MOBILITY ANALYSIS - AN APPROPRIATE TECHNIQUE IN SMALL AREA DEMOGRAPHY

Arnold R. Post, Fels Center, University of Pennsylvania

I wish to acknowledge that this paper derives from conversations with one Robert Reams. He was a supervisor of special censuses for the Census Bureau in the late 1950's and had developed a rule of thumb for estimating current population which he found to be quite accurate. In a variety of areas, 1958 populations were approximately equal to 90 percent of their 1950 populations plus about 3.75 times their increase in dwelling units since 1950.

Actually, it may be observed that populations behave with considerable regularity for given changes in their numbers of households (or housing supply). Multiple linear regressions for collections of areas in Pennsylvania indicate that their censused populations may be reliably estimated as deriving from the populations previously censused and subsequent changes in number of households. More than 99 percent of the variance in final population is 'explained,' and the coefficients of the two independent variables are both very highly significant. Standard errors and constant terms tend to be relatively small, and residuals appear to be reasonably well distributed according to normal expectations.

Several analyses have been done on the following sets of areas:

A. A collection of counties in metropolitan areas at least partly in Pennsylvania and having at least 3,000 nonwhite residents, but including only those counties or parts of Philadelphia with at least 97 percent of population in households (28 cases);
B. A collection of 20 randomly selected Pennsylvania counties; and
C. The collection of 427 census tracts (as of 1950) in the Philadelphia Standard Metropolitan Statistical Area (SMSA), but outside the city proper.

Some of the results are shown in Table 1, where the figures in parentheses show the t-values of the coefficients of the estimating equations. A t-value greater than 3.0 usually fosters belief that a real-world dependency has not been disproved by the data at hand. As can be seen, these t-values range from 5.57 to 205.3.

The random sample of counties and the suburban census tracts contained predominantly white populations resident mostly in single family dwellings. Among the counties were several which lost population during the 1950's, while the tracts were characterized mostly by rapid growth. The metropolitan counties are associated with higher percentages of nonwhite population and apartment dwellers.

Relations 3,5 appear to be distinct from each other, perhaps reflecting the different types of areas they have been drawn from. Both, however, can be regarded as within the range of random deviation about relation 4, which was derived from the probability sample.

The coefficients of relation 4 have reasonable values. The coefficient of initial population is 89.8 percent, which is not far off the tenth power of .99. Janet Abu-Lughod and Mary Mix Foley

Table 1
Linear Regression Equation Parameters For Estimating Population After Ten Years, Given Initial Population and Subsequent Change in Households
(t-values in parenthesis)

Data Set	Decade	Initial Population	Change in Households	Constant Term	Mean Value Est. Pop. ($\bar{Y}$)	Standard Error No.	Standard Error % of $\bar{Y}$	Relation Number
A	1930-40	.9053 (41.4)	3.3425 (5.57)	3,753	221,006	8,803	4.0 %	(1)
	1940-50	.9116 (126.4)	3.6951 (25.1)	1,398	240,652	4,295	1.8 %	(2)
	1950-60	.9234 (80.3)	3.8495 (20.0)	516	267,822	12,612	4.7 %	(3)
B	1950-60	.8982 (53.6)	3.9583 (20.2)	2,065	86,736	2,410[1]	2.8 %	(4)
C	1950-60	.8858 (145.9)	4.1010 (205.3)	28	5,293	343[2]	6.5 %	(5)

(1) 65% of residuals less than one standard error, 10 positive (max. 5,425), 10 negative (min. -3,341)
(2) 95% of residuals less than one standard error, about 60% positive.

