REPORTING OF HOUSEHOLD INCOME: COMPLETE VERSUS INCOMPLETE RESPONSE

Thesia I. Garner and Laura A. Blanciforti. U.S. Bureau of Labor Statistics

ABSTRACT

The nonreporting of household income is a common and pervasive problem in surveys. Accurate income information is a valuable resource for the research analyst examining expenditure behavior. Economic research extending from utility analysis relies on the relationship between spending and income. In performing research using surveys, the deletion of observations lacking income information leads to a problem called sample selection bias. But, the availability of large data bases containing a vide range of demographic and economic information allows for the examination of the statistical linkage of particular attributes that any affect the reporting of income.

The purpose of this study was to examine and statistically link particular socioeconomic attributes that affect the probability of reporting income information. The approach examined both complete income reporters and incomplete income reporters. The socioeconomic variables included in the model were age, race, sex, education, and occupation of the reference person. Additional variables included whether the household (consumer unit) owned or rented its place of residence, lived in a central city, and region of residence. Binomial logit analysis was used to model the probability of income response completeness. Data from the Interview portion of the 1983 U.S. Consumer Expenditure (CE) Survey were analyzed. Results from this study have important implications for research, income imputation procedure development for the CE Survey, and for improving CE Survey data quality.

KEYWORDS

U.S. Consumer Expenditure Survey; income; probability of income response completeness; socioeconomic variables

INTRODUCTION

Income is an important variable in economic analyses. However, when socioeconomic and income data are to be collected from survey respondents, incomplete responses or nonresponses frequently result. For example, a household may report having received income from employment but any not report the amount. The usefulness of survey data to economic researchers is sensitive to how missing data are treated. Standard analyses of incomplete data assume that missing data are missing at random. In other words, that the missingness depends upon the values of factors in the design but not upon the values of the outcome variable itself (Little and Rubin 1983). Little and Rubin (1983) have stressed the importance of checking the plausibility of this assumption. However, to do this, it is necessary first to identify factors which are related to the incompleteness or nonreporting of data in surveys. Results from such examinations can be used to select procedures for analysis which account for nonrandom missingness, to develop data imputation methods which account for the potential nonrandomness of missing data, and to develop improved data collection procedures.

The purpose of this study was to examine and statistically link particular socioeconomic attributes to the probability of re porting income information. The approach examined both complete income reporters and incomplete income reporters. This study used data from the Consumer Expenditure (CE) Survey which collects a large amount of information, both demographic and economic, on U.S. households or consumer units

Our concern in this research was partial or item nonresponse, not total or interview unit nonresponse. Examining this issue for the CE Survey is of interest because of three primary

1/ A consumer unit comprises either : (1) all members of a particular household who are related by blood, marriage, adoption, or other legal arrangements; (2) a person living alone or sharing a household with others or living as a roomer in a private home or lodging house or in permanent living quarters in a hotel or motel, but who is financially independent; or (3) two or more persons living together who pool their income to make joint expendi care decisions. Financial independence is determined by the three major expense categories: ture food, and other living expenses. To be considered financially independent, at least two of the three major expense categories must be provided by the respondent.

In the majority of cases, there is one consumer unit per household (U.S. Department of Labor 1986).

reasons. First, the CE Survey is the only national survey which collects detailed information on total consumer expenditures. Consumer unit characteristics are also collected so that expenditures can be related to the characteristics. As such, the data set is a rich source of information available to researchers to conduct economic analysis. Therefore, it is important that researchers are familiar with the data base and the underlying assumptions concerning the distribution of missing values. Second, since income is an important variable used for Bureau of Labor Statistics (BLS) tabulations and research, there is interest in developing an income imputation procedure for the CE Survey. An important goal in formulating an imputation method is to minimize nonresponse bias. And third, devising improved data collection procedures to overcome the nonreporting of income is a constant goal of the Bureau to improve data quality.

