummy variable, (P), was used as a dependent variable to classify their admission or denial. A value of one was assigned to an admission family, a value of zero to the denial family. The independent variables were the age of household head (A), employment status (N = 1 if employed, N = 0 otherwise), marital status (M = 1 if still living with their spouses, M = 0 otherwise), monthly income (Y, in dollars), and education (E, in years). This regression formulation can be considered as a discriminant function. Thus, the coefficients of these independent variables reflect, if the sign is positive, the probability of the truth being told. On the other hand, if the sign is negative, the coefficients indicate the probability of lying.

One Statistical problem in the estimation of the zero-one dependent variable is that the error term is heteroskedastic [3]; thus the ordinary classical least-squares technique is no longer efficient, although it is still unbiased. To overcome this problem, the estimated $\hat{P}$ from ordinary least-squares was used to construct a variable $[\hat{P}\,(1 - \hat{P})]^{1/2}$ as the weights (W) for each variable in the specified function.

After multiplying the weights for each variable, the model was re-estimated, using ordinary least-squares. The estimated coefficients are the results of the weighted regression and are unbiased and efficient. These results are as follows:

$$P = \underset{(0.04)}{0.09} + \underset{(0.004)}{0.009A} + \underset{(0.16)}{0.19N} - \underset{(0.166)}{0.007M} - \underset{(0.0001)}{0.0002Y} + \underset{(0.004)}{0.002E}$$

$$R^2 = 0.12 \qquad N = 186$$

The values in parentheses are the standard errors of coefficients. N is the sample size. In this equation, the age and income variables are statistically significant at the 5 per cent level, one-tailed test. The older the household head, the more likely he is to admit his participation under welfare programs. According to the income coefficient, the higher the level of household income, the less likely the interviewee is to admit that the household was under welfare programs. The probability of admitting welfare program participation is reduced by 2 per cent with $100 increase in monthly income. Education and employment variables have positive signs in relation to the dependent variable, although they are not statistically significant.

IV. Concluding Remarks

This study found that the disposable income reported in a household interview was about 250 per cent more than the amount recorded at the county welfare agency. This could be because the welfare families tended to underreport their income to the county agency. The level of income of the household was shown to be the most important variable in explaining the difference of income between that reported from the interview and that recorded at the welfare agency. The higher the level of income, the larger the difference.

It was also found that about 27 per cent of the welfare families denied participating in welfare programs. This could be due to the desire to "save face" in front of interviewers or a short time period under the program that the families may have forgotten. The level of household income was found to be the most important variable to predict the probability of denying their welfare experience. The higher the level of income, the larger probability of denying welfare participation.

The findings in this study shows that the magnitude of the difference between the income and welfare information obtained from household interview and recorded at the county welfare agency is much larger than the estimate given in the David study [2]. I suggest that the welfare agency should be more careful in checking recipients' income information so that the welfare rolls could be reduced. On the other hand, I suggest that when researchers analyze welfare information (excluding that dealing with disposable income), they should rely on welfare agency records rather than household survey information.

References

[1] Borus, Michael E., "Response Error in Survey Reports of Earnings Information," Journal of the American Statistical Association, Vol. 61, 1966, pp. 729-38.

[2] David, Martin, "The Validity of Income Reputed by a Sample of Families Who Received Welfare Assistance During 1959," Journal of the American Statistical Association, Vol. 57, 1962, pp. 680-685.

[3] Goldberger, Arthur S., Econometric Theory, New York: Wiley and Sons, Inc., 1964 pp. 248-255.

[4] Hu, Teh-wei; Cohn, Elchanan; Fox, Thomas G.; Kaufman, Jacob J., A Cost-Effectiveness Study of Child Health and Welfare Programs, Institute for Research in Human Resources, The Pennsylvania State University, University Park, Pennsylvania, 1971.

*This Study is supported in part by a grant from the Governor's Office, Commonwealth of Pennsylvania and the Central Fund for Research, The Pennsylvania State University. I wish to thank Irwin Feller for helping with the collection of the data and Norman Kalber for the computation.

A SIMPLE LINEAR REGRESSION MODEL WITH A NON-CONSTANT VARIANCE

Howard S. Kaplon, Towson State College

In the classical theory of regression assumptions of normality are generally made and these in turn lead to the assumption of constancy of variance. In many situations the assumption of constancy of variance is not satisfied. This paper is concerned with the study of such a model in simple linear regression. Computational programs have been obtained to give estimates of parameters in case the standard deviation is assumed to be linear.

Simple linear regression of Y on X is usually defined by the equation $E[Y|X=x] = \alpha+\beta x$ and $Var[Y|X=x] = \sigma^2$ for all x. Here α and β are the regression coefficients and σ^2 is the assumed constant variance. The statistical problem of the investigator here is to estimate values of α and β with $\hat{\alpha}$ and $\hat{\beta}$ computed from a random sample. These estimates are of use in predicting values of the dependent statistical variable Y for observed values of the independent mathematical variable X. In addition, since in most situations the investigator would want to calculate a confidence band for his predicting equation, an estimator $\hat{\sigma}^2$ of σ^2 is also needed. The three estimators are usually derived using the method of maximum likelihood.

Let (X_1,Y_1), (X_2,Y_2), ... , (X_n,Y_n) be a random sample taken from the population of concern (assumed to be normal and $Var[Y|X=x]$ is constant for all x). To estimate the parameters α, β and σ^2 the likelihood function

$$L=\left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}[y_i-(\alpha+\beta x_i)]^2\right\} \quad (1)$$

is set up and we take the three partial derivatives

$$\frac{\partial(\log L)}{\partial\alpha},\ \frac{\partial(\log L)}{\partial\beta} \text{ and } \frac{\partial(\log L)}{\partial\sigma^2}$$

and set them equal to 0. This procedure yields

$$\sum_{i=1}^{n}(y_i-\alpha-\beta x_i) = 0, \quad (2)$$

$$\sum_{i=1}^{n}x_i(y_i-\alpha-\beta x_i) = 0 \quad (3)$$

$$\text{and } n\sigma^2 = \sum_{i=1}^{n}(y_i-\alpha-\beta x_i). \quad (4)$$

Upon solving the normal equations (2) and (3) for $\hat{\alpha}$ and $\hat{\beta}$ we obtain the desired estimators

$$\hat{\alpha} = \frac{(\Sigma y_i)(\Sigma x_i^2)-(\Sigma x_i)(\Sigma x_i y_i)}{n\Sigma x_i^2-(\Sigma x_i)^2} \quad (5)$$

$$\text{and } \hat{\beta} = \frac{n\Sigma x_i y_i-(\Sigma x_i)(\Sigma x_i y_i)}{n\Sigma x_i^2-(\Sigma x_i)^2} \quad (6)$$

then by using these values (5) and (6) in equation (4) $\hat{\sigma}^2$ can easily be found.

As an example of the above model consider the situation where an engineer wishes to determine the effect of heat and cold on the expansion and contraction of a certain metal (with a specified temperature) to be used in the construction of a bridge. He takes a sample of known length of this metal at a known temperature (the specified temperature) and heats it to a new, known temperature, and he measures the new length of the sample. He repeats this experiment as often as he feels is necessary, each time heating or cooling a sample of the same length from the same temperature. However, due to the measuring device, the time elapsed between the time when the metal has reached the desired temperature and the time it is measured, and human error, he does not expect his measurements to be exactly correct. Further he assumes that the errors in the measurements are normally distributed and independent of each other. By allowing the lengths of the pieces of metal to be represented by the random variable Y and the temperature by the variable X, the preceding model seems to fit this situation. The engineer can compute $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\sigma}^2$, set up prediction equation $\hat{Y} = \hat{\alpha} + \hat{\beta}X$, calculate a desired confidence band, and construct the bridge accordingly.

However, consider an example of the of the public health official who needs to predict the hours of health care that will be needed per year by the adults at each age over 21 years of age. He takes a stratified random sample and obtains a pair of values for each individual: X - age and Y - hours of medical care needed per year. The official might very well choose the following as his model: Let $Y_i = \alpha + \beta X_i + e_i$ where e_i are normal distributed independent errors with a mean of 0 and at least two of the x_i are distinct. In this model the assumption that all σ_i^2 are equal is not a logical one. It would seem that as age increases, not only would the average need for medical care increase, but so would the variation of this need. Also, just because of the increase in the values of the random variable with the increase in the age - X, the variances σ_i^2 tend to increase. For these reasons the official could pick his model for regression as

$$E[Y_i|X_i=x_i] = \alpha+\beta x_i$$

and $\mathrm{Var}[Y_i|X_i=x_i] = \sigma_i^2$

where $\sigma_i = \gamma+\delta x_i$.

Now as before, the statistical problem is to estimate the regression coefficients and the variances, i.e. to find $\hat{\alpha}$, $\hat{\beta}$, $\hat{\gamma}$ and $\hat{\delta}$. However, with this model the computation of these estimators is not as simple as in the previous model. It is the purpose of this paper to offer a method of deriving these estimators. The general theory used here would seem to generalize to models with non-linear regression and standard deviation equations, as well as multiple regression models, but the computational procedures are bound to be more complicated.

Given a set of data $\{(X_1,Y_1), (X_2,Y_2), \ldots, (X_n,Y_n)\}$ that fulfill the requirements of our new model, the density function of Y_i is given by

$$f(y_i;\alpha,\beta,\gamma,\delta) = \frac{\exp\left\{-\frac{1}{2}\frac{[y_i-(\alpha+\beta x_i)]^2}{(\gamma+\delta x_i)^2}\right\}}{2\pi(\gamma+\delta x_i)} \qquad (7)$$

The method of estimation will again be that of maximum likelihood. The likelihood function now becomes

$$L(\alpha,\beta,\gamma,\delta) = \frac{\exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\frac{[y_i-(\alpha+\beta x_i)]^2}{(\gamma+\delta x_i)^2}\right\}}{(2\pi)^{\frac{n}{2}}\prod_{i=1}^{n}(\gamma+\delta x_i)} \qquad (8)$$

and
$$\log(L) = -\frac{n}{2}\log 2\pi - \sum_{i=1}^{n}\log(\gamma+\delta x_i) - \frac{1}{2}\sum_{i=1}^{n}\frac{[y_i-(\alpha+\beta x_i)]^2}{(\gamma+\delta x_i)^2}. \qquad (9)$$

Taking the first partial derivative of (9) with respect to each of α, β, γ and δ and setting these equal to 0 we have the following system of four equations in the four unknowns α, β, γ and δ:

$$\frac{\partial\log(L)}{\partial\alpha} = \sum_{i=1}^{n}\frac{[Y_i-(\alpha+\beta x_i)]}{(\gamma+\delta x_i)^2} = 0$$

$$\frac{\partial\log(L)}{\partial\beta} = \sum_{i=1}^{n}\frac{x_i[y_i-(\alpha+\beta x_i)]}{(\gamma+\delta x_i)^2} = 0$$

$$\frac{\partial\log(L)}{\partial\gamma} = -\sum_{i=1}^{n}\frac{1}{\gamma+\delta x_i} + \sum_{i=1}^{n}\frac{[y_i-(\alpha+\beta x_i)]^2}{(\gamma+\delta x_i)^3} = 0 \qquad (10)$$

$$\frac{\partial\log(L)}{\partial\delta} = -\sum_{i=1}^{n}\frac{x_i}{\gamma+\delta x_i} + \sum_{i=1}^{n}\frac{x_i[y_i-(\alpha+\beta x_i)]^2}{(\gamma+\delta x_i)^3} = 0.$$

$\hat{\alpha}$, $\hat{\beta}$, $\hat{\gamma}$ and $\hat{\delta}$ are the solutions to these four equations and are our desired estimators of α, β, γ and δ respectively.

Before continuing, let's consider two special cases: one where $\delta = 0$ and one where $\gamma = 0$. If $\delta = 0$ ($\gamma \neq 0$) then the system of equations (10) reduces to equations (2), (3) and (4), the case with a constant variance, and a closed form solution exists. If $\gamma = 0$ ($\delta \neq 0$) then system (10) becomes

$$\frac{\partial\log(L)}{\partial\alpha} = \frac{1}{\delta^2}\sum_{i=1}^{n}\frac{[y_i-(\alpha+\beta x_i)]^2}{x_i^2} = 0$$

$$\frac{\partial\log(L)}{\partial\beta} = \frac{1}{\sigma^2}\sum_{i=1}^{n}\frac{[y_i-(\alpha+\beta x_i)]}{x_i} = 0 \qquad (11)$$

$$\frac{\partial\log(L)}{\partial\delta} = -\frac{n}{\delta} + \frac{1}{\delta^3}\sum_{i=1}^{n}\frac{[y_i-(\alpha+\beta x_i)]^2}{x_i^2} = 0.$$

Upon solving the first two of these for $\hat{\alpha}$ and $\hat{\beta}$ we obtain the closed form solution

$$\hat{\alpha} = \frac{n\sum\frac{y_i}{x_i^2} - \sum\frac{1}{x_i}\sum\frac{y_i}{x_i}}{n\sum\frac{1}{x_i^2} - \left(\sum\frac{1}{x_i}\right)^2} \qquad (12)$$

and
$$\hat{\beta} = \frac{\sum\frac{1}{x_i^2}\sum\frac{y_i}{x_i} - \sum\frac{1}{x_i}\sum\frac{y_i}{x_i^2}}{n\sum\frac{1}{x^2} - \left(\sum\frac{1}{x_i}\right)^2}. \qquad (13)$$

Then by using these values from (12) and (13) in the third equation of system (11) we get

$$\hat{\delta} = \left[\frac{1}{n}\sum\frac{[y_i-(\hat{\alpha}+\hat{\beta}x_i)]^2}{x_i^2}\right]^{\frac{1}{2}}. \qquad (14)$$

However if both $\gamma \neq 0$ and $\delta \neq 0$, then a simple closed form solution of the system of equations (10) does not exist. An approach to this problem can be taken through the Newton-Raphson iteration method of approximating the solution set for α, β, γ and δ. This method is based on the Taylor series expansion. Consider the system of equations (10) as the following system

$$\begin{aligned} f_1(\alpha, \beta, \gamma, \delta) &= 0\\ f_2(\alpha, \beta, \gamma, \delta) &= 0\\ f_3(\alpha, \beta, \gamma, \delta) &= 0\\ f_4(\alpha, \beta, \gamma, \delta) &= 0 \end{aligned} \qquad (15)$$

Expanding these four functions in a Taylor series we get for i = 1, 2, 3, 4

$$0=f_i(\alpha,\beta,\gamma,\delta)=f_i(\alpha_k,\beta_k,\gamma_k,\delta_k) + (\alpha-\alpha_k)f_{i\alpha}(\alpha_k,\beta_k,\gamma_k,\delta_k) + (\beta-\beta_k)f_{i\beta}(\alpha_k,\beta_k,\gamma_k,\delta_k) + (\gamma-\gamma_k)f_{i\gamma}(\alpha_k,\beta_k,\gamma_k,\delta_k) + (\delta-\delta_k)f_{i\delta}(\alpha_k,\beta_k,\gamma_k,\delta_k) + \cdots \quad (16)$$

where $f_{i\alpha}$, $f_{i\beta}$, $f_{i\gamma}$ and $f_{i\delta}$ are respective first partial derivatives of f_i. By replacing α, β, γ, δ with α_{k+1}, β_{k+1}, γ_{k+1}, δ_{k+1}, neglecting the non-linear terms in $(\alpha_{k+1}-\alpha_k)$, $(\beta_{k+1}-\beta_k)$, $(\gamma_{k+1}-\gamma_k)$, $(\delta_{k+1}-\delta_k)$ of the expansion and dropping the middle term of the double equality (16), these equations for i = 1, 2, 3, 4 become

$$-f_i(\alpha_k,\beta_k,\gamma_k,\delta_k)=\Delta\alpha_k f_{i\alpha}(\alpha_k,\beta_k,\gamma_k,\delta_k)+ \Delta\beta_k f_{i\beta}(\alpha_k,\beta_k,\gamma_k,\delta_k) + \Delta\gamma_k f_{i\gamma}(\alpha_k,\beta_k,\gamma_k,\delta_k) + \Delta\delta_k f_{i\delta}(\alpha_k,\beta_k,\gamma_k,\delta_k) \quad (17)$$

where $\Delta\alpha_k=\alpha_{k+1}-\alpha_k$, $\Delta\beta_k=\beta_{k+1}-\beta_k$, $\Delta\gamma_k=\gamma_{k+1}-\gamma_k$, $\Delta\delta_k=\delta_{k+1}-\delta_k$.

We solve these four linear equations for $\Delta\alpha_k$, $\Delta\beta_k$, $\Delta\gamma_k$ and $\Delta\delta_k$; and knowing the values of α_k, β_k, γ_k and δ_k, we now have values for α_{k+1}, β_{k+1}, γ_{k+1} and δ_{k+1} which are better estimates of actual solutions α, β, γ and δ, than α_k, β_k, γ_k and δ_k are. Then increase k by one and continue this process until the maximum of $\Delta\alpha_k$, $\Delta\beta_k$, $\Delta\gamma_k$ and $\Delta\delta_k$ tends toward zero. We stop the iteration process when max $[\Delta\alpha_k, \Delta\beta_k, \Delta\gamma_k, \Delta\delta_k] < \varepsilon$ where ε is a predetermined small number, and for this value of k we use $\alpha^*=\Delta\alpha_k+\alpha_k$, $\beta^*=\Delta\beta_k+\beta_k$, $\gamma^*=\Delta\gamma_k+\gamma_k$ and $\delta^*=\Delta\delta_k+\delta_k$ to serve as values for $\hat{\alpha}$, $\hat{\beta}$, $\hat{\gamma}$ and $\hat{\delta}$. In order to complete this process, one needs to have initial estimates α_0, β_0, γ_0 and δ_0. Each step in this process, that is, the system (17), is solved easiest by setting up the matrix system

$$\begin{bmatrix} f_{1\alpha} & f_{1\beta} & f_{1\gamma} & f_{1\delta} \\ f_{2\alpha} & f_{2\beta} & f_{2\gamma} & f_{2\delta} \\ f_{3\alpha} & f_{3\beta} & f_{3\gamma} & f_{3\delta} \\ f_{4\alpha} & f_{4\beta} & f_{4\gamma} & f_{4\delta} \end{bmatrix} \cdot \begin{bmatrix} \Delta\alpha_k \\ \Delta\beta_k \\ \Delta\gamma_k \\ \Delta\delta_k \end{bmatrix} = \begin{bmatrix} -f_1 \\ -f_2 \\ -f_3 \\ -f_4 \end{bmatrix} \quad (18)$$

It seems easier to estimate α_0 and β_0 separately from γ_0 and δ_0 instead of estimating all four in one step. When σ does not depend on x, the method of maximum likelihood for obtaining the estimates of α and β does not involve the variance, σ^2. Our new model assumes a linear change in the standard deviation of normal populations, but since the normal distribution is symetric, it seems reasonable to use the same method even in our case where $\sigma = \gamma + \delta x$.

To find the initial estimates, α_0 and β_0, we solve the normal equations (2) and (3) as before. Since we originally assumed that there are at least two values of x_i that are different, we see that the normal equations have a unique solution. Hence the values we use for α_0 and β_0 are the right sides of equations (5) and (6) respectively.

Because of our assumption, $\sigma = \gamma + \delta x$, the original method of estimating σ_0 will not work. However, the fact that σ is a linear function of x is very helpful. When x is small, we naturally expect σ to take on different values than when x is large. In the case when we have several values of y for each value of x, one way of finding initial estimates for γ and δ is to estimate the population of the distribution of y's for each x, and then a least-squares line is fitted to the points of the estimated standard deviations for the various x's. The coefficients of this straight line would then serve as the initial estimates γ_0 and δ_0. This method of obtaining initial estimates would not work if there were only one value of y for each value of x since there would be no way of calculating the standard deviation with only one value. (If only a few x's had multiple y's, this method would produce poor and erratic results.) In that case, and in general, we can group some of smaller x values together and assume that the standard deviation of the y distributions, for each of the grouped x values, is the same. We form two more groups collecting middle values of x and larger values of x together. The standard deviations, s, for the y's and the means, $\bar{x}$, are obtained for each of these three groups. The graph of $(\bar{x},s)$, with these three points on it, gives a least-squares line. This line may be regarded as an estimate of the equation $\sigma = \gamma + \delta x$ thus allowing us to use the coefficients of this calculated line as our values for γ_0 and δ_0. By using standard least-squares techniques our estimates are

$$\delta_0 = \frac{\sum_{i=1}^{n}(\bar{x}_i-\bar{\bar{x}})(s_i-\bar{s})}{\sum_{i=1}^{n}(x_i-\bar{\bar{x}})^2} \quad (19)$$

and $\gamma_0 = \bar{s} - \delta_0\bar{\bar{x}}$ (20)

where $\bar{\bar{x}} = \frac{1}{3}(\bar{x}_1+\bar{x}_2+\bar{x}_3)$ and $\bar{s} = \frac{1}{3}(s_1+s_2+s_3)$, the means and standard deviations of the three groups.

The importance of good estimates α_0, β_0, γ_0 and δ_0 is that if these initial values are close to the values of α^*, β^*, γ^* and δ^*, then the Newton-Raphson iteration produces solutions that converge to $\hat{\alpha}$, $\hat{\beta}$, $\hat{\gamma}$ and $\hat{\delta}$ quickly.

There is one major limitation to the Newton-Raphson iteration method of solving the system of equations (17) or equivalently (18). When the Jacobian determinant

$$J=\begin{vmatrix} f_{1\alpha} & f_{1\beta} & f_{1\gamma} & f_{1\delta} \\ f_{2\alpha} & f_{2\beta} & f_{2\gamma} & f_{2\delta} \\ f_{3\alpha} & f_{3\beta} & f_{3\gamma} & f_{3\delta} \\ f_{4\alpha} & f_{4\beta} & f_{4\gamma} & f_{4\delta} \end{vmatrix}$$

vanishes at or near any of the points $(\alpha_k, \beta_k, \gamma_k, \delta_k)$ in the process, slow convergence, or especially, divergence of the iteration may be expected. This can easily be seen in the case of one variable. In this case equations (16) become

$$0=f(\alpha)=f(\alpha_k)+(\alpha-\alpha_k)f'(\alpha_k)+\frac{(\alpha-\alpha_k)^2\, f''(\alpha_k)}{2}+\cdots \qquad 22)$$

and thus equations (17) reduce to

$$-f(\alpha_k)=(\alpha_{k+1}-\alpha_k)f'(\alpha_k). \qquad (23)$$

Upon solving for α_{k+1} we obtain

$$\alpha_{k+1}=\alpha_k+\frac{f(\alpha_k)}{f'(\alpha_k)} \qquad (24)$$

where now the Jacobian is $J = \left|f'(\alpha_k)\right|$. If at any time this determinant, $\left|f'(\alpha_k)\right|$, vanishes, we see that the resulting solution for α_{k+1} in equation (24) does not make sense. When we have four variables instead of one, the situation is more involved, but the idea is essentially the same. If the value of J vanishes at or (in the case of many variables) near the point $(\alpha_k, \beta_k, \gamma_k, \delta_k)$, then the resulting solution for $(\alpha_{k+1}, \beta_{k+1}, \gamma_{k+1}, \delta_{k+1})$ does not make sense.

If this situation occurs, as an alternative to using the unobtainable α^*, β^*, γ^* and δ^* for $\hat{\alpha}$, $\hat{\beta}$, $\hat{\gamma}$ and $\hat{\delta}$, it appears that α_0, β_0, γ_0 and δ_0 may serve as acceptable substitutes.

In conclusion, consider two examples, For both of these examples the values of the parameters α, β, γ and δ were set, and eight different values of X were chosen. For each value of X, five values of Y were derived using a table of normal, zero-one, random deviates - z - and letting $y=(\gamma+\delta x)(z)+(\alpha+\beta x)$.

Example 1:

TABLE I

x.	y				
1.0	-3.64	11.96	11.00	-1.48	11.60
2.0	15.76	44.88	4.28	-17.28	21.22
3.0	9.68	-3.28	45.36	26.32	-14.00
4.0	24.04	16.10	2.60	-1.72	13.22
5.0	24.20	23.80	33.20	8.40	27.40
6.0	16.84	-0.54	30.48	4.52	13.98
7.0	73.52	41.81	29.36	31.04	67.28
8.0	9.50	51.36	24.06	3.24	14.70

The preset values of the parameters were $\alpha = 5.0$, $\beta = 3.0$, $\gamma = 10.0$ and $\delta = 2.0$. With $\varepsilon = 0.0001$ the values computed by an IBM 7094 computer are $\hat{\alpha} = \alpha^* = 3.16$, $\hat{\beta} = \beta^* = 3.42$, $\hat{\gamma} = \gamma^* = 12.46$ and $\hat{\delta} = \delta^* = 1.14$.

Example 2:

TABLE II

x.	y				
1.0	-1.58	11.93	-6.13	6.33	-5.57
2.0	19.07	18.89	0.02	6.38	7.01
5.0	27.05	39.05	51.35	33.05	15.80
6.0	9.63	15.07	25.61	16.09	16.77
9.0	39.12	-27.35	33.60	36.36	54.30
10.0	62.75	-1.75	30.75	-35.75	35.25
12.0	50.47	108.96	88.46	53.08	39.45
13.0	3.73	56.81	4.66	68.28	57.12

The preset values of the parameters were $\alpha = 2.0$, $\beta = 3.0$, $\gamma = 5.0$ and $\delta = 2.0$. With $\varepsilon = 0.0001$ the computed values are $\hat{\alpha} = \alpha^* = 0.488$, $\hat{\beta} = \beta^* = 3.83$, $\hat{\gamma} = \gamma^* = 4.26$ and $\delta = \delta^* = 2.52$.

When the values of $\hat{\alpha}$, $\hat{\beta}$, $\hat{\gamma}$ and $\hat{\delta}$ are substituted for α, β, γ and δ in the left sides of equations (15), the maximum value of $f_i(\hat{\alpha}, \hat{\beta}, \hat{\gamma}, \hat{\delta})$ for $i = 1, 2, 3, 4$ is 0.00000021 in example 1 and 0.00000014 in example 2.

And finally, in support of the adequacy of α_0, β_0, γ_0 and δ_0 in case the Newton-Raphson iteration method diverges, the values of these initial estimates in example 1 are $\alpha_0 = 3.0$, $\beta_0 = 3.46$, $\gamma_0 = 11.22$ and $\delta_0 = 1.13$.

THE INDEX OF INCOME CONCENTRATION IN THE 1970 CENSUS OF POPULATION AND HOUSING

Joseph J. Knott, Bureau of the Census*

Introduction

Publications showing results of the 1970 Census of Population will contain the Index of Income Concentration (also known as the Gini Index of Inequality) for families, unrelated individuals and for persons. They will be available for areas or cities with population over 50,000, counties, States, and for the United States. The primary purpose of this paper is to outline the procedure used to compute the Index so that the procedure may be duplicated by interested users. Also presented are results of the research undertaken to determine the effect of the various assumptions used in the estimation technique.

Section I outlines the procedure used to compute the Index of Income Concentration (or Index). Section II analyzes some of the effects of various assumptions and constraints used in developing the Index. It is divided into six parts: (A) The overall effect on the Index from using estimated means, (B) use of the midpoint of an income interval as the estimated mean of the income interval, (C) use of the Pareto formula to estimate the mean of the open-end interval, (D) assumption involved in splitting larger $2,000, $3,000, and $5,000 income intervals into $1,000 income intervals, (E) choice of the size of the open-end income interval, and (F) the range of acceptable Indexes. Section III summarizes key findings.

Procedure for Computing the Index of Income Concentration

The Index is defined in terms of the Lorenz curve, and may be represented as the ratio of the area between the diagonal and the Lorenz curve to the area under the diagonal. 1/ The computation of the Index uses an approximate integration technique and requires the percent distribution of units and the percent distribution of aggregate income both by income classes.

The 1970 Census publications show selectively income size distribution of the number of families, unrelated individuals, and persons. A percent distribution is obtained from a numerical distribution by dividing the units in each income class by the total number of units covered in the distribution. It is the computation of the percent distribution of aggregate income which usually presents problems in computing the Index. The Census publications do not show aggregate income by each income class and consequently the aggregate income for each income class must be estimated by multiplying the number of units by the assumed mean for each income class.

In general, in the computation of the Index, the midpoint of an income class is assumed to be the mean of the income interval. This is true for income intervals ranging between $1,000 to $15,000. For "less than $1,000," $500 is assumed to be the mean. For the $15,000 to $19,999 and $20,000 to $24,999 intervals, $17,000 and $22,000, respectively, are assumed to be the means. The Pareto formula is usually used to estimate the mean of the open-end interval.

In order to lessen the error associated with the linearity assumption applied in the approximate integration technique, larger income intervals are divided into smaller income intervals by relating the logarithm of units by the logarithm of income within the class interval. For example, the family income distributions contained in the Census detailed publications show the income interval $12,000 to $14,999. This composite interval is subdivided into three $1,000 intervals. (See table 1.)

Table 1.--INCOME SIZE DISTRIBUTION RELATIONSHIPS FOR SPLITTING THE $12,000-$15,000 INCOME INTERVAL INTO THREE $1,000 INTERVALS

Ratio of frequency of above $12,000 interval to frequency of above $15,000 interval	Percent of $12,000 to $15,000 interval	Percent of $12,000 to $13,000 interval	Percent of $13,000 to $14,000 interval	Percent of $14,000 to $15,000 interval
Under 1.5..................	100	40	33	27
1.5 to 2.5.................	100	44	32	24
2.5 to 3.5.................	100	49	31	20
3.5 and over...............	100	53	28	19

The above table is used as follows:

1. Compute the number of units with income over $15,000 (or F_{15+}). For example, F_{15+} = 349 units.

2. Compute the number of units with income over $12,000 (or F_{12+}). For example, F_{12+} = 425 units.

3. Compute the ratio $\frac{F_{12+}}{F_{15+}}$ or $\frac{425}{349} = 1.218$

4. Find the proper line in the above table for 1.218 (or line 1 above) and apply the percentages to the number of units in the $12,000 to $14,999 interval to get the frequency within the three $1,000 income intervals.

* Comments by Dr. Murray S. Weitzman, Assistant Division Chief for the Economic Statistics Programs, and staff members of the Consumer Income Statistics Branch, Population Division are gratefully acknowledged.

There are two open-end intervals ($15,000 and over; and $25,000 and over) used in the calculation of the Index. In most cases, the mean computed by using the Pareto Formula (the Pareto estimate) of the open-end is used. The Pareto estimate of the $25,000 and over open-end **income interval is computed.**

First derive the slope in the formula:

$$\text{Slope} = \frac{\log_{10}\left[\frac{F_{25+} + F_{15-25}}{F_{25+}}\right]}{\log_{10}\left[\frac{25,000}{15,000}\right]} = \frac{\log_{10} \frac{F_{15+}}{F_{25+}}}{.22185}$$

Where F_{25+} = Number of units with income over $25,000

F_{15+} = Number of units with income over $15,000

F_{15-25} = Number of units with income in the range $15,000 - $24,999

From the above, the Pareto estimate (of the $25,000 and over interval) is derived:

$$\frac{\text{Slope}}{\text{Slope (minus) } 1.0} \times \$25,000 = \text{Pareto estimate}$$

If the frequency in the $15,000 to $24,000 interval is zero, the Pareto estimate cannot be calculated and $36,000 is used as the estimated mean of the open-end interval. Also, if the Pareto estimate is outside the range of $25,000 to $75,000, it is not used and $36,000 is used as the mean of the open-end. 2/ This range constraint is seldom used, and is usually associated with a distribution having a very small base.

The Pareto estimate of the $15,000 and over income interval is computed similarly except that the acceptable range is $15,000 to $40,000. If the Pareto estimate falls outside of this range then the estimate of $23,000 is used.3/

When the percent distribution of units (P_i) and the accumulated percent distribution of aggregate income (A_i) are obtained on the expanded interval distribution (by the above method), the Index is then computed as follows:

$$\text{Index} = 1 - \sum_{i=1}^{n} \left(P_i\right) \times \left(A_{i-1} + A_i\right)$$

P_i = Percent of units in the ith income interval

A_i = Cumulative percent of aggregate income in the ith income interval (when i = 0, A_i = 0)

n = Number of income classes

Assumptions Used in Computing the Index

A. Overall Effect on the Index in Using Assumed Interval Means versus Tabulated Means

The problem is to determine the effect on the Index of using assumed interval means (midpoints) rather than tabulated interval means. The findings show that with relative few income intervals, the use of midpoints as interval means tends to result in estimates about as good as estimates of the Index using tabulated means.

To investigate this problem the Index was computed on a distribution with 190 income intervals using tabulated interval mean values. This is the "Perfect" Index in the sense the "bias" introduced by using the approximate integrated technique is greatly reduced. The smaller (19) interval distributions used to calculate the Index are simply collapses of the 190 interval distribution data. It should be noted that by definition, the number of intervals has an effect on the value of the Index in that a reduction in the number of intervals tends to bias the Index downward. (See table 2).

Table 2.--INDEX OF INCOME CONCENTRATION FOR FAMILIES AND UNRELATED INDIVIDUALS BY AGE BY THREE COMPUTATION METHODS IN 1969

AGE	"PERFECT" Index (190 intervals)	Tabulated Means (19 intervals)	Census Estimation Procedure 1/
Families......................Total	.349	.346	.346
14 - 24	.300	.298	.296
25 - 34	.274	.272	.270
35 - 44	.301	.298	.296
45 - 54	.323	.318	.323
55 - 64	.367	.363	.367
65 and over	.434	.432	.439
Unrelated Individuals.........Total	.480	.475	.469
14 - 24	.454	.447	.426
25 - 34	.370	.368	.343
35 - 44	.404	.401	.406
45 - 54	.428	.425	.429
55 - 64	.438	.434	.432
65 and over	.471	.458	.469

1/ The Census estimation procedure as detailed in the first part of this paper uses 14 tabulated income intervals expanded to 19 with assumed means used to compute the percent aggregate income distribution.

Source: Bureau of the Census, Current Population Survey

As compared with the "Perfect" Index, the Census estimation procedure based on assumed means approximates it fairly well. The slight overestimate of the interval means compensates for the underestimate of the Index caused by the reduction in the number of income intervals.

B. Midpoints as Means of Income Classes

The problem here is to test whether or not midpoints represent good estimates of the actual interval means. For income intervals between $1,000 and $15,000, the midpoint of the interval was used as the mean of the interval. For the under $1,000 interval, $500 was used and for the $15,000 to $19,999 and $20,000 to $24,999 intervals, $17,000 and $22,000, respectively, were used as the means. The use of the midpoint as mean of an income interval is supported by an Internal Revenue Service (IRS) tabulation of adjusted gross income (AGI) by AGI class. The mean AGI of the intervals from $1,000 to $10,000, all fell within $18 of the midpoint. (See table 3) The mean of the "under $1,000" class is not relevent because persons with AGI under $600 are not required to file a tax return.