Source: U.S. Censuses of Population for 1930,1940,1950, and 1960. Private families, occupied dwelling units, and occupied housing units have been treated as if synonymous.

have reported (1) an estimate to the effect that in 1955, about 1 percent of the population engaged in household formation. Presumably, then, 99 percent did not; and over the course of the decade, assuming constancy, about 89 percent would be resident in the same number of households as at the beginning of the decade. The coefficient of household change (3.96) lies near the mid-point of a range defined by the average size of all husband-wife households (3.7 persons) and the average size of such households with husband under 45 years of age (4.3 persons.) (2) The coefficient is therefore consistent with the notion that gains in households are principally the work of young people, as has been well established by Ned Shilling. (3)

These relations in Table 1 were derived from the period 1930-60. Considering that this period spans economic extremes of depression and boom, the similarities among them are notable. The coefficients of initial population are less than 1.0 and those of household change are not unwieldy as average household sizes. Both have proper sign.

I can't, at this time, present statistical findings for the nation's large metropolitan areas, but I have drawn some pictures to show population contours for some of them. For the very largest metropolitan areas (1950 SMA's, generally) the populations of 1950 are shown in Figure 1 as a joint function of 1940 population and net gain in households. Similar information is shown in Figure 2 for 1960 populations of SMA areas of between 200-700 thousand persons as of 1950 and with household gains of less than 105,000. The data have been scaled assuming 4 persons per marginal household. Under the assumed hypothesis, the contours should be straight, equally spaced, and parallel with a slope of -45 degrees. In Figure 1, the contours can be seen to approach the vertical axis at values greater than their own, as should be the case.

Now, given such a well-defined and appropriate collection of regression equations which seem to be broadly applicable, it is only reasonable to suppose that they index a process. Since the coefficient of initial population presumably indicates a fraction of population apt to remain in a stable community (one characterized by no net change in number of households), it can be termed a persistence rate. The coefficient of household change can be considered a rate of marginal population change since it would indicate the average size of marginal households. Accordingly, we would have two concepts, persistence and marginal change, which are somewhat analagous to the traditional concepts of natural increase and net migration, except that we are here dealing with a "natural increase in housing need" and the mobility occasioned to satisfy it. More concretely, a population is assumed to generate additional family or household heads in the course of a decade, and they will export themselves to areas where the desired quarters are available, if they can manage it, whether across the street or across the nation. They will also take with them wives and potential children, among others, as they see fit.

Commonly accepted demographic algorisms rely on concepts of natural increase (surplus of births over deaths) and net migration; and it has been asserted many times that population change in a given place derives either from natural increase or net migration. This assertion carries with it an implication that net-migrants are both sterile and immortal. Indeed, the implication appears explicitly in some matrix models of population growth. As a migrant, myself, currently, I take exception.

Traditional application of these concepts is often called "cohort survival analysis." The initial population of a given place in this context, provides a source of people likely to bear and survive anywhere. The concept of net migration re-introduces boundaries very much as an afterthought. Migrants are, after all, only ordinary mortals who have been reclassified. Some attrition among cohort survivors has to be allowed for somewhere in any realistic accounting of the two classes of people, that is, today's migrants are yesterday's naturalized residents of some other place.

Algorisms relying on the concepts of persistence and marginal change have been categorized as dealing in "mobility analysis." These concepts relate nicely to population trends as defined in geographic partitions. For each area, conditional allowance is made both for departures (as a function of initial population) and arrivals or additional departures (as a function of household change.) The initial population in this context provides a source of people apt to survive and remain (on net) in a given place; and the additional households, if any, provide accommodation for new families or families moving about. To my mind, mobility analysis is richer in spatial context and logically more of a piece than cohort survival analysis and has much to recommend it. Statistically significant results have also been obtained from age-specific data.

Mobility analysis, however, is not any easier to apply than cohort survival analysis; and it does require independent estimates of postcensal or future changes in households to explore likely patterns of population development.

For current estimates of household change, my preference is to rely on school enrollment data, interpreting them with the aid of census cross-tabulations of 5 to 19 year-olds by grade of school in which enrolled, using three-year grade groups and five-year age groups. With the aid of the cross-tabs, one can develop one operator to generate an age distribution from enrollment data and another to generate an enrollment distribution from age data. (The cross-tabs are published by single years of age and grade among the detailed statistics for each state.)

Current school enrollment data will not automatically be available for each municipality since public school districts sometimes don't coincide with municipal boundaries and private school service areas may be quite independent of them. The data, however, are worth scratching for, it being important to secure the equivalent of pupil counts as of April.