For this study of the probability of complete income response, consumer units were identified as complete or incomplete income reporters. This designation vas used since the BLS separates complete income reporters from incomplete income reporters for the purpose of classifying consumer units by income. In general, complete income reporters were defined to be consumer units that reported a non-zero amount for a major source of income or a non-zero amount for other sources of income, while refusals or don't know responses identified consumer units as incomplete income reporters (a more specific definition is presented later). This approach vas followed since missing income in the CE Survey is a multivariate problem. For example, a consumer unit may provide non-zero values for interest earned but may not report wages and salaries. For the CE Survey, the income question covers four broad types of income: wages and salaries; self-employment income; retirement income; and all other income. The components which are missing can vary across nonrespondents.

We hypothesized that the probability of income response completeness was a function of various socioeconomic characteristics. The socioeconomic variables included in our model were the age, race, sex, education, and occupation of the reference person. These characteristics were selected on the basis of a priori notions of the factors that might influence income response behavior. Reference person characteristics were used for the analysis since the consumer unit's characteristics are most often identified by the reference person's. Additional variables included whether the consumer unit owned or rented its place of residence, lived in a central city or outside a central city, and lived in the Northeast, South, Northcentral, or heat. Binomial logit analysis was used to model the probability of income response. Data from the Interview portion of the 1983 CE Survey were analysed (see the Methods and Procedures section for a description of the Survey).

This paper is organized into four remaining sections. In the Background section, the issue of nonresponse and its implications for estimation are briefly discussed. The Methods and Procedures section includes a description of the data source, dependent variable, independent variables and hypotheses, and estimation procedure. The Results section is divided into three parts: distribution of variables, logit results, and the probability of income response completeness and impact of changing characteristics. In the last section, the Summary, findings from the study are reviewed and implications for future research are noted.

BACKGROUND

Nonresponse is an issue of concern in virtually all large-scale household surveys, primarily because of its impact on estimation results. As noted by Champney and Bell (1982), nonresponse can be handled for analysis in one of three ways: calculations can be restricted to only those respondents with complete data; calculations can be restricted to only those variables with reported values; or a missing value imputation procedure can be applied to estimate what nonrespondents would have reported had they answered the question. For each approach, a number of assumptions are implicitly made concerning the nature of the nonresponse. Violations of the assumptions can obviously affect one's results. If complete cases are not a random subsample of the original population, restricting the analysis to only these cases would involve loss of efficiency and may yield biased results. This is frequently a problem in panel surveys (for references see Little and Schluchter 1985). If missing values are imputed, standard multivariate analyses of the imputed data are liable to distortions and these distortions may be difficult to assess (Little and Schluchter 1985).

The majority of imputation procedures in use currently are for the case in which the missing values are missing at random (David, et al. 1983). An example is the "hot deck" procedure used by the U.S. Bureau of the Census to impute income for nonrespondents in the Current Population Survey (CPS). For this procedure it is "mused that there are correlations between income and observable characteristics in the population. These characteristics are used to match income respondents and nonrespondents. In general, nonrespondents of income are assigned the income of statistically matched respondents. Thus, the "procedure involves the assumption that the reason income is missing

2/ The reference person is the first member of the consumer unit mentioned by the respondent when asked to "Start with the name of the person or one of the persons who owes or rents the home." (U.S. Department of Labor 1986).

may be associated with any number of income correlates but has nothing to do with income itself"(Lillard, Smith. and belch 1986, p. 490).

In many item nonresponse cases, the assumption that the data are missing at random does not seem justified. Even after adjusting for available covariate information, several researchers (see Greenlees, Reece, and Zieschang 1982; Lillard, Smith, and Welch 1982) found that the probability of nonresponse to wage and salary items and earnings questions in the CPS was likely to depend upon unrecorded income values. Thus, the distribution of income among respondents and nonrespondents was likely to differ in some unknown ray even after taking account of covariates.

Results from previous studies were used as a basis for identifying variables to include in the probability of income response model. These studies included examinations of the probability of reporting earnings or wages and salaries for the CPS (Greenless, Reece. and Zieschang 1982; Lillard, Smith, and Welch 1986), income imputation procedures (Little and Samuhel 1983), nonresponse rates in the Survey of Income and Program Participation, SIPP (Coder and Feldman 1984), and factors affecting survey data quality (Andrews and Herzog 1986). Previous income reporting studies focused primarily on the reporting of earnings or wages and salaries of individuals. Often the samples were restricted to the working age population and to males. An exception vas provided by Coder and Feldman (1984) in their examination of income nonresponse for SIPP; these researchers produced nonresponse rates for several sources of income for individuals aged 15 and older. In contrast to previous studies, our unit of analysis vas the consumer unit, with all consumer unit members aged 14 and over requested to provide income information.