As data in table 3 show, the CPS tabulated mean within each interval between $2,000 and $15,000 consistently falls below their midpoint in each income interval. This is contrary to what would be expected of a right skewed income frequency distribution. As the units increase in frequency from one interval to another it would seem logical the same increasing frequency would be found within the interval. However, this is not the case. A tabulation by $100 and $250 intervals clearly shows that there is a high frequency in the $100 or $250 interval which contains the even $1,000 amount. Attachment 1 is a bar graph showing the number of families tabulated by small income intervals. The high frequency in the intervals containing the even $1,000 amount is quite evident. This tendency is shown in total family income which is the sum of eight separate income questions per family member and more than one person. This apparent reporting bias is being studied further.

Table 3.--MEAN AGI AND TOTAL MONEY INCOME IN 1969 BY SIZE CLASS

Size Class	Mean Adjusted Gross Income 1/	Mean Total Family Income 2/
Total	$7,959	$10,577
Under $1,000	946 3/	51
$1,000 to $1,999	1,491	1,543
$2,000 to $2,999	2,493	2,475
$3,000 to $3,999	3,488	3,486
$4,000 to $4,999	4,502	4,475
$5,000 to $5,999	5,495	4,457
$6,000 to $6,999	6,497	6,436
$7,000 to $7,999	7,495	7,453
$8,000 to $8,999	8,490	8,443
$9,000 to $9,999	9,495	9,447
$10,000 to $11,999	12,134	10,876
$12,000 to $14,999		13,280
$15,000 to $19,999	17,013	18,284
$20,000 to $24,999	22,093	
$25,000 and over	46,132	35,786

1/ Preliminary Statistics of Income, 1969, "Individual Income Tax Return," Internal Revenue Service, Table 4, page 22.

2/ U.S. Bureau of the Census, Current Population Reports, Series P-60, No. 75, "Income in 1969 of Families and Persons in the United States," Table 1, page 19.

3/ Not comparable since persons with Adjusted Gross Income below $600 are not required to file a tax return.

C. Use of Pareto Formula to Compute the Mean of the Open-End Income Interval

This analysis shows that the use of the Pareto Formula tends to overestimate the mean of the open-end if compared with the tabulated mean of the open-end income interval.

Table 4 shows the Pareto estimate of the mean of the open-end interval and the actual tabulated value from the March 1970 CPS. The Pareto estimate of open-end income interval of $25,000 and over is clearly better for families, than it is for unrelated individuals. The difference between the Pareto estimate and the tabulated **means** indicates that the Pareto estimate should be used carefully. Unfortunately the tabulation of means by income interval is expensive in terms of computer core space and if tabulated means are not available, the use of the Pareto estimate is the most feasible alternative for estimating the mean of the open-end. It should also be noted that the tabulated means from the CPS are slight underestimates of the Census means since CPS income data by type cannot be coded above $99,900, while the Census items can be coded to $990,000.

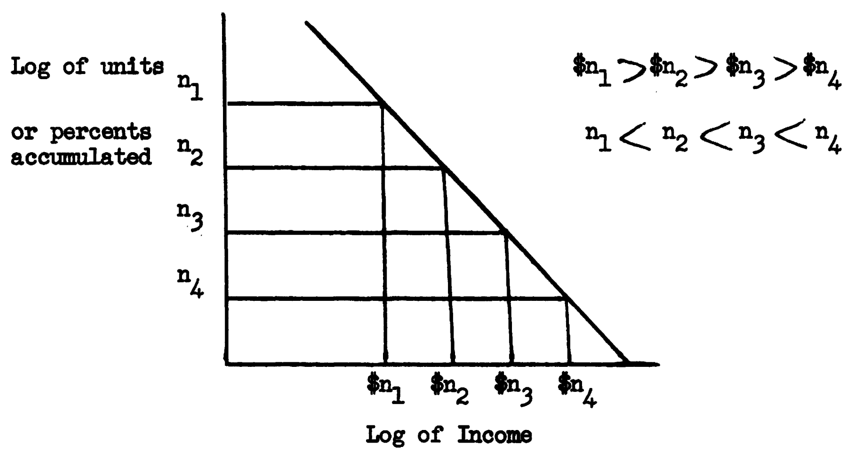
Table 4.--Pareto Estimates and Tabulated Mean Values of the \$25,000 and Over and \$15,000 Open-End Income Intervals for Families and Unrelated Individuals by Selected Characteristics for 1969

Selected Characteristics	\$25,000 and over			\$15,000 and over		
	Pareto	Tabulated	Percent Difference	Pareto	Tabulated	Percent Difference
All families	\$35,975	\$35,786	+0.5	\$25,650	\$21,625	+18.6
All unrelated individuals	39,500	38,480	+2.7	21,750	22,791	- 4.6
Negro and other races						
Families	33,000	31,117	+6.1	23,100	19,681	+17.4
Unrelated individuals	34,950	30,342	+15.2	19,800	16,717	+18.4

Source: Bureau of the Census--Estimates derived from data in the Current Population Survey.

D. Splitting Income Intervals

The assumption of a log-log relationship on which the broad intervals are split is a good assumption to use for the above \$10,000 interval on almost all distributions. This is clearly shown by graphing distributions on log-log paper and observing the linear relationship. From about \$6,000 or \$7,000 to \$10,000 the graph curve shows a shift from log-log to more of a log-normal relationship. The log-normal relationship is also clearly shown on log-normal graph paper. The tables for splitting six different income intervals are given in Attachment 2. These tables are constructed from the following formula.



$$\frac{\log n_1 - \log n_2}{\log \$n_2 - \log \$n_1} = \frac{\log n_1 - \log n_4}{\log \$n_4 - \log \$n_1}$$

$$\log n_2 = \log n_1 - \frac{(\log n_1 - \log n_4)\ (\log \$n_2 - \log \$n_1)}{(\log \$n_4 - \log \$n_1)}$$

Percent or number of units with income over n_2 = Antilog ($\log n_2$)

The tables were constructed by computing the values of the n_2 (all intermediate points desired) for various values of the ratio $\frac{n_1}{n_4}$ of curve (under 1.5, 1.5 to 2.5, 2.5 to 3.5, and 3.5 and over). The percent proportions of n_1 to n_2, n_2 to n_3, n_3 to n_4 to the n_1 to n_4 class were then computed for the midpoint of the 1.5 to 2.5 and 2.5 to 3.5 ranges; and for the under 1.5 and 3.5 and over income interval, 1.5 and 3.5 were used.

E. Choice of the Size of the Open-End Income Interval

For the computation of the Index for family income distributions, the $25,000 and over open-end income interval is used, and for unrelated individuals and persons, $15,000 is used in the 1970 Census. The choice of the open-end is important because it determines the relative importance of the Pareto estimate. Different open-end intervals were used for families and unrelated individuals because they make the Index more comparable in terms of the percent of units in the open-end interval. This gives more equal weight to the Pareto estimate.

Table 5.--ACCUMULATED PERCENT OF UNITS FOR FAMILIES AND UNRELATED INDIVIDUALS FOR SELECTED INCOME CLASSES

Total money income	Percent of units over the specified income level	
	Families	Unrelated Individuals
Over $12,000	32.9	4.9
Over $15,000	19.2	2.4
Over $25,000	3.6	0.6

Source: Bureau of the Census, Current Population Reports, Series P-60, No. 75, Table 16.

As the table shows, 3.6 percent of all families had incomes above $25,000, but only 0.6 percent of unrelated individuals was in the same interval. This difference would result in the Pareto mean having six times the weight for family distributions relative to unrelated individual distributions. This disparity is reduced by using the $15,000 and over interval as the open-end interval for unrelated individuals, and the $25,000 and over interval for families (i.e., 2.4 percent for unrelated individuals relative to 3.6 percent for families).

F. Range of Published Indexes

For publication purposes, only Indexes within the range of .200 to .650 will be published. An Index outside this range will be suppressed and three dots will be shown (...). Indexes outside this range, for the most part, represent Indexes computed on very small bases. In any case, users can compute Indexes, if desired, for these distributions by using the technique outlined in this paper.

Summary

In summary, the estimation technique used to compute the Index of Income Concentration from the Census publications appears to give good results in most cases. It is interesting to note that (when compared to an Index computed on the basis of 190 intervals), the estimation procedure results in estimates about as good as estimates of the Index produced by using tabulated number and aggregate income for 19 size income intervals.

The tendency for respondents to report estimated income to the nearest $1,000 is an interesting phenomenon which is being analyzed further.

Findings showed that the various assumptions used to compute the Index do not invalidate the relative accuracy of the Index. The assumption of the midpoint as the mean of the income interval is supported by AGI data, but CPS income data suggest that midpoints are too high. The use of the Pareto formula also tends to overestimate the mean of the open-end interval, but not uniformly. Furthermore, data show that the number of intervals used to compute the Index makes a difference. Any comparison of Indexes requires that they be computed using the same number of income intervals.

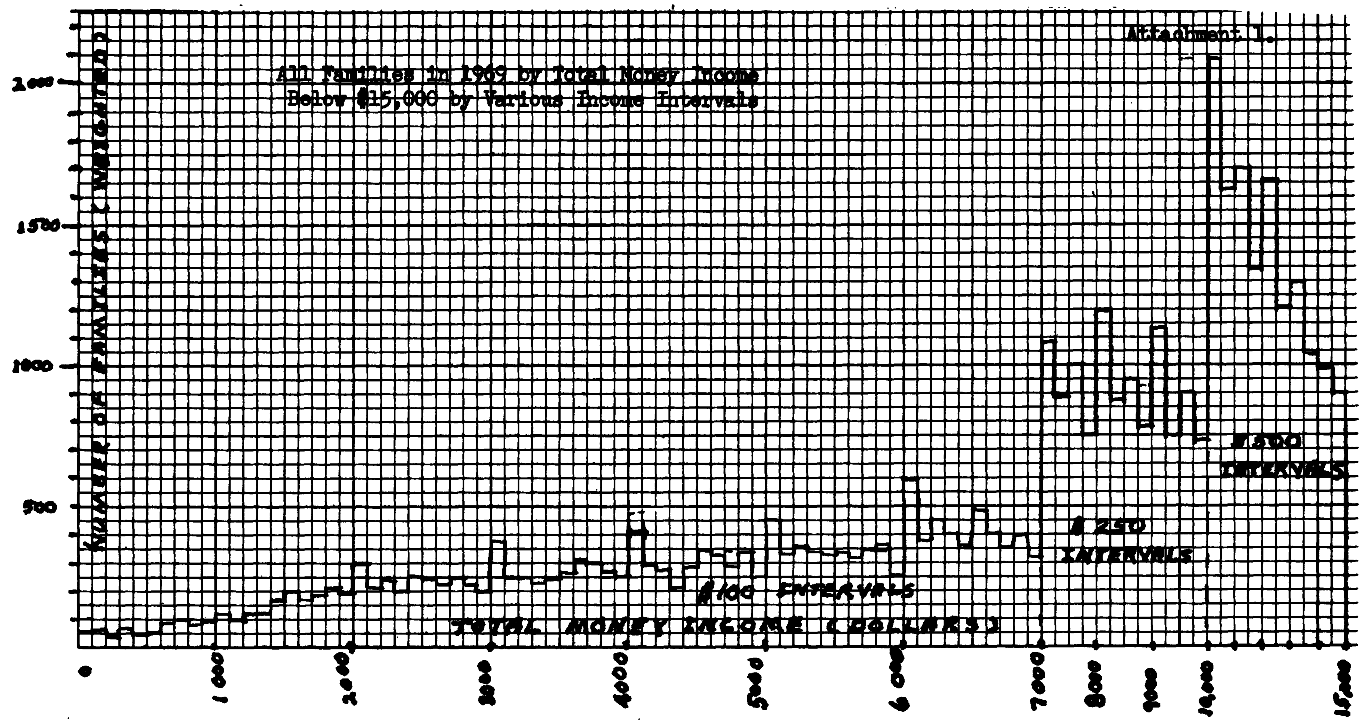
FOOTNOTES

1/ An expanded discussion of the geometric interpretation of the Index of Income Concentration may be found in: Rich Man, Poor Man, by Herman P. Miller, Thomas Y. Cromwell Co., New York, 1971 appendix B, pp. 274 - 279.

2/ Implicit in this constraint is a ratio of $F_{25+}/F_{15+} = 2.15$. The value of $36,000 is obtained from CPS income data.

3/ Implicit in this constraint is a ratio of $F_{15+}/F_{12+} = 1.60$. The value of $23,000 is obtained from CPS income data.

Attachment 1.



Attachment 2.

Proportions to be Used to Split Broad Income Intervals into Smaller Income Intervals

A. $6,000 - $7,999

F6+/F8+	F6-8	F6-7	F7-8
Under 1.5	100	56	44
1.5 - 2.5	100	62	38
2.5 - 3.5	100	66	34
3.5 and over	100	68	32

B. $8,000 - $9,999

F8+/F10+	F8-10	F8-9	F9-10
Under 1.5	100	56	44
1.5 - 2.5	100	61	39
2.5 - 3.5	100	66	34
3.5 and over	100	68	32

C. $10,000 - $14,999

F10+/F15+	F10-15	F10-11	F11-12	F12-13	F13-14	F14-15
Under 1.5	100	27	23	19	17	14
1.5 - 2.5	100	30	24	19	15	12
2.5 - 3.5	100	34	25	18	13	10
3.5 and over	100	36	25	17	13	9

D. $10,000 - $11,999

F10+/F12+	F10-12	F10-11	F11-12
Under 1.5	100	55	45
1.5 - 2.5	100	61	39
2.5 - 3.5	100	66	34
3.5 and over	100	67	33

E. $12,000 - $14,999

F12+/F15+	F12-15	F12-13	F13-14	F14-15
Under 1.5	100	40	33	27
1.5 - 2.5	100	44	32	24
2.5 - 3.5	100	49	31	20
3.5 and over	100	53	28	19

F. $15,000 - $24,999

F15+/F25+	F15-25	F15-20	F20-25
Under 1.5	100	60	40
1.5 - 2.5	100	65	35
2.5 - 3.5	100	69	31
3.5 and over	100	71	29

F_{m+} - Number of units with income over $m,000

F_{m-n} - Number of units with income from $m,000 to $n,000

A STOCHASTIC PROCESS MODEL OF CONJUGAL HISTORY*

P. Krishnan
Cornell University and University of Alberta

1. INTRODUCTION

1.1 Interest in the population problems of developing countries, has led scientists to develop and construct mathematical models as an aid in understanding the dynamics of population growth and change. Research in model-building in the field of human fertility was initiated by Alfred Lotka. While Lotka was concerned with deterministic models, most of the recent work in fertility models deals with stochastic or probability models. A good review of the attempts in stochastic model-building can be found in Sheps et al.[1] The role of social-structural factors in influencing fertility has been noted by several researchers (Davis[2], Lauriat[3], Pool[4], Stycos[5]). But in the stochastic process models developed to date, most of the vital social-structural factors are not considered at all, or considered insofar as they modify biological functions. For social policy decisions, especially in view of the forces of modernization impinging on developing nations, one has to know the role and contribution of social factors to fertility. To develop a sociologically sound model of fertility, one has to study conjugal history in detail. Even a model of conjugal history per se, is of interest to family sociologists for a deeper understanding of the different forces at work on marriage and its disruption.

1.2 A female in any society, ignoring childhood and adolescent days, will be in any one of the following states.

A_1' - Not in union by not entering into union

A_2' - In union

A_3' - Not in union by divorce

A_4' - Not in union by separation

A_5' - Not in union by desertion

A_6' - Not in union by death of husband

1.3 The classification of conjugal states as given above is most general. But in many cases, data may not be available with regard to some of these states. In such cases we shall have to switch back to the usual way of classifying females as unmarried (state A_1), married (state A_2), widowed (state A_3), and divorced (state A_4). We shall confine the discussion to this simple and common classification of conjugal states. The results derived can be extended to the most general case easily.

1.4 To get a more realistic model of conjugal history, 'death' state has to be brought explicitly into the picture. This is not attempted here due to the following reasons: (a) Most of the data that we have are cross-sectional, and retrospective marital and childbearing histories are collected only for those who are living and are sampled or interviewed at the time of the survey (b) Mortality is not, comparatively speaking, a significant force to be reckoned with especially in the age span of 15-49 years, which is of interest as far as human fertility is concerned. A model of conjugal history incorporating mortality explicitly, is suggested by the author elsewhere.[6] The model suggested in this paper, is thus conditioned on the survival of the female. Unconditional results can be obtained by integrating these results over the relevant probability space assuming a realistic probability distribution for survival.

1.5 A diagrammatic representation of the conjugal history of a female, conditional on survival, is given below (Figure 1).

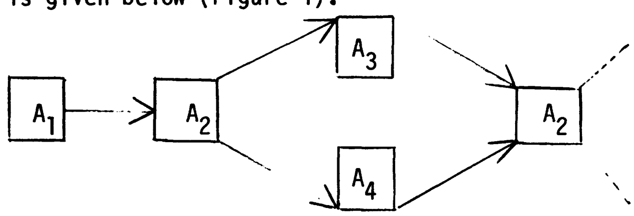


Figure 1 Chart of Conjugal History

Legend: A_1 Unmarried State A_3 Widowed State

A_2 Married State A_4 Divorced State

An unmarried female after a random length of time is pushed to the married state (A_2). Conditional on the survival of the female, there are two forces acting on the female who has moved to A_2. The female runs the risk of moving to the widowed state A_3, or to the divorced state A_4, or she can remain in A_2. If she makes a transition to A_3 or A_4, she may remarry and thus move back to A_2. Remarriage depends on many factors. But it is clear that the conjugal history of a female can be completely characterized by the knowledge of the sequence in which these states are visited and the length of time spent in each state at each visit.

---

*Based on the doctoral dissertation submitted to the Graduate School at Cornell University. The author is highly indebted to Professors Robert McGinnis, J. Kiefer, N.U. Prabhu, D.I. Pool and H.M. Taylor, members of the special committee, for their valuable suggestions and advice at every stage of his graduate training and research. The research reported here was supported by a fellowship under the H.S.S. Program.

2. SOCIOLOGICAL CONSIDERATIONS

2.1 Various sociological considerations make one view the length of stay in each state as a random variable. The frequency of dating, the availability of males in the appropriate age groups as potential mates, the stigma attached to remaining "single", the social recognition of married couples in preference to "singles", etc. are all social factors which affect the annual chance of marriage of an unmarried female. Thus the time spent in unmarried state can be considered as a random variable. The length of stay in married state also is a chance variable in view of the sex-differential in the expectation of life, the husband-wife age differences, educational status of the spouses, participation in labor force, the tabu attached to extra-marital relations, etc., in addition to the psychological compatibility of the spouses.

2.2 After becoming divorced or widowed, a female can hope to remarry, if the society under consideration has no tabu on such remarriages. The female thus may return to married state. There are several factors, such as age, beauty, number of children the female has begotten, the availability of females in relation to males, which affect the chance of remarriage. Over and above this, if the female stays in unmarried state, or divorced, or widowed state for a long time, her chances of marriage/remarriage diminish. That is to say the principle of cumulative inertia[7] applies here, due to obvious reasons. Thus the length of stay in divorced or widowed also is a random variable. After remarriage, the female again becomes exposed to the risks of widowhood or divorce and the process repeats itself. Hence the whole conjugal or marital history can be viewed as a stochastic process.

2.3 The probabilities of transition from state A_i to state A_j (i, j = 1, 2, 3, 4) are sociologically interpretable. A female who has moved to A_2 cannot go back to A_1. Hence in the development of the model here we shall be concerned with the conjugal history of ever married females.

3. A MARKOV RENEWAL/SEMI-MARKOV PROCESS

3.1 From the various considerations stated above, the conjugal history of an ever married female can be characterized as a Markov Renewal/Semi-Markov Process[8], if we further assume that the parameters of the various probability distributions and the transition matrix do not change over time. This assumption may not be strictly true in view of the fact that after remarriage, the transition matrix and the various distributions are likely to be different from those of the first marriage. In such a case we can characterize the conjugal history as a General Markov Renewal Process[9]. As a first approximation of the process, M.R.P. comes as handy. Since the number of states is just three, the results can be derived from fundamental probability considerations without recourse to renewal theory. Derivations are not given here for lack of space but are available elsewhere.[10]

3.2 We use the standard notation.

States B_1 - Married state

B_2 - Widowed state

B_3 - Divorced state

Transition Matrix $P - (p_{ij})$ (i, j = 1, 2, 3)

p_{ij} - Annual probability of transition from state B_i to state B_j. p_{32} and p_{23} will be taken as zero.

$F_{ij}(t)$ - Distribution function of length of stay in B_i before transition to B_j.

$\overset{*}{\mu}_{ij}$ - mean of $F_{ij}(t)$

$G_{ij}(t)$ - Distribution function of the first passage time from B_i to B_j

m_{ij} - mean of $G_{ij}(t)$

$\mu_i - \sum_j p_{ij} \overset{*}{\mu}_{ij}$

3.3 Results that are of sociological importance are given below.

3.3.1 Conditional on survival, the mean (first passage) time of a divorced female to remarriage is

$$m_{31} = \frac{p_{31} \overset{*}{\mu}_{31} + p_{33} \overset{*}{\mu}_{33}}{p_{31}}$$

3.3.2 Conditional on survival, the mean (first passage) time of a widowed female to remarriage is

$$m_{21} = \frac{p_{21} \overset{*}{\mu}_{21} + p_{22} \overset{*}{\mu}_{22}}{p_{21}}$$

3.3.3 Conditional on survival, the mean (first passage) time of a married female to widowhood is

$$m_{12} = \frac{p_{31} \mu_1 + p_{13} \mu_3}{p_{31} p_{12}}$$

3.3.4 Conditional on survival, the mean (first passage) time of a married female to divorce is

$$m_{13} = \frac{p_{21} \mu_1 + p_{12} \mu_2}{p_{13} p_{21}}$$

3.3.5 Conditional on survival, the mean (first passage) time of a widowed female to divorce is

$$m_{23} = \frac{(1-p_{11})\,\mu_2 + p_{21}\,\mu_1}{p_{21}\,p_{13}}$$

3.3.6 Conditional on survival, the mean (first passage) time of a divorcee to widowhood is

$$m_{32} = \frac{(1-p_{11})\,\mu_3 + p_{31}\,\mu_1}{p_{12}\,p_{31}}$$

3.3.7 Conditional on the survival of the female, the mean first recurrence time of married state is given by

$$m_{11} = p_{12}\,m_{21} + p_{13}\,m_{31} + \mu_1$$

3.3.8 Conditional on the survival of the female, the mean first recurrence time of widowed state is

$$m_{22} = \mu_2 + p_{21}\,m_{12}$$

3.3.9 Conditional on the survival of the female, the mean first recurrence time of divorce state is

$$m_{33} = \mu_3 + p_{31}\,m_{13}$$

3.3.10 It may be noted that if any of the waits is defective, the mean first passage or recurrence time(s) involving the defective distributions will become infinite.

4. EXPECTED NUMBER OF MARRIAGES, DIVORCES, AND WIDOWHOODS IN (o,t] SUBJECT TO SURVIVAL

4.1 Let $N_i(t)$ be the random number of times a female enters B_i $(i = 1, 2, 3)$. Then $N_1(t)$ is the number of marriages contacted, $N_2(t)$ the number of widowhoods experienced, and $N_3(t)$ the number of divorces undergone. We will compute the mean values of $N_i(t)$ $(i = 1, 2, 3)$ given that the female starts from B_1, married state (i.e. after first marriage). Following Smith[11], the sequence of recurrence times of B_i forms a renewal process. If the first passage from j to i is added to this sequence, we have a general renewal process. Using Murthy's[12] approximation, we get

a) expected number of remarriages in (o,t]

$$= \frac{t}{m_{11}} + \frac{m_{11}^{(2)}}{2m_{11}^2} - 1$$

b) expected number of divorces in (o,t]

$$= t/m_{33} + \frac{m_{33}^{(2)}}{2m_{33}^2} - \frac{m_{13}}{m_{33}}$$

c) expected number of widowhoods in (o,t]

$$= t/m_{22} + \frac{m_{22}^{(2)}}{2m_{22}^2} - \frac{m_{12}}{m_{22}}$$

where $m_{ii}^{(2)}$ is the second moment of the first recurrence time of B_i $(i = 1, 2, 3)$.

4.2 A few remarks are in order now. A family sociologist will be interested to know how much percent of married life time is lost due to disrupting forces such as divorce and widowhood. To gauge the intensity of such disrupting forces, an index of marital disruption can be developed. The transition matrix and the waits are likely to be age dependent. This problem can be solved to some extent by considering age cohorts, assuming that the female population of each cohort is homogeneous with respect to the characteristics under consideration.

5. ILLUSTRATION

5.1 The survey of economic opportunity (S.E.O.) conducted by the U.S. Bureau of the Census in 1967 for the Office of Economic Opportunity (O.E.O.) contains information on the marital history of ever married females of the U.S. canvassed in the sample. The geographical coverage of the study, the size of the sample, the sampling procedure, and the method of estimation are all discussed in a Census Bureau paper.[13]

5.2 The following results have been obtained ignoring age (i.e. all ever married females are lumped together) and race. Race has been ignored because the S.E.O. had an overrepresentation of the Negro population. Differential with respect to color is well established. Our aim is not to study differentials. Also we have omitted the group of females who had experienced 26-55 years of married life in the estimation of the elements of the transition matrix, and the mean waits in "married" state. It was seen that the mean waits were not underestimated even when truncation was taken into account.

5.3 Considering first marriage alone, we have the following estimates.

$$\hat{P} = \begin{bmatrix} M & W & D \\ .9865 & .0048 & .0087 \\ .0265 & .9735 & 0 \\ .1447 & 0 & .8553 \end{bmatrix}$$

$$\hat{\mu}^*_{11} = 12.46 \qquad \hat{\mu}^*_{21} = 6.92 \qquad \hat{\mu}^*_{31} = 6.38$$

$$\hat{\mu}^*_{12} = 19.79 \qquad \hat{\mu}^*_{22} = 9.82 \qquad \hat{\mu}^*_{33} = 9.80$$

$$\hat{\mu}^*_{13} = 9.56$$

5.4 We have the estimates of the mean first passage and the first recurrence times as follows:

$$\hat{m}_{11} = 14.8 \quad \hat{m}_{21} = 367.7 \quad \hat{m}_{31} = 64.3$$
$$\hat{m}_{12} = 2732.3 \quad \hat{m}_{22} = 82.2 \quad \hat{m}_{32} = 2797.0$$
$$\hat{m}_{13} = 1640.1 \quad \hat{m}_{23} = 2008.6 \quad \hat{m}_{33} = 246.6$$

5.5 A few comments are in order now. The absolute values of the mean first passage times etc. have no significance. Large values of the mean first passage times etc. have ensued from the defective nature of the distribution functions of waits. Relative values are meaningful (cf. Kemeny and Snell[14]). Thus, if a divorced female is considered as waiting for one unit of time to remarry, the widowed female, on an average, will take 5.7 units of time for remarriage. If the time from "married" state to "divorced" state is taken as one unit of time, the time to "widowed" state is two units. Another interesting result that emerges from this model is that the mean time from "divorced" state to "widowed" state and the mean time from "married" state to "widowed" state are nearly of the same order of magnitude. Also, if the mean recurrence time of "married" state is taken as one unit of time, the mean recurrence time of "widowed" state is nearly 5.6 units and of "divorced" state 16.8 units. This last result seems to be highly unrealistic. The probable reason for this may be ignoring the age factor. Older females are more likely to become widowed earlier than becoming divorced.

FOOTNOTES

(1) Mindel C. Sheps, Jane A. Menken, and Annette P. Radick, "Probability Models for Family Building: An Analytical Review", Demography, 6 (1969), 161-83.

(2) Kingsley Davis, "Human Fertility in India", American Journal of Sociology, 52 (1946), 243-54.

(3) Patience Lauriat, "The Effect of Marital Dissolution on Fertility", The Journal of Marriage and the Family, 31 (1969), 484-93.

(4) D.I. Pool, "Conjugal Patterns in Ghana", Canadian Review of Sociology and Anthropology, 5 (1969), 241-53.

(5) J.M. Stycos, Human Fertility in Latin America: Sociological Perspectives (Ithaca, N.Y: Cornell University Press, 1968), Chapter 13.

(6) P. Krishnan, Some Stochastic Models in the Study of Human Conjugal History and Fertility. Doctoral Dissertation Submitted to the Graduate School, Cornell University, Ithaca, New York, 1971.

(7) Robert McGinnis, "A Stochastic Model of Social Mobility", American Sociological Review, 35 (1968), 712-22.

(8) R. Pyke, "Markov Renewal Processes: Definition and Preliminary Properties", Annals of Mathematical Statistics, 32 (1961), 1231-42.

(9) A.M. Kshirsagar and Y.P. Gupta, "Some Results in Markov Renewal Processes", Calcutta Statistical Association Bulletin, 18 (1969), 61-72.

(10) P. Krishnan, ibid.

(11) W.L. Smith, "Renewal Theory and its Ramifications", Journal of the Royal Statistical Society, Series B, 20 (1958), 243-302.

(12) V.K. Murthy, "On the General Renewal Process", Mimeograph Series No. 293, Institute of Statistics, University of North Carolina, Chapel Hill, N.C., 1961.

(13) U.S. Bureau of the Census, "Probabilities of Marriage, Divorce, and Remarriage", Current Population Reports, Special Studies, Series P-23, No. 32, p. 2 (July 29, 1970).

(14) John G. Kemeny, and J. Laurie Snell, Finite Markov Chains (Princeton, N.J.: D. Van Nostrand Company Inc.,1967), p. 193.

THE USE OF MORTALITY DATA IN EVALUATING SYNTHETIC ESTIMATES

Paul S. Levy, National Center for Health Statistics

1. Background and Introduction

For the past 15 years, the National Health Survey through its Health Interview, Health Examination and other surveys has been gathering data on a considerable range of health parameters. The statistics obtained from these surveys are expressed generally as estimates for the United States as a whole or for each of the four broad geographical regions. Estimates for smaller areas such as States cannot be readily obtained because (1) the sample sizes are not adequate for areas this small and (2) the sampling design uses strata which cut across State lines, and this makes it difficult to combine the estimates for strata into estimates for States.

Because of the increasing need for small area estimates of health parameters the National Center for Health Statistics (NCHS) has been exploring methodology for using National Health Survey data to produce estimates for these small areas. This research has resulted in a method called *synthetic estimation* whereby State estimates are produced by using census data on the distribution of a State's population into demographic domains (which we will subsequently refer to as "a-cells") along with the National Health Survey estimates for these domains for the entire United States. An NCHS publication [1] describes this synthetic estimation methodology and uses it to produce synthetic estimates of disability for each State from Health Interview Survey (HIS) estimates for the United States.

Since there is an ever increasing need for small area estimates of health parameters for purposes of local planning and since good local population data essential for synthetic estimation, are now available from the 1970 Decennial Census, an NCHS program to produce small area estimates based on synthetic methodology is seriously being considered. It is first necessary, however, to evaluate the accuracy of the estimates produced by this methodology, and this report presents the results of a study designed to gain insight into the accuracy of synthetic estimates as well as a method which would use ancillary data to obtain an improved synthetic estimate.

2. The NCHS Synthetic Estimate

While the NCHS publication [1] presents the methodology in greater detail, we will outline the methodology below. The synthetic estimate is constructed in two stages with the first stage having the form

$$\approx{X}_s = \sum_{a=1}^{K} p_{sa} \bar{X}_a \qquad (1)$$

Where

$\approx{X}_s$ = the first stage synthetic estimate of X-characteristic for the s-th State

P_{sa} = proportion of the population in the s-th State belonging to the a-th demographic cell

$\bar{X}_a$ = the unbiased probability estimate of the X-characteristic for the a-th demographic cell for the U.S. as a whole

and

K = the number of demographic cells.

The final estimate has the form

$$\tilde{X}_s = \approx{X}_s \left(\frac{\bar{X}'_r}{\sum_{s=1}^{R} \approx{X}_s P_{rs}} \right) \qquad (2)$$

where $\tilde{X}_s$ = the final synthetic estimate of X-characteristic for State s,

$\bar{X}'_s$ = the unbiased probability estimate of X characteristic for the r-th geographical region

P_{rs} = the proportion of the population of region r belongs to the s-th State

and

R = the number of States in region r.

The basic feature of the first stage synthetic estimator is that for each demographic cell (a-cell) it uses a probability estimate specific to that demographic cell over the entire United States in conjunction with the proportion of the State's population falling into that cell (from census data) to obtain the contribution of that cell to the synthetic estimator. The final synthetic estimate is simply the initial estimate times a post-stratification factor which makes the sum of the synthetic estimates for all States in a geographical region agree with the probability estimate for that region.

Synthetic estimates applied to National Health Survey data have been difficult to evaluate because valid unbiased estimates produced by ordinary probability survey methods are not available for States. Indirect methods such as observing their consistency from one year to another [1] or comparing the synthetic estimates for the largest SMSA's with the HIS unbiased estimators [2] have been used but these have not been altogether convincing.

Since mortality statistics from U.S. Vital Statistics Annual Volumes [3] are available by cause of death for all States by age, sex, and race, a study was planned to compute synthetic estimates for each State for several causes of death using the U.S. mortality data for age, race, and sex a-cells and the corresponding census data on the distribution of each State into these a-cells. If the synthetic estimates should agree well with the true deaths, it would be evidence that the synthetic procedure might produce valid estimates. If not, then some insight might still be gained into the pattern of discrepancy between the synthetic estimate and the true value. This evaluation study is described in the next section.

3. Evaluation Study

3.1 *Methods and Materials*

Mortality data. Deaths from motor vehicle accidents (E810-E835), major cardiovascular renal diseases (330-334, 400-468, 592-594), suicide (E963, E970-E979) and tuberculosis (001-019) for each of forty-nine States (with the District of Columbia and Maryland combined into one "State") were transcribed onto IBM cards from the 1960 U.S. Vital Statistics Annual Volumes [3]. New Jersey was not included because deaths by race were not available for that State. For each State, the deaths were transcribed for the following 40 age-sex-race groups:

Age (under 1 year, 1-4 years, 5-9 years
10-14 years, 15-19 years, 20-24 years, 25-44 years
45-64 years, 65-74 years, 75+)

Race (white, nonwhite)
Sex (male, female).

