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.