METHODS AND PROCEDURES

Data Source

The data used in this study were drawn from the interview portion of the 1983 Consumer Expenditure (CE) Survey. The CE Survey data are collected by the Bureau of the Census under the auspices of the BLS. The Survey is the most comprehensive source of data available on the expenditures, income, and corresponding socioeconomic and demographic characteristics of American consumer units. The interview sample, selected on a rotating panel basis, is targeted at 4,800 consumer units per quarter. Each quarter one-fifth of the sample is new to the survey. After being interviewed for five consecutive quarters, each panel is dropped from the survey. Detailed income data are collected during interviews two and five only. For the purpose of this study, the sample vas defined as all consumer units participating in a second interview during 1983. Because consumer units living in rural areas outside Standard Metropolitan Statistical Areas (SMSA's) were not surveyed in that year, they were not part of the study sample.

Dependent Variable

Income reporting was defined in terms of the completeness of income information obtained from consumer units. The distinction between a complete and an incomplete income reporter, used in this analysis (and by BLS in its publications of CE Survey data), was based on whether the respondent provided values for various sources of income. Sources were grouped into two categories: major sources of income and other sources of income.

Major sources of income include: wages and salaries income from non-farm business partnership or professional practice income or loss from own farm Social Security or Railroad Retirement Supplemental Security Income.

Other sources of income include[3]: unemployment compensation workmen's compensation and veteran's payments including educational benefits public assistance or welfare receipts including income from job training grants such a Job Corps interest received on savings accounts or bonds regular income received from dividends, royalties, estates, or trusts income received from pensions or annuities from private companies, military, or government

[3]/Additional income, not included in "Other Sources of Income" for the purpose of identifying consumer units as complete or incomplete income respondents, includes money received from the care of foster children, cash scholarships and fellowships or stipends not based on earnings and the annual value of food stamps. Income from these sources, however, is added to the income received from the sources noted in the text to calculate a consumer unit's income before taxes. Income before taxes is the income variable used in CE Survey publications to classify consumer units.

income or loss received from roomers or boarders
income or loss received from payments from other rental units
regular contributions received from alimony or child support and other sources combined.

A consumer unit was defined as a complete income reporter if:

the reference person had a non-zero mount reported for a major source of income, any entry for a major source of income for at least one other consumer unit member vas recorded, and any entry for other sources of income was recorded; or

consumer unit member(s) other than the reference person had a non-zero mount reported for a major source of income, valid zeros or valid blanks were recorded for all the major sources of income for the reference person, and any entry was recorded for other sources of income; or

consumer unit members) had a non-zero amount reported for at least one other source of income and valid zeros or valid blanks were recorded for all major sources of income for all members.

A valid zero or valid blank was recorded if a consumer unit member did not receive income from a source. For example, if the reference person did not receive income from wages and salaries, a valid blank would be recorded in the data base for this person for wages and salaries. Given this definition, it was possible even for complete income respondents not to have provided a full accounting of all income from all sources. For example, if Social Security recipients had the the value of their Social Security checks reported, the consumer unit was considered a complete income reporter, although the individuals neglected to report interest earned on savings. Consumer units with other combinations of entries to the income questions were considered to be incomplete income respondents. In the extreme case of across-the-board zero income, the response was considered invalid and also constituted an incomplete income report.

Independent Variables and Hypotheses

Eight socioeconomic factors were included in the model as independent variables. Five were characteristics of the reference person in the consumer unit: age, race, sex, education, and primary occupation. The other three were consumer unit residence characteristics: housing tenure, degree urban, and region. All independent variables except those for age entered the model as categorical variables. Previous research (Andrews and Herzog 1986; Coder and Feldman 1984; Greenlees, Reece, and Zieschang 1982; Lillard, Smith, and Welch 1986; Little and Samuhel 1983) was used to provide a framework upon which to build the model and to develop hypotheses. A variable could influence income response completeness independently of the value of income or the value of income could indirectly affect response completeness through its affect on the independent variable. However, no attempt was made in this study to isolate the direct effects of the independent variables on response completeness probabilities. Definitions of all variables included in the model are presented in Table 1. Each variable and hypothesized effect are discussed.