Population data. The populations in each of the above mentioned age-sex-race groups were transcribed onto IBM cards from the 1960 U.S. Census Volumes for each State in the study.

Data Processing. A program was written in Fortran IV and the analysis was carried out on a remote entry terminal to the UNIVAC 1108 at the National Bureau of Standards in Gaithersburg, Maryland. The basic output was the synthetic estimate, $\tilde{X}_s$ of the number of deaths by specific cause for each State in 1960.

3.2 *Results*

The agreement between the synthetic estimate and the true number of deaths is expressed by the *percentage absolute difference* defined as

$$\left| \frac{X_s - \tilde{X}_s}{\tilde{X}_s} \right| \times 100 \qquad (3)$$

where

X_s = the true number of deaths (by specific cause) for the s-th State

and

$\tilde{X}_s$ = the second stage synthetic estimate of these deaths (obtained from the 40 age-sex-race a-cells).

The results are summarized in Table 1 which gives the frequency distribution of these percentage absolute differences along with the median and mean percentage absolute difference, and the relative variances (V^2 = variance/mean^2) of the percentage absolute differences for each of the four causes of death considered.

The accuracy of the synthetic estimates as measured by the percentage absolute difference varied considerably among the four causes of death examined. The median percentage absolute difference was 5.9% for major cardiovascular-renal deaths, 9.8% for suicides, 15.9% for deaths from motor vehicle accidents, and 24.3% for tuberculosis deaths. Likewise, the variability among States as expressed by the relative variance of the percentage absolute difference was highest for tuberculosis with $V^2 = 1.11$ and lowest for major cardiovascular diseases ($V^2 = 0.57$). The relative variance for deaths from motor vehicle accidents was equal to 1.00 and for suicide was 0.65.

3.3 *Discussion*

From the results presented above and shown in Table 1, it is clear that the accuracy of the synthetic estimates as summarized by the median percentage absolute difference varied considerably among the four causes of death considered. While the agreement between the synthetic estimate and the true value was generally good for major cardiovascular-renal diseases and fairly good for suicide, it was generally poor for motor vehicle accidents and very poor for tuberculosis deaths.

In order to get some insight into the effectiveness of the synthetic estimator against possible competing estimators, we compared it to a regionally adjusted estimator, X'_{as} obtained for each State, S, by multiplying the population of the State by the crude death rate in the geographical region wherein the State lies. The percentage absolute difference between X'_{as} and the true number of deaths X_s was calculated for each State and the median percentage absolute difference over all States was obtained and is shown below next to the comparable figure for the synthetic estimator, $\tilde{X}_s$.

Median Percentage Absolute Difference

Estimator	Motor Vehicle Accidents	Major C.V.R. Diseases	Suicides	Respiratory T.B.
$\tilde{X}_s$	15.9	5.9	9.8	24.3
X'_{as}	14.6	6.3	10.2	32.0

The results above seem to imply that except for respiratory T.B., the synthetic estimate did little or no better than the estimator X'_{as} which is obtained from very crude regional information.

Table 1. Cumulative Percentage Distribution of Percentage Absolute Differences Between the Synthetic State Estimate and the True Number of Deaths ($100 \times | X_s - \tilde{X}_s | / \tilde{X}_s$) for Each of the Four Causes of Deaths Investigated

Percentage Absolute Difference $100 \times \left\| \frac{X_s - \tilde{X}_s}{\tilde{X}_s} \right\| \%$	Motor Vehicle Accidents f	Motor Vehicle Accidents cum. %	Major C.V.R. Diseases f	Major C.V.R. Diseases cum. %	Suicides f	Suicides cum. %	Tuberculosis f	Tuberculosis cum. %
0- 8.0%	16	32.7	26	53.1	20	40.8%	13	26.5
8.1-16.0%	9	51.0	20	93.9	13	67.3%	6	38.7
16.1-24.0%	9	69.4	3	100.0	6	79.5%	4	46.9
24.1-32.0%	6	81.6			6	91.8%	4	57.1
32.1-40.0%	3	87.8			3	98.0%	9	73.4
40.1%+	6	100.0			1	100.0%	13	100.0
Total	49		49		49		49	
Median % Absolute Difference		15.9		5.9		9.8		24.3
Mean % Absolute Difference		20.2		6.9		13.7		31.6
Relative Variance (V^2)		1.00		0.57		0.65		1.11

4. Improvement by Regression on Ancillary Data

4.1 *Background and motivation*

One of the basic limitations on the synthetic estimator, $\tilde{X}_s$ is that it is adjusted only for the specific set of demographic cells (or a-cells) taken into consideration. If the parameter being estimated is influenced by factors other than those taken in account by the a-cells, then the synthetic estimate will not reflect this influence. Often it is not possible to include in the a-cell array all the variables thought to be of importance in estimating the variable because of the unavailability of data on these variables. For example, even though the risk of a person's dying from a motor vehicle accident may be a function of the amount of time he spends in motor vehicles, there is no way of creating a-cells to reflect this.

A second type of limitation in the synthetic estimator, $\tilde{X}_s$, is that it may not reflect local conditions which are highly related to the parameter being estimated. For example, the probability of a person's dying from a motor vehicle accident is known to be generally higher in States which have lower population densities [4]. Since these types of variables are often available for local areas such as States, it might be possible that a modified synthetic estimate can be constructed which takes into account these variables.

In the following sections, we propose a method of using local variables which might be related to the parameter being estimated in conjunction with the synthetic estimator $\tilde{X}_s$ to produce an improved estimator of the parameter. It is felt that this method will be especially applicable to small area estimates using data from the complex, highly stratified multi-stage nationwide probability surveys such as the Health Interview Survey.

4.2 *Method of estimation*

The method presented below uses the a-cell adjusted synthetic estimate $\tilde{X}_s$ in conjunction with an ancillary variable Z_s to produce an improved estimator. In particular, the following model is assumed:

$$Y_s = \alpha + \beta Z_s + \epsilon_s \tag{4}$$

where

Z_s = the value of the Z variable for the s-th State

$$Y_s = \frac{X_s - \tilde{X}_s}{\tilde{X}_s} \times 100,$$

$\tilde{X}_s$ = the synthetic estimate of the X-characteristic for the State s,

X_s = the true value of the X-characteristic for State s,

ϵ_s = a term representing random error,

and

α, β = regression parameters to be estimated.

If estimates $\hat{\alpha}$ of α and $\hat{\beta}$ of β were available and substituted into the right hand side of equation (4) with ϵ_s omitted, manipulation of the expression would give us the following estimator $\hat{X}_s$ of X_s:

$$X_s = \tilde{X}_s \ (1 + \frac{\hat{\alpha} + \hat{\beta} Z_s}{100}) \tag{5}$$

Equation (4) merely states that the percentage difference, Y_s, between the synthetic estimate, $\tilde{X}_s$, and the true value X_s is a linear function of some variable Z_s. For example, Z_s might be the population density of State s, $\tilde{X}_s$ the synthetic estimate of deaths from motor vehicle accidents and X_s the true number of deaths for that State. Equation (4) would then state that except for random variation, the percentage difference between the true and synthetic estimates of deaths from motor vehicle accidents for a State is a linear function of its population density.

4.3 *Estimation of* α and β

Since Z_s is assumed to be available for each State, the regression coefficients α and β could be estimated if corresponding values of Y_s were available. The percentage difference, Y_s, however, is a function of the true parameter, X_s, which is not known, as well as the synthetic estimate $\tilde{X}_s$ which can be obtained. If some estimate, X_s' of X_s were available, however, it could be substituted for X_s into the expression for Y_s and estimates of α and β could be obtained from least squares.

One of the problems, however, is that estimates X_s' of X_s are not available for States from the National Health Surveys. One can obtain, however, estimates X_c' of X_c where X_c is the value of characteristic X for the c-th ***strata combination.*** A ***strata combination*** is defined here as any unit that can be constructed by combining strata. Since unbiased estimates are available for strata, unbiased estimates X_c' can be obtained for strata combinations. Also, since strata are generally counties or groups of counties, the ancillary variable Z_c can be readily obtained. Likewise, the synthetic estimate, $\tilde{X}_c$ can be obtained for each strata combination. Thus, if we divide the United States into C strata combinations and obtain X_c', Z_c' and $\tilde{X}_c$ in the usual way, we can estimate α and β by least squares from the data pairs (Z_c', Y_c') where

$$Y_c' = \frac{X_c' - \tilde{X}_c}{\tilde{X}_c} \times 100, c = 1, \ \ldots, C \tag{6}$$

Once, the estimates $\hat{\alpha}$ and $\hat{\beta}$ are obtained, they can be substituted into equation (5) and estimates $\hat{X}_s$ of X_s can be obtained for each State.

An example of this estimation procedure was constructed from the mortality data on deaths from motor vehicle accidents discussed in Section 3. For the ancillary variable, we let Z represent population density and divided the United States into the 14 State combinations shown in Table 2. From these we obtained the appropriate $\tilde{X}_c$, X_c and Y_c'. The least squares estimates of α and β obtained from these 14 State combinations were $\hat{\alpha} = 12.0626$ and $\hat{\beta} = -.0660$. The correlation between Z and Y as estimated from the 14 sample points was $r = -.6034$. Having obtained $\hat{\alpha}$ and $\hat{\beta}$, the estimated deaths $\hat{X}_s$ were computed for each State and the distribution of percentage absolute differences is shown in Table 3 alongside that for $\tilde{X}_s$. Clearly $\hat{X}_s$ is an improvement over $\tilde{X}_s$ in the sense that the median percentage absolute difference was 10.0 for $\hat{X}_s$ as compared with 15.9 for $\tilde{X}_s$.

5. Some Comments

While the scope of this evaluation study was not large enough to make any final conclusions, about the value and accuracy of synthetic estimation, some extrapolations might prove valuable in planning further studies:

(1) The estimator, $\hat{X}_s$, might be especially suitable to estimate health parameters from National Health Survey Data. Without loss of generality, Z_s might be a vector of ancillary data and the estimator $\hat{X}_s$ would be a multiple regression type estimator. The problem would be to find a set of variables, Z, which might be related to the health characteristic being estimated and different health characteristics would require different sets. There is a wealth of variables available for small areas from the 1970 U.S. Census which might be related to health variables, and this method of estimation could make use of these Census data.

(2) Since there was much variability in the agreement of synthetic estimates with true values not only among States for each cause of death, but also among causes of death with respect to median and mean percentage absolute differences, one might generalize that

Table 2. State Clusters Used to Obtain Regression Coefficients

State Cluster	Population Density (Persons/Mile^2)	Motor Vehicle Deaths		
		Synthetic Estimate $\tilde{X}_c$	True Value X_c	Percentage Difference Y_s $100 \times (X_c - \tilde{X}_c)/\tilde{X}_c$
1. Maine, N. H., Vt.	39.88	267.6	375	40.13
2. Mass., R. I., Conn.	617.98	1154.6	991	-14.17
3. N. Y., Pa.	302.34	3800.8	3857	1.48
4. Ohio, Illinois	204.20	4267.7	3648	-14.52
5. Indiana, Michigan	133.95	2662.2	2884	8.33
6. Wisconsin, Mo., Iowa, Minnesota	55.58	3133.8	3385	8.02
7. N. D., S. D., Neb., Kansas	16.10	1073.4	1220	13.65
8. Del., Md., D. C.	361.90	1071.7	700	-34.68
9. Va., N. C., Fla.	94.13	3378.8	3249	- 3.84
10. W. Va., S. C., Ga., Ky., Tenn., Alabama., Louisiana	73.39	5216.4	5492	5.28
11. Ark., Okl., Texas, Miss.	36.79	3912.2	4140	5.82
12. Alaska, Nevada, Ariz., Montana, Idaho, Wyoming, New Mexico, Utah	4.02	1373.4	1809	31.7
13. Colorado, Oregon, Washington	23.90	1680.7	1556	- 7.42
14. California, Hawaii	100.32	4355.2	4044	- 7.15

$\alpha = 12.0626$

$\beta = -.0660$

Table 3. Distribution of Percentage Absolute Differences for $\hat{X}_s$ and $\tilde{X}_s$ with Respect to Motor Vehicle Accident Deaths 1960

Percentage Absolute Difference	Frequencies (f) and Cumulative Percentages (cp) $\hat{X}_s$ f	$\hat{X}_s$ cp	$\tilde{X}_s$ f	$\tilde{X}_s$ cp
0- 4.0%	10	20.4	7	14.3
4.1- 8.0%	10	40.8	9	32.7
8.1-12.0	9	59.2	6	44.9
12.1-16.0	3	65.3	3	51.0
16.1-20.0	1	67.3	6	63.3
20.1-24.0	5	77.6	3	69.4
24.1-28.0	2	81.6	4	77.6
28.1-32.0	2	85.7	2	81.6
32.1-36.0	0	85.7	2	85.7
36.1-40.0	1	87.8	1	87.8
40.1-44.0	3	93.9	1	89.8
44.1-48.0	1	95.9	0	89.8
48.1+	2	100.0	5	100.0
Total	49		49	
Median % absolute difference	10.0		15.9	
Mean % absolute difference	16.1		20.2	
Relative Variance	1.03		1.00	

great care should be taken in the interpretation of synthetic estimates.

(3) Much more work is needed in the development of methodology to produce estimates of health characteristics for small areas.

REFERENCES

1. U.S. Department of Health, Education, and Welfare: Synthetic State Estimates of Disability – PHS Publication No. 1759. U.S. Government Printing Office (1968).

2. National Center for Health Statistics. Unpublished data (1968).

3. U.S. Department of Health, Education, and Welfare: Vital Statistics of the United States 1960. Volume II – Mortality Part B. U.S. Government Printing Office (1963).

4. Fuchs, V. R. and Leveson, I.: Motor Accident Mortality and Compulsor Inspection of Vehicles, Journal of the American Medical Association, *201*, pages 87-91 (1967).

TWO APPROACHES TO HEALTH INTERVIEW SURVEYS

Regina Loewenstein, Columbia University

Introduction

Planning health programs and legislation for health insurance requires evaluation of different health delivery systems and various methods of payment. The necessary information about utilization and charges for personal health services are frequently obtained from household health interview surveys.

Different types of questions have been used in surveys to ask people about utilization and charges for personal health services. In some studies, specific conditions and services are first identified, and in other studies specific facilities used are first identified. Other studies use a combination of these approaches, no probe questions or different types of probe questions.

The survey described here was designed to compare interviews with the conditions and facilities approaches with regard to reported amounts of services and charges, and also with regard to accuracy of data. In interviews with the conditions approach, probe questions asked about specific illnesses and types of health care during the past year; and with the facilities approach, probe questions asked about specific types of facilities used in the same time period.

Comparisons of the volume and accuracy of reporting in interviews with the two approaches were done with three types of data:

1. Interviews with two matched samples of the community: to compare amounts of reported services and charges during the previous year.

2. Record checks with facilities reported used by persons from the community samples: to compare under and over-reporting of numbers of services, and over-reporting of use of specific facilities.

3. Interviews with matched samples of known users of different types of facilities: to compare under-reporting of use of specific facilities.

The area covered by this survey was the Washington Heights Health District, a lower and middle-class community in the northernmost section of Manhattan of New York City. When interviews were done in 1965-1966, the population was about 270,000.

The study was supported by Grant CH 00128 from the National Center for Health Services Research and Development and was one of the collaborating projects of the Master Sample Survey supported by the Health Research Council of New York City under Contract U-1053.

Community samples

Two matched probability area samples, each consisting of 1000 housing units, were chosen so that each sample was representative of the health district. Families in one sample were randomly assigned to be interviewed with schedules using the conditions approach, and those in the other sample with the facilities approach. In July 1965 to March 1966, interviews were obtained from 759 families with 1862 persons using the conditions approach, and from 781 families with 1943 persons using the facilities approach.

Interview schedules were designed to compare information about all types of health services, except dental care, reported in response to probe questions with each approach. The first 28 of 54 probe questions for conditions cases were almost identical to the list used at that time by the National Health Interview Survey. In the schedule for facilities cases, the first 2 of 50 probe questions asked about family doctor and general practitioner, and other questions asked about specific specialists and clinics. Whenever possible, there were parallel probe questions, such as, heart trouble and heart specialist.

For each conditions person with one or more reported attended condition or service during the year, questions were asked about services and charges during one year from each facility used for each condition. Whenever a specific facility was reported more than once, questions were asked to obtain information not previously reported. For each facilities person with one or more reported types of facilities, similar questions were asked to obtain names of facilities and unduplicated services and charges.

Persons and families covered by interviews with the two approaches did not differ significantly in 90 of 92 personal characteristics. Random assignments of cases to interviewers minimized interviewer-respondent interaction, and assured completion of approximately the same number of conditions and facilities cases by each interviewer. Therefore, observed differences between the two groups of cases can be assumed to be due only to the different approaches.

Differences of the two schedules with regard to reported services and charges to various types of facilities by all persons and by specific demographic subgroups were analyzed by tests of homogeneity. Under the assumption that more reported services and charges indicate better recall, these detailed findings may be used to select the best approach for a survey of a community or group with known characteristics.

For all persons, the facilities approach yielded better recall of clinics with charges, and the conditions approach more reporting of private doctors and free facilities. (Table 1.) With the conditions approach, recall was better by males, women 65 years or older, and persons of low income. More services and charges were reported with the facilities approach by persons of higher income level.

Among almost all demographic subgroups there were fewer persons with unknown amounts of services and charges among the facilities than the conditions cases. In selecting the type of approach for a specific survey, one would try to maximize reported services and charges and to minimize unknown amounts.

To compare the amount of work used for each approach, a unit of reporting was defined as a condition-facility for the conditions approach and as a facility for the facilities approach, whether or not additional services or charges were reported. The mean number of such units was 15 percent less for facilities cases than for conditions cases.

In summary, the facilities technique required less work, had fewer persons with unknown services and charges, and yielded about the same reported amounts of services and charges.

Record checks

One method of studying relative accuracy of the two types of schedules was comparisons of services reported in interviews and in records of specific facilities mentioned. These record checks measured over-reporting of the use of facilities and accuracy of reported numbers of services.

Doctors returned more than 700 or 67 percent of 1100 questionnaires about persons who reported ambulatory care by them. Arrangements were made to search for records of 150 persons who reported use of a large out-patient department of a hospital, and 111 or 74 percent were located. In addition, records of 96 members of a prepaid comprehensive group practice were made available to the study staff.

The results of these record checks of ambulatory medical care are summarized in Table 2. Differences among types of facilities were greater than differences between the two approaches.

Over-reporting of the use of facilities was less than 10 percent for private medical doctors, slightly more than 10 percent for prepaid care, and 20 percent for clinic care. Mean visits per person-facility per year were .4 higher in interviews than records for private doctors and 2 visits higher for clinic care. For prepaid care, the means from interviews were less than the means from records, but for facilities cases the means were almost equal.

The correlation coefficient of visits in interviews and records was about .7 for conditions cases and .6 for facilities cases for both private doctors and clinics. But for prepaid care, the correlation coefficient of .48 for conditions cases was significantly less than .74 for facilities cases.

For clinic cases, the high percent of so-called over-reporting and the considerably larger mean visits from interviews than records may reflect incomplete records as well as over-reporting by respondents. Some persons may have reported as visits days spent in waiting rooms when no services were received, or trips to make appointments or to purchase medicines.

In summary, the largest difference between approaches revealed by the record checks showed more accuracy with the facilities schedule in the reporting of prepaid care. For private doctors and clinics, with most of the reported visits, there were no differences in the relative accuracy of the two approaches.

Samples of known users

In order to compare under-reporting of the use of different types of facilities with the two approaches, interviews were also sought about persons known to have used specific facilities during one year prior to the interview. Samples of known users of facilities in the study area were limited to persons living there when services were received and also when interviews were completed.

Through contacts in the community, seven samples were selected. A sample of adult private patients was obtained from four doctors and a sample of private patients who were children was selected from school records of pre-school examinations. School records also gave samples of children examined and/or treated by school doctors without charges and by clinics. In addition, there was a sample of persons who received free care in health department clinics and there were two samples of persons who had free or low-cost care for psychiatric problems.

Interviews with families of these patients were done at the same time and by the same staff as those from the community samples, and interviewers had no information about the type of case. Random assignments were also used for these cases.

For each type of facility, reporting of the use of the known facility was slightly, but not significantly, higher with the facilities approach. (Table 3.) There were more variations in reporting by type of case than by approach.

Private doctors were reported for about two-thirds of each of the two samples. Free care by school doctors and health department clinics was reported for 30 to 40 percent of persons with known services. The specific psychiatric clinic was reported for more than half

OF THE PATIENTS.

FURTHER ANALYSES SHOWED THAT RECALL OF THE USE OF A SPECIFIC FACILITY WAS BETTER NOT ONLY WHEN THERE WERE CHARGES BUT ALSO WHEN THERE WERE MORE VISITS DURING THE SURVEY YEAR.

ESTIMATED AMBULATORY CARE BY PRIVATE MEDICAL DOCTORS

SINCE PRIVATE MEDICAL DOCTORS WERE INCLUDED IN BOTH THE RECORD CHECKS AND INTERVIEWS ABOUT KNOWN USERS, A MATHEMATICAL MODEL WAS DEVELOPED TO APPLY THESE RESULTS TO ESTIMATE MEAN PERSON-DOCTORS AND MEAN VISITS TO PRIVATE DOCTORS PER PERSON PER YEAR. IN THE APPENDIX ARE THE DERIVATION OF THIS MODEL AND THE RESULTS OF COMPUTATIONS OF ESTIMATED MEANS.

ABOUT 80 PERCENT OF THE ESTIMATED TOTAL NUMBER OF PERSON-DOCTORS WAS REPORTED WITH BOTH TYPES OF SCHEDULES. BUT THE UNDER-REPORTING OF VISITS WAS SOMEWHAT LESS FOR THE FACILITIES APPROACH THAN FOR THE CONDITIONS APPROACH. THE RATIO OF THE MEAN REPORTED VISITS TO ESTIMATED MEAN VISITS WAS .88 FOR CONDITIONS CASES AND .93 FOR FACILITIES CASES.

IMPLICATIONS

BECAUSE OF THE SLIGHT DIFFERENCES IN YIELD AND ACCURACY OF THE TWO APPROACHES, AND BECAUSE THE FACILITIES APPROACH IS EASIER AND CHEAPER TO EXECUTE, IT IS SUGGESTED THAT THE FACILITIES APPROACH BE USED. BASED ON THE MOST EFFECTIVE QUESTIONS IN BOTH APPROACHES, AND STATING ALL QUESTIONS AS FACILITIES, THE FOLLOWING LIST OF 14 PROBE QUESTIONS IS PROPOSED:

- HOSPITAL STAYS
- DOCTORS WHO GAVE CARE IN HOSPITALS
- SURGEONS
- FAMILY OR REGULAR DOCTORS
- CLINICS
- EYE DOCTORS, OPTOMETRISTS, OR PLACES TO GET GLASSES
- OBSTETRICIANS AND GYNECOLOGISTS
- PEDIATRICIANS
- OTHER DOCTORS
- OTHER PERSONS, SUCH AS, FOOT DOCTORS, CHIROPRACTORS, ETC.
- ANY OTHER PLACE AND ANYONE ELSE FOR:
  - PHYSICAL EXAMINATIONS
  - X-RAYS, TESTS, SHOTS
  - COLDS, VIRUSES, ETC.
  - ANYTHING ELSE

SPECIFIC QUESTIONS CAN BE ADDED OR DELETED ACCORDING TO THE PURPOSES OF THE SURVEY AND THE AGE-SEX COMPOSITION OF THE PERSONS IN A PARTICULAR STUDY.

TABLE I

REPORTED AMBULATORY VISITS FOR HEALTH SERVICES* DURING YEAR FOR COMMUNITY SAMPLES, BY TYPE OF FACILITY AND TYPE OF SCHEDULE

TYPE OF FACILITY	PERCENT WITH VISITS		MEAN VISITS PER PERSON	
	CONDITIONS	FACILITIES	CONDITIONS	FACILITIES
NUMBER OF PERSONS	1862	1943	1862	1943
ANY FACILITY	77.8	73.8	5.3	5.4
PAID				
PRIVATE MEDICAL DOCTOR	48.9	47.9	3.0	2.7
CLINIC	12.4	14.4	.6	.7
PRIVATE NON-MEDICAL PRACTITIONER*	14.7	15.1	.4**	.4**
PREPAID	6.3	7.0	.3	.4
FREE	24.8	22.0	1.0	1.1
TYPE UNKNOWN	9.0	4.7	.0	.1

*EXCLUDES DENTAL CARE.
**EXCLUDES VISITS TO OPTOMETRISTS.

TABLE 2

RECORD CHECKS OF REPORTED AMBULATORY MEDICAL CARE DURING YEAR, BY TYPE OF FACILITY AND TYPE OF SCHEDULE

	Private doctors		Clinics		Prepaid centers	
	Conditions	Facilities	Conditions	Facilities	Conditions	Facilities
Reported use of facility						
Number of person-facilities	336	362	51	60	47	49
Percent with care in interviews and not in records	8.3	9.7	19.6	20.0	19.2	8.2
Percent with care in records and not in interviews	*	*	*	*	12.8	10.2
Reported number of visits						
Number of person-facilities	319	330	51	60	47	49
Mean number of visits in year						
Reported in interviews	4.9	3.9	8.3	6.8	3.4	4.0
From records	4.5	3.5	6.3	4.8	4.5	4.2
Difference of means	.4	.4	2.0	2.0	- 1.1	- .2
Correlation coefficient of number of visits reported in interviews and found in records	.72	.62	.76	.64	.48	.74

*Data not available.

TABLE 3

REPORTED USE OF AMBULATORY HEALTH FACILITIES KNOWN TO HAVE BEEN USED DURING YEAR, BY TYPE OF CASE AND TYPE OF SCHEDULE

Type of Case and Schedule	Number of cases	Total percent	Reported used	Not reported used	Uncertain if reported used*
Private doctors seen by adults					
Conditions	94	100.0	67.0	30.9	2.1
Facilities	84	100.0	67.9	31.0	1.2
Private doctors seen by school children					
Conditions	70	100.0	58.6	31.4	10.0
Facilities	63	100.0	74.6	17.5	7.9
Psychiatric clinic					
Conditions	49	100.0	49.0	46.9	4.0
Facilities	54	100.0	61.1	38.9	.0
Schizophrenic cohort					
Conditions	27	100.0	25.9	59.3	14.8
Facilities	25	100.0	52.0	40.0	8.0
Health Department clinics					
Conditions	129	100.0	36.4	47.3	16.3
Facilities	131	100.0	42.0	35.1	22.9
Clinics used by school children					
Conditions	20	100.0	35.0	50.0	15.0
Facilities	21	100.0	38.1	57.1	4.8
School doctors					
Conditions	59	100.0	30.5	59.3	10.2
Facilities	61	100.0	34.4	63.9	1.6

*Primarily due to reporting of a vaguely defined location that had several facilities.

APPENDIX

ESTIMATED AMBULATORY CARE BY PRIVATE MEDICAL DOCTORS

ASSUMPTIONS

1. THE RECALL OF PERSON-DOCTORS WITH RECORD CHECKS IS REPRESENTATIVE OF THE RECALL OF ALL PERSON-DOCTORS DURING THE YEAR.
2. THE RECALL OF PERSON-DOCTORS IN INTERVIEWS WITH SAMPLES OF KNOWN USERS IS REPRESENTATIVE OF THE RECALL OF ALL PERSON-DOCTORS USED DURING THE YEAR.
3. THE NUMBERS OF VISITS REPORTED BY DOCTORS IN QUESTIONNAIRES ARE MORE ACCURATE THAN THE NUMBERS REPORTED IN INTERVIEWS.

NOTATION

LET N = NUMBER OF PERSONS COVERED IN INTERVIEWS WITH COMMUNITY SAMPLES

D = NUMBER OF PERSON-DOCTORS FOR ONE YEAR REPORTED IN INTERVIEWS WITH COMMUNITY SAMPLES

P = PROPORTION OF PERSON-DOCTORS OVER-REPORTED (BASED ON RECORD CHECKS)

U = PROPORTION OF PERSON-DOCTORS NOT REPORTED (BASED ON INTERVIEWS WITH SAMPLES OF KNOWN USERS)

V = NUMBER OF VISITS REPORTED FOR N PERSONS BY D PERSON-DOCTORS IN ONE YEAR

K = ADJUSTMENT FACTOR FOR MEAN VISITS PER PERSON-DOCTOR PER YEAR

$$= \frac{\text{MEAN VISITS PER PERSON-DOCTOR FROM RECORDS}}{\text{MEAN VISITS PER PERSON-DOCTOR FROM INTERVIEWS}}$$

DERIVATION

UD = ESTIMATED NUMBER OF PERSON-DOCTORS UNDER-REPORTED

PD = ESTIMATED NUMBER OF PERSON-DOCTORS OVER-REPORTED

$D + UD - PD = D(1 + U - P) = D'$ = ESTIMATED NUMBER OF PERSON-DOCTORS FOR N PERSONS IN ONE YEAR

V/D = MEAN REPORTED VISITS PER PERSON-DOCTOR PER YEAR

$K(V/D)\ D(1 + U - P)$ = ESTIMATED NUMBER OF VISITS TO ESTIMATED NUMBER OF PERSON-DOCTORS IN ONE YEAR = ESTIMATED TOTAL VISITS BY N PERSONS IN ONE YEAR

$K(V/N)\ (1 + U - P) = V'/N$ = ESTIMATED MEAN VISITS PER PERSON IN ONE YEAR

SUMMARY OF MODEL

TO OBTAIN A BETTER ESTIMATE OF MEAN VISITS PER PERSON, MEAN REPORTED VISITS IS MULTIPLIED BY TWO FACTORS. ONE FACTOR IS A CORRECTION FACTOR OF MEAN VISITS PER PERSON-DOCTOR BASED ON RECORD CHECKS. THE SECOND FACTOR IS BASED ON THE ESTIMATE OF UNDER-REPORTING OF PERSON-DOCTORS BASED ON INTERVIEWS WITH SAMPLES OF KNOWN USERS AND ON THE ESTIMATE OF OVER-REPORTING OF PERSON-DOCTORS BASED ON RECORD CHECKS.

RESULTS OF COMPUTATIONS

	CONDITIONS	FACILITIES
D/N = MEAN REPORTED PERSON-DOCTORS IN ONE YEAR	.66	.70
D'/N = ESTIMATED MEAN PERSON-DOCTORS IN ONE YEAR	.82	.86
D/D' = PROPORTION OF PERSON-DOCTORS REPORTED	.80	.81
V/N = MEAN REPORTED VISITS IN ONE YEAR	3.0	2.7
V'/N = ESTIMATED MEAN VISITS IN ONE YEAR	3.4	2.9
V/V' = PROPORTION OF VISITS REPORTED	.88	.93

Table 2 - Parameters of the quadratic approximation of log $\overset{0}{e}_x$ for a few model life tables for the males.

Mortality level (1)	a (2)	$b\times10^2$ (3)	$c\times10^3$ (4)
10	1.5984	-.433	-.102
25	1.6467	-.313	-.119
40	1.6922	-.251	-.127
55	1.7375	-.253	-.128
70	1.7644	-.194	-.134
85	1.7942	-.193	-.134
100	1.8233	-.223	-.130

The estimated values of c are sufficiently small as these were expected to be. The values are negative, show a slowly declining trend and a sign of increase at the end. b, which is also small and negative, appears to be an oscillating but diminishing function of the levels. The increasing trend of 'a' is, however, more regular. Obviously, the estimates of these parameters, being dependent upon the model life tables, are subjected to the defects of those tables and hence should not be regarded as final. For a thorough analysis the parameters may be allowed to vary within tolerable limits and the results may be checked for consistency by comparisons with other life table functions. So far as the model life tables are concerned, the life expectancies reproduced from the estimated parameters compare favorably with the actual values. (Appendix 1)

3. Derivation of other life table functions

It is apparent from the results presented earlier that an alternative method for constructing model life tables can be formulated if it can be shown that the information about $\overset{0}{e}_x$ is sufficient to generate other life table functions. Fortunately, T_x can be determined directly from $\overset{0}{e}_x$, when l_x can be obtained from $l_x = T_x/\overset{0}{e}_x$ and hence the entire life table can be completed. This is so because

$$\int_{x_1}^{x_2} \frac{dx}{\overset{0}{e}_x} = \int_{x_1}^{x_2} \frac{l_x dx}{T_x} = \log_e (T_{x_1}/T_{x_2}) \quad (8)$$

Since $T_0 = l_0 \overset{0}{e}_0$ is known for a given model table, any T_x can be solved from (8) by putting $x_1 = 0$ and $x_2 = x$ provided the integral on the left hand side of (8) can be evaluated.

The values of $\overset{0}{e}_x$ can, however, be used to solve the above equation by numerical integration. It has been found that the approximation by trapezoidal rule even for five year intervals beyond age 5 is quite satisfactory. Thus

for $x_1 \geq 5$ and $x_2 - x_1 = 5$,

$$\log (T_{x_1} / T_{x_2}) =$$

$$\frac{1}{2} (x_2 - x_1) (1/\overset{0}{e}_{x_1} + 1/\overset{0}{e}_{x_2}) \quad (9)$$

For the first age interval (0,5), in which $\overset{0}{e}_x$ generally assumes its maximum value at $\hat{x}$, the recommended procedure is to use a quadratic approximation of $1/\overset{0}{e}_x$ for each of the two subintervals $(0, \hat{x})$ and $(\hat{x},5)$, with or even without any correction for equalizing the derivatives at $\hat{x}$. In each of these two cases, the quadratic is assumed to produce a minimum value at $\hat{x}$, thereby reducing the number of parameters from three to two. The parameters can therefore be estimated from the given values of $\overset{0}{e}_x$ at the two boundaries. Thus writing the quadratic as

$$1/\overset{0}{e}_x = m + nx + px^2 \quad (10)$$

and subjecting the equation to the condition that the minimum value is assumed at $\hat{x}$, the requirement for which is

$$n = -2p\hat{x} \quad (11)$$

the integral in (8) can be written as

$$\int_0^{\hat{x}} 1/\overset{0}{e}_x \, dx = \int_0^{\hat{x}} [m - px(2\hat{x} - x)] \, dx \quad (12)$$

for the interval 0 to $\hat{x}$. Since $\overset{0}{e}_0$ and $\overset{0}{e}_{\hat{x}}$ are known, the parameters can be solved from (10), and (12) can be evaluated. The same procedure may be applied to the interval $(\hat{x},5)$.

The equations (8) to (12) have been examined in some detail to verify the utility of this approach. For that, the model life table for males corresponding to level 70 ($\overset{0}{e}_0 = 53.6$) has been selected and the results shown in Table 3. This table has a life expectancy (combined for the two sexes) of 55 years which seems to be quite close to the value of that index in India at the present time.