Figure 1

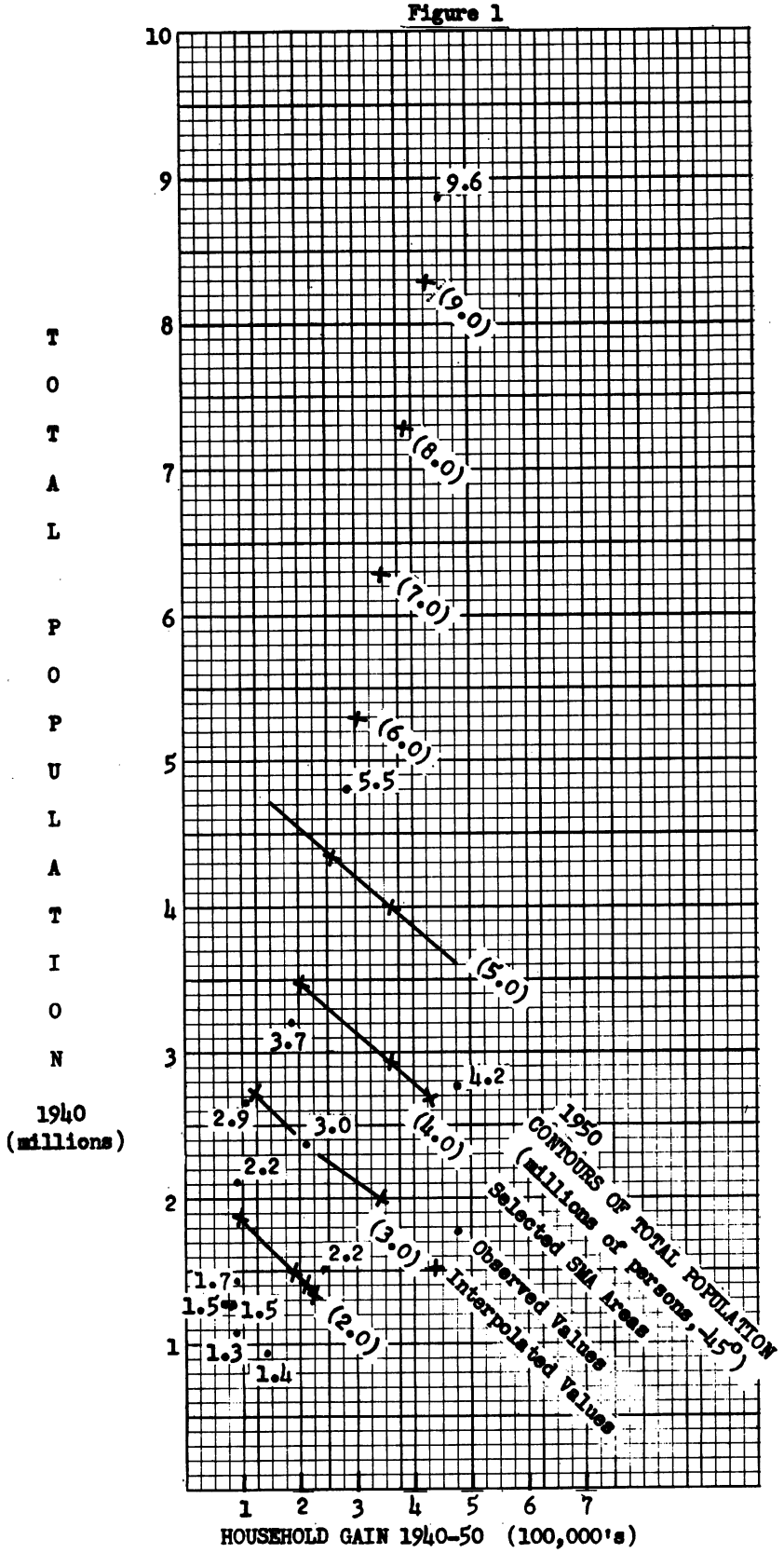