Age. It was hypothesized that consumer units with elderly reference persons would be less likely to be complete income reporters. This my be related to the values and types of their incomes or to their general response to the survey. For example, consumer units with older reference persons may be less willing to divulge income amounts or they may not be willing or able to produce dollar amounts as necessary to be identified as complete income reporters. As the interview progresses, individuals may tire and refuse to answer the income questions which are placed at the end of the CE Survey Interview questionnaire. Previous researchers (Greenlees, Reece, and Zieschang 1982), analyzing the CPS data, reported that older employed individuals were less likely to respond to CPS wage and salary questions than were the younger. Andrews and Herzog (1986), conducting research on survey measures, found that age was the demographic factor associated with the largest differences in the measurement of the quality of survey data. They suggested that elderly individuals may answer with less precision than do younger individuals in surveys. The researchers also stated that elderly individuals may be more influenced by the particular format used for presenting questions or eliciting answers. To account for possible nonlinear effects of age, an age squared term was included in the model.

Race. Race was expected to be an important variable in the response model. For this study race was defined in terms of two groups: (1) black and (2) white and other. The other group comprises such

Table 1. Definitions of Variables

Dependent Variable

Response: Unity if consumer unit responded to income questions which identified the unit as a complete income reporter; zero if incomplete income reporter.

Independent Variables

Age of Reference Person: Age in years.

Age Squared: Age of reference person squared.

Race of Reference Person: Unity if black; zero if white or other.

Sex of Reference Person: Unity if male; zero if female.

Education of Reference Person

No school: Unity if never attended school; zero otherwise.

Elementary: Unity if 1-8 years of schooling completed or less than a high school graduate; zero otherwise.

College: Unity if a college graduate (4 years); zero otherwise.

Postgraduate: Unity if more than 4 years of college completed; zero otherwise.

(Omitted category was High School: reference person is a high school graduate or has completed some years of college.)

Primary Occupation of Reference Person

Sales: Unity if person received the most earnings in the past 12 months from employment in a technical, sales, or administrative support occupation; zero otherwise.

Services: Unity if person received the most earnings in the past 12 months from employment in a services, farming, forestry, or fishing occupation or in the armed forces; zero otherwise.

Laborer: Unity if person received the most earnings in the past 12 months from employment as an operator, fabricator, or laborer; zero otherwise.

Craft: Unity if person received the most earnings in the past 12 months from employment in a precision production, craft, or repair occupation; zero otherwise.

Self-employed: Unity if person received the most earnings in the past 12 months from self-employment; zero otherwise.

Retired: Unity if person was retired in the past 12 months; zero otherwise.

Not Working or other: Unity if person was not working in the past 12 months or did not respond to occupation question; zero otherwise.

(Omitted category was Salaried Professional or Manager: reference person received the most earnings in the past 12 months from employment in a managerial or professional specialty occupation.)

Housing Tenure: Unity if consumer unit owned home; zero if consumer unit rented.

Degree Urban: Unity if consumer unit resided in the central city of a SMSA; zero if consumer unit resided inside a SMSA in other places of 50,000 or over, places less than 50,000 and other urban territories, urban places of 2,500 to 50,000 outside an urbanized area, rural non-farm areas, or on a rural farm, or outside a SMSA in urbanized areas and urban places of 2,500 to 50,000 outside an urbanized area.

Region

Northeast: Unity if consumer unit resided in the Northeast region; zero otherwise.

Northcentral: Unity if consumer unit resided in the Northcentral region; zero otherwise.

West: Unity if consumer unit resided in the West region; zero otherwise.

(Omitted category was South: consumer unit resided in the South region.)