Table 3. Graduated values of l_x compared with model values for model life tables level number 70 for males.

Age x (1)	$\overset{0}{e}_x$ (2)	1000 T_x Graduated (3)	1000 l_x Graduated (4)	1000 l_x Model (5)
0	53.6	53,565	1,000	1,000
1	58.8	52,644	895	882
5	56.4	49,186	872	840
10	53.9	44,920	833	828
15	50.7	40,821	805	819
20	47.0	36,844	787	806
25	42.9	32,958	768	787
30	38.5	29,138	757	767
35	34.0	25,370	746	748
40	29.7	21,668	730	726
45	25.4	18,052	711	699
50	21.5	14,564	677	665
55	17.8	11,266	633	620
60	14.6	8,249	564	561
65	11.8	5,624	477	484
70	9.4	3,488	371	387
75	7.3	1,898	260	274
80	5.6	862	154	162

It may be pointed out that the life expectancies were computed by rounding off at the first decimal digit and accordingly there is no sense in carrying out the computations of l_x with a radix of 100,000. A radix of 1,000 has therefore been chosen for which the two sets of figures demonstrate considerable closeness. It seems certain that the figures would be a lot closer if the computations of life expectancies were carried out to a few more significant digits and the numerical integrations were based on intervals shorter than five years.

4. Summary and Discussion

The model life tables prepared by the United Nations were based on a study that showed that the shape of the mortality curve is retained at all mortality levels and the infant mortality rate alone is generally sufficient for generating the entire mortality curve. The study reported here is based on the finding that the life expectancy can be regarded as an exponential function of age and for all practical purposes the logarithm of the former can be approximated by a quadratic equation of the latter variable namely, age, for the entire range except the childhood interval of less than 5 years. This age interval (0-4) also includes the age at which expectation of life assumes its maximum value and that age approaches the age 0 with increase in life expectancy. The model life tables were used to determine this age, the maximum life expectancy and the parameters of the quadratic equation for a number of levels, and the results were quite encouraging. Finally, it has been shown that a set of life expectancies is theoretically sufficient to generate the entire life table. The conclusion can therefore be drawn that while the expectation of life at birth, from the point at view of its definition uses the entire information of the life table, it can also be manipulated, under certain empirical conditions, to release all the information that it used with virtually little or no loss in that process.

Footnotes

1. Thus the force of mortality $\mu_x = Bc^x$

2. Makeham, W. M. "On the Law of Mortality and Construction of Annuity Tables" Journal of the Institute of Actuaries, Vol. 8, p. 301 (1860) Makeham wrote $\mu_x = A + Bc^x$ which by means of integration can also be written as $l_x = ks^x g^{c^x}$

3. Perks, W. "On Some Experiments in the Graduation of Mortality Statistics" Journal of the Institute of Actuaries, Vol. 63, p. 12 (1932) Perks modified the Makeham-Gompertz law by writing $\mu_x = (A+Bc^x)/(1+Dc^x)$

4. Hyrenius, Hannes. "Life Table Technique for the Working Ages" Demography, Vol. 7, p. 393. (1970). The assumption is $m_x = a_0 + a_1 x + a_2 x^2 + \ldots$

5. Reed, L. J. and Merrell, M. "A Short Method for Constructing an Abridged Life Table" American Journal of Hygiene, Vol. 30, p. 33 (1939) The formula for age interval of length n years, is

$$ {}_nq_x = 1 - e^{-n\,{}_nm_x - an^3\,{}_nm_x^2} $$

6. Greville, T. N. E., "Some Methods of Constructing Abridged Life Tables" The Record of the American Institute of Actuaries, Vol. 32, p. 29 (1943)

7. United Nations "Manuals on Methods of Estimating Population - Manual III, Methods for Population Projections by Sex and Age", ST/SOA/Series A, Population Studies, No. 25 (1956)

8. Coale, A. J. and Demeny P., Regional Model Life Tables and Stable Populations, Princeton University Press, Princeton, New Jersey (1966)

9. Mitra, S. "A Few properties of the Expectation of Life $\overset{0}{e}_x$" Presented at the Second World Population Conference in Belgrade, 1965

Appendix 1. Values of $\overset{0}{e}_x$ computed from model life table of males compared with those obtained by fitting a second degree curve to log $\overset{0}{e}_x$.

Note: M=Model, G=Graduated

Values of $\overset{0}{e}_x$ corresponding to model number

Age	10		25		40		55		70		85		100	
x	M	G	M	G	M	G	M	G	M	G	M	G	M	G
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
0*	24.8	24.8	31.9	31.9	39.2	39.2	46.4	46.4	53.6	53.6	61.5	61.5	68.5	68.5
5	38.1	37.6	43.7	42.5	49.1	47.5	54.0	52.7	58.6	56.4	62.6	60.4	66.3	64.3
10	35.8	35.1	40.7	40.1	45.7	45.1	50.2	50.0	54.4	53.9	58.2	57.8	61.6	61.3
15	32.3	32.4	37.0	37.4	41.6	42.3	45.9	46.8	49.9	50.7	53.5	54.3	56.8	57.6
20	29.1	29.6	33.5	34.4	37.9	39.0	42.0	43.2	45.7	47.0	49.1	50.3	52.1	53.3
25	26.2	26.7	30.5	31.2	34.6	35.5	38.4	39.3	41.8	42.9	44.8	45.9	47.5	48.5
30	23.4	23.8	27.4	27.9	31.2	31.8	34.7	35.2	37.8	38.5	40.5	41.2	42.8	43.5
35	20.7	21.0	24.4	24.6	27.8	28.1	31.0	31.0	33.7	34.0	36.1	36.5	38.2	38.6
40	18.1	18.3	21.4	21.5	24.5	24.5	27.2	27.0	29.7	29.7	31.8	31.8	33.6	33.5
45	15.8	15.8	18.6	18.4	21.2	21.0	23.6	23.1	25.7	25.4	27.5	27.2	29.1	28.8
50	13.6	13.4	15.9	15.6	18.1	17.8	20.1	19.5	21.9	21.5	23.4	23.0	24.8	24.3
55	11.6	11.2	13.4	13.0	15.2	14.8	16.8	16.2	18.3	17.8	19.6	19.1	20.8	20.2
60	9.7	9.4	11.0	10.8	12.4	12.1	13.7	13.3	15.0	14.6	16.0	15.7	17.0	16.6
65	7.9	7.7	8.9	8.8	9.9	9.8	11.0	10.8	11.9	11.8	12.8	12.6	13.6	13.4
70	6.2	6.3	7.0	7.0	7.8	7.8	8.5	8.6	9.3	9.4	10.0	10.0	10.6	10.6
75	4.8	5.0	5.4	5.6	6.0	6.2	6.5	6.7	7.1	7.3	7.6	7.8	8.0	8.4
80	3.6	4.0	4.0	4.3	4.5	4.8	4.9	5.2	5.3	5.6	5.6	6.0	5.9	6.4

(*Graduated $\overset{0}{e}_o$ has been assumed to be the same as model $\overset{0}{e}_o$)

CURRENT DEVELOPMENTS ON COLLECTING INCOME DATA IN THE CURRENT POPULATION SURVEY*

Mitsuo Ono, Bureau of the Census

Introduction

Since 1947, the Bureau of the Census has been publishing annually in the P-60 Current Population Report Series, a publication entitled Consumer Income containing data on size distribution of money income for families, unrelated individuals and persons cross-classified by various social and economic characteristics. The basic data for this report are collected in the March supplement of the Current Population Survey (CPS), a sample survey currently covering 50,000 households. Because of the increasing use and the need for better statistics on income size distribution data, not only by governmental agencies, but also by business and academic organizations, the Bureau has instituted a long-range program to improve both the quality and quantity of income data available from the CPS.

The purpose of this paper is to report some of the progress made in this work and to outline further plans. It is divided into six parts. The first part provides a framework of analysis. The second part summarizes key results in attempting to reduce the income nonresponse rate in the CPS. The third part outlines some of the results found in expanding the availability of income data compiled in the CPS. The fourth part covers efforts being made to improve the income allocation procedure. The fifth part highlights some of the developments in processing and publishing income data. The last part notes overall directions being taken in this program.

Framework of Analysis

It is a truism that the collection, processing and publication of social and economic statistics from household surveys must be treated as an information system, with both inputs and outputs related together in a production function. It is essentially an economic problem since the primary goal is to find the "best" mix of inputs to generate relevant, accurate, and timely information.

This information system consists of various operations, including field collection, processing, tabulation, publication and analysis. Within each of these operations, there are many activities which are functionally related with one another. Obviously, the overall effectiveness of this or any other system is determined by its weakest link in the production process.

Hence, it is quite possible that the results of a well-designed field operation can be negated by a shoddy processing operation. Obviously, any information system to be efficient requires not only adequate managerial controls at every phase of the production process but also priorities to strengthen problem areas. With this in mind, an operational plan has been devised to improve further not only the compilation but also the publication of income data. The first topic relates to efforts being made to reduce income nonresponses in the field collection operation.

Developments to Reduce Income Nonresponses

As noted in reference (1), one of the more persistent problems encountered in collecting CPS money income data in the field has been the nonresponse problem. Nonresponses refer to partial or complete failures to obtain income information from respondents when contacted because of refusals, insufficient knowledge or other reasons.1/ A more detailed account of this problem is covered in references (1), (2), and (3). Only additional results are noted here.

When the CPS income improvement program was first instituted in the March 1969 CPS, the family income nonresponse (hereafter designated as NA) rate increased slightly from the previous year, 17 to 19 percent. However, in the next year, the NA rate dropped from 19 percent to 14 percent. (See table 1). Modifications instituted in the March 1970 CPS included the following:

(1) Advance letters requesting respondent's cooperation in reporting income and specifying the income questions to be asked were mailed to all households in continuing rotation groups in March. Households being interviewed for the first time or returning to the sample after an eight-month lapse received the regular letters asking for their cooperation but not mentioning the income questions.

(2) There was a one-week extension of the interview period for three-fourths of the sample. Wherever possible, followup calls and separate questionnaires were used to secure income data when initial interviews were either unsuccessful or incomplete.

(3) Both work experience and income data were collected simultaneously for the full sample. In previous years, work experience information had been collected separately in February and April rather than in March.

Although the drop in the family income nonresponse rate between the 1969 and 1970 CPS was very gratifying, there was still some doubt as to the enduring effect of changes in the field collection procedures on the nonresponse rate. This uncertainty was associated with the idea that the publicity generated for the 1970 Census of Population and Housing and its proximity to the March 1970 CPS interview period might have induced more responses than otherwise.

*Comments by Dr. Murray S. Weitzman, Assistant Division Chief for Economic Statistics Programs, members of the Consumer Income Statistics Branch, Population Division and the assistance of Ralph Bailey and Dorcas Graham, Sampling Systems Branch, Statistical Methods Division are acknowledged.

This question was settled by the results of the 1971 CPS which repeated the basic improvements implemented in the 1970 CPS plus a few additional modifications. The one-week followup period was extended to cover all sample units. Also, another three days were allowed for the regional offices to institute telephone followups on interviews which interviewers were unable to complete, excluding outright refusals. The 1971 CPS results indicate that the income improvement program had indeed stopped the trend in higher NA rates. The family income NA rate for the 1971 CPS (14.6 percent) was essentially the same as that for the 1970 CPS (14.3 percent). The person's income NA rate was 10.8 percent in the 1971 CPS, not significantly different from the 10.3 percent in the 1970 CPS. Also, there was no significant change in the allocation rates for different types of income. The allocation rate covering only persons with income was 7 percent; 10 percent for persons with wages and salaries; 18 percent for persons with nonfarm self-employment income; and 11 percent for persons with "other income" types. 2/

Although the 1971 CPS results indicate a leveling off in the NA rates, there is still the need to further reduce the NA rate. For this effort, an analytical framework has been developed in which the interviewing process in the CPS is envisioned as a two-way communication process between the enumerator and the respondent. The enumerator is viewed as representing the input or demand side (for information) interacting with the latter representing the output or supply side (of information). The type of interview, i.e., telephone or personal interview, represents the channel of communication. As discussed in reference 1, detailed information has been compiled regarding the supply side (i.e., kind of respondents with high NA rates, type of income involved, etc.). Also, data have been compiled to examine the "structure" of income nonresponses. Preliminary analysis of the various combinations of income items not answered by nonrespondent persons and family heads using 1970 CPS data show that about 43 percent of all nonrespondents (36 percent of family heads) did not respond to any of the income items. About 37 percent of nonrespondents (40 percent of family heads) had only one income item missing, primarily the wage and salary item. About one percent of nonrespondents (one percent also for family heads) did not respond to any of the five "other income" types. (See table A on this page)

Based upon these and other findings, it was hypothesized that nonresponses can be viewed as a continuum ranging from "hard-core" refusal cases (with its frictional NA rates) to "softer" situations that could possibly be resolved by using different interviewing techniques, e.g., leaving a mail-in form to obtain income information when the respondent did not have the information available. To examine this hypothesis, a pilot study was conducted in the March 1971 CPS reinterview procedures (conducted in April 1971) whereby reinterviewers probed for additional information from income nonrespondents in the March 1971 CPS. The question asked was as follows: Are there conditions other than a personal visit (or telephone call) under which you would have provided income information to an interviewer?

If the respondent answered "yes" and provided a condition (or conditions) under which he would have provided income information, this information was recorded. If the respondent answered that income information would not be provided under alternative survey conditions, his explanation for his refusal was recorded. A summary of results obtained from this study is found in table B.

Table A.--Distribution of Nonrespondents, by Nonresponse (NA) Patterns: March 1970 CPS

Combinations of income items not answered	Persons 14 years old and over 1/			Family heads
	All persons	Male	Female	
Total.....(percent)	100.0	100.0	100.0	100.0
All (8) income items NA..........	43.4	39.4	48.2	36.3
One item NA......................	37.2	38.7	35.3	39.8
Wage/salary income.............	20.2	21.1	19.1	18.3
Property income................	6.9	6.8	7.1	8.4
Two items NA.....................	8.4	10.5	5.8	12.9
Wage/salary and property income.	3.0	4.2	1.6	5.0
Three items NA...................	6.8	7.2	6.2	6.7
Wage/salary, nonfarm, and farm self-employment income........	5.2	5.1	5.2	4.5
Four items NA....................	1.1	1.2	1.0	1.3
Five items NA....................	1.7	1.3	2.0	1.5
All "other income" types.......	1.4	1.0	1.8	1.1
Six items NA.....................	0.7	0.8	0.6	0.8
Seven items NA...................	0.8	0.8	0.8	0.7

1/ Inclusive of family heads

Table B.—Comments by Income Nonrespondents Obtained In March 1971 CPS Reinterview

	Number	Percent
Nonrespondent would provide income data under certain conditions which are....................	35	30.0
1) Rectifiable under present procedures (e.g. use of personal instead of telephone interviews)............................	25	21.0
2) Not rectifiable under present procedures (e.g. would respond only if legally required)..............................	10	9.0
Nonrespondent would not provide income data under any circumstances................................	82	70.0
Total.........	117	100.0%

Based on the above data, it appears that it may be possible to reduce the present family income non-response rates by about 20 percent or 2 to 3 percentage points from a family income rate of 14 percent if current procedures are modified without regard to additional costs. However, these results also indicate that even with additional improvements, the family income NA rate as presently defined will not go below 10 percent because of the presence of "hard-core" refusal cases. 3/

Further analysis of NA rates has been made on one aspect of the communication channel, i.e., by type of interview. A hypothesis was presented that NA rates covering telephone interviews are higher than rates for personal interviews. To test this hypothesis, family NA rates by both personal and telephone interviews were analyzed from the March 1970 CPS. The results were as follow:

Type of Interview	Family Income NA Rate	Families No.(thous)	Families Percent
Total	14.3%	51,238	100.0%
Personal	11.6	31,678	61.8
Telephone (Regular)	17.6	14,960	29.2
Telephone (Call Backs)	21.2	4,487	8.8
Others (e.g., Individual questionnaire returns)	28.3	113	0.2

The above finding indicates that telephone interviews (consisting of about 40 percent of all family interviews) tend to have higher NA rates than that for personal interviews.

Since information regarding the demand or the enumeration side is almost nonexistent, research plans have been developed to expand the availability of this type of information. This includes finding possible answers to questions such as these:

(1) How do income NA rates differ by enumerators, regional offices, and geographical areas?

(2) How do personal characteristics of enumerators associate with NA rates?

(3) Can field procedures be developed to take into account these personal factors?

Research findings from these studies will be published as soon as they become available. Eventually, data obtained from both the demand and supply sides will be joined together to develop a theory of income nonresponse behavior.

The next topic deals with some of the results found in trying to expand the availability of detailed types of income information collected in the CPS when the questionnaire space is very limited.

Developments in Expanding Availability of Information on Income Types Collected in the CPS

One possible method to obtain more information when space is limited would be to devise a cycling procedure whereby different types of income are collected over a period of time.

Thus, if only five "other income" questions could be asked at any single point in time, three of the five questions could be used to ask individual types of property income, e.g., dividends, interest and net rentals, in one year while in another year, four of the five questions could be allocated to asking individual questions on public transfer payments such as unemployment compensation, workmen's compensation, veteran's payments, and government employee pensions, with the last question on "all others." One of the major drawbacks of this cycling procedure is that information on annual changes for specific types of income would not be available since there would be a break in the time series.

Another possible method to expand the availability of income information would be to collect within one question, individual income types with a "yes-no" circle and with an income amount box used to record the composite amount received for all of the separate income types reported under the specific question. For example, in the March 1971 CPS, the following questions (partial) were used in the income supplement:

48. <u>During 1969, did...receive any money from</u>:

48b.	48d.	48e.
-Dividends? Yes 0 No 0	-Unemployment Compensation? Yes 0 No 0	-Private Pension or annuities? Yes 0 No 0
-Interest on savings accounts or bonds? Yes 0 No 0	-Workmen's Compensation? Yes 0 No 0	-Alimony? Yes 0 No 0
-Net rental income or income from estates? Yes 0 No 0	-Government employee pensions? Yes 0 No 0	-Regular contribution from persons not living in this household? Yes 0 No 0
	-Veteran's payments? Yes 0 No 0	-Anything else? Yes 0 No 0
How much altogether? $	How much altogether? $	How much altogether? $

The strategy here is to find ways to relate uniquely the amount reported to a specific income type. Under this procedure, useful information could be derived if income types reported were mostly single items, with a one-to-one relationship between the income type reported and the dollar amount reported. A study was conducted using data from the March 1971 CPS to examine this relationship. Tabulations were made using unedited "yes-no" circles restricted to persons who were fully reported with some income amounts, i.e., all "yes-no" circles marked with some income amount reported for a specified question. The results of this study showed that 89 percent of all persons with the types of income covered in question 48d were fully reported (as defined previously). For question 48b, the comparable rate was 85 percent and for question 48e, it was 86 percent.

Among "fully reported" persons in question 48d, the results as shown below indicated that 96 percent of them received only one of the four types covered in this question or 85 percent (96% of 89%) of all persons with the types of income covered in question 48d.

Table C.--Combinations on Reporting Detailed Types of Income for Question 48d - March 1971 CPS

Combinations	Number (thousands)	Percent
Total	9,867	100.0
Only one "yes" circle marked	9,442	95.7
Unemployment compensation	3,703	37.5
Workmen's compensation	787	8.0
Government employee pensions	1,358	13.8
Veteran's payments	3,594	36.4
All "yes" circles marked	none	none
Other combinations	425	4.3

Based on the above, it can be concluded that income size distributions for "fully reported" persons reporting single amounts could be developed for types of income covered under question 48d. Income information could also be cross-classified by other social and economic characteristics as may be needed.

Findings from this study showed also that for question 48b, 73 percent of "fully reported" persons had unique property income types. For question 48e, 97 percent of "fully reported" persons had a unique income type reported. The above results showed that the use of "yes-no" circles can significantly expand the analytical value of income data collected in the CPS without enlarging the questionnaire space allocated for income questions.

The next topic covers efforts taken to improve the CPS income allocation procedures.

Developments to Improve Income Allocation Procedures Using Control Card Information

One stage in the processing operation involves the editing and allocation procedures in which a missing CPS income item is imputed either as a "none" or a positive or negative dollar amount. A preliminary study was conducted to determine if information obtained from CPS household sample control cards could provide family income values that would raise the accuracy of the allocated data. In this study, a sample of 400 randomly selected cards was matched to March 1970 CPS questionnaires and their income responses compared. In order to obtain independent control figures, 200 questionnaires with no income allocations were also selected. CPS income allocations were classified under three categories: (1) All income amounts allocated, (2) income partially allocated and partly reported, and (3) any allocations of "none."

The key finding of this study showed that for the first group, income information obtained from household control cards would improve the accuracy of the allocation procedure by using the income interval of the amount on the control card as the range in which allocated amounts on total family income are accepted. Thus, in this procedure, any allocated amount that falls outside of income limits (as picked up from information shown on the control cards) would be rejected and reallocated until an amount is found which falls within the control card limits. One of the unresolved problems of this procedure, however, is that since only total family income information is included on control cards, a way to allocate family income to each family member must be found. This problem, in addition to resolving problems of matching and processing, is still under investigation and results will be reported as soon as they become available.

The next topic relates to developments in the use of computer technology to expedite processing and publication of income data.

Improvements in Computer Processing of Income Data

Significant advances have been made, not only in developing better managerial control over the computer processing of income data, but also in developing greater computer capabilities to generate more timely and varied income information.

With the staffing of personnel versed in both computer technology and income analysis, there has been an accelerated effort to obtain much needed information for managerial control during different stages of computer processing of CPS income data. These "early warning" tabulations provide advance information on the relative quality of the data at different production intervals. This control enables income technicians to catch programming or processing errors fairly early in the production process without having an extensive delay in the final production of the data. More important, this tighter managerial control in the production process results in eliminating the high costs of reruns and reprocessing of computer tape files. Also, this operation provides advance information of income data to be generated in the final tabulations.

Another development in this endeavor has been the use of computer technology to produce final reproduction tables which are used in the printing of published reports. This has eliminated the need for reproduction table typing and tedious table proofing and it is now possible to publish final income reports much more expeditiously than before.

The increasing use of computer technology has also made it possible to show much more information than had been previously published in the income reports. For example, in the final 1969 and 1970 income reports, time series data on selected social and economic characteristics of families located within different income and quintile groupings were shown. This expansion in computer applications will make possible the publication of much richer and more varied types of cross-classified information in future reports.

Along similar lines, computer programs have been written to produce more varied data, e.g., projected income distributions, Gini Indexes, standard errors of medians, etc., to be used not only for publication purposes but also to check data prior to publication.

These are only a few highlights of this important development in the use of computer technology to improve the processing, publication, timing, and analytical value of the data presented in the Consumer Income Reports. High priority has been placed to expand this work.

Directions for Further Improvement

The long-range goal of the income improvement program is to attain an efficient information system which provides to users relevant, accurate and timely information on consumer income. An important step taken in this endeavor is that we are currently conducting a detailed and systematic audit of each phase of the data production process in order to determine what parts need strengthening and how they can be improved. We have set up, as a prime target, plans to develop more detailed technical documentations of the income processing procedure so that interested users may be able to evaluate, and to suggest further improvements. High priority work has been initiated to compile publication designated the "Income Operations Manual" which puts together in very detailed fashion, all of the instructions and procedures used to collect, process, and to disseminate CPS income information.

Work is also being undertaken to publish a technical paper which would include key findings of various research studies which have been or are being undertaken with respect to problems on collecting and processing income data in the CPS. It is hoped that these documentations will be helpful to other technicians who are also involved in compiling and publishing income statistics.

There are other aspects of the income improvement program which have not been mentioned, i.e., plans for record check and reverse record check studies. As these results become available, they will be reported in future publications.

Table 1.--INCOME NONRESPONSE RATE IN THE CPS (1948, 1958-1971) FOR FAMILIES, UNRELATED INDIVIDUALS AND PERSONS 14 YEARS OLD AND OVER, BY COLOR AND SEX, FOR THE UNITED STATES

Survey year	Families and unrelated individuals						Persons					
	Families			Unrelated individuals			Male			Female		
	Total	White	Negro and other	Total	White	Negro and other	Total	White	Negro and other	Total	White	Negro and other
1971 CPS (1107)	14.6	14.8	13.5	11.6	12.1	8.5	12.5	12.6	11.4	9.3	9.5	7.7
1970 CPS (1107)	14.3	14.4	12.8	11.4	11.6	9.7	12.0	12.1	11.1	8.7	8.7	7.0
1969 CPS (1107)	19.0	19.2	17.4	14.5	14.8	11.9	15.0	15.2	13.3	12.2	12.4	10.3
1968 CPS (1107)	17.2	17.3	16.2	12.6	13.1	9.9	12.7	12.9	11.4	9.7	9.7	9.4
1967 CPS (1107)	21.9	22.3	18.0	15.5	16.3	10.1	16.2	16.6	12.6	11.9	12.2	9.6
(1105)	19.0	19.5	15.1	17.2	17.6	14.4	15.2	15.6	11.3	9.3	9.5	7.3
1966 CPS (1105)	14.8	15.0	12.8	15.4	15.5	14.7	(NA)	(NA)	(NA)	(NA)	(NA)	(NA)
1965 CPS (1105)	14.0	14.2	12.7	13.9	14.0	13.5	10.9	(NA)	(NA)	6.9	(NA)	(NA)
1964 CPS (1105)	(NA)	(NA)	(NA)	(NA)	(NA)	(NA)	10.5	(NA)	(NA)	6.6	(NA)	(NA)
1963 CPS (1105)	(NA)	(NA)	(NA)	(NA)	(NA)	(NA)	9.1	(NA)	(NA)	5.6	(NA)	(NA)
1962 CPS (1105)	(NA)	(NA)	(NA)	(NA)	(NA)	(NA)	8.9	(NA)	(NA)	5.3	(NA)	(NA)
1961 CPS *	9.0	9.1	7.3	11.8	11.6	13.0	7.7	7.8	6.9	4.4	4.6	2.9
1960 CPS *	10.5	10.4	10.8	12.5	12.3	13.8	9.0	8.9	9.6	5.0	5.0	5.3
1959 CPS *	10.9	11.0	10.4	13.0	12.2	17.7	9.3	9.2	9.4	5.0	5.0	5.0
1958 CPS *	11.2	11.4	8.6	12.0	12.2	10.7	9.2	9.5	7.2	4.9	5.0	4.1
1948 CPS *	7.5	7.9	5.3	7.3	7.6	5.2	6.7	(NA)	(NA)	3.8	(NA)	(NA)

*Prior to March 1962 CPS income nonresponses were not allocated. NA - Not available.

Note: Beginning from the March 1968 CPS, use of more advanced (1107) electronic equipment enabled the Bureau of the Census to introduce improved income editing and allocation procedures. The data from the March 1967 CPS were processed using both the new (1107) and old (1105) procedures in order to bridge the two sets of information.

REFERENCES

1. "Characteristics of Income Nonrespondents in the Current Population Survey," by E. Spiers, J. Coder, and M. Ono, presented at the 1971 Annual Meeting of the American Statistical Association.

2. "The Effects of 'Improved' Methods of Collecting Income Statistics on Nonresponse Rates-1969 CPS Experiments," by J. Chai, Proceedings of the Social Statistics Section, American Statistical Association, 1970.

3. "Income Nonresponses in the Current Population Survey," by M. Ono and H.P. Miller, Proceedings of the Social Statistics Section, American Statistical Association, 1969.

FOOTNOTES

1/ Nonresponses do not necessarily mean that entries should have been positive or negative amounts. Many of the nonresponses can be "nones."

2/ The base of each percentage covers only persons with the particular type of income.

3/ Income information is obtained by interviewers asking eight different questions for each person 14 years old and over living in the sample unit. Failure to obtain information on any one of these eight questions makes him a nonrespondent. If any family member is a nonrespondent, his family is considered to be a nonrespondent family.

RELIABILITY OF A GUTTMAN SCALE SCORE *

Charles H. Proctor, North Carolina State University

A probability model that seems to reflect the spirit of the Guttman scale model was used in an earlier paper [Proctor, 1970] to furnish an analysis of item response data. By using the estimates for the probability model provided there and by assigning integer scores to the true types, it becomes possible, as this note will describe, to calculate a reliability for these scale scores. The calculation involves the use of both the estimates of the proportions of the underlying true types as well as of the, so called, misclassification parameter. This is a conditional probability with the following meaning. If a subject belongs to a given true type then his responses to each item can be anticipated and the misclassification parameter is the probability that his response to a given item is opposite from that anticipated. By assigning equal probabilities to all true types an alternative reliability, called "flat" reliability, can be calculated. It is quite a bit simpler to associate a standard error to the flat reliability than to the scale reliability, and the quantity may also be more intrinsically interesting.

The assignment of integer scores to the true types, to some extent runs counter to the spirit of the ordered category, rather than numerical, nature of the true types. In data handling practice, scores may be preferred to just the category assignment, particularily for use in a regression computation, consequently some measure of reliability would be welcome for correcting regression coefficients and multiple correlation coefficients for attenuation [Cochran, 1970].

The scoring may be described as follows. A response pattern will be represented, as usual, as a string of plusses and minuses. The integer scores 0, 1, ..., K will correspond to the true type patterns (namely -- ...-, -- ...+, ..., -+...+, ++...+) with that number of plus responses. A non-scale response pattern will be scored for that true type which maximizes its (the non-scale response pattern's) posterior probability. The prior or underlying probabilities of the true types (namely, θ_0, θ_1, ..., θ_K), will be estimated by the maximum likelihood scoring method along with the probability of misclassification, α. These estimators were described in the earlier paper [Proctor, 1970]. The posterior probability of the ℓ^{th} true type for a given, say the i^{th}, response pattern, is obtained by multiplying $\hat{\theta}_\ell$ by $\hat{\alpha}^{D_{i\ell}}(1-\hat{\alpha})^{K-D_{i\ell}}$, where "$D_{i\ell}$ is the number of item responses that need to be changed to modify true type ℓ into observed response pattern i." The hat notation signals the use of estimates.

The true type scores may be denoted by τ and the response pattern scores by X. The correlation between these two random variables could reasonably be referred to as the index of Guttman scale score reliability while its square will be called the scale reliability and written SR. [See Lord and Novick, 1968, p. 61, for the definition of reliability.] The joint distribution of τ and X is fully specified by the parameters θ_0, ..., θ_K, and α and by the scheme for scoring. Having point estimates of these parameters, it is a routine matter to calculate the scale reliability ($\hat{SR}$, say) as if these were the parameter values. This produces a consistent estimate of scale reliability.

In some respects it is unfortunate that the reliability of a scale should depend on the underlying distribution of the true scores. If as a standard distribution of true scores one takes the uniform (each τ-value has probability 1/K+1) and if the response patterns are scored by the number of plusses, then the correlation between true score and observed score depends only on the misclassification parameter. The square of this correlation will be called flat reliability--"flat" in honor of the uniform distribution of τ. The following results point out how the formula for flat reliability is derived.

The observed score X is now the sum of item zero-one scores, say $X = X_1 + X_2 + \cdots + X_K$. Here $X_1 = 1$ whenever a true type $\tau = 1$ appears (which appearance has probability 1/K+1) and the response is a consistent one, or when other true types appear (each with probability 1/K+1) and the response is not consistent. Upon recalling that the quantity α is the probability of an inconsistent (or "misclassified") response while $1-\alpha$ is the chance of a response consistent with the true type one obtains:

$$E(X_1) = [K\alpha + (1-\alpha)]/(K+1).$$

Similarily,

$$E(X_2) = [(K-1)\alpha + 2(1-\alpha)]/(K+1).$$

Finally,

$$(1) \qquad E(X) = K/2.$$

By slightly heavier algebra one can find:

$$(2) \qquad E(X^2) = K/2 + K(K-1)[1 - \alpha(1-\alpha)]/3,$$

and

$$(3) \qquad V(X) = E(X^2) - [E(X)]^2$$

$$= \frac{K(K+2)}{12}\left[1 - 4\alpha(1-\alpha)\frac{K-1}{K+2}\right].$$

Incidentally, the fact is often used that the sum of the squares of first K integers is $K(K+1)(2K+1)/6$. Their sum is $K(K+1)/2$. Formula (1) was obtained by squaring $X_1 + X_2 + \cdots + X_K$, using the fact that $E(X_i^2) = E(X_i)$ for these indicator variates, and then by finding $E(X_iX_j)$ as the sum of 3 parts--one from those true types where $X_i = 1$ and $X_j = 1$ arises from two consistent responses [with probability $(1-\alpha)^2$], another where $X_iX_j = 1$ is produced by two inconsistent responses [with probability α^2] and the third case where $X_i = 1$ and $X_j = 1$ reflects one inconsistent and the other a consistent response [with probability $\alpha(1-\alpha)$].

By similar steps one finds that:

$$(4) \qquad E(\tau) = K/2$$

and

$$(5) \qquad E(\tau^2) = (2K^2 + K)/6 ,$$

while

$$(6) \qquad V(\tau) = \frac{K(K+2)}{12} .$$

$$(7) \qquad E(\tau X) = [K(2K+1) - \alpha K(K+2)]/6$$

and

$$(8) \qquad Cov(\tau,X) = K(K+2)(1-2\alpha)/12.$$

The final step is to square $Cov(X,\tau)$ and then divide by $V(X)$ and by $V(\tau)$ to get:

$$(9) \qquad FR = (1 - 2\alpha)^2/\left[1 - 4\alpha(1-\alpha)\frac{K-1}{K+2}\right],$$

the formula for flat reliability of a Guttman scale. It is apparent that a probability of misclassification of over 0.5 will cause reliability to go below zero. That is, "guessing" will cause reliability to decrease.

As an estimate of FR one would replace α by $\hat{\alpha}$ as obtained from the maximum likelihood estimation calculations. Since that computational routine also provides a standard error for $\hat{\alpha}$ an approximate one can be furnished for $\hat{FR}$. By taking the derivative of FR with respect to α, substituting $\hat{\alpha}$ for α in that expression and then multiplying its absolute value into the standard error of $\hat{\alpha}$ this standard error becomes:

$$(10) \qquad S.E.(\hat{FR}) = \frac{12(1 - 2\hat{\alpha})[S.E.(\hat{\alpha})]}{(K+2)\left[1 - 4\hat{\alpha}(1-\hat{\alpha})\frac{K-1}{K+2}\right]^2} .$$

A set of data that showed $\hat{\alpha} = .0780$ with $S.E.(\hat{\alpha}) = .00764$ would thus show $\hat{FR} = .852$ with $S.E.(\hat{FR}) = .016$. This set of data is No. IV in Table 1. The scale reliability was estimated as $\hat{SR} = .872$ for those data. As might be expected there is quite close numerical agreement between SR and FR over various sets of data (See Table 1). There would seem to be some advantages to FR since it is standardized for the distribution of true types, and this uniform distribution could be seen as a somewhat ideal distribution for Guttman scaling purposes. In cases where the underlying distribution is not close to uniform (as for data Set I in Table 1), there SR may be different from FR (in fact $\hat{FR} = .66$ while $\hat{SR} = .72$ for those data). Perhaps the full range of integer scores would not be entirely appropriate for these data.