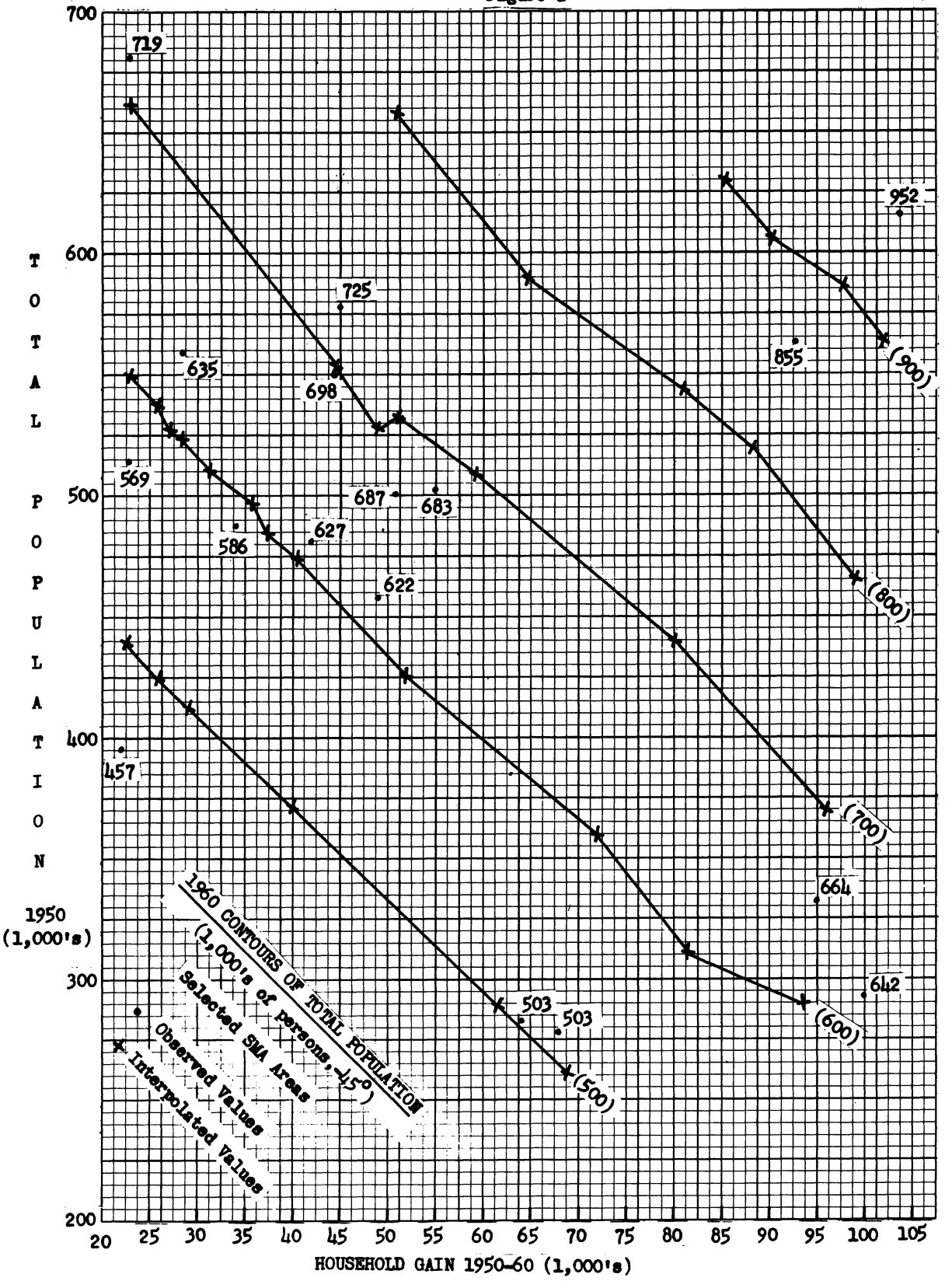