Constant: Unity for all observations.

races a American Indiana, Alaskan natives, Asians, and Pacific Islanders. Race is a variable commonly used by labor economists in earnings equations, along with age, sex, education, experience, and region of residence (Lillard, Smith, and Welch 1986; Greenlees, Reece, end Zieschang 1982). The CPS hot deck procedure for imputing income includes race, among other variables, u a covariate to classify nonrespondents (Little and Samuhel 1983). However, Greenlees, Reece, and Zieschang (1982) reported that race was not significant in their probability of wage response equations, tie made no specific hypothesis with respect to race due to possible interactions with other variables in the model.

Sex. Like race, sex was expected to be an important variable in the response model. Be: vas defined as that of the reference person. In the CE Survey, the reference person could be male or female. Differentials in income, particularly earnings, exist for males and females with females, on average, earning less than miles. Differences in consumer wait income response completeness may result from differences in the income of sale and female reference persons. However, male sad female reference person units may differ in their responses due to varying degrees of awareness, within the units, of the income of family members. One could also hypothesize that consumer units with reference persons of a certain sex are more adept at recordkeeping or are more willing to cooperate with survey interviewers than are units with reference persons of the other sex. Unlike the majority of studies examining the probability of reporting earnings (e.g., Greenlees, Reece, and Zieschang 1982; Lillard, Smith, and Welch 1986) which focus primarily on the reporting of sales, we were interested in determining whether differences in reporting probabilities of completeness exist for consumer units with sale or female reference persons. Therefore, we made no hypothesis concerning whether consumer units with male or female reference persons mould be more likely to be complete income reporters.

Education. It vas hypothesized that consumer units with more educated reference persons would be less likely to be complete income reporters than would consumer units with less educated reference persons. Education of the reference person was defined in terms of five levels of education: (I) never attended school; (2) one to eight years of schooling (elementary school) or some high school completed; (3) high school graduate or same years in college completed; (4) four year college graduate; and (5) more than four years of college. In previous studies (Greenlees, Reece, and Zieschang 1982; Lillard, Smith, and Welch 1986) of the probability of reporting earnings, increases in the years of education completed lead to decreases in the response probability. Lillard, Smith, and Welch (1986) specifically stated that attending high school or beyond may have led to a decrease in the reporting of earnings. The importance of including education as a differentiating variable for respondents and nonrespondents of income vas highlighted by these researchers (Lillard, Smith, and Welch 1986) in their review of Census earnings allocations:

> During the 1968-75 period ... By not using schooling [in the imputation algorithm], we overstate the incomes of nonreporters who have little schooling and understate the incomes of nonreporters with high levels of schooling completed. (p.500)

Although consumer units with educated reference persona may be better equipped to answer detailed income questions, there may be reasons why no response or partial responses result. The more educated may have both higher incomes and more varied sources of income which require sore detail in their income response. In addition, the more educated my place a higher value on their time and privacy than do the less educated. Thus, we would expect increases in education to be negatively related to the probability of complete income response.

Occupation. Occupation was expected to be an important socioeconomic variable in the complete income response model. Occupation was defined in terms of the reference person's employment and earnings status during the past 12 months. Occupations were divided into eight categories. Working individuals were grouped into one of six different categories depending upon the type of employment from which they received the most earnings in the past 12 months. These categories were: sales, services, laborer, craft, self-employed, and salaried professional or manager. A detailed description of each employment category is also presented in Table 1. Two additional occupational categories were included, one for retirees and the other for reference persons not working or not responding to the occupation question. It vas hypothesized that consumer units with self-employed reference persons would be the least likely to be complete income respondents. Coder and Feldman 01984), when examining data from the 1983 SIPP, reported an income nonresponse rate of 14 percent for self-employed individuals. In contrast, they reported a 6.2 percent income nonresponse rate for individuals with monthly wage or salary income. When examining the proportion of nonreporting white males by type of earnings using 1980 CPS data, Lillard, Smith, and Welch (1986) found that 25.7 percent of the self-employed individuals in non-farm occupations and 24.7 percent of the self-employed individuals in farming occupations were nonreporters of earnings, compared to 16.3

percent of wage and salary workers. The researchers noted that there are several occupations in which nonreporting is considerably higher than the average, and that these occupations share one or both of the following characteristics: "they are among the highest income occupations, or considerable ambiguity surrounds the calculation of net income from receipts and expenses for income tax purposes" (Lillard, Smith, and Welch 1986, p.492). Lawyers and judges, dentists, and doctors fit both criteria and approximately one-third of each group does not respond to income questions. Farmers and private household workers, although rich lower incomes on average, also experience tome difficulty in calculating their income for tax purposes. Approximately one-fourth of the individuals in each of the groups (lawyers and judges, dentists, doctors, farmers, and private household workers) noted by Lillard, Smith, and Welch (1986) were nonreporters of income. Individuals in these groups are expected to account for a large majority of the self-employed. Thus, since earnings account for a major portion of income, on average we would expect consumer units with self-employed reference persons to be the least likely to be complete income reporters.