FOOTNOTES

* Computer time was made available for this work from a National Science Foundation grant to the Triangle Universities Computer Center.

REFERENCES

[1] Cochran, W. G. Some effects of errors of measurement or multiple correlation. Journal of the American Statistical Association, 1970, 65, 22-34.

[2] Hayes, D. G. and Borgatta, E. F. An empirical comparison of restricted and general latent distance analysis. Psychometrika, 1954, 19, 271-279.

[3] Proctor, C. H. A probabilistic formulation and statistical analysis of Guttman scaling. Psychometrika, 1970, 35, 73-78.

TABLE 1. Guttman Scale Reliability and Flat Reliability for Six Sets of Data *

	True Type Proportions						Misclassification	Scale Reliability	Flat Reliability
Set	θ_0	θ_1	θ_2	θ_3	θ_4	θ_5	$\hat{\alpha}$	$\widehat{SR}$	$\widehat{FR} \pm S.E.(\widehat{FR})$
I	.17	.01	.08	.09	.07	.59	.163 ± .014	.720	.661 ± .035
II	.05	.15	.23	.25	.13	.19	.038 ± .006	.925	.932 ± .012
III	.11	.12	.27	.26	.20	.03	.074 ± .009	.818	.861 ± .018
IV	.24	.15	.22	.13	.08	.18	.078 ± .008	.872	.852 ± .016
V	.14	.10	.14	.30	.21	.11	.028 ± .005	.951	.951 ± .010
VI	.33	.11	.14	.13	.17	.13	.051 ± .010	.932	.907 ± .019

* Source [Hayes and Borgatta, 1954].

ADD-ONS: OBTAINING DESIRED STRATUM SIZES WITH POST-STRATIFICATION

Martin Rosenzweig, University of Pennsylvania

INTRODUCTION

The use of post-stratification may be dictated by basically two situations in practice: First, a reasonably useful population list, or frame, exists, but it is not organized for stratification. Second, the frame may not list the characteristic to be used to stratify the sample.

In the first case, the standard procedure is to re-organize the frame into strata, edit the new frame, and sample this new frame according to some specified sampling plan. However, resource constraints - lack of time, lack of money - may make this unfeasible, particularly for a small study.

On the other hand, if the sample is to be stratified by age or income, for example, no frame exists. Of course, census data does exist for these, and other, characteristics which suggest post-stratification may lead to gain in precision.

The difficulty with post-stratification is that with large samples, it often leads only to modest gains over random sampling (say 25% or so), and with smaller samples, you face the possible embarrassment of empty strata. In view of this, an alternative approach is suggested--

"ADD-ONS"

If there are L strata, and n_h^* is the desired allocation for stratum h, then the procedure is (1) take a series of independent random samples, sampling until

$$n_h \geq n_h^* \qquad h = 1, 2, ---, L$$

where n_h is the number identified as belonging to stratum h, and (2) if

$$n_h > n_h^*$$

sub-sample the n_h members of stratum h to achieve the desired stratum size n_h^*. Then (3), to estimate the mean, for example, if

$$n_h^* = n_{h1} + n_{h2} + .. + n_{hk},$$

where n_{hi} is the sample number in stratum h in the ith sample of the series, in each stratum we use

$$\bar{y}_h = (n_{h1}\bar{y}_{h1} + n_{h2}\bar{y}_{hr} + .. + n_{hk}\bar{y}_{hk}) \mid n_h^*$$

(Note the above is an identity, we need only compute the final mean). Then we see

$$V(\bar{y}_h) = \frac{S_h^2}{n_h^*} \text{ (Ignoring the finite correction)}$$

which is exactly the variance of the mean of a sample of size n_h^* where S_h^2 is the stratum variance. Also, we can now construct our stratified estimator

$$\bar{y}_{st} = \sum_{h=1}^{L} w_h \bar{y}_h$$

where $w_h = n_h^* \mid n$ is the allocation we seek. So that we have

$$V(\bar{y}_{st}) = \Sigma\, w_h^2\, V(\bar{y}_h)$$

the usual variance for a stratified estimator.

Where the characteristic with which we want to stratify is not directly available, the above procedure may be modified, for example, as follows: Draw a random sample of size n_1, then screen (perhaps, by phone) to find the n_{h1}, that is, the number in each stratum. If $n_h < n_h^*$ for any stratum, repeat this procedure. Continue until $n_h \geq n_h^*$ h = 1, 2, .., L, as before.

It is not difficult to show with 2 strata of sizes N_1 and N_2, $N_1 + N_2 = N$, and desired allocation n_1^* and n_2^*, respectively, $n_1^* + n_2^* = n$, that

$$P(n = n_0) = \frac{n_1^*}{n_0} \frac{\binom{N_1}{n_1^*}\binom{N_2}{n_0 - n_1^*}}{\binom{N}{n_0}} + \frac{n_2^*}{n_0} \frac{\binom{N_1}{n_0 - n_2^*}\binom{N_2}{n_2^*}}{\binom{N}{n_0}} .$$

However, with increasing number of strata this formula quickly becomes awkward. Even in the 2 strata case it is not convenient for calculating E(n). If we were able to find E(n), then

$$R.E. = \frac{n_{PS}}{E(n)} \quad ,$$

where n_{PS} is the sample size required for the standard post-stratified estimator to yield the same variance as this estimator, would give a measure of efficiency.

Some preliminary numerical results indicate that in the favorable case where N_1 and N_2 do not differ radically and S_1 and S_2 are known, substantial gains in efficiency are possible.

CONCLUSION

We have now a technique for post-stratification situations which: (1) enables us to achieve <u>any</u> pre-selected allocation, hence achieve any precision required, and (2) is simple to apply. In the first type of case, the additional costs of this technique are trivial, and in the second case, will probably be relatively modest.

WAGE RATE DIFFERENCES AMONG THE WORKING POOR

John R. Shea and Richard J. Emerine, II*
Center for Human Resource Research, The Ohio State University

1. INTRODUCTION

The Nixon administration's welfare reform proposals would extend for the first time on a broad scale income transfers to the working poor. According to the most recent version of the Family Assistance Plan (FAP), an estimated 13.9 million adults and children would be eligible for benefits in fiscal year 1973 under the Opportunities for Families Program, the component of FAP designed for households in which at least one person is employable [11].

If public policy is to assist such families in escaping from poverty, it is important to understand the factors that determine low-income status. Several recent studies have attempted to account for variation in the labor supply of adult family members, and major social experiments are being carried out in New Jersey and elsewhere to ascertain the likely consequences for work effort of welfare schemes embodying features of a negative income tax [4,5,7,13].

Clearly, the earned portion of income is the product of two components: (1) time worked and (2) a rate of pay per unit of time. This paper seeks to complement recent efforts to illuminate the determinants of low-income status by examining this second component of earnings: the wage rate. In the next section of this paper, the conceptual framework guiding the analysis is presented. In Section 3, the data source is described and a basic model of relationships is specified. Section 4 provides background information concerning labor force status and the magnitude of poverty in the population groups that constitute the data base. In Section 5, the regression results are presented. Finally, Section 6 contains a brief discussion of the findings.

2. CONCEPTUAL FRAMEWORK

Only infrequently have personal differences in pay rates (or, average hourly earnings) been subjected to careful analysis [3,6,8]. While there has been considerable empirical work over the years on the process of wage determination, most efforts have either (1) sought to account for interoccupational, interindustry, or interarea variation in rates of pay, or have (2) analyzed the apparent effects of organizations (e.g., unions) and legal constraints (e.g., the Fair Labor Standards Act) on the wage structure. At the same time, studies of the factors that influence an individual's income have generally ignored the separate components of earnings. In some cases, failure to decompose the analysis into (1) hours worked per year and (2) dollars per hour rests on the absence of suitable data. In other instances, researchers have not been especially interested in whether factors such as more education and improved health enhance income primarily through the channel of increased labor supply or through a higher average rate of pay. Yet, for many purposes it is important to know how the factors that influence earned income affect each of these two components. For example, in interpreting the relationship between low educational attainment or poor health and low earned income, social intervention strategies would likely differ depending on whether such personal variables were associated primarily with reduced annual hours of work or with low hourly earnings.

Human capital theory offers a useful framework within which to examine differences in hourly wage rates. In their analyses of the factors that influence a person's annual earnings, human capital theorists in the Becker-Schultz tradition have generally concentrated attention on productivity-increasing effects of human capital formation. An individual's wage rate (assumed to reflect marginal productivity) is viewed as the result of labor demand conditions and of natural and acquired abilities, as measured by educational attainment, health status, years of work experience, and so forth.

In addition to these measures, there is good reason to include several other variables in an analysis of wage rates. For several reasons, including deficiencies in the quality of educational opportunity and the existence of pervasive discrimination in employment, race is generally related systematically to earned income, controlling for the influence of other variables. Region of the country and size of place of residence (i.e., degree of urbanization) are also important correlates of income, especially of earned, money income [2]. This is the case for at least two reasons. First, there are differences among areas and regions in the cost of living, with consumer prices higher than average in the North and West and in larger cities. Second, the historic migration of families from rural areas (and, small towns) to the city, and from the South to the North and West suggests that real wage rates have been in disequilibrium.

Because of constraints on geographic mobility, many women may, on occasion, be subject to monopsonistic pressures in local labor markets. We have used marital status as a proxy for this kind of immobility. While frequently not available in other data sources, we have added two other variables to the analysis: years of service with present employer and, in the case of women, percentage of years since leaving school that the individual has worked at least six months. In addition to on-the-job training (or, on-the-job learning) as measured by exposure to the labor market, job tenure may reflect the acquisition of valuable specific human capital and the existence of valuable job rights stemming from the seniority system. For these reasons we would anticipate, *ceteris paribus*, a positive relationship between job tenure and the wage rate, and between labor force exposure and the dependent variable.

3. DATA, VARIABLES, AND THE MODELS

Individuals selected for analysis in this paper represent subsets of respondents in two of the National Longitudinal Surveys (LGS). Specifically, the analysis uses first-round interview data for poverty and nonpoverty groups of 45-to-59-year-old men interviewed in 1966, and of 30-to-44-year-old women surveyed a year later.[1] The cohorts contain approximately 5,000 individuals each, and represent national probability samples of the civilian noninstitutional population in these age categories. We are fortunate in having a large number of both blacks and whites in each sample. Blacks and other nonwhites were overrepresented by a 3-to-1 ratio relative to whites in order to permit reasonably confident inferences concerning differences between the races in labor market experiences. In other words, of the approximately 5,000 sample cases in each cohort, nearly 1,500 are blacks and other nonwhites.[2]

Unfortunately, we do not have a measure of natural ability for either the men or the women. On the other hand, we do have measures of most of the other variables hypothesized to influence hourly rate of pay.[3] We could have included occupational assignment as an explanatory variable, but we chose to use educational attainment instead. Of course, these two variables are highly intercorrelated, and the influence of education on earnings is mediated through occupational assignments. We take the view, however, that individuals generally settle into the occupational structure at places which maximize their hourly earnings consistent with (1) individual preferences concerning the nonpecuniary aspects of particular jobs and (2) the existence of discrimination in the labor market.[4]

Regression coefficients were estimated for several models of the following form:

$$Y = a + b_1X_1 + b_2X_2 + \ldots + b_nX_n + e.$$

The variables in the several models are described in Table 1. Following a brief discussion of the extent of poverty in the two cohorts, the results of the regression analysis will be presented.

4. POVERTY STATUS

Since this paper emerges from a larger study of potential recipients of family assistance payments--and, because family composition is a defining characteristic--we have not examined the wage rates of employed men and women living in families without children. Thus, automatically, approximately one-sixth of the women and half of the men are excluded from our concern here. Of the remainder, over half of the black women and approximately two-fifths of the black men would have qualified for payments had the Family Assistance Plan as described in H.R. 16311 been in effect at the time they were interviewed [12]. Among whites, approximately one-sixth of the women and one-eighth of the men would have qualified.[5] A woman's participation in the labor force reduces the chance that her family will be in poverty whether or not she has a husband. Among women 30 to 44 years of age living in households with at least one child, a smaller percentage of the poor than of the nonpoor were in the labor force when interviewed in 1967. Among the poor, 58 percent of the blacks and 36 percent of the whites were in the labor force at the time of the 1967 survey. Comparable percentages of the nonpoor were 75 and 45 percent. Participation in the labor force is a less important factor in accounting for the poverty position of families headed by older men. Nine-tenths of the black men in poverty and 86 percent of the white were in the labor force when interviewed in 1966. Of course, the participation rates of nonpoor black and white men were even higher: 97 and 99 percent, respectively.

5. RESULTS OF THE ANALYSIS

Tables 2 and 3 present the basic regression results for the women and men employed as wage or salary workers when first interviewed. In each case, separate regressions were run for each poverty status and racial group. In the case of men, size of place of residence, race, region, job tenure, and educational attainment show up as important determinants of hourly wage rate. Of course, there are exceptions to this statement, and the relative importance of the variables differs somewhat according to color groups and poverty category. Consider job tenure. In the case of the nonpoor, each year of service with a given employer seems to add \$0.01 to \$0.03, on average, to the individual's hourly wage rate. Among the poor, the coefficient is less than \$0.01 and not significantly different from zero. The reason for this difference by poverty status may be that the poor are more likely to be in those types of jobs in which neither on-the-job training nor institutional influences would operate to increase the wage rate with increasing service.[6]

City size and region are also significant factors--in this case for both the poor and the nonpoor. Controlling for the influence of other variables, living in areas with 25,000 or more inhabitants increases the expected wage rate by anywhere from \$0.28 to \$0.53 per hour. Living in the South as opposed to other areas of the country reduces the "expected" rate of pay by a roughly comparable amount, but the influence of region appears to be more powerful among the poor than the nonpoor. Another variable that shows a strong relationship to the hourly wage rate is highest year of school completed, but the relationship is not consistent in the case of poor men.

Ignoring the group in poverty for a moment, it is instructive to compare nonpoor blacks and whites. Nonpoor black men with some high school averaged \$0.22 more per hour than those with fewer than nine years of school, the omitted category, and those with 12 or more years of school earned an average of \$0.96 more per hour. Comparable increments for nonpoor white men were \$0.71 and \$2.51 per hour compared to those in the omitted group.

In Table 3, many of the same relationships between the wage rate and other variables are evident in the case of employed women. Since region of residence could not be entered into the regressions, city size and race may have picked up some of the influence of region. Years of service with present employer is again a significant variable for the nonpoor, and our direct measure of past work experience--percentage of years since school in which the respondent worked six months or more--is salient for the same group. Once again, it would appear that work experience pays off for the nonpoor but not for the poor. Educational attainment is important, especially among the nonpoor. It is also worth mentioning that race (being black) is inversely associated with the wage rate; and, while not significantly different from zero, the coefficient of the health limitations variable posseses the proper sign. Our measure of geographic immobility, however, did not perform according to expectations; marital status does not bear a consistent relationship to the hourly wage rate.

6. DISCUSSION

It is quite clear that race, region, city size, job tenure, and years of schooling strongly influence a person's wage rate. In addition, the number of years of past work experience is an important variable for the women. Nevertheless, with the possible exception of region (for men) and race (for women), the impact of these variables on the wage rate appears to be greater for the nonpoor than the poor. This is undoubtedly, in part, a consequence of how poverty status is defined, since low wage rates are an important factor in accounting for the inclusion of employed individuals in the poverty category. Thus, for this group, there is relatively little variation in the dependent variable. At the same time, we are inclined to believe that natural ability (for which we lack a direct measure) and possible underlying interactions among variables (e.g., low educational attainment and lack of job tenure) may also be important determinants of the low wage rates of substantial numbers of respondents, especially in the case of the poor. The uniformly lower constant term for poor men and women compared to their nonpoor counterparts hints at the possibility of important interactions not captured in the linear models presented here.

FOOTNOTES

*This paper is an outgrowth of a special study entitled "Analysis of Characteristics of Potential Recipients of Family Assistance Through Use of Longitudinal Surveys Data." The National Longitudinal Surveys project is sponsored by the Manpower Administration of the U.S. Department of Labor, under the authority of the Manpower Development and Training Act. Data are collected by the U.S. Bureau of the Census. Interpretations and viewpoints expressed in this paper do not necessarily represent the position or policy of the Department of Labor. We wish to thank M. Borus, S. Kim, A. Kohen, G. Nestel, H. Parnes, and R. Roderick for helpful suggestions on an earlier draft of the paper.

[1]Several summary reports on the two cohorts are available [9,10].

[2]Since we are more interested here in labor market behavior than in universe estimates of personal characteristics, unweighted sample cases are used in the regression analysis. While combined in some cases, regressions have been run separately on blacks and whites, with other races excluded. Deliberate "oversampling" of blacks relative to whites implies that regressions using combined, unweighted observations reflect the black experience more than would be the case if the sampling ratios had been the same.

[3]An exception is region of residence in the case of the women, where--through oversight--the variable was not added to the initial data tape; it is being added to the updated data files covering the 1969 survey.

[4]Ivar Berg [1] examined differences in earnings by educational attainment within occupations; as might be expected he frequently found little difference attributable to education.

[5]Exact percentages are impossible to determine because of the failure of some respondents (approximately 10 percent of those with children) to report fully on their income, assets, and liabilities.

[6]It is perhaps worth noting that fully two-fifths of the poor men report 20 or more years of service with their present (1966) employer. This fraction is still considerably lower, however, than for their nonpoor counterparts: three-fifths.

REFERENCES

[1] Berg, Ivar, Education and Jobs: The Great Training Robbery, New York: Praeger, 1970.

[2] Fuchs, Victor R., Differentials in Hourly Earnings by Region and City Size, 1959, Occasional Paper 101, New York: National Bureau of Economic Research, 1967.

[3] ________, "Male-Female Differentials in Hourly Earnings," paper presented at the Annual Meetings of the Allied Social Science Association Detroit, December 1970.

[4] Greenberg, David H., and Kosters, Marvin, Income Guarantees and the Working Poor: The Effects of Income Maintenance Programs on the Hours of Work of Male Family Heads, Santa Monica, Calif.: Rand, 1970.

[5] Hill, Charles R., "The Economic Determinants of Labor Supply of the Urban Poor," unpublished Ph.D. thesis, University of Minnesota, August 1970.

[6] Johnson, George E., and Stafford, Frank P., "Social Returns to Quantity and Quality of Schooling" paper presented at the Annual Meetings of the Allied Social Science Associations, Detroit, December 1970.

[7] Kalachek, Edward D., and Raines, Fredric Q., "Labor Supply of Low Income Workers," Technical Studies, The President's Commission on Income Maintenance Programs, Washington, D.C., 1970.

[8] Morgan, James, and Sirageldin, Ismail, "A Note on the Quality Dimension of Education," Journal of Political Economy (September-October 1968).

[9] Parnes, Herbert S., et al., The Pre-Retirement Years, Manpower Research Monograph No. 15, Vols. I and II, Washington, D.C.: U.S. Government Printing Office, 1970 and 1971.

[10] Shea, John R., et al., Dual Careers, Manpower Research Monograph No. 21, Vol. I, Washington, D.C.: U.S. Government Printing Office, 1971.

[11] United States House of Representatives Committee on Ways and Means, Revised Press Release No. 5: Summary of Provisions of H.R. 1, "The Social Security Amendments of 1971," as amended and ordered reported to the House of Representatives, May 17, 1971, 34.

[12] United States Senate, Committee on Finance, H.R. 16311, "The Family Assistance Act of 1970," Administration Revision, June 1970.

[13] Watts, Harold W., "Adjusted and Extended Preliminary Results from the Urban Graduated Work Incentive Experiment," Discussion Paper, Madison: Institute for Research on Poverty, University of Wisconsin, June 10, 1970.

Table 1 Variables and Omitted Categories Used in the Models

Cohort and variable symbol	Description
Dependent variable	
$/hour	Hourly wage rate, in dollars: Continuous variable[a]
Explanatory variables	
Men and women:	
Rc	Race: 1 if black; 0 if white[b]
H	Health: 1 if health limits kind of work; 0 otherwise
T	Tenure in current job (years): Continuous variable
S	Size of place of residence: 1 if area of 25,000 or more; 0 otherwise
Men only:	
E_i	Educational attainment, a series of dummy variables of highest year of school completed, specifically, depending on poverty status:
	E_{0-4} 1 if 0-4; 0 otherwise--an omitted category
	E_{0-7} 1 if 0-7; 0 otherwise--an omitted category
	E_{5-7} 1 if 5-7; 0 otherwise
	E_8 1 if 8; 0 otherwise
	E_{9+} 1 if 9 or more; 0 otherwise
	E_{9-11} 1 if 9-11; 0 otherwise
	E_{12+} 1 if 12 or more; 0 otherwise
A_j	Age, a series of dummy variables, specifically:
	A_{45-49} 1 if 45-49; 0 otherwise
	A_{50-54} 1 if 50-54; 0 otherwise
	A_{55-59} 1 if 55-59; 0 otherwise--an omitted category
Rg	Region of residence: 1 if South; 0 otherwise
Women only:	
M	Marital status: 1 if married; 0 otherwise
W	Work experience, expressed as percentage of years since leaving school that respondent worked six months or more to nearest percent: Continuous variable
E_k	Educational attainment, a series of dummy variables of highest year of school completed, specifically, depending on poverty status:
	E_{0-7} 1 if 0-7; 0 otherwise
	E_{0-8} 1 if 0-8; 0 otherwise
	E_8 1 if 8; 0 otherwise
	E_{9-11} 1 if 9-11; 0 otherwise
	E_{12} 1 if 12; 0 otherwise
	E_{12+} 1 if 12 or more; 0 otherwise--an omitted category
	E_{13} 1 if 13 or more; 0 otherwise--an omitted category

a Respondents were asked how much they earned on their current jobs, and if not reported as an hourly rate, hourly equivalents were calculated on the basis of usual hours worked per week.

b Nonblack-nonwhites (e.g., American Indians, Orientals) were excluded from the analysis.

Table 2 Average Rate of Pay ($ per hour): Estimated Regression Coefficients for Men 45 to 59 Years of Age Employed as Wage or Salary Workers at Time of Survey, 1966[a]

(Standard errors in parentheses)

Variables and statistics	Blacks and whites			Blacks			Whites		
	Poor and nonpoor	Poor	Nonpoor	Poor and nonpoor	Poor	Nonpoor	Poor and nonpoor	Poor	Nonpoor
S (1 = 25,000+)	.52	.29	.51	.51	.28	.37	.51	.35	.53
	(.10)*	(.10)*	(.11)*	(.10)*	(.12)*	(.14)*	(.12)*	(.23)	(.13)*
Rc (1 = Black)	-.84	-.10	-.78						
	(.12)*	(.11)	(.14)*						
Rg (1 = South)	-.34	-.52	-.23	-.52	-.65	-.29	-.26	-.39	-.20
	(.11)*	(.14)*	(.12)	(.10)*	(.18)*	(.11)*	(.14)	(.26)	(.15)
H (1 = Health limits kind of work)	.04	.10	-.03	.14	.12	.08	.04	.16	-.06
	(.12)	(.11)	(.14)	(.11)	(.14)	(.15)	(.15)	(.21)	(.16)
T (years)	.031	.004	.030	.021	.007	.014	.033	.003	.032
	(.005)*	(.005)	(.005)*	(.004)*	(.006)	(.001)*	(.006)*	(.012)	(.006)*
A_{45-49} (1 = 45-49)	.03	-.23	.07	-.04	-.21	.03	.09	-.22	.09
	(.10)	(.10)*	(.11)	(.09)	(.12)	(.12)	(.13)	(.23)	(.14)
A_{50-54} (1 = 50-54)	-.02	.13	-.03	-.15	.21	-.28	.03	-.36	.02
	(.14)	(.14)	(.16)	(.12)	(.16)	(.16)	(.18)	(.35)	(.19)
E_{5-7} (1 = 5-7 years)		.09			.13			-.21	
		(.12)			(.13)			(.31)	
E_8 (1 = 8 years)	.13	-.21	.12	.05	-.20	-.01	.32	-.16	.31
	(.17)	(.15)	(.20)	(.13)	(.19)	(.18)	(.24)	(.29)	(.27)
E_{9+} (1 = 9+ years)		.25			.31			-.07	
		(.14)			(.16)			(.36)	
E_{9-11} (1 = 9-11 years)	.54		.49	.34		.22	.75		.71
	(.14)*		(.16)*	(.10)*		(.12)	(.22)*		(.24)*
E_{12+} (1 = 12+ years)	2.30		2.22	1.16		.96	2.60		2.51
	(.16)*		(.18)*	(.17)*		(.19)*	(.23)*		(.25)*
Constant term	2.19	1.75	2.39	1.61	1.64	2.06	2.00	1.78	2.16
	(.27)*	(.24)*	(.32)*	(.24)*	(.30)*	(.34)*	(.41)*	(.48)*	(.39)*
# of observations	1,657	208	1,449	424	156	268	1,233	52	1,181
$\bar{R}^2$	.28	.20	.21	.35	.19	.16	.20	.10	.18
F	64.01*	6.20*	39.57*	26.12*	5.14*	6.64*	34.32*	1.78	30.61*
Dependent variable:									
Mean	$3.38	$1.59	$3.64	$2.28	$1.52	$2.73	$3.75	$1.80	$3.85
S.D.	$2.16	$0.72	$2.17	$1.03	$0.73	$0.92	$2.26	$0.69	$2.32

* Significant at .05 level.

a Restricted to the "definitely poor" and "nonpoor" living in families with at least one child; see Section 4 for a definition of poverty status.

b The omitted category is A_{55-59}.

c The omitted category for the poor is E_{0-4}; for the nonpoor, and for the poor and nonpoor combined, E_{0-7}.

d Excludes respondents for whom information on one or more variables was not ascertained.

Table 3 Average Rate of Pay ($per Hour): Estimated Regression Coefficients for Women 30 to 44 Years of Age Employed as Wage or Salary Workers at Time of Survey, 1967[a]

(Standard errors in parentheses)

Variables and statistics	Blacks and whites			Blacks			Whites		
	Poor and nonpoor	Poor	Nonpoor	Poor and nonpoor	Poor	Nonpoor	Poor and nonpoor	Poor	Nonpoor
S (1 = 25,000+)	.34 (.04)*	.28 (.06)*	.30 (.05)*	.58 (.09)*	.33 (.07)*	.70 (.15)*	.23 (.05)*	.10 (.15)	.23 (.05)*
M (1 = married)	.11 (.05)*	-.04 (.06)	.02 (.07)	.14 (.08)	-.07 (.07)	.01 (.14)	.06 (.07)	.04 (.14)	.01 (.07)
Rc (1 = black)	-.33 (.05)*	-.37 (.07)*	-.20 (.06)*						
H (1 = health limits kind of work)	-.07 (.07)	-.06 (.09)	-.07 (.08)	-.11 (.13)	-.03 (.11)	-.21 (.20)	-.04 (.08)	-.15 (.19)	-.01 (.09)
T (years)	.081 (.016)*	-.036 (.023)	.089 (.019)*	.068 (.028)*	-.039 (.024)	.108 (.043)*	.086 (.020)*	-.052 (.058)	.087 (.021)*
W (percentage of years)	.004 (.001)*	.001 (.001)	.005 (.001)*	.003 (.001)*	.000+ (.001)	.006 (.002)*	.005 (.001)*	.004 (.003)	.005 (.001)*
E_{0-7} (1 = 0-7 years)		-.52 (.09)			-.45 (.10)*			-.55 (.22)*	
E_{0-8} (1 = 0-8 years)	-1.15 (.07)*		-1.00 (.09)*	-1.51 (.13)*		-1.08 (.19)*	-.94 (.09)*		-.95 (.10)*
E_8 (1 = 8 years)		-.44 (.11)*			-.36 (.12)*			-.31 (.31)	
E_{9-11} (1 = 9-11 years)	-.95 (.07)*	-.22 (.08)*	-.89 (.07)*	-1.39 (.12)*	-.14 (.09)	-1.25 (.16)*	-.75 (.08)*	-.38 (.15)*	-.74 (.08)*
E_{12} (1 = 12 years)	-.65 (.06)*		-.63 (.06)*	-1.02 (.12)*		-.86 (.15)*	-.52 (.07)*		-.55 (.07)*
Constant term	2.00 (.09)*	1.70 (.09)	2.02 (.11)*	1.96 (.18)*	1.31 (.11)*	1.55 (.28)*	1.93 (.11)*	1.71 (.19)*	1.99 (.12)*
# of observations	1,352	236	1,115	431	166	266	920	69	850
$\bar{R}^2$	.31	.34	.24	.41	.29	.33	.22	.11	.21
F	67.13*	14.27*	40.08*	38.86*	9.57*	17.20*	32.98*	2.01	28.71*
<u>Dependent variable</u>:									
Mean	$1.95	$1.26	$2.09	$1.71	$1.13	$2.07	$2.06	$1.58	$2.10
S.D.	$0.90	$0.56	$0.90	$1.02	$0.49	$1.10	$0.82	$0.57	$0.83

* Significant at .05 level.

\+ Rounded to nearest tenth of a cent.

a Restricted to the "definitely poor" and "nonpoor" living in families with at least one child; see Section 4 for a definition of poverty status.

b The omitted category for the poor is E_{12+}; for the nonpoor, and for the poor and nonpoor combined, E_{13+}.

c Excludes respondents for whom information on one or more variables was not ascertained.

MEASURING THE QUALITY OF HOUSING

Ko Ching Shih, U.S. Department of Housing & Urban Development*

I. BACKGROUND

In the decennial censuses of 1940, 1950 and 1960, the Bureau of the Census had collected information on the structural condition of housing by direct observation and overall rating. A post-census evaluation study (16) in 1967, however, resulted in the rejection of much of this data on statistical grounds; the collection of this type of data was discontinued starting with the 1970 Census (17).

In 1968, the National Commission on Urban Problems found that both the definitions and the supporting data relating to substandard housing in most Federal urban programs were inadequate and in many cases inconsistent. The Commission recommended the following definition of a substandard unit:

> "... any dwelling unit in which there is a substantial departure from accepted minimum housing code provisions..." (12).

Various U.S. Housing Acts have required local communities who undertook Federal urban programs to provide decent, safe and sanitary housing (18,20).

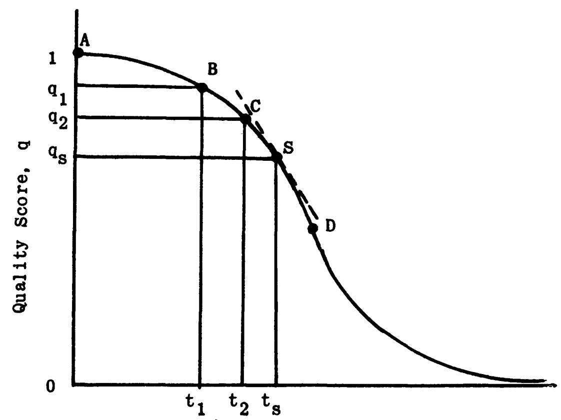
A study of any local housing code would show that there are actually many elements and variables involved in a quality measurement. Not only many variables of the housing structure must be considered, but also the number of residents, and neighborhood environmental factors. Because of the complex nature of such measurements, the task would be simplified if data collection facilities already in existence in a community could be utilized.

It is the purpose of this paper to present some theoretical concepts and the appropriate methods and procedures for measuring the quality of housing as a by-product of a local community's established housing inspection program.

II. THE HYPOTHESIS - A MICRO-MODEL

A micro-model of the deterioration process of a housing structure is shown in Figure 1:

Figure 1 A Micro-Model of the Deterioration Process of a Housing Structure with m Constant



With all other variables constant, a quality score q can be defined as a function of age t and maintenance m, i.e.

$$q = f(t,m) \qquad (1)$$

where q ranges from zero to one.

For new structures with all aspects in good order, q at t = 0 is one, or perfect.

At a point B after a time t_1, assuming a constant level of maintenance, the quality score q has declined to q_1 or $1 - \Delta q_1$, where Δq_1 is a penalty score. This could be called the point of minor deficiency. Similarly, C could be called the point of major deficiency, D the point of critical deficiency, and so on.

Point S could be set as the substandard point, that is, the critical point in terms of housing code. Then all housing units with quality scores falling below S would be classified as substandard.

Looking again at the micro-model curve, it can be seen that a tangent at point S is easily drawn. That is to say, if time-series data were collected at pre-determined intervals, the rate of change of the quality coefficient at S can be estimated by taking the partial differential of f with respect to t, i.e.

$$r = \frac{\partial f}{\partial t} \qquad (2)$$

* The methods and precedures presented here are not necessarily official interpretations of regulations of the U.S. Department of Housing & Urban Development. The author is solely responsible for any errors made.

r is truly the rate of substandardization, the critical statistic for the estimation of the level of substandard housing in a given stratum of a community.

In practice, the quality score at the substandard point S is computed by means of the following equation:

$$q_S = 1 - \sum_{i=1}^{S} \Delta q_i \qquad (3)$$

The quality score can be readily converted to a quality coefficient which can be interpreted flexibly and meaningfully.

Coefficients for crowdedness, environment, transportation, and other factors can also be computed with appropriate variables.