TOTAL POPULATION 1940 (millions)
HOUSEHOLD GAIN 1940-50 (100,000's)
1950 CONTOURS OF TOTAL POPULATION (millions of persons, 45)
Selected SMA Areas
Observed Values
Interpolated Values
(9.0)
(8.0)
(7.0)
(6.0)
(5.0)
(4.0)
(3.0)
(2.0)
9.6
5.5
4.2
3.7
3.0
2.9
2.2
2.2
1.7
1.5
1.5
1.3
1.4

Figure 2


700
600
500
400
300
200
TOTAL POPULATION 1950 (1,000's)
719
725
635
698
569
687
683
586
627
622
457
503
503
952
855
(900)
(800)
(700)
664
642
(600)
(500)
1960 CONTOURS OF TOTAL POPULATION (1,000's of persons, 45°)
Selected SMA Areas
Observed Values
Interpolated Values
20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100 105
HOUSEHOLD GAIN 1950-60 (1,000's)

With enrollment data at hand, one attacks them from two directions. (1) On the basis of the last census, a persistent population of school-age is estimated as if there had been no change in the number of households. An average number of these children per household is also calculated. Then a disaggregated estimate of their enrollment by age and grade group is prepared to provide an index of the age distribution in the grade groups currently. (2) The actual enrollments by grade group are assumed to have the same relative age distribution as enrollments from the persistent population. The age groups are then summed and blown up to indicate the age distribution which generated the actual enrollments. This expansion is required since not all school-age children will actually be in school.

The difference between the presumed actual and the persistent populations (5-19) can then yield an estimate of net household change according to which the total population and its age distribution can be estimated.

At this point, after the direction of net migration can be determined, useful reference to vital statistics can be made. The argument, however, has to be interpretive. Abu-Lughod and Foley in the work cited, indicate that changes in household size appear to be a main factor associated with a fimily's taking new quarters. Birth statistics, then, would overstate preschoolers for a population supporting out-migration and understate them for a population augmented by in-migration, i.e., babies seem to be a cause of mobility and they travel with their parents.

The above paragraphs sketch some of the procedures available to make mobility analysis operational, and difficulties have scarcely been mentioned. The real world and the future continue to present problems that no amount of arithmetic can render precise. I think it is fair to say that the past is about as indefinite as the future is improbable, and the two are related by an uncertain, though actual, present. In the late 1950's, I, for one, did not anticipate the development of near revolutionary conditions in many of our major cities; and some of my estimates are doubtless less apposite for that. In terms of the present discussion, I would expect these conditions to affect rates of persistence and marginal change by influencing especially the mobility of mature families. Although I did anticipate the declining trend in births in fair degree, I didn't foresee the growth of the public share of total enrollments.

Accepting such matters as part of the rules of the game, we can still suppose some regularities exist. Glancing again at Table 1, it may be noted that between 1930 and 1960 (relations 1,2,3), rates of persistence and marginal change seem to be directly related with both rates rising. My own hunch is that some explanation is offered by the concurrently rising trend in large area rates of natural increase. What with reduced rates of natural increase now prevailing, it may be that current rates of persistence and marginal change are also now at lower levels. It may also be noted, however, that as regards relations 3,5, their coefficients are inversely related with the greater spread being characteristic of the richer populations. I incline to consider this as evidence that richer populations can form new households at greater convenience. Although I can't quantify these matters, I anticipate that the greater mobility of mature families, if any, will tend to make both rates less well defined in the present decade, that with lowered natural increase, the rates will define a somewhate shorter vector, and that with continued prosperity, the spread between them will tend to remain wide.

As for internal matters, Table 2 exhibits a good deal of variety among ten-year rates of persistence and marginal change as derived from analysis of age-specific data for the metropolitan counties. There are very low rates of marginal change for 15-29 year olds in 1960; and their level may relate to the scarcity of persons born between 1930 and 1944 in the total population, suggesting a future avenue of research towards a lagged model. Ten-year age-specific rates were also derived from the random sample of 20 Pennsylvania counties; and these rates may be more appropriate in suburban and rural areas. They are shown in Table 3.