III. A MACRO-MODEL

Let us employ the following notation:

- j stratum number
- V aggregate total housing inventory $= V' + V''$
- V' aggregate standard inventory
- V'' aggregate substandard inventory
- N number of new housing units completed or remodeled
- S number of units with deficiencies rated as slight
- M number of units with deficiencies rated as minor
- J aggregate number of units with deficiencies rated as major $= J' + J''$
- J' number of units with deficiencies rated as major, but above the substandard level
- J'' number of units with deficiencies rated as major that are below the substandard level
- C number of units with deficiencies rated as critical
- D number of units scheduled to be demolished or in the process of demolition

The macro-model can then be expressed by the following equations:

$$V_{ij} = \sum_j [V'_{ij} + V''_{ij}] \qquad (4)$$

or

$$V_{ij} = \sum_j [(N_{ij} + S_{ij} + M_{ij} + J'_{ij}) + (J''_{ij} + C_{ij} + D_{ij})] \qquad (5)$$

or

$$V_{ij} = \sum_j [N_{ij} + V_{(i-1)j} - D_{(i-1)j} + D_{ij}] \qquad (6)$$

And

$$V_{(i-1)j} = \sum_j [N_{(i-1)j} + S_{(i-1)j} + M_{(i-1)j} + J'_{(i-1)j}] + [J''_{(i-1)j} + C_{(i-1)j} + D_{(i-1)j}] \qquad (7)$$

$V'_{ij} = \sum_j [N_{ij} + S_{ij} + M_{ij} + J'_{ij}]$, but is not necessarily equal to

$$\sum_j [N_{ij} + N_{(i-1)j} + S_{(i-1)j} + M_{(i-1)j} + J'_{(i-1)j}],$$

because the coefficient of each individual housing unit depends greatly on the maintenance efforts. In other words,

$$M_{(i-1)j} \neq M_{ij} \text{ and } J'_{(i-1)j} \neq J_{ij}.$$

For this reason,

$$V''_{ij} = \sum_j [V'_{(i-1)j} - D_{(i-1)j}] + rV'_{(i-1)j} \qquad (8)$$

$rV'_{(i-1)j}$ is the portion of $V'_{(i-1)j}$ which fell below the substandard level during the time period between (i-1) and i.

Since $V'_{(i-1)j} = \sum_j [N_{(i-1)j} + M_{(i-1)j} + J'_{(i-1)j}]$,

$$\text{thus } rV'_{(i-1)j} = r \sum_j [N_{(i-1)j} + M_{(i-1)j} + J'_{(i-1)j}] \qquad (9)$$

In the short run, $rN_{(i-1)j} \simeq N_{ij}$, i.e. it is unlikely any unit of $N_{(i-1)j}$ became substandard during a relatively short period of time; it is also true that $rM_{(i-1)j} \simeq M_{ij}$.

In other words, it is likely that most new substandard units came from $J'_{(i-1)j}$.

Approximately,

$$V''_{ij} = V''_{(i-1)j} - D_{(i-1)j} + rJ'_{(i-1)j} \qquad (10)$$

Thus, $J'_{(i-1)j}$ is the critical stratum in estimating the parameters of the current substandard housing inventory (V''_{ij}), and r is a critical estimator.

IV. PROCEDURES

Based on the Housing Code of the City of Rock Island, Illinois (14), quality conditions of a housing structure are classified as either sound, minor, major, or critical. Definitions of each of the four conditions together with the codes and penalty score weightings are shown in Table 1.

Table 1 Definition, Code and Penalty Score Weighting, Rock Island, Ill.

CODE	QUALITY CONDITION	PENALTY SCORE WEIGHTING	DEFINITION
1	SOUND	0	ELEMENT SOUND, NO REPAIRS NEEDED
2	MINOR	1	TO A MINOR DEGREE, DEFECTS THAT CAN ORDINARILY BE CORRECTED IN THE COURSE OF NORMAL MAINTENANCE.
3	MAJOR	4	TO A MAJOR DEGREE, DEFECTS THAT REQUIRE SUBSTANTIAL REPAIR OR REPLACEMENT BUT WHICH ARE NOT SERIOUS STRUCTURAL FAILURES.
4	CRITICAL	16	TO A CRITICAL DEGREE, DEFECTS OF THE PRIMARY STRUCTURAL COMPONENTS ONLY AND ARE SERIOUS STRUCTURAL FAILURES.

Twenty-one variables in five structural elements (primary components, secondary components, related components, system components and availability of plumbing facilities) were selected for computing the housing structural quality coefficient q. Two variables (persons per room and persons per block) were selected for computing the crowdedness coefficient c, and modified quality coefficient q'. Fifteen variables in five environmental factors (space, street conditions, utilities, atmospheric conditions and general conditions) were selected for computing the environmental coefficient v and modified quality coefficient q''.

Table 2 shows the classification of quality rating variables and penalty scores for computing quality coefficients q, q' and q''.

Table 3 shows the overall quality rating of a housing structure together with the range of penalty scores and structural quality coefficients q.

Table 3 Overall Quality Rating of A Housing Structure, Rock Island, Ill.

OVERALL QUALITY RATING	RANGE OF PENALTY SCORES p	RANGE OF QUALITY COEFFICIENT q
N	0	1.00
S	0 - 14	1.00 - 0.90
M	14 - 22	0.90 - 0.85
J	22 - 46	0.85 - 0.68
J'	22 - 32	0.85 - 0.78
J''	32 - 46	0.78 - 0.68
C	46 - 72	0.68 - 0.50
D	72 - 144	0.50 - 0.00

The computing equation for q, using the simplified questionnaire, is

$$q = 1 - \frac{1}{144}\left[\sum_{k=1}^{5} A_k + \sum_{k=6}^{21} B_k\right] \qquad (11)$$

where 144 is the maximum penalty score, A_k is one of the primary components, and B_k is one of the other components as listed in Table 2.

The first modified quality coefficient q' is simply the average of q and the crowdedness coefficient c. The second modified quality coefficient q'' is found by averaging in v, the environmental coefficient given by:

$$v_j = 1 - \frac{1}{56}\sum_{i=1}^{14} v_i \qquad (12)$$

where 56 is the maximum penalty score.

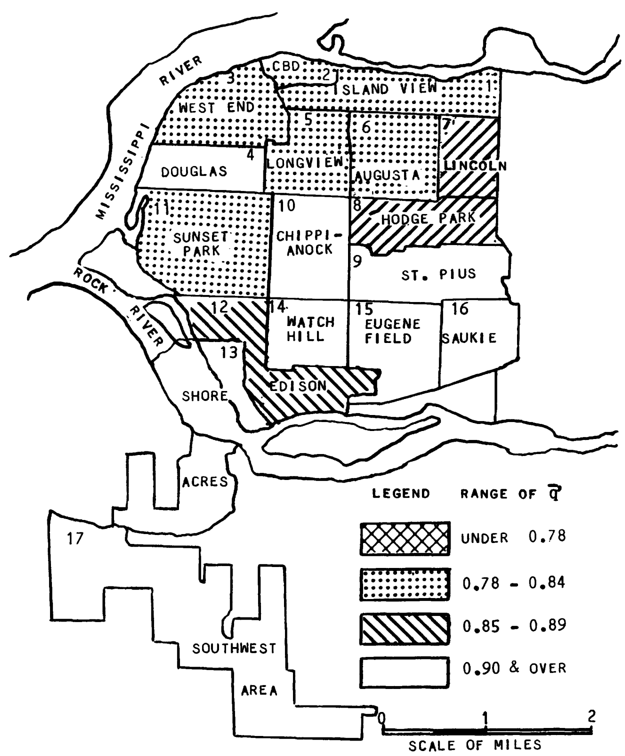
By means of a series of one-page precoded questionnaires and a low-cost, simple to operate portable data terminal - the IBM Information Recorder, all of the measurable variables can be logically recorded during the course of a housing inspector's daily routine. The inspectors are highly trained, and they can observe and record the appropriate scores for each variable according to predetermined standards.

V. EMPIRICAL RESULTS

These are some of the highlights of the empirical results:

(1) According to the urban growth theory, many American cities have grown concentrically or on a neighborhood by neighborhood basis. The origin is usually called the central business district, and growth spreads out to the peripheries - the new neighborhoods. Statistically, these neighborhoods are ideal strata. Figure 2 shows the distribution of modified quality coefficient $\bar{q}'$ by neighborhood.

Figure 2 Distribution of Modified Quality Coefficient $\bar{q}'$ by Neighborhood Rock Island, Ill.



In conducting housing quality surveys, stratification of sampling units by neighborhood is essential in order to allocate the samples efficiently. These statistics are certainly useful for code enforcement and urban renewal programs.

(2) The survey also shows that $\bar{q}'$ of owner-occupied housing (0.91) is significantly higher than for renter-occupied units (0.82) and low rent housing (0.74), and that the variances of the quality coefficients and their distribution patterns are valuable indicators when used in the decision-making process of housing planning.

(3) The elementary sampling units (ESU's) of the environmental survey are block faces. In conjunction with the transportation study which the city is planning to conduct in the near future and a law enforcement study which is in progress, the end results of the quality measurement can be more effectively utilized and expanded.

(4) In the June 1970 JASA, Kain and Quigley concluded that "the quality of a bundle of residential services has at least as much effect on its price as such quantitative aspects as number of rooms, number of bathrooms, and lot size. . ." (7). In actual practice in the determination of the market price of a housing structure, the element of quality is one of the independent variables (16). Specifically,

$$P_k = f\ (q_k, C_k) \qquad (13)$$

where P_k is the price of house k;

q_k is the quality coefficient of house k;

C_k are the characteristic variables of house K.

C_k is given by a regression equation developed by Musgrave (11).

Hence, Musgrave's equation may be modified as:

$$P_k = q_k \left[b_o + b_1C_{1k} + b_2C_{2k} + \ldots + b_{21}C_{21k} + e_k \right] \qquad (k=1,\ 2,\ \ldots\ n) \qquad (14)$$

or:

$$P_k = q_k \left[b_o + \sum_{j=1}^{21} b_kC_{jk} + e_k \right] \qquad (k=1,\ 2,\ 3,\ \ldots\ n) \qquad (15)$$

where b_o is the constant term in the regression; b_1, b_2,... b_{21} are the regression coefficients corresponding to C_{jk} = 1 if house k is in category j and = 0 otherwise; e_k is the "error" or "residual" term in the regression equation. If a house is brand new, q=1, thus,

$$P_k = b_o + \sum_{k=1}^{n} b_kC_{jk} + e_k \qquad (16)$$

The quality coefficient of each characteristic variable may vary significantly. Thus, equation (14) many be written as

$$P_k = b_o + q_1b_1C_{1k} + q_2b_2C_{2k} + \ldots + q_{21}b_{21}C_{21k} + e_k \qquad (17)$$

or

$$P_k = b_o + \sum_{j=1}^{21} q_jb_jC_{jk} + e_k \qquad (k=1,\ 2,\ 3,\ \ldots n) \qquad (18)$$

If a house is occupied or was previously occupied, its quality coefficient q will be a significant independent variable for its price determination.

(5) By means of a Total Housing Information System (THIS), Rock Island, Illinois, currently maintains a nearly perfect sampling frame of PSU's in terms of local real property units. Thus, housing inspection samples are ideally stratified and randomly selected. One stratum consists of non-residential structures, which, however, do contain living quarters, for example, housing units above a grocery store. This stratum is the missing inventory component that usually does not show up in many reports. The quality coefficients of this missing component, in combination with other statistics such as vacancy rates, are significant values for urban relocation applications.

VI. CONCLUSIONS

(1) For the time being, it is not feasible to measure the quality of housing on a national scale. There are several reasons: 1) Not many communities are ready to operate a continuous program of this kind. 2) The enumerators must be trained housing inspectors. 3) Many local housing codes are far from standardized. 4) Local needs and capabilities vary too greatly. 5) Too often politics is involved.

(2) For the purposes of maximum benefit and generality, measuring the quality of housing must be a basic subsystem of an integrated total housing inventory system. In other words, the project should be dynamically operated on a long-term basis.

(3) For HUD and the Census Bureau, it is ideal to select several communities of various sizes in each region of the nation, and develop various types of housing inventory systems, including housing inspection as a basic subsystem. Thus a relatively small number of samples can be used to determine certain basic variables for the purposes of estimation, projection, and further development.

Table 2 Classification of Quality Rating Variables and Penalty Scores

Quality Coefficient	Target Population	Element Variable	Penalty Scores 1 (0)	2 (1)	3 (4)	4 (16)	Maximum Penalty Scores
q	Housing Structure [PSU]	All Structural Variables	0	21	84	80	144
		a. Primary Components:	0	5	20	80	80
		1. Foundation Walls	0	1	4	16	16
		2. Exterior Walls	0	1	4	16	16
		3. Roof & Roof Structure	0	1	4	16	16
		4. Floor & Floor Structure	0	1	4	16	16
		5. Bearing Walls & Columns	0	1	4	16	16
		b. Secondary Components:	0	6	24	-	24
		6. Nonbearing Walls	0	1	4	-	4
		7. Interior Stairs & Railings	0	1	4	-	4
		8. Porch & Steps	0	1	4	-	4
		9. Windows & Window Units	0	1	4	-	4
		10. Doors & Door Units	0	1	4	-	4
		11. Chimney & Cornices	0	1	4	-	4
		c. Related Components:	0	4	16	-	16
		12. Lighting & Ventilation	0	1	4	-	4
		13. Adequacy of Floor Space	0	1	4	-	4
		14. Entrances & Exits	0	1	4	-	4
		15. Grounds	0	1	4	-	4
		d. System Components:	0	3	12	-	12
		16. Plumbing System	0	1	4	-	4
		17. Electrical System	0	1	4	-	4
		18. Heating System	0	1	4	-	4
	Housing Unit [ESU]	e. Availability of Plumbing Facilities:	0	3	12	-	12
		19. Kitchen Sink	0	1	4	-	4
		20. Flush Toilet	0	1	4	-	4
		21. Bathtub or Shower	0	1	4	-	4
q'	Persons & Households	1. Persons per Room	NA	NA	NA	NA	NA
		2. Persons per Block	NA	NA	NA	NA	NA
q"	Block-face	All Environmental Factors	0	14	42	-	56
		a. Space	0	1	4	-	4
		1. Open Space	0	1	4	-	4
		b. Street Conditions:	0	5	20	-	20
		2. Street Pavement	0	1	4	-	4
		3. Street Width	0	1	4	-	4
		4. Sidewalk	0	1	4	-	4
		5. Street Lighting	0	1	4	-	4
		6. Offstreet Parking	0	1	4	-	4
		c. Utilities:	0	3	12	-	12
		7. Water Supply	0	1	4	-	4
		8. Sewage Disposal	0	1	4	-	4
		9. Drainage	0	1	4	-	4
		d. Atmospheric Conditions:	0	3	12	-	12
		10. Noise	0	1	4	-	4
		11. Air Pollution	0	1	4	-	4
		12. Odor	0	1	4	-	4
		e. General Conditions:	0	2	8	-	8
		13. Safety	0	1	4	-	4
		14. General Conditions	0	1	4	-	4

PENALTY SCORE CODES: HOUSING STRUCTURE - 1 - SOUND, 2 - MINOR DEFECT, 3 - MAJOR DEFECT, 4 - CRITICAL DEFECT, PLUMBING FACILITIES - 1 - COMPLETE, 2 - PARTIAL, 3 - NONE. ENVIRONMENTAL FACTORS - 1 - ADEQUATE OR NORMAL, 2 - MARGINAL OR ACCEPTABLE,, 3 - NEGLECTED OR UNACCEPTABLE. NA - NOT APPLICABLE

REFERENCES

[1] Adelman, Irma and Griliches, Zvi, "On an Index of Quality Changes," *Journal of the American Statistical Association*, 56 (1961), 535-48.

[2] American Public Health Association, *An Appraisal Method for Measuring the Quality of Housing: A Yard-stick for Health Officials and Planners*, New York: 1945. Also Twichell, Allan A., "An Appraisal Method for Measuring the Quality of Housing," *The American Sociological Review*, 13 (1948), reprinted in *Urban Housing*, William L. C. Wheaton, Grace Milgram, and Margy E. Meyerson (editors), New York: The Free Press, 1966.

[3] Committee on Banking and Currency, House of Representatives, 90th Congress, 1st Session, *Basic Laws and Authorities on Housing and Urban Development*, Washington, D. C.: U. S. Government Printing Office, 1967.

[4] Hoyt, Homer, "The Structure and Growth of Residential Neighborhoods in American Cities," Federal Housing Administration, 1939, reprinted in *Urban Housing*, William L. C. Wheaton, Grace Milgram, and Margy E. Meyerson (editors), New York: The Free Press, 1966.

[5] Intergovernmental Task Force on Information Systems, *The Dynamics of Information Flow: Recommendations to Improve the Flow of Information with and among Federal, State, and Local Governments*, Washington, D. C., 1968.

[6] Johnson, Ralph J., Williams, Huntington, and McCaldin, Roy O., "The Quality of Housing 'Before' and 'After' Rehabilitation," *The American Journal of Public Health*, 45 (1945), reprinted in *Urban Housing*, William L. C. Wheaton, Grace Milgram, and Margy E. Meyerson (editors), New York: The Free Press, 1966.

[7] Kain, John F. and Quigley, John M., "Measuring the Value of Housing Quality," *Journal of the American Statistical Association*, 64 (1970), 532-48.

[8] Kristof, Frank S., *Urban Housing Needs through the 1980's: Analysis and Projection, Measuring Substandard and Overcrowded Housing*, National Commission on Urban Problems, Research Report No. 10, Washington D. C.: U. S. Government Printing Office, 1968.

[9] Manvel, Allen D., *Local Land and Building Regulation*, National Commission on Urban Problems, Research Report No. 6, Washington, D. C.: U.S. Government Printing Office, 1969.

[10] Manvel, Allen D., *Housing Conditions in Urban Poverty Areas*, National Commission on Urban Problems, Research Report No. 9, Washington, D. C.: U.S. Government Printing Office, 1969.

[11] Musgrave, John C., "Measurement of Price Changes in Construction," *Journal of the American Statistical Association*, 64 (1969), 771-86.

[12] National Commission on Urban Problems, *Building the American City*, House Document No. 91-34, 91st Congress, 1st Session, Washington, D. C.: U.S. Government Printing Office, 1968.

[13] Ratcliff, Richard U., "Housing Standards and Housing Research," *Land Economics*, 28 (1952), reprinted in *Urban Housing*, William L. C. Wheaton, Grace Milgram, and Margy E. Meyerson (editors), New York: The Free Press, 1966.

[14] City of Rock Island, Illinois, *Housing Code*, Rock Island, Ill.: 1970.

[15] Rodwin, Lloyd, "The Theory of Residential Growth and Structure," *The Appraisal Journal*, 18 (1950), reprinted in *Urban Housing*, William L. C. Wheaton, Grace Milgram, and Margy E. Meyerson (editors), New York: The Free Press, 1966.

[16] U. S. Bureau of the Census, *Measuring the Quality of Housing: An Appraisal of Census Statistics*, Working Paper No. 25, Washington, D. C.: U. S. Bureau of the Census, 1967.

[17] U. S. Bureau of the Census, "Proposed Procedure for Estimating Substandard Housing in 1970," (preliminary), unpublished staff paper, July, 1970.

[18] U. S. Department of Housing and Urban Development, *Determining Local Relocation Standards*, Technical Guide No. 9, Urban Renewal Administration, Washington, D. C.: U. S. Department of Housing and Urban Development, 1961.

[19] U. S. Department of Housing and Urban Development, *Urban & Regional Information Systems: Support for Planning in Metropolitan Areas*, Washington, D. C.: U. S. Government Printing Office, 1958.

[20] U. S. 91st Congress, 2nd Session, *Uniform Relocation Assistance and Real Property Acquisition Policies Act of 1970*, House Report No. 91-1656, Washington, D. C.: U. S. Government Printing Office, 1970.

ALTERNATIVE MODELS OF INDIVIDUAL MOBILITY

Alden Speare, Jr., Population Studies and Training Center, Brown University

The decision to move is usually a complex one which involves several factors and is influenced by various characteristics of the mover and his place of residence. Any model which attempts to represent this decision making process must involve several variables. A simple type of model is one in which mobility, represented by a 0,1 variable, is viewed as a linear function of independent variables such as age, education, occupation, home ownership and duration of previous residence (see [3] and [4]). In an earlier paper [8], I tried to expand this model to include intervening variables such as residential satisfaction and a previously expressed desire to move which in turn depend on individual and residence characteristics.

The parameters of these models can be determined through ordinary multiple regression by defining the mobility variable as equal to 1 for movers and 0 for non-movers. For any subgroup of the population the mean value of the mobility variable lies between 0 and 1 and can be interpreted either as the proportion of movers in that subgroup or the probability that an individual in that subgroup will be a mover. The unstandardized regression coefficients which are obtained from the multiple regression can be interpreted as the relative contributions of the independent variables to mobility if the effects can be assumed to be linear and additive. This interpretation is especially clear in the case where all independent variables are discrete variables (for a discussion of the use of discrete variables in regression see Suits, [9] and Goldberger, [2], pp. 218-227). The expected probability that a person with given characteristics will move is simply the sum of the regression coefficients that correspond to the characteristics he possesses plus the constant:

$$\hat{y}_i = a + \Sigma\, b_k\, X_{ki}$$

where $X_{ki} = 1$ if he possesses the K^{th} characteristic and 0 otherwise. The ability to divide up the probability of moving into separate components which at least crudely represent the independent effects of different variables makes multiple regression an attractive technique for the study of individual mobility. However, the use of a dichotomous dependent variable presents at least two problems.

First, the assumption of homoscedasticity is violated since the mobility variable has a binomial distribution with asymtotic variance equal to PQ. Goldberger ([2], pp. 249-250) has suggested that this problem can be solved by introducing weights inversely proportional to the variance or $(PQ)^{-1}$. Unfortunately the true value of P is unknown and must be estimated from the observed value. When the numbers in some subgroups of the sample are small, the estimates are subject to considerable error. This is especially a problem when P is near 0 or 1 for the weight gets very large and small deviations of the observed P from the true P can have large effects on the weight used.

Another problem with this model is that for some combinations of values of the independent variables, the expected value, $\hat{y}_i$, may be either greater than unity or less than zero. Anyone accustomed to evaluating the validity of a model by its behavior at extremes would be inclined to reject this model since an event cannot have a negative probability of occurrence or a probability greater than unity. Others who are willing to mentally make the conversion of a negative probability to a zero probability might still raise the objection that to use $(y_i-\hat{y}_i)^2$ provides an unnecessary addition to the variance when $\hat{y}_i$ is less than 0 or greater than 1.

Several alternative models have been suggested for dealing with this problem. One suggestion is to use a logit transformation to limit the expected value of the dependent variable to the 0 to 1 interval (see figure 1). Thiel [10] has discussed this model at length for the case where the independent variables take on a limited number of discrete values. For 3 independent variables the proportion of cases with y_i=1 for each subgroup defined by a combination of values on the independent variables, f_{jkl}, is estimated by P_{jkl}, where

$$P_{jkl} = (1+e^{-L_{jkl}})^{-1}$$

and $$L_{jkl} = a + b_1 X_{1j} + b_2 X_{2k} + b_3 X_{3l}$$

P_{jkl} equals .5 when $L_{jkl} = 0$, equals 0 when $L_{jkl} = -\infty$ and 1 when $L_{jkl} = +\infty$. The regression coefficients can be determined approximately by calculating:

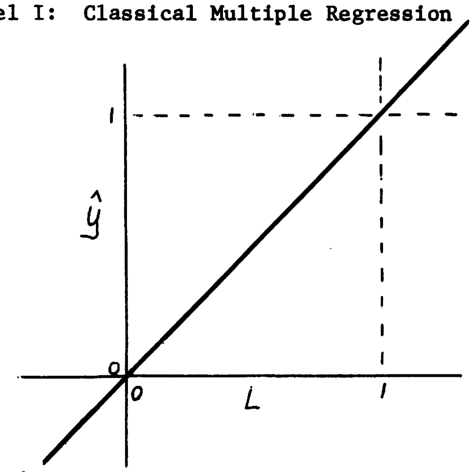
$$L_{jkl} = \ln \frac{f_{jkl}}{1-f_{jkl}}$$

and regressing this on the independent variables. Unfortunately L_{jkl} is undefined when f_{jkl} equals either 0 or 1. Thiel recommends dropping these cases from the analysis by giving them a weight equal to 0. Other procedures, such as changing f_{jkl} to what it would be if a half a case had the opposite value, are discussed by Gart and Zweifel [17]. The fact that none of these procedures provide an unbiased treatment of these cases is a weakness of this approach.

A third model is a simple modification of the classical multiple regression equation in which $\hat{y}$ is truncated at 0 and 1 (see figure 1). This can be accomplished in the computation stage by the instructions:

Figure 1
Three Alternative Multiple Regression Models

Model I: Classical Multiple Regression

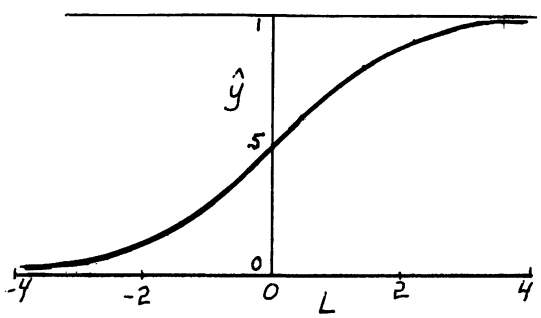


$\hat{y} = L_{jkl}$

$L_{jkl} = a + b_1 X_{1j} + b_2 X_{2k} + b_3 X_{3l}$

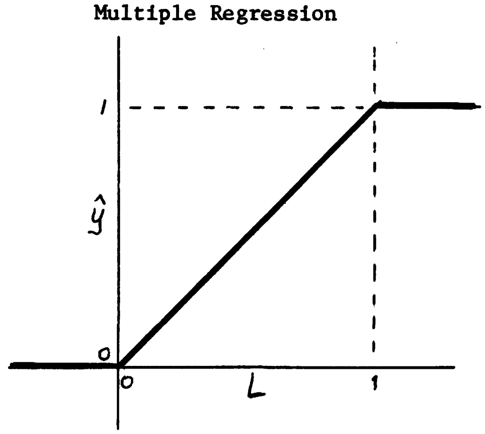
(same for all models)

Model II: Regression of Logit



$\hat{y} = (1 + e^{-L_{jkl}})^{-1}$

Model III: Truncated Classical Multiple Regression



$\hat{y} = 0$ for $L_{jkl} < 0$

$\hat{y} = L_{jkl}$ for $0 \leq L_{jkl} \leq 1$

$\hat{y} = 1$ for $L_{jkl} > 1$

if $\hat{y} < 0$, set $\hat{y} = 0$

if $\hat{y} > 1$, set $\hat{y} = 1$

Since y no longer has a continuous first derivative, the least-squares equations can not be solved directly. However, a solution can be obtained through iteration.

The choice of a model should depend on at least three criteria: (1) the ease of performing the computations involved; (2) the goodness of fit to the data, and (3) the appropriateness of the model to the theoretical assumptions about the relationships of the variables. While some complex model might provide the best goodness of fit, a simpler model might be preferable if the fit were almost as good and the model seemed more clearly in line with the theoretical assumptions.

All three of the models described above seem to be appropriate for the study of individual mobility. The classical multiple regression equation states that the probability of moving is simply a weighted sum of the values of the independent variables. This model is simple and seems reasonable in the lack of information indicating more complex relationships. The truncated multiple regression equation is similar with the exception that once the sum of factors is sufficient to predict either mobility or immobility (i.e., $\hat{y} = 1.0$ or $\hat{y} = 0$) the addition of other factors favorable to mobility (or immobility) makes no difference. The logit model is very similar to the linear model near the center. However, as one moves away from the center each additional factor has less effect on the probability of moving. The extremes of 0 and 1 can never be reached which is probably realistic in that there are always some factors which dispose a person to stay or move which are not included in any particular model.

THE DATA

The data come from interviews taken in the 1969 round of the Rhode Island Health Study and a telephone follow-up interview one year later (see [6]). The original survey included 1081 respondents who were representative of the Rhode Island population aged 21 and over and the married population of all ages. A sub-sample of 724 respondents was selected which contained all those who had ever been married, who were under 65 years of age, who were either the head of the household or spouse of the head, and who were not currently serving in the military. The age and marital status restrictions were deemed necessary because of the large variation in mobility rates with these variables (see Speare, [7]). A small number of respondents who were neither the head or spouse of the head were excluded because these persons might not be involved in the decision to move. The military were excluded because it was felt that their movement might not be entirely voluntary.

In the original interview, respondents were asked questions about their characteristics, the characteristics of their residence, and their satisfaction with various aspects of their housing and geographical location. They were also asked whether they had any wish to move or plans to move within the next year.

Approximately one year later, these same respondents were contacted by telephone (or field interview where necessary) and were asked if they had moved. Every effort was made to obtain the follow-up interview and 95 percent of the respondents in our sub-sample were reinterviewed. In cases where the follow-up interview could not be obtained, the interviewers tried to ascertain whether or not the person had moved. Of the 724 respondents who met the criteria for this study, movement was ascertained for 711 (10 persons had died or entered institutions during the year and 3 refused to be reinterviewed.

THE ANALYSIS

A special program called NDIMA was written to perform the multiple regression allowing any of the three models to be selected. The program first tabulated the data in an N-dimensional matrix where each of N-1 independent variables and the dependent variable represents a dimension. The size of each dimension is 2, which restricted the analysis to all dichotomous variables, although variables with 3 or more categories could be handled by dividing them into two or more dichotomous variables. The least square equations were set up from the data matrix. Appropriate transformations and weights were calculated and the equations were solved through matrix inversion to yield estimates for the regression coefficients.

For models one and three, weights directly proportional to the number of cases and indirectly proportional to the estimated variance were used with the exception that when the observed proportion, f_{jkl}, was less than .05 or greater than .95 the variance for f_{jkl} = .05 was used. This kept the weights from getting too large and provided a finite weight for cases where f_{jkl} = 0 or f_{jkl} = 1.0. For model two, a modified form of the logit was used which was defined at f_{jkl} = 0 and f_{jkl} = 1.0 as described in Gart and Zweifel [1], p. 181:

$$L_{jkl} = \ln \frac{R_{jkl} + .5}{T_{jkl} - R_{jkl} + .5}$$

where R_{jkl} = No. of movers in cell
T_{jkl} = Total number in cell

An appropriate weight for this model is:

$$W_{jkl} = \frac{(R_{jkl} + .5)\ (T_{jkl} - R_{jkl} + .5)}{T_{jkl} + 1}$$

which is similar to the weight NPQ suggested by Thiel ([10], p. 109) when N is large, but has less bias when N is small and is defined at f_{jkl} = 0 and f_{jkl} = 1.

After a solution was obtained, new weights were calculated based on the estimated probabilities and the equations were solved again. For model three further iteration was required. A revised sum of squares was calculated by setting the estimated probability equal to 0 whenever it was negative and 1 when greater than 1. The regression coefficients were then successively incremented and decremented by a small amount and a test was made to see if the sum of squares was reduced. This process was repeated until convergence was obtained.

The following independent variables were chosen for the analysis based on previous research with the classical multiple regression model (see [8]). All independent variables were defined as of the original interview in 1969:

1. Age of the head of household: 0 = Ages 18-34; 1 = Ages 35-64.
2. Owner/Renter status: 0 = Owner; 1 = Renter or Other.
3. Duration of Residence: 0 = 0 to 4 years; 1 = 5 or more years.
4. Friends and Relatives Index. An index representing the proportion of one's friends and relatives who live in the immediate neighborhood or the same section of town. 0 = Relatively low proportion of friends and relatives; 1 = Relatively high proportion.
5. Index of Residential Satisfaction. An index made up of the weighted sum of the expressed level of satisfaction with each of eight items dealing with aspects of housing, neighborhood, and residential location. The item weights were proportional to the relative importance attributed to each item by all respondents. 0 = Relatively low satisfaction; 1 = relatively high satisfaction.
6. Wish to Move: Based on response to the question "Do you have any wish to move within the next year?" 0 = No; 1 = Yes.

THE RESULTS

The results of stepwise multiple regression for the three models are shown in Table 1. These results are based on 678 cases for which there was complete information for all of the variables. The order in which the variables are added is arbitrary although it approximates the common procedure of adding those variables which account for the greatest increase in the "explained" sum of squares first. A crude measure of the goodness of fit for each of the models is the coefficient of determination obtained by taking the ratio of explained sum of squares to the total sum of squares based on deviations of each case

TABLE 1

Results of Three Different Multiple Regression Models for Predicting Residential Mobility

	Classical Multiple	Logit Model[a]	Truncated Classical Model
A. With Three Independent Variables			
Constant	.069	-.662	.069
Wish to Move	.221	.419	.221
Age	-.056	-.226	-.056
Owner or Renter	.101	.269	.101
Coef. of Determination[b]	.168	.178	.168
B. With Four Independent Variables			
Constant	.063	-.674	.019
Wish to Move	.219	.373	.259
Satisfaction Index	.021	.121	.046
Age	-.051	-.219	-.146
Owner or Renter	.100	.246	.180
Coef. of Determination[b]	.169	.185	.187
C. With Five Independent Variables			
Constant	.073	-.650	.028
Wish to Move	.220	.390	.260
Satisfaction Index	.031	.119	.056
Age	-.046	-.158	-.096
Owner or Renter	.093	.223	.168
Duration of Residence	-.022	-.139	-.077
Coef. of Determination[b]	.171	.186	.191
D. With Six Independent Variables			
Constant	.079	-.560	.019
Wish to Move	.209	.358	.299
Satisfaction Index	.026	.096	.106
Age	-.050	-.191	-.110
Owner or Renter	.080	.227	.190
Duration of Residence	-.018	-.088	-.068
Proportion of Friends and Relatives in the Neighborhood	.020	-.123	-.110
Coef. of Determination[b]	.171	.192	.196

[a] All Coefficients have been multiplied by .25, the value of $\Delta p/\Delta \cdot L$ at $p = .05$

[b] Adjusted for degrees of freedom

from its expected value. This is equivalent to the correlation ratio discussed by Neter and Maynes [5]. Since any particular case must either be a mover or a non-mover, the deviations are typically large. Using this measure of goodness of fit, the logit and truncated regression models are potentially superior because it is possible to generate expected probabilities near 0 and 1 for many combinations of the independent variables.

The results are generally in agreement with these expectations. The logit model provides a better fit to the observed data for all four runs. The truncated model is the same as the classical model for three variables because none of the expected probabilities fall outside the 0 to 1 interval. However, it is superior to the classical model for four or more variables where some combinations of the independent variables require truncation.

In general, the logit model and the truncated classical model assign larger effects to the independent variables than the classical model does. This is most apparent if one compares the relative size of the last variable to be added in each run. For instance, duration of residence which is added for the five variable run has a regression coefficient of only -.022 for the classical multiple regression, but a coefficient of -.139 for the logit model and -.077 for the truncated classical model.

The three models also differ in the decisions that are made about whether or not to add an additional variable to the model. For the classical model, none of the additions beyond three variables is statistically significant on the basis of an F-test of the increment to the explained variance. On the other hand, all of the additions to the truncated classical model, up to six independent variables, were statistically significant. Additions to the logit model were also statistically significant with the exception of duration of residence.

In summary, both the logit model and the truncated classical model are superior to the classical multiple regression model for the analysis of individual mobility. They both tend to allow for more independent variables in the model and to assign larger effects to these variables. In comparing the two models, the logit model has the advantage of providing a continuous function which can be solved exactly whereas the truncated model requires iteration. On the other hand, the results of the truncated model can be interpreted directly as components of the probability of moving attributable to different independent variables and these may be summed simply subject only to the simple transformation at the extremes. All three models encounter problems when subgroup sizes are small because of the difficulty in estimating the variance for these subgroups which is used to calculate weights. Further work is needed to establish efficient procedures for small samples.

ACKNOWLEDGMENT

This paper was prepared for the Annual Meeting of the American Statistical Association in Fort Collins, Colorado, August 23-26, 1971. The work was supported by United States Public Health Grant HS-00246 from the National Center for Health Services Research and Development. The author is grateful for the assistance of Dr. Harold Organic who directed the field study and Dr. James Sakoda who helped with the computer programming and commented on an earlier version of the paper.

REFERENCES

[1] Gart, John J., and James R. Zweifel, "On the Bias of Various Estimators of the Logit and its Variance with Application to Quantal Bioassay," Biometrika, 54, 1967, pp. 181-187.

[2] Goldberger, Arthur S., Econometric Theory, New York: John Wiley and Sons, Inc., 1964.

[3] Lansing, John B. and Eva Mueller, The Geographic Mobility of Labor, Ann Arbor: Institute for Social Research, The University of Michigan, 1967.

[4] Morrison, Peter A., "Chronic Movers and the Future Redistribution of Population: A Longitudinal Analysis," Demography, Vol. 8, 1971, pp. 171-184.

[5] Neter, John and E. Scott Maynes, "On the Appropriateness of the Correlation Coefficient with a 0,1 Dependent Variable," Journal of the American Statistical Association, Vol. 65, 1970, pp. 501-509.

[6] Organic, Harold N., and Sidney Goldstein, "The Brown University Population Research Laboratory: Its Purposes and Initial Progress," in Irving I. Kessler and Morton L. Levin (editors), The Community as an Epidemiologic Laboratory: A Casebook of Community Studies, Baltimore: The Johns Hopkins Press, 1970.

[7] Speare, Alden, Jr., "Home Ownership, Life Cycle Stage, and Residential Mobility," Demography, Vol. 7, 1970, pp. 449-458.

[8] Speare, Alden, Jr., "Residential Satisfaction as an Intervening Variable in Residential Mobility." Paper presented at the Annual Meeting of the Population Association of America, Washington, D. C., April 22-24, 1971.

[9] Suits, Daniel B., "Use of Dummy Variables in Regression Equations," Journal of the American Statistical Association, Vol. 52, 1957, pp. 548-551.

[10] Thiel, Henri, "On the Estimation of Relationships Involving Qualitative Variables," American Journal of Sociology, Vol. 76, 1970, pp. 103-154.

CHARACTERISTICS OF INCOME NONRESPONDENTS IN THE CURRENT POPULATION SURVEY*

Emmett Spiers, John Coder, and Mitsuo Ono
U.S. Bureau of the Census

Introduction

One of the significant problems encountered in household surveys which depend upon public cooperation to obtain information on social and economic characteristics is that of missing information. Missing information is the result of either noninterviews or nonresponses. Noninterviews are failures either to contact sample units or to obtain their cooperation after they are contacted. Nonresponses are partial or complete failures to obtain information for particular items from respondents.

Income nonresponses have been an important problem in collecting income data from the March Supplement of the Current Population Survey (CPS). In this survey, interviewers ask eight income questions for each person 14 years old or over living in the sample household. A person is designated a nonrespondent if he fails to answer one or more of the eight questions. A family is considered a nonrespondent if any member of the family 14 years old or over does not answer one or more of the eight income questions.

Attempts to reduce nonresponse rates have resulted in lower rates over the past 3 years.[1/] The income improvement program has reduced the family nonresponse rate from 19.0 percent in the March 1969 CPS to 14.3 percent in the March 1970 CPS. In the March 1971 CPS, it was 14.6 percent, about the same as that for the previous year.

In order to maximize the amount of income information available from the March CPS, missing income information is imputed or allocated to nonrespondents using "fully reported" income information obtained from respondents with the same social and economic characteristics as those of nonrespondents. This allocation procedure is necessary since nonrespondents usually have different social and economic characteristics from respondents. If this were not the case, missing income information could be adjusted uniformly to control totals.

The purpose of this paper is to describe the characteristics, such as age, education, occupation, work experience, and family relationship of nonrespondents and to show the effect of the allocation procedure on the income data.

This paper is divided into four parts. The first part outlines major characteristics of income nonrespondents. The second part describes the income allocation method and traces some of the impact of income allocation on the income data. The third part notes the difference in the NA rate when a person for whom the information is being obtained is interviewed instead of interviewing a "proxy" family member. The fourth part summarizes the general findings covered in this paper.

Characteristics of Nonrespondents

The income nonresponse (NA) rates for white families were generally higher than rates for families of Negro and other races. This statement also held true for unrelated individuals. The NA rates for white families and families of Negro and other races were about three and five percentage points higher than those for unrelated individuals, respectively.

The NA rate for men was usually higher (about three percentage points in the March 1971 CPS) than the rate for women. Among males, white persons had higher NA rates than persons of Negro and other races (about one percentage point higher in the March 1971 CPS). Among females, white persons had an NA rate of 9.5 percent compared with 7.7 percent for persons of Negro and other races. The difference in the NA rate by sex can be attributed to several reasons. A larger proportion of men have income. A larger proportion of men are self-employed. Also, a larger proportion of men have property income. It is more difficult to estimate self-employment income and property income than other types of money income.

Overall, an income nonrespondent is more likely to have the following characteristics: White male who is self-employed, a male between the ages of 45 and 64, a person who completed 12 or more years of school, a year-round full-time worker, or a male who lives in a large metropolitan area with a population of a million or more persons.

Because these characteristics are also those of persons with higher than average incomes, incomes of nonrespondents, after allocation, are usually higher than those for respondents. Thus, the mean income of male nonrespondents (from the March 1970 CPS) was $8,013, about 13 percent higher than the mean for male respondents ($7,087). However, since male nonrespondents represented only 12 percent of all male persons 14 years old and over, the overall mean income ($7,202) of male persons was only about 2 percent higher than the income level for male respondents.

The NA rate for self-employed workers was about twice that for wage and salary workers. As shown in table A, the NA rate for self-employed workers not in agriculture was 25 percent as compared with 12 percent for wage and salary workers in the same classification. Table A also shows the NA rate by type of earnings. This information shows that the NA rate for self-employment income for nonagricultural self-employed workers was about 19 percent. On the other hand, the NA rate for wage and salary income for nonagricultural wage and salary workers was about 10 percent. It is clear from these data that reporting of self-employment income presents more problems than reporting for wage and salary income. This is

* Statistical assistance was provided by Robert W. Cleveland, Consumer Income Statistics Branch, Population Division.

understandable since self-employment income is more difficult to account for than wage or salary income, which is regularly received and recorded for tax accounting purposes. Furthermore, self-employed workers are more independent and less likely to release personal income data for business reasons.

Table A shows that the professional and managerial self-employed workers in nonagricultural industries not only had higher NA rates, but also higher mean earnings than other self-employed workers in the same industrial classification. Thus, the NA rate for the professional and managerial workers was 28 percent as compared with the NA rate of 19 percent for other self-employed workers. The mean earnings value of professional and managerial self-employed workers ($11,478) was almost double the mean earnings value of other self-employed workers ($5,527.)2/ The NA rate for workers in agricultural industries was lower than the NA rate for workers in nonagricultural industries.

The NA rates among occupation groups varied from a high of 19 percent for managerial workers to 8 percent for farm laborers.3/(See table B.) When self-employed workers were excluded, the NA rate for managerial workers dropped to 16 percent. The overall NA rate for salaried professional workers was 12 percent and it varied from 22 percent for physicians and surgeons to 10 percent for primary and secondary school teachers. The only other occupation having an NA rate above the overall average for male workers (13 percent) was for sales workers (15 percent).

Other variables associated with income NA rates were age and education and residence.

The NA rate for men in different age categories varied from 7 percent for men 14 to 19 years old to about 16 percent for men between 45 and 64 years old. (See table C-1.)

For men 25 years old and over, the NA rate by years of school completed varied from 10 percent for men with less than 8 years of school, 14 percent for men who completed high school only, to 16 percent for men who completed college. (See table C-1.)

The NA rate also varied with age when the educational level was kept constant. For college graduates, the NA rate rose monotonically from 10 percent for men age 25 to 34 years to 26 percent for men 65 years old and over (presumably associated with the high NA rates for self-employed professional and managerial workers). For high school graduates, the NA rate peaked at age 55 to 64 years and for elementary school graduates, the NA rate peaked at 45 to 54 years old.

The proportion of year-round full-time workers was disproportionately greater for nonrespondents who were high school graduates and college graduates in the age groups having the peak NA rates compared to respondents in the corresponding groups.4/ Thus, for nonrespondents who were high school graduates, 55 to 64 years old, 87 percent were year-round full-time workers as compared with 77 percent for respondents. Within this age and education grouping, the NA rate for year-round full-time workers was 20 percent compared with 11 percent for other workers. As shown in table C-2, similar results were obtained for college graduates 65 years old and over.

The NA rate also differed by residence and region. Thus, the NA rate for men residing in metropolitan areas was 13 percent as opposed to 10 percent in nonmetropolitan areas. Moreover, the NA rate in larger metropolitan areas (population of 1 million or more) was higher than in small metropolitan areas. The NA rate, however, was not significantly different between males residing inside central cities and outside central cities. The NA rate in the Northeast region was about 15 percent. This compared to about 11 percent in each of the other three Census regions. (Table D.)

Impact of Allocation on Income

The next topic is concerned with the income allocation method and its impact on income data. The income allocation procedure is designed to allocate an amount (a positive or negative dollar amount or "none") for each income item that was unanswered by the respondent. The allocated amount is derived from a reported income value of an income respondent with similar social and economic characteristics who precedes the nonrespondent. That is, the income amount(s) assigned to a nonrespondent is the income amount(s) stored from the last respondent with similar characteristics.

There are eight separate income questions on the March CPS Supplement schedule.5/ These income questions can be grouped into two categories: (1) Earnings and (2) income other than earnings.

The socioeconomic characteristics used in the earnings allocation procedure are: (1) Number of weeks worked last year, (2) occupation of longest job last year, (3) family relationship, (4) sex, (5) age, (6) race (white or Negro and other races), and (7) class of worker. These seven characteristics are combined in various ways to create 235 mutually exclusive classes. A person who failed to answer a particular earnings item is allocated the earnings amount from the last person in that class who reported on all three earnings items.

Two important items considered in creating these 235 classes were: (1) Characteristics for each class should be correlated with earnings, i.e., the classes should be homogeneous, and (2) the number of persons in each class should be large enough to avoid having many persons' earnings allocated from the same person's amount.

Income types other than earnings are allocated based on the following socioeconomic characteristics: (1) Total earnings, (2) family relationship, (3) sex, (4) worker-nonworker status, (5) age, and (6) race (white or Negro and other races). These six characteristics are grouped to form 286 mutually exclusive classes. As in the earnings allocation procedure, nonrespondents in a particular class are allocated amounts from the last person within the class that reported the missing item.6/

In order to analyze the impact of the CPS income allocation procedures, analytical data were tabulated separately for income respondents and income nonrespondents. As indicated earlier, nonrespondents tend to be persons with social and economic characteristics similar to those for income respondents who have higher than average incomes, e.g., college graduates, middle-aged persons, self-employed professional and self-employed managerial workers, or residents of large metropolitan areas. Therefore, the general effect of the allocation procedure is to shift the income size distribution upward and to raise the overall level of income above that of respondents.

Table E shows income size distribution data obtained from the March 1970 CPS for (1) Respondents, (2) nonrespondents, and (3) respondents and nonrespondents combined, by male and female persons.

The effect of allocation on the income size distribution was most noticeable for incomes of $15,000 or more. About 11.6 percent of the male nonrespondents had incomes of $15,000 or more while only 7.4 percent of the respondents had incomes above that amount. The effect of allocating income to the nonrespondents was to increase the percent of males with incomes above $15,000 to 7.9 percent.

The allocation process affects the mean of the income distribution to a greater extent than the median. This is to be expected since the mean is affected more by high income amounts than the median, which is a positional value. For males, the difference between the medians for nonrespondents and respondents was 6.4 percent. The corresponding difference in the mean was 13.1 percent. The effect of allocating income to the nonrespondents, however, was to increase the final median above the reported median by only 0.8 percent and increase the final mean above the reported mean by 1.6 percent.

The percent of persons with allocated income by income intervals is shown in table E. The percent of persons with allocated income (the NA rate) rose sharply for the two highest income intervals. The NA rates were 17 and 22 percent for the $15,000 to $24,999 interval and the $25,000 and over interval, respectively, for male persons.

Family Relationship and Type of Respondent Queried

The problem here is to determine to what extent the NA rate differed by type of respondent queried. Overall, 49 percent of all males and 67 percent of all females were persons who answered for themselves in the March 1970 CPS. As expected, data showed that NA rates were lower for persons responding to the question for themselves (designated as self) as compared to persons for whom questions were answered by another member (designated proxy) of the household. For all males, the NA rate for "self" responses was 7.5 percent and for "proxy" responses, the rate was 14.7 percent. The corresponding rates for females were 5.9 percent and 8.7 percent, respectively.

Family heads who answered for themselves had lower nonresponse rates than for heads whose answers were given by some other member of the family. Approximately 46 percent of all family heads answered for themselves. These heads had a nonresponse rate of 7.5 percent. This compared to a 12.1 percent nonresponse rate for "proxy" responses.

Seventy-seven percent of the wives of family heads answered for themselves in the March 1970 CPS. These "self" respondents had a nonresponse rate of 4.8 percent. The nonresponse rate for "proxy" respondents who answered for the wife of head was 8.5 percent.

Twenty-two percent of other family members responded for themselves. For family members who responded for themselves, the nonresponse rate was 5.2 percent. For family members who did not respond for themselves, the nonresponse rate was 8.3 percent.

Eighty-six percent of all unrelated individuals answered for themselves. The nonresponse rate for these unrelated individuals was 9.4 percent. However, the NA rate for unrelated individuals who did not answer for themselves was 26 percent, the highest of NA rates among the different categories noted above.

These data show that by directly contacting the person for whom information is being obtained, the NA rate could be reduced. It is noted that 46 percent of family heads answered for themselves in the March 1970 CPS, and since the family head is generally the main source of income for the family, direct interviews with family heads may be a way to reduce family income nonresponse rates.

Summary

In summary, findings indicate that income nonrespondents tend to have the following characteristics: White males, self-employed workers, men between the ages of 45 and 64 years, men who completed 12 years of school or more, year-round full-time workers, or men who live in metropolitan areas with populations of a million or more persons. Since these characteristics are associated with persons with higher than average incomes the level of NA rates is directly related with the level of income. Hence, income allocations tend to raise the average level of income and to shift the overall income size distribution upward. Moreover, the impact of the allocations is greater on the mean than on the median. With respect to the income NA rate, it is lower for "self" respondents than for "proxy" respondents.

Table A.--NONRESPONSE RATES AND MEAN EARNINGS IN 1969, OF CIVILIAN MALES 14 YEARS OLD AND OVER BY CLASS OF WORKER, IN THE UNITED STATES

Class of worker	Total (thousands)	Nonresponse rate	Approximate nonresponse rate on major type of income1/	Mean earnings of respondents2/
Total worked last year3/	55,111	12.7	(X)	$7,261
Wage and salary workers	49,137	11.7	9.4	7,146
Self-employed workers	5,974	21.1	15.9	7,446
In agriculture	3,594	11.5	(X)	3,266
Wage and salary workers	1,676	9.1	7.9	2,012
Self-employed workers	1,918	13.7	9.5	4,414
Not in agriculture	51,517	12.8	(X)	7,542
Wage and salary workers	47,461	11.8	9.5	7,430
Self-employed workers	4,056	24.5	18.9	9,082
Professional and managerial	2,544	27.9	(NA)	11,478
Other self-employed workers	1,512	18.8	(NA)	5,527

X Not applicable.
NA Not available.
1/ It is assumed that the major type of earnings is the type consistent with the class of worker.
2/ Mean earnings were used because mean income was not available. The latter would have been preferred because persons are grouped by total income nonresponses rather than earnings nonresponses.
3/ Excludes unpaid family workers.

Table B.--INCOME NONRESPONSE RATES AND MEAN INCOME IN 1969, FOR CIVILIAN MALES 14 YEARS OLD AND OVER BY OCCUPATION OF LONGEST JOB, IN THE UNITED STATES

Occupation	Total (thousands)	Nonresponse rate	Mean income		
			Total	Respondents	Non-respondents
Total worked last year	55,700	12.6	$7,779	$7,654	$8,641
Professional, technical, and kindred workers, total	7,193	14.1	12,047	11,796	13,585
Self-employed, total	690	30.0	21,447	22,045	20,030
Physicians and surgeons	133	33.8	33,563	40,546	(B)
Other self-employed	557	29.1	18,531	17,908	20,064
Salaried, total	6,500	12.4	11,054	10,928	11,948
Engineers, technical	1,207	12.4	13,632	13,700	13,150
Physicians and surgeons	101	21.8	14,661	14,699	(B)
Teachers, primary and secondary	846	9.5	9,857	9,732	11,054
Other salaried workers	4,348	12.7	10,488	10,319	11,649
Farmers and farm managers	1,841	13.4	5,048	5,190	4,125
Managers, officials and proprietors, exc. farm, total	7,053	19.1	12,542	12,526	12,609
Self-employed, total	1,854	27.2	9,425	9,324	9,697
In retail trade	820	28.4	8,713	8,426	9,435
Other self-employed	1,034	26.2	9,991	10,015	9,924
Salaried	5,199	16.3	13,652	13,518	14,341
Clerical and kindred workers	4,068	11.7	6,850	6,881	6,619
Sales workers, total	3,262	15.0	8,132	7,821	9,877
In retail trade	1,467	15.5	5,326	4,898	7,632
Other sales workers	1,795	14.6	10,407	10,164	11,824
Craftsmen, foremen, and kindred workers, total	10,667	11.3	8,094	8,109	7,973
Foremen	1,451	10.3	9,760	9,823	9,214
Craftsmen	9,216	11.4	7,831	7,836	7,796
Operatives and kindred workers	10,842	10.0	6,251	6,260	6,167
Private household workers	110	27.3	1,857	1,546	(B)
Service workers, except private household	3,995	12.4	4,536	4,514	4,689
Farm laborers and foremen	1,805	8.1	1,746	1,638	2,901
Laborers, except farm and mine	4,861	10.3	3,596	3,593	3,619

B Base less than 75,000.

Table C-1.--INCOME NONRESPONSE RATES AND MEAN INCOME IN 1969, FOR MALES 14 YEARS OLD AND OVER BY AGE AND YEARS OF SCHOOL COMPLETED, IN THE UNITED STATES

Characteristic	Total (thousands)	Non-response rate	Mean income		Proportion of civilians with income working year round full time	
			Respondents	Non-respondents	Respondents	Non-respondents
AGE						
Total	69,027	12.0	$ 7,087	$ 8,013	58.3	63.7
14 to 19 years	11,125	7.4	1,051	1,587	6.0	8.1
20 to 24 years	7,067	10.1	4,287	2,828	41.1	37.3
25 to 34 years	12,045	9.6	8,333	8,798	78.7	80.8
35 to 44 years	11,087	12.9	9,995	10,360	84.0	84.5
45 to 54 years	11,081	16.0	9,827	10,113	80.6	81.6
55 to 64 years	8,561	15.4	8,143	9,828	68.5	76.5
65 years and over	8,062	13.1	4,200	5,004	13.5	18.7
YEARS OF SCHOOL COMPLETED 1/						
Total	50,835	13.3	8,330	9,084	68.1	71.2
Elementary, total	14,352	11.3	8,330	9,084	68.1	71.2
Less than 8 years	7,575	10.3	4,064	5,780	41.0	49.4
8 years	6,778	12.5	5,699	6,573	53.0	55.4
High school, total	23,632	13.1	8,262	8,496	75.2	75.4
1 to 3 years	8,171	12.3	7,236	7,586	67.2	65.5
4 years	15,461	13.5	8,811	8,932	79.6	80.1
College, total	12,851	15.7	12,520	12,293	80.0	79.9
1 to 3 years	5,548	14.9	10,351	10,592	78.7	78.4
4 years or more	7,303	16.4	14,200	13,462	81.1	81.0

1/ Limited to persons 25 years old and over.

Table C-2.--INCOME NONRESPONSE RATES AND MEAN INCOME IN 1969, FOR MALES 14 YEARS OLD AND OVER BY AGE AND WORK EXPERIENCE, BY YEARS OF SCHOOL COMPLETED, IN THE UNITED STATES

Characteristic	Total (thousands)	Non-response rate	Mean income		Proportion of civilians with income working year round full time	
			Respondents	Non-respondents	Respondents	Non-respondents
ELEMENTARY SCHOOL GRADUATES BY AGE						
25 to 34 years	572	11.4	$ 6,389	7,710	71.1	67.9
35 to 44 years	1,051	14.6	6,849	6,374	74.7	75.9
45 to 54 years	1,370	12.3	7,146	7,369	76.7	73.3
55 to 64 years	1,665	13.6	6,328	8,179	65.2	67.1
65 years and over	2,120	11.1	3,549	4,226	13.4	13.8
HIGH SCHOOL GRADUATES BY AGE						
25 to 34 years	4,807	9.4	8,148	7,985	83.5	81.9
35 to 44 years	3,848	12.3	9,573	9,724	87.1	84.6
45 to 54 years	3,696	16.7	9,760	9,577	84.7	86.6
55 to 64 years	2,081	18.2	9,094	9,418	76.5	86.7
65 years and over	1,029	16.2	5,218	5,700	19.8	23.1
COLLEGE GRADUATES BY AGE						
25 to 34 years	2,413	9.9	10,946	11,784	79.7	84.9
35 to 44 years	1,934	14.5	15,831	17,460	92.2	95.3
45 to 54 years	1,457	21.3	17,934	13,965	91.4	89.9
55 to 64 years	860	23.0	16,719	13,852	80.6	81.2
65 years and over	640	26.3	10,763	7,735	26.1	34.9

Table D.--NONRESPONSE RATES AND MEAN INCOME IN 1969, FOR MALES 14 YEARS OLD AND OVER, BY RESIDENCE AND REGION, IN THE UNITED STATES

Residence/region	Total (thousands)	NA rate	Mean income		
			Total	Respondents	Nonrespondents
RESIDENCE					
Nonfarm	65,181	12.1	$7,352	$7,236	$8,163
Farm	3,846	10.6	4,658	4,599	5,133
In metropolitan areas	44,820	13.1	7,901	7,829	8,358
1,000,000 or more	24,272	14.0	8,390	8,352	8,612
Less than 1,000,000	20,548	12.1	7,326	7,228	8,010
In central cities	20,108	13.3	7,080	7,007	7,680
Outside central cities	24,712	13.0	8,541	8,484	8,911
Outside metropolitan areas	24,207	9.8	5,911	5,771	7,155
REGION					
Northeast	16,791	15.0	7,536	7,431	8,104
North Central	19,449	11.4	7,561	7,457	8,341
South	20,938	10.7	6,223	6,075	7,398
West	11,849	11.0	7,857	7,796	8,341

Table E.--INCOME RESPONDENTS, NONRESPONDENTS, AND NONRESPONSE RATES BY TOTAL MONEY INCOME IN 1969, BY SEX, IN THE UNITED STATES

Income	Male				Female			
	Total	Respond-ents	Non-respond-ents	Percent with allo-cated income	Total	Respond-ents	Non-respond-ents	Percent with allo-cated income
Total	69,028	60,756	8,272	(X)	76,277	69,639	6,638	(X)
Percent with income	92.5	92.1	95.9	(X)	65.8	64.4	80.5	(X)
Percent with no income	7.5	7.9	4.1	(X)	34.2	35.6	19.5	(X)
Percent with income	100.0	100.0	100.0	(X)	100.0	100.0	100.0	(X)
$1 to $999 or loss	10.9	10.9	10.5	12.1	29.1	29.8	23.2	8.5
$1,000 to $1,999	8.6	8.6	8.3	11.9	19.0	19.0	19.3	10.8
$2,000 to $2,999	7.5	7.5	7.4	12.3	12.0	12.0	12.4	11.0
$3,000 to $3,999	6.6	6.9	5.7	10.6	10.9	10.9	10.9	10.7
$4,000 to $4,999	6.2	6.3	5.4	10.6	8.8	8.9	8.9	10.7
$5,000 to $5,999	7.0	7.1	6.9	12.1	6.9	6.8	8.1	12.4
$6,000 to $6,999	7.6	7.6	7.5	12.4	4.8	4.7	6.2	13.7
$7,000 to $9,999	21.6	21.8	19.8	11.4	5.9	5.6	7.5	13.4
$10,000 to $14,999	16.2	16.1	17.0	13.0	1.9	1.9	2.5	13.9
$15,000 to $24,999	6.1	5.8	8.3	17.0	0.4	0.3	0.6	18.5
$25,000 and over	1.8	1.6	3.3	22.0	0.1	0.1	0.3	23.0
Median	$6,429	$6,378	$6,788	(X)	$2,132	$2,090	$2,554	(X)
Mean	7,202	7,087	8,013	(X)	2,945	2,891	3,402	(X)

X Not applicable.

FOOTNOTES

1/ For a more detailed discussion of the CPS income improvement program, see paper by Mitsuo Ono entitled "Current Developments on Collecting Income Data in the Current Population Survey," presented at the 1971 annual meeting of the American Statistical Association. Nonresponse rates cited below and in the first two paragraphs of the next section were obtained from table 1 of Dr. Ono's paper.

2/ Mean earnings were used because mean income was not available. The latter would have been preferred because persons are grouped by total income nonresponses rather than earnings nonresponses.

3/ Excluding the small number of private household workers.

4/ There are large sampling errors within the age-education-work experience groups. Hence, these estimates are subject to further analysis.

5/ In the March 1970 CPS the eight questions covered the following: Earnings--(1) Money wages or salary; (2) net income from nonfarm self-employment; (3) net income from farm self-employment; Other income--(4) Social Security; (5) dividends, interest, rent, income from estates or trusts; (6) public assistance or welfare payments; (7) unemployment compensation, workmen's compensation, government employee pensions, or veterans' payments; (8) private pensions, annuities, alimony, regular contributions from persons not living in this household, royalties, and other periodic income.

6/ For a more detailed discussion of the income allocation procedures used in the March CPS, see a paper by E. Spiers and J. Knott entitled, "Computer Method to Process Missing Income and Work Experience Information in the Current Population Survey," Proceedings of Social Statistics Section, American Statistical Association, 1969.

WHO ARE THE USERS OF THE 1970 CENSUS SUMMARY TAPES?

Marshall L. Turner, Jr., U. S. Bureau of the Census

Data from the 1970 Census of Population and Housing are now available for every town, village, city, county and State in the country. At this writing, 100-percent census tract data for over three-quarters of the Standard Metropolitan Statistical Areas (SMSA's) are also completed. In addition to the printed census reports, data are being disseminated widely on computer summary tapes. The Census Bureau alone has sold over 1,300 reels of the first count summary information. Approximately 140 different organizations purchased these materials; 28 percent of these purchasing organizations were summary tape processing centers, and as such, represent likely sources for the further proliferation of summary tape copies. Feedback indicates that three of these centers alone have produced over 2,500 copy reels of first count data for their customers. Based on these statistics, a minimum of about 3,800 first count summary tapes have been distributed around the U. S.

With respect to the major types of organizations that have purchased summary tapes from the Bureau, our records identify 38 percent of these groups as State and local government agencies or regional and local planning commissions. Colleges and universities account for another 32 percent of the Bureau's purchasers. About 17 percent were classified as computer and related services groups. The remaining 13 percent was made up of retail organizations, newspapers and magazines, and Federal government agencies.

At this writing, the Bureau has orders from 87 organizations for approximately 6,800 reels of data from the second through the sixth counts. Almost 30 percent of these organizations are classified as State and local government agencies or regional and local planning commissions. Of the remaining organizations that have placed advance orders, universities represent 32 percent; Federal government agencies account for another 8 percent; 10 percent are identified as computer and related services groups; and retail organizations, newspapers and magazines make up the final 20 percent.

EXAMPLES OF USERS OF SUMMARY TAPE DATA

The planning department of one State has embarked upon a program aimed at meeting the needs of many State and local planning groups. Using a widely available software retrieval system, this State planning department is offering courses in the maintenance and utilization of the first count summary tapes to local planners who have access to computers. For users without computer capability, this organization is making available printed publications containing extracted summary tape statistics for each county and county subdivision in the State, and in some cases, maps which graphically portray this information.

In addition to these educational and data dissemination functions, this State planning department is serving as a clearing house for information about census tabulations available from various city and regional planning departments throughout the State. This clearing house effort is being carried out through seminars and newsletters.

As another example, a regional medical and social services program in one State is already making extensive use of the first count summary data in the form of special printed reports. Through the services of a local summary tape processing center, the State-wide staff of this program has been supplied with special publications for each county showing measures of "dependent" population groups based on age distributions; estimates of the incidence of "broken homes" based on marital status and household relationship statistics; and general demographic profiles based on age, race, and sex distributions. Using the second count summary tapes, this summary tape center plans to prepare similar types of statistical measures for each minor civil division and census tract in the State. These small area data will be used to perform a more detailed analysis of the medical and social services needs of the local areas.

From the perspective of the academic users, one large university has provided us with a description of how they are using the first count summary tapes. Utilizing several custom-designed computer programs, this university has developed a file maintenance and retrieval system which allows terminal access to the first count data from various academic departments on the main campus and from terminals located on the campuses of a number of affiliated schools throughout the State.

In terms of substantive work with the first count data, this university has prepared printed reports on social indicators for each county and city and in the State. These reports, which provide measures of housing conditions and demographic characteristics in various areas, are being made available to public libraries and local government agencies.

These data will be used also in sociology and political science courses to be taught in the forthcoming school year. To this end, this social indicator information will be stored "on-line" in the university's computer system. The students will be given assignments concerning the relationships of the demographic characteristics of various areas in the State to the sociological implications of these characteristics. Using a "conversation" mode of communication with the computer, the students will then be able to retrieve specific area summary statistics relevant to their assignments.

MICROFILM AND PAPER COPIES OF CENSUS DATA

In addition to the first count summary tapes themselves, the Bureau has received a significant number of orders for these data in the form of 16mm microfilm and paper printouts of the microfilm. As of July 1, we had filled over 330 orders for paper copies of first count data for selected counties, county subdivisions and enumeration districts throughout the country. Approximately 150 orders for microfilm copies of the first count data had also been filled.

Copies of the 100-percent census tract tabulations (census tract report tables P-1, H-1, and for tracts with at least 400 Negroes, table H-3) are also becoming available. By late July, the Central Users' Service had distributed about 275 copies of the tables available for 175 SMSA's. This distribution included copies provided to the census tract key person in each SMSA as well as the copies purchased directly from the Bureau by other users.

A CRITIQUE OF THE DELPHI TECHNIQUE

Gordon Welty, American University

The Delphi technique is a recent RAND Corporation development in long-range forecasting. A group of experts are polled for their opinions on a given forecasting problem. The opinions are aggregated and fed back for a second round of opinion formation. Hence, the exercise is iterative, and polling, aggregation and feedback continue until consensus develops /cf. 3/.

In particular, a percentage of the respondents giving the most extreme individual forecasts each round, usually the upper and lower quartiles, are requested to reconsider the forecast they gave in a given round, in succeeding rounds, in light of their deviance from the group norm. This labelling or selection procedure supposedly hastens the development of consensus.

In several papers, we have examined a number of problems associated with structural aspects of the Delphi technique, such as aggregation of expert opinions /9/ and the selection of experts /10/. In this paper we shall study a further problem related to Delphi, namely that of the robustness of the Delphic exercise in withstanding deliberate manipulation of judgment and deceitful opinion formation.

It is important for our purposes to differentiate <u>risk bearing</u> from <u>confronting uncertainty</u>. If instead of a known world, we are talking about "unique events," etc., then known frequencies will not apply, since there won't be any sequences upon which the frequency can be based. Consider the distinction between a mechanical "one-armed bandit" with its known risk and pari-mutuel gambling schemes with its uncertainty. In the case of the mechanical schemes, there are fixed odds (known perhaps only to the house), while in the case of pari-mutuel betting, the odds change constantly, as a function of the social psychologically determined behavior of the bettors.

Following Frank Knight /5/, the relative frequencies in the known or measurable case is called <u>risk</u> and one's judgments in the case of unmeasurable circumstances will be judgments of <u>uncertainty</u>. For the latter, Knight notes that there are two fundamental methods of dealing with uncertainty, based respectively upon "reduction by grouping" and upon "selection of men" to control uncertainty /5, p. 239/. <u>Grouping</u> consists in categorizing the world (or its attributes) which confront men, while <u>selection</u> consists of categorizing the men who confront the world. For grouping, Knight has emphasized that nothing in the universe of experience is absolutely unique any more than any two things are absolutely alike. Consequently, it is always possible, for Knight, to form classes if a loose enough interpretation of similarity is accepted /5, p. 227/.

Knight goes on to point out several illustrative social institutions which deal with uncertainty. For <u>grouping</u>, the best-known sort of institution is that of insurance. The best-known institutions which select men to control uncertainty are those of <u>speculation</u>. These include future markets, produce and security exchange, etc. In elaborating upon the distinction between grouping and selecting men to control uncertainty, Knight points out that in the former case, the institutions don't serve to lessen "real risk" but merely spread it around. In the second case, however, the institutions <u>do</u>, by a process of elimination, lessen the real risk. "There is better management, greater economy in the use of economic resources, as well as a mere transformation of uncertainty into certainty" /5, p. 259/.

As Knight points out, "The problem of meeting uncertainty . . . passes inevitably into the general problem of management, of economic control" /5, p. 259/. This is the point at which we seek to specify a problem to be expected in a Delphi exercise. Since, in Delphi, the control of the exercise is <u>centralized</u>, in the person of the Delphi manager, we have the institutional analogue of the monopolistic institution rather than the competitive (decentralized) institution. As is well-known in elementary economic analysis, a monopolistic institution frees one parameter to function variably (e.g. "price"). Since this parameter can function variably, it can be "fixed" at a socially suboptimal level by the monopolist. The same, we will suggest, may be the case for the institution of the Delphic exercise.