Since rates of community development may change from one pentad to the next (I don't like to say quinquennium, except to refuse to say it), it is desirable to work in terms of five-year intervals. Conjectures, however, as to the values of five-year rates are simply conjectures at this time. A special census was done of the Cleveland SMSA in 1965, and these data may provide a springboard.

A comparison of persistence and survival rates is given in an article scheduled to appear in the November (1969) issue of the Journal of the American Institute of Planners. A comparison is also offered there of age distributions of marginal changes and net migratory increments. In addition, it is noted that relation 5, Table 1, derived from census tract data, yields good results with respect to the population of the whole Philadelphia SMSA and to that of each of its constituent counties, including Philadelphia, itself, whose census tracts did not contribute to the estimating equation. The bias is on the order of -1.5 percent. (Note: Some may wish to investigate these relationships in their own regions of interest. If so, it is necessary to examine an area large enough to provide residence for by far the greater part of both its initial and final populations. All the census tracts in one suburban county may not be sufficient if the county, say, has doubled its population in the period under review. In such a case, the distribution of initial population by tract is not apt to be significant.)

Mobility analysis has been employed as the basis for a computerized routine to yield estimates of the short run future for a variety of suburban and rural school districts. Analysis done by hand for Philadelphia indicates (1) that the city, with a gain of only 20,000 households, is continuing to lose population. A population of about 1.9 million is anticipated for 1970, about 5 percent less than 1960's

Table 2
Ten Year Age-Specific Rates, Metropolitan Counties

Terminal Age Group	Persistence 1930-40	Persistence 1940-50	Persistence 1950-60	Marginal Change 1930-40	Marginal Change 1940-50	Marginal Change 1950-60
0-4	.1053[1]	.1510[1]	.2047[1]	.3695	.5714	.5323
5-9	.1117[2]	.1260[2]	.1882[2]	.4020	.4193	.5176
10-14	.9389	.9016	.8187	.1393	.2149	.3457
15-19	.8805	.8373	.8054	.2439	.1883	.1812
20-24	.8874	.8265	.7662	.2454	.2390	.1128
25-29	.8397	.7766	.8006	.3216	.4169	.2406

(1) Generation rates based on initial population 5-39.
(2) Generation rates based on initial population 10-44.
Source: See note, Table 1.

Table 3
Ten Year Age-Specific Rates,
20 Pennsylvania Counties, 1950-60

Terminal Age Group	Persistence	Marginal Change
0-4	.091[1]	.665
5-9	.096[1]	.560
10-14	.818	.356
15-19	.751	.307
20-24	.575	.359
25-29	.635	.376

(1) Generation rates based on total initial population.
Source: See note, Table 1.

2.0 million; (2) that the proportion of population Negro appears to be increasing more rapidly than during the 1950's and will reach 37 percent by 1970; and (3) that the public share of a nearly constant total enrollment seems apt to increase by about 7 percent on a base of 410,000 by 1980.

The Philadelphia estimates do not represent a consensus. Other analysts see a small gain for the city population, slower growth of the proportion Negro, and less than a 28,000 gain in public school enrollments in the coming decade. The 1970 census should lay some doubts to rest. Although my goal as regards accuracy is to be most accurate among those who overestimate, I should think that a census of 1.95 million or less total population would indicate that some skill has been achieved in the application of mobility analysis.

Overall, the regularities that have been observed do encourage me to believe that locally standard patterns of population growth and development have become a little more apparent.

REFERENCES

(1) Janet Abu-Lughod and Mary Mix Foley, "Consumer Strategies," Part II of Nelson Foote, et al., Housing Choices and Housing Constraints (New York: McGraw-Hill, 1960), Table 19, p.99 and p.100 (footnote).
(2) U.S., Bureau of the Census, Current Population Report, P-20, No. 173 (June 25,1968), Tables 1 and 5.
(3) Ned Shilling, "Net Household Formation - A Demographic Analysis," (unpublished Master's Essay, Columbia University, 1955), cited by Louis Winnick, American Housing and Its Use (New York: Wiley and Sons, 1957), p.81.