The Delphi manager may deliberately misrepresent the outcome of one Delphi round to the participating experts in the next round, in an attempt to influence the outcome of the entire exercise. This influence would operate along the same lines as the group influence observed in Sherif-Asch type social psychological experiments /1;8, Ch. 7/. Being an "expert" and supposing the rest of the experts disagree with oneself may be sufficient for opinion change. Evidence for such an hypothesis can be provided by an experiment with a group of Delphi participants, wherein an attempt is made deliberately to mislead them. Are the Delphi participants susceptible to such influence and deception?

The Greek historians tell the story

of King Athanus of Alus in Thessaly who had two wives, first Nephele and then Ino. Ino was jealous of her step-children, and planned their death. A famine occurred after Ino convinced the local women to secretly roast the seed grain, and no crops grew. Athanus sent a messanger to the Oracle of Delphi to find the cause of the famine. Ino bribed the messenger to lie on his return. The messenger consented and said falsely the Oracle proclaimed the famine would cease only when Nephele's children were sacrificed to Zeus. Thus, we find deception.

The children, however, escaped to Colchis in Asia Minor, on the Golden Ram whose fleece was later retrieved by Jason and the Argonauts. Then an Oracle proclaimed that Athanus must be sacrificed for the country. The wicked Ino and her children met unhappy ends, and the king went insane and left the country. Thereafter the eldest male heir in each generation of the family of Athanus was sacrificed, since Athanus' sacrifice never properly occurred /4, pp. 161-163/. The point we would like to emphasize is that the Oracle, once deceived, was apparently unable to rectify its utterances. While this tale has an element of myth about it, nevertheless a basic point we wish to address is clear.

The anonymity of the experts who participate in a Delphi exercise may have a unique effect, precisely the same as we saw in the Greek tale of the deception of the Oracle, that warrants consideration. With the Delphi technique, not only is rectification of erroneous assumptions perhaps not possible, as we have seen, but the mechanism may actually facilitate reaching erroneous conclusions. Indeed, Dean Cyphert and Dr. Gant have presented experimental evidence that suggests that this is the case /2/.

In an attempt further to examine this circumstance, we replicated a substantial portion of Professor Rescher's Delphi-like study of anticipated changes in American values by the year 2000 A.D. /7/. While the questions posed were identical to those of Rescher's generic or "primary" Question 2, and comparable procedures were used, the respondents were radically different. Instead of high status scientists, such as used by Rescher /6, p. 21/, 192 sophomore engineering students were selected.

We were able to gather demographic information from 168 of the student participants. 78% reported they were sophomores, and 61% were 19 years of age. 24% reported that they resided within cities of 50,000 or more persons; 32% resided in suburban areas; 35% resided within towns of 2,000 to 50,000 persons; the remainder of the participants reported they resided in rural areas. Father's occupation provides a rough and ready measure of socio-economic status. 8% of the student participants reported their father's occupation as "professional," 10% as "managerial," 13% as technical and engineering, and 12% reported their father's occupation as sales, services, or bureaucratic. In addition, 21% reported their father's occupation as a craft (skilled blue collar) and 25% as unskilled labor. The remainder of the responses were scattered across several self-employment categories.

A preliminary null hypothesis was that there is no difference in the ability to forecast changes in values by high status scientists and university sophomores.

For each of Rescher's 37 items or "secondary questions," representing a value of American society in the year 2000 A.D. /cf. 7, p. 140, p. 145/, an opinion of the probable change in emphasis was elicited on a five point scale (ranging from 1 = greatly increased emphasis to 5 = greatly decreased emphasis). Each item mean was computed and compared by means of the F-test with the (rescaled) item means reported in Rescher's study. Since the covariance structure of Rescher's data was unknown, it was not possible to compute a single multivariate F ratio for this comparison. Instead, we computed a univariate F for each of the 37 items. For 21 of the items, we found no significant difference at p = 0.05 between the means of Rescher's distribution and our distribution. 8 items showed that Rescher's subjects expected more emphasis in the future upon the value itemized, and the other 8 items showed that the students expected more emphasis on the value itemized. A tabular display of this analysis, including item means and standard deviations, is available in Table I following.

Hence, we concluded that there was no significant difference in the ability to forecast values and their changes of the two groups of respondents. Thus, we felt confident, at least for forecasting subject-matter of values and their changes, that we could use the student Delphi participants as the subjects of a further study of the Delphi technique.

We can then turn to our major hypothesis. Let us suppose of a Delphi exercise that the median of the first round of forecasts indicates a specific central tendency in the expert's judgments. If the manager of the Delphi exercise chooses arbitrarily and deceitfully to change the forecast, as did Ino, and feeds back a median value for the secound round substantially different from the "true" value, then there are two interesting alternatives to consider. On the one hand, it might be supposed that the expertise of the respondents would permit their immediate recognition of the deception, in which case they would seek to reestablish the "true" value, refuse to participate in

TABLE I

Item No.	Mean (Rescher)	S. D. (Rescher)	Mean (Student)	S. D. (Student)	Significance of F (df = 1,246; p = 0.05)
1	2.02	0.94	1.79	0.86	
2	1.80	0.80	1.65	0.80	
3	1.96	1.01	2.39	1.08	R
4	1.70	0.69	2.00	0.94	R
5	2.68	1.01	2.65	0.99	
6	2.27	1.12	2.19	1.12	
7	3.59	1.08	3.16	1.15	S
8	2.70	0.97	2.48	1.10	
9	2.95	1.12	2.81	1.19	
10	2.41	1.14	2.54	1.09	
11	2.68	1.18	2.73	1.09	
12	2.70	1.16	3.02	1.21	
13	3.27	1.20	2.21	1.24	S
14	3.27	1.45	2.66	1.30	S
15	1.71	0.80	1.67	0.76	
16	2.91	0.94	2.50	1.00	S
17	2.05	0.86	2.43	3.02	
18	2.43	0.89	2.27	0.93	
19	3.02	0.96	3.34	1.13	R
20	3.43	1.01	3.43	0.98	
21	3.32	0.96	2.69	1.02	S
22	2.54	0.99	2.29	1.15	
23	2.59	0.93	2.69	1.17	
24	2.79	1.16	2.66	1.10	
25	2.27	1.07	1.69	1.04	S
26	2.29	0.95	2.73	1.09	R
27	2.04	0.95	2.38	0.97	R
28	2.59	0.87	2.92	1.01	R
29	2.23	1.01	2.65	1.01	R
30	1.84	0.65	1.88	0.92	
31	3.48	1.04	3.42	0.99	
32	2.52	1.10	2.79	0.98	
33	2.09	0.64	2.10	0.88	
34	1.95	0.96	1.53	0.86	S
35	1.70	0.83	2.35	1.02	R
36	2.43	1.08	2.23	0.95	
37	2.86	1.09	2.18	1.15	S

Legend: We have included above the 37 item means and standard deviations for the responses as given by Rescher's experts and the student participants. Also, we indicate that the differences were not significant (blank), that Rescher's experts expected more emphasis on the item in the future (R), or that the student participants expected more emphasis (S). Recall that the lower the mean value, the higher the expected emphasis on that value. The items are listed in 7, p. 140.

an obviously corrupt exercise, etc. In the Greek tale we gave, this would have been illustrated by the Oracle's refusal to prescribe the sacrifice of Athanus.

On the other hand, it might also be expected that the experts would not return, in subsequent rounds, to the "true" value. Under this alternative, they would more likely reflect the arbitrarily chosen median in the second, etc. rounds, as the Greeks relate the Oracle in fact did, and even rationalize their first round "deviance" from what they suppose to be a group judgment. Hence, on the second alternative, the experts in the Delphi exercise are deceived.

The second alternative might be expected because the expert who has been deceived in this fashion will not have prepared a rationale for his deviance and status of "minority of one," as would have an "intellectual maverick." After all, the deceived was (and still is) part of the deceived majority. Hence, he will be inclined to change his "deviant" judgment to accord with what he supposes to be the group judgment, rather than generate a rationalization for an unanticipated iconoclasm. Once he has made this accommodation, he can then rationalize his new estimate by denouncing his earlier assumptions.

As we have noted, Cyphert and Gant conducted an experiment which bore on this problem. While undertaking a Delphic exercise on the goals for the School of Education of the University of Virginia at Charlottesville, they introduced a "bogus goal" which was initially rated as having a low priority among all goals considered by the Delphi participants. The consensus was distorted and reported in later rounds as positive, and the final consensus showed the bogus goal rated considerably above the average /2, p. 13/. They concluded that "the hypothesis that the /Delphi/ technique can be used to mold opinion as well as to collect it was supported" /2, p. 14/.

We varied Cyphert and Gant's experimental procedure somewhat in our examination of the effectiveness of influence processes in causing shifts in group opinion. We fed back information to the subjects in the second round that labelled, for each of the 37 items noted above, various percentages of the respondents deviant. The range of percentages was from seven and a half percent, as a lower limit, up to eighty-six percent, as an upper limit. Thus, percentage labelled deviant per item was the independent variable, and would appear on the face of it to be a somewhat weaker intervention than that given in Cyphert and Gant's experiment. The dependent variable was the distance moved from the first to the second round. The mode of analysis was the product moment correlation.

If the correlation was negative, we could suppose that the first alternative, that the participants were not susceptible to influence and opinion formation, was true. If the correlation had been positive, we could suppose that the second alternative, that for Delphi participants to be arbitrarily labelled "deviant" could cause substantial shifts in the group opinion being formed by a Delphi exercise, was true. In fact, the correlation was a healthy -0.53; hence, we found evidence (at $p < 0.01$ for $df = 35$) that arbitrary labelling of deviants did not have an effect upon opinion formation to be expected on the basis of Cyphert and Gant's experiment; cf. scattergram in Fig. I.

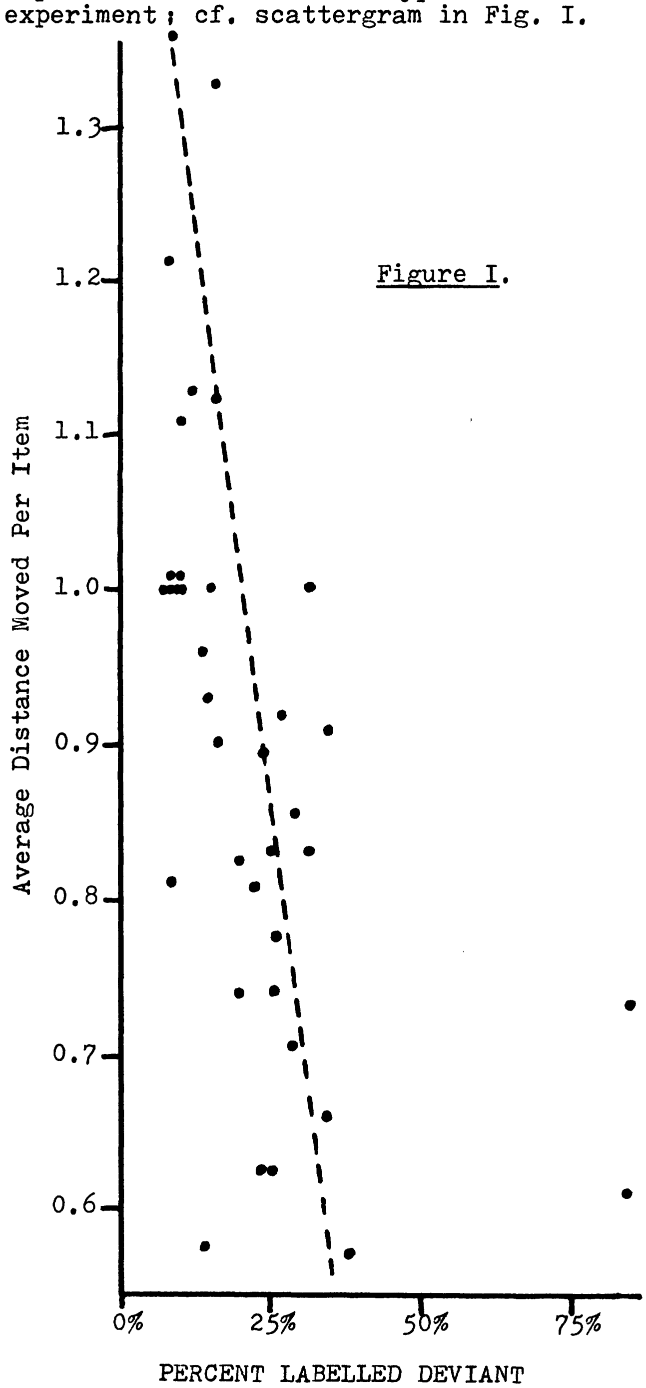


Figure I.

The data from our experiment is available in Table II below.

The Delphi technique appears, on the basis of our research, to be more powerful an institution in resisting wilful and arbitrary manipulation than we might have been lead to believe on the basis of Cyphert and Gant's work. At least one point is clear: further research on the structure of the Delphi exercise is called for.

SUMMARY

The Delphi exercise as a long-range forecasting technique can be considered, in Frank Knight's terms, a mechanism for coping with uncertainty. This permits us to bring to bear on Delphi the corpus of microeconomic theory. As such, the exercise can be treated as an institutional analogue to the monopolistic market of economic analysis. Such an anology allows us to note that, as in the case of the monopoly, the institution of Delphi and its manager can fix one variable of the analytical scheme at a socially suboptimal level.

The Greek historians relate an event in the kingdom of Alus where such a monopolistic practice occurred. The Oracle of Delphi's pronouncements were deliberately distorted, and the oracles were unable to recti y this distortion. The socially suboptimal level of functioning of the oracles was maintained.

Some recent evidence suggests that

TABLE II

Item No.	Deviations 0	1	2	3	4	Average Distance	No. Deviants
1	5	6	3	1	0	1.000	15
2	4	7	2	1	0	1.000	14
3	19	35	9	2	0	.908	65
4	13	16	3	2	0	.824	34
5	22	14	10	0	0	.739	46
6	8	5	6	2	0	1.095	21
7	12	8	6	3	0	1.000	29
8	32	23	10	0	0	.662	65
9	12	9	9	0	0	.900	30
10	86	44	20	7	3	.731	160
11	10	10	8	0	0	.929	28
12	8	7	13	1	1	1.333	30
13	14	4	11	2	1	1.125	32
14	27	8	6	7	1	.918	49
15	3	6	4	1	0	1.214	14
16	27	12	12	0	0	.706	51
17	18	13	6	1	0	.737	38
18	22	18	5	0	0	.622	45
19	23	19	12	4	1	1.000	59
20	20	9	12	0	0	.805	41
21	6	7	3	0	0	.813	16
22	10	6	9	0	0	.960	25
23	98	37	16	5	3	.604	159
24	20	23	10	1	0	.852	54
25	6	7	4	2	0	1.105	19
26	9	6	6	0	2	1.130	23
27	24	15	5	1	0	.622	45
28	4	2	7	1	0	1.357	14
29	18	15	9	2	0	.886	44
30	14	9	3	0	0	.577	26
31	23	11	11	2	0	.830	47
32	7	4	5	1	0	1.000	17
33	36	32	3	1	0	.569	72
34	8	8	1	2	1	1.000	20
35	24	22	14	0	0	.833	60
36	22	18	8	0	1	.776	49
37	5	8	2	2	0	1.059	17

Legend: We have included above the 37 items and the frequency distribution of distances moved between rounds one and two of the exercise. The last two columns give the average distance moved per item, and the number of participants who were labelled deviant and thus were respondents for a given item. The data of the last two columns were the basis of the correlation of -0.53, as well as the scattergram displayed overleaf.

similar distorting effects can be induced in a Delphic exercise and are not amenable to rectification within the exercise. We have undertaken a study which bears on the problem of the deliberate distortion of the responses of a Delphi exercise, and conclude on the basis of our data that the structure of the Delphi technique is more robust than one might have concluded on the basis of the earlier work.

ACKNOWLEDGMENT

We would like to thank Mr. Harris N. Miller for his helpful comments and criticisms of this study. Of course, the author alone is responsible for any errors or opinions remaining.

REFERENCES

/1/ Asch, Solomon, "Studies of Independence and Conformity: I," Psychological Monographs, Vol. 70, Whole Number 416 (1956).

/2/ Cyphert, Frederick and William Gant, "The Delphi Technique." Presented to the American Educational Research Association, Annual Meeting, Minneapolis, Minnesota (February, 1970), Mimeo.

/3/ Dalkey, Norman, "The Delphi Method." Presented to the American Statistical Association, 131st Annual Meeting, Fort Collins, Colorado (August 23, 1971).

/4/ Frazer, James, The Golden Bough: Part III, The Dying God, London: Macmillan (1923).

/5/ Knight, Frank, Risk, Uncertainty and Profit, Boston: Houghton, Mifflin (1921).

/6/ Rescher, Nicholas, "Study of Value Change," Journal of Value Inquiry, Vol. 1 (1967).

/7/ Rescher, Nicholas, "A Questionnaire Study of American Values by 2000 A.D." in K. Baier and N. Rescher eds. Values and the Future, New York: Free Press (1969).

/8/ Sherif, Muzafer, An Outline of Social Psychology, New York: Harper and Brothers (1948).

/9/ Welty, Gordon, "Note sur l'optimisme des futurologues," Revue internationale des sciences sociales, Vol. 22 (1970).

/10/ Welty, Gordon, "A Critique of Some Long-range Forecasting Developments," Proceedings of the 38th Session of the International Statistical Institute (1971).

THE PRODUCTION OF HEALTH BY A RURAL POPULATION

Henry W. Zaretsky, American Medical Association

I. Introduction

The purpose of this study is to propose a model of the production of health and apply it to data obtained in a rural health survey. For our purpose the "production of health" will be defined as the process whereby individuals combine medical care, other commodities (diet, recreation, etc.) and their own time to maintain their health status. This framework permits empirical measurement of the contribution of medical care relative to that of other factors in the maintenance of health. In Section II the production model will be developed and the estimation procedure set forth. In Section III the data used in fitting the model will be described. In Section IV the results will be discussed. This will be followed by a conclusion (Section V).

II. A Model of the Production of Health

It is assumed that the typical consumer engages in three activities -- work, consumption, and health maintenance or production.[1] He supplies labor on the labor market to earn wages which he combines with his non-wage income, if any, to purchase consumption commodities and commodities used in the production of health. The production of health involves the combination of medical care, other non-consumption commodities, and the consumer's own time. The consumer has to distribute his time, as well as his income, among the three activities. He is assumed to maximize the expected value of a utility (preference) function subject to a budget constraint, a time constraint, and a production constraint.[2]

In this study we will concentrate on the production side of the model. It is assumed that an individual inherits a stock of health capital.[3] This stock is assumed to depreciate with age after some stage of the life cycle and is subject to further deterioration, largely random. To the extent that the actual stock of health falls short of the desired stock, the individual will consider health production to restore or maintain his health stock.

The production function is a mathematical statement of the technological relationship between the output of a process and the inputs. The major purpose of the production function is to present the possibilities of substitution between the inputs (factors of production) to achieve a given output. For any set of inputs, the production function is interpreted to define the maximal output realizable therefrom.[4] In the application here the output of the production process is defined as improvements in health status (gross additions to health stock) and the inputs are defined as medical care, other non-consumption commodities, and the consumer's own time.

The relative amounts of time and medical care input into the production process depend on their relative productivities and their relative prices. It is expected that persons with high earnings rates would use relatively less of their own time and relatively more medical care in the production of health than persons with low earnings rates. The earnings rate is assumed to be closely related to the individual's perception of the "price of his own time." Moreover, efficiency in production varies from person to person. The more efficient can produce a given amount of health with less input -- time and medical care -- than the less efficient.[5] It has generally been observed that more educated persons are more efficient producers of money earnings than less educated persons. Since education improves market productivity it is reasonable to expect that it improves nonmarket productivity as well. This implies a positive relationship between education and the health production process. Thus an increase in education would increase the amount of health produced from given amounts of medical care and time. Since earnings are related to production efficiency, it would appear that the more efficient (educated) while using less of both inputs, for a given output, would use relatively less of their own time in the production of health.

Consider the following production function:

$$(1) \quad H = A_1 e^{a_1 E} M^{\alpha_1} T^{\alpha_2} u_1$$

where H is the amount of health produced, E stands for the education level, and M and T are the medical care and time input, respectively. The term A_1 is a constant, a_1 is the education coefficient, α_1 is the elasticity of health with respect to medical care (proportional change in health production resulting from a proportional change in medical care input), α_2 is the elasticity of health with respect to time, and u_1 is a random error term. If it is further assumed that $\alpha_1 + \alpha_2 = 1$, or $\alpha_2 = 1-\alpha_1$, the implication is that if medical care and time are increased, say 10 per cent, then output is also increased 10 per cent. This is known as constant returns to scale. The form of production function depicted in (1) is called Cobb-Douglas.[6,7] The non-consumption commodities input is not included in (1) as it is difficult to measure and since this input largely reflects life styles and environmental factors, its effect is likely to be absorbed by the education variable.

The primary purpose of this paper is to estimate α_1. Direct estimation of equation (1) by ordinary least squares would result in biased estimates owing to the simultaneous nature of the health production process. The medical care coefficient can be interpreted as a measure of the effect of medical care on health. But the feedback of H on M and T must be considered. Medical care and time are not exogenous but are influenced by the current level of health stock. Thus the medical care coefficient could also be interpreted as the effect of health on the demand for medical care. In order to deal with the simultaneity, estimates are obtained by using two-stage least squares.

The full model, referred to above,[8] suggests that the demand for medical care and the demand for the time input each depend on income and education. Income reflects the economic determinants and education reflects production efficiency and attitudinal variables. The following demand specification is proposed:

$$(2) \quad T/M = A_2 e^{a_2 E + a_3 I} u_2,$$

where I is income, A_2 is a constant, and u_2 is the error term. If the production coefficients in equation (1)--α_1 and α_2-- are restricted to sum to unity, both equations (1) and (2) would be exactly identifiable and estimation by indirect least squares, two-stage least squares and quasi-information maximum likelihood would all yield the same consistent and efficient estimates.[9]

Noting the restriction on α_1 and α_2, dividing (1) through by M, and converting (1) and (2) into logarithms, we obtain the following two equation system;

$$(3a)\quad \log H/M = \log A_1 + (1-\alpha_1)\log(T/M) + a_1E + b_1S_1 + b_2S_2 + b_3S_3 + \log u_1$$

$$(3b)\quad \log T/M = \log A_2 + a_2E + a_3I + c_1S_1 + c_2S_2 + c_3S_3 + \log u_2$$

The data are obtained from a four strata sample of households. To allow for possible shifts in intercept among the strata, dummy variables are included in each equation; S_i is set equal to unity if the observation is from stratum i and set to zero otherwise.

III. The Data

The data used to test the above model were obtained in the Yolo County Health Survey.[10] Data were gathered on 1100 households (3400 individuals)--a four percent sample. A two-stage stratified sampling procedure was employed. The study area was divided into four strata--Davis (Stratum 1), Woodland (Stratum 2), East Yolo (Stratum 3), and Rural Yolo (Stratum 4). Stratum 4 is the most rural of the strata. The other strata are characterized by higher population density and relatively less agricultural employment.

Definitions of Variables

1) M: Gross personal medical expenditures. This includes annual (1969-70) out-of-pocket expenditures by individuals for services of physicians, dentists and other health manpower; hospital care; nursing home care; x-ray and laboratory tests; and medical appliances. To this sum, "Adjusted Insurance Premiums" is added. This is an approximation of the individual's share of the household's expense for health insurance. "Adjusted Insurance Premiums" is calculated by dividing total annual household health insurance premiums by the number of equivalent adults in the household. This term is calculated by counting all children under 12 and the second adult as one-half. All other adults are counted as one.[11,12] Due to the form of equations (3a) and (3b), the analysis is restricted to persons for whom M is not zero.

2) I: Adjusted income. This is total household income for 1969, before taxes, per equivalent adult.

3) T: Time input. Because of recall problems with this item, it was decided to use work-loss days for the expenditure period as a proxy. Thus the entire analysis is restricted to employed adults. It should be noted that this is a poor measure of time devoted to health production and is strongly influenced by economic factors.[13]

4) E: Education. This is an eight-point scale. The values are shown in Table 1.

Table 1. Education Intervals

Highest Level of Education	Value
No formal school	0
Some grade school	1
Completed grade school	2
Some high school	3
Completed high school	4
Vocational training	5
Some college	6
Completed college	7
Post graduate study	8

5) H: Health Status. This is a weighted sum of medical conditions checked by respondents.[14] The conditions and their corresponding weights are shown in Table 2. An index value of "zero" implies "perfect" health while a value of 128 is the poorest. The major difficulty with the variable is that it is a proxy for health stock and not the amount of health produced.

IV. Results

The two stage least-squares estimates are shown below;

$$(3'a)\quad \log H/M = 2.5199 - 0.0341E + 1.3733\ T/M^{*} - 0.4445\ S_1^{*} - 0.1220\ S_2 - 0.0284\ S_3 \qquad R^2 = 0.175,\ N = 569$$

$$(3'b)\quad \log T/M = -3.2639 - 0.0745\ E^{*} - 0.0001\ I^{**} - 0.0649\ S_1 - 0.0258\ S_2 - 0.1850\ S_3 \qquad R^2 = 0.065,\ N = 569$$

where (*) and (**) indicate statistical significance at the .05 and .01 levels respectively. If u_2 is normally distributed with zero mean and constant variance, the t tests are valid for (3'b) as all the regressors are exogenous. In (3'a) the t test is not appropriate but the t values are shown to indicate the relative sizes of standard errors.

Equation (3'b) indicates that education and income each have a negative effect on T/M. Thus persons with more education and persons with more income have a less "time intensive" production process than do their less educated and lower income counterparts. This is a reflection of the relative value of their own time.

Since our proxy for health status is an inverse measure, the coefficients stated in (1) and (3a) must be reinterpreted. The elasticity of health with respect to medical care is to be interpreted as $-\alpha_1$ and that with respect to time as $\alpha_1 - 1$, both summing to -1 instead of 1, as they did previously. From (3'a) the estimate of $-\alpha_1$ is -2.3733. Thus a 10 percent increase in medical care results in a 24 percent decrease in our inverse health status variable or a 24 percent <u>increase</u> in health status. This implies that med-

Table 2. Conditions and Weights Used in Construction of "Health Status Index" (Adults)

Condition	Weight	Condition	Weight
High blood pressure	4	Frequent cramps in legs	4
Heart condition	2	Pain in heart or tightness in chest	4
Stroke	4	Trouble breathing or shortness of breath	4
Bronchitis	2	Swollen ankles or feet	4
Asthma or hay fever	2	Pains in the back or spine	2
Arthritis or rheumatism	4	Repeated pains in stomach	2
Epilepsy	4	Frequent headaches	2
Sugar diabetes	4	Constant coughing or frequent heavy chest colds	2
Cancer or tumor	4	Blurred, haziness or clouding of vision	4
Tuberculosis	4	Stiffness, swelling, or aching in any joint or muscle	4
Emotional or mental illness	4	Getting very tired in a short time	4
Stomach or duodenal ulcer	4	Blind spots in vision	2
Gall bladder trouble	4	Seeing double	4
Liver trouble	4	Episode of fainting	5
Hernia or rupture	4	Feelings of lightheadedness or dizziness	5
Kidney trouble	4	Trouble hearing	2
Back trouble	2	Now unable to carry on normal activities	4
Trouble passing urine	4	Injury in past 12 mo. restricting normal activities	6

ical care makes a strong positive contribution while time makes a negative contribution (1.3733) --a 10 percent increase in time devoted to health maintenance results in a 14 percent decrease in health status. It is possible that the data inadequacies, noted in Section III, could account for this disturbing result.[15] The result could also indicate that time and medical care are really not substitutable, on the average, and that a certain minimum amount of medical care is necessary and without this minimum amount of care, time has a negative effect on health status. This would indicate that the longer one attempts to treat a medical problem by remaining at home but not obtaining necessary medical care, the more aggravated the problem will become.

The education coefficient in (3'a) has the expected sign, indicating that more educated persons are more efficient producers of health than their less educated counterparts. It was expected that rural individuals, owing to lower earnings and inaccessibility of medical services, would have a more time intensive production process and a less efficient production process than urban individuals. The coefficients of the stratum dummy variables did not confirm this.[16] Furthermore, an analysis of variance of T/M among strata showed no significant differences.

V. Conclusion

A model of the production of health was tested on data obtained from a rural health survey. The results imply that, in general, time and medical care are not substitutable in treating a medical condition and that medical care is the most productive of the two inputs. Production efficiency is positively affected by education. Individuals with high income and individuals who are more educated have a less time intensive production process than their less educated and lower income counterparts. The time-medical care ratio does not significantly differ between the urban and rural regions of the study area.

NOTES

[1]This method of viewing consumer theory is similar to that developed by G.S. Becker, "A Theory of the Allocation of Time," Econ. Jour., 1965; and K. Lancaster, "A New Approach to Consumer Theory," Jour. of Polit. Econ., 1966. This is an application of a method set forth by R.R. Wilson, "The Theory of Consumer Behavior: Production and the Allocation of Time," Winter Meeting (1969), Econometric Society, New York.

[2]The complete model, including the derivation of demand functions, is presented in H.W. Zaretsky, "The Demand for Health Care," Ph.D. Dissertation in progress, Department of Economics, University California, Davis, 1970.

[3]See M. Grossman, "The Demand for Health: A Theoretical and Empirical Investigation," National Bureau of Economic Research, 1970.

[4]See R.W. Shepard, Theory of Cost and Production Functions, Princeton, 1970.

[5]M. Grossman, op.cit.

[6]R. Auster, I. Leveson, and D. Saracheck in their, "The Production of Health, An Exploratory Study," Jour. of Hum. Resources, (Fall 1969), used a Cobb-Douglas production function with constant returns to scale to explain variations in mortality rates across states.

[7]There are two major difficulties inherent in this form of production function: (1) If any input is zero, output must be zero. (2) This form requires the "elasticity of substitution" to be unity. The implication is that if the ratio of the price of time to the price of medical care would increase by 10 percent, the ratio of time to medical input would decrease by this same amount. Alternative forms are being considered for further study. There has developed a substantial literature on production functions and their estimation. A useful survey can be found in A. Walters, "Production and Cost Functions: An Econometric Survey," Econometrica (January-April, 1963).

[8]H.W. Zaretsky, op.cit.

[9]See E. Malinvaud, Statistical Methods of Econometrics, Amsterdam, 1966.

[10]This was a comprehensive health survey conducted during June 1970 by the Department of Community Health, School of Medicine, University of California, Davis.

[11]This method was used in W.J. McNerney, et al., Hospital and Medical Economics, Chicago, 1962.

[12]In conducting the survey we were not permitted the luxury of verifying medical expenses. Since consumer recall or even knowledge of that portion of the medical bill paid by insurance or other third parties was poor, only the net or out-of-pocket expenditures could be used. To approximate the individual's total expenditure, his expense for prepayment was added.

[13]See M. Silver, "An Economic Analysis of Variations in Medical Expenses and Work-Loss Rates," Empirical Studies in Health Economics, H. E. Klarman ed., Baltimore, 1970.

[14]This is a modification of an index developed by A.I. Kisch, J.W. Kovner, F.J. Harris, and G. Kline in "A New Proxy Measure for Health Status," Health Services Research (Fall, 1969).

[15]It should be noted that the simple correlations between H and T and between H and M are each positive (each about .1). There is reason to suspect that M and T are each alternative measures of health status. Furthermore, as indicated above, the variable H should measure amount of health produced (change in health status) while our measure of H is a proxy for current health status or level.

[16]The coefficient of S_1 (Davis) in (3'a) is absolutely larger than the coefficients of the other dummy variables and has the expected sign. This would indicate that Davis residents, with the education level held constant, are relatively more efficient producers of health. Although the standard error of the coefficient in question is less than half the size of the coefficient, it must be noted again that the t tests are not appropriate here.

MINUTES OF THE ANNUAL MEETING OF THE SOCIAL STATISTICS SECTION

Fort Collins, Colorado, August 25, 1971

The meeting was opened by Otis Dudley Duncan, the Chairman, at 5:30 p.m. As a result of the recent elections, the list of officers for 1972 is:

Chairman	Eva Mueller
Chairman-Elect	Theodore D. Woolsey
Vice-Chairman	Charles B. Nam
Vice-Chairman	Sidney Goldstein
Secretary	Regina Loewenstein
Section Representative on the Board of Directors	John D. Durand
Section Representative on the Council	Eli S. Marks
Publications Liaison Officer	Karl E. Taeuber

The Chairman reported that per the request of this Section the office of the American Statistical Association computed the percentages of members and Fellows of the Association that were from the Social Statistics Section. The proportion of Fellows from this Section was the same as the proportion of members. Continued interest by members of the Section in nomination of Fellows was encouraged.

The recent report of the Board of Directors of the American Statistical Association entitled A Study of Future Goals of ASA was discussed. Recommendation 40 is:

> "Examine ASA's possible role in assisting with the development of better ways of evaluating social projects and programs and with the preparation of social indicators. Direct participation and assistance of the leadership in the Social Statistics Section should be sought also on this project."

The Section concurs with this recommendation and suggests that the President of the American Statistical Association appoint an ad hoc committee to work with Federal governmental agencies and the Social Science Research Council on this problem, and then report to the American Statistical Association about the status of the work. Members of the Section were asked to send names of suggested members of this Committee to John D. Durand. The Program Committee for the 1972 Meeting might include on the program one or more sessions about social indicators, including the functions of different government agencies in development of social indicators.

Recommendation 39 of this Report is:

> "To determine ways in which statistical thought and/or methodology could be used to help accelerate progress in solving such problems as a quality environment. Do we, for example, have adequate statistics on conditions of the environment such as air, water and soil pollution? Or, do we have methodology that can contribute to the analysis of pertinent data and construct meaningful statistical indicators of environmental conditions?"

The Section expresses interest in this recommendation and hopes to participate in work on this recommendation in the future.

Charles Nam requested consideration of a special mailing to members of the Social Statistics Section with a request for suggested topics and speakers for the 1972 Meeting. Proposed topics were: (1) Measurement of discrimination toward women, black, young, etc., (2) evaluation and improvement of quality of international statistics published by the United Nations, and (3) standards for published periodic statistical series.

Karl Taeuber was thanked for acting as Program Chairman for the 1971 Meeting when that function was not part of his usual responsibilities.

1971 Officers of the Social Statistics Section

Chairman	Otis Dudley Duncan
Chairman-Elect	Eva Mueller
Vice-Chairman	Philip C. Sagi
Vice-Chairman	Charles B. Nam
Secretary	Regina Loewenstein
Section Representative on the Board of Directors	John D. Durand
Section Representative on the Council	Hubert Blalock
Publications Liaison Officer	Karl E. Taeuber
Editor of Proceedings	Edwin D. Goldfield