Housing Tenure. Housing tenure was defined as whether a consumer unit owned or rented its place of residence. Homeownership was hypothesized to be negatively related to the probability of being a complete income reporter. This is based on the assumptions that high income consumer units are most likely to be incomplete income reporters and that high income consumer units are also more likely to be homeowners than they are to be renters.

Degree of Urbanization. The degree of urbanization was expected to be an important variable in the response probability model. Degree of urbanization was defined in terms of whether the consumer unit resided within a central city or outside a central city. As noted earlier, consumer units living in rural areas outside SMSA's were not surveyed in 1983; thus, they were not included in the outside a central city category, tie would expect consumer units living outside the central city to be less likely to be complete income reporters if consumer units in these areas have higher incomes than do units living in central cities. This is based on the assumption that completeness of income response is influenced by the sine of income. Greenlees, Reece, and 2ieschang (1982) reported that the earnings of individuals in the suburbs of SMSA's were higher than the earnings of individuals residing in central cities. However, our sample vas not restricted to wage and salary workers. In addition, complete income reporters in this study included consumer units with members receiving retirement or public assistance income. Consumer units with these sources of income may be, on average, more prevalent in the central city. Thus, no hypothesis was made concerning the relationship of the degree of urbanization wish reporting propensities.

Region. Region was the final variable included in the model which was expected to be related to complete income response. Each consumer unit was identified as living in one of four regions: Northeast, Northcentral, South, or West. Previous researchers (Lillard, Smith, and Welch 1986; Greenlees, Reece, and Zieschang 1982), using Census data, have noted the importance of region in determining the reporting propensities of earnings. Lillard, Smith, and Welch (1986) reported that living in the South had a strong positive independent effect on reporting propensities, even after controlling for the variable's influence through earnings.

Greenlees, Reece, and Zieschang (1982) found that individuals living in the South or West were most likely to answer the CPS wage and salary questions. Consumer units may differ by region in their willingness to provide income information. This difference may be due to a more general willingness on the part of consumer units within certain regions to respond to income questions. It was hypothesized that consumer units residing in the Northcentral or in the Northeastern regions of the country would be less likely to be complete income reporters than consumer units living in the South and that those living in the West would be more likely.

In summary, we identified several socioeconomic characteristics thought to be related to the probability of complete income response. These included the age, race, sex, education, and principal occupation of the reference person, whether the consumer snit owns or rents its place of residence, lives within a central city or outside a central city, and lives in the Northeast, Northcentral, South, or West. Factors we hypothesized to be characteristic of consumer units least likely to be complete income reporters included: older age, sore education, and self-employment of the reference person; sad homeownership and consumer unit residence in the Northeasters or Northcentral regions of the country.

Estimation Procedure

The statistical analysis of the probability of complete income response was based upon a binomial logic model.[4] Binomial logit analysis is a statistical technique wed to analyse discrete choice

[4] There are numerous references in the literature to logic analysis. See Domencich and McFadden (1975), Judge et al. (1982), Maddala (1977), Maddala (1983), Pindyck and Rubinfeld (1981).

problems of a binary format. Binomial logit involves the basic principles of regression analysis, namely fitting the slope and intercept of a regression line with maximum likelihood estimation, but adjusts for the respecification of the dependent variable into a binary form.

In this study the model under consideration was

$$P_i = \text{Prob}(Y_i = 1) = F(X_i\beta), \text{ and} \qquad (1)$$

$$1-P_i = \text{Prob}(Y_i = 0) = 1 - F(X_i\beta) \qquad (2)$$

where the $F(X_i\beta)$ is a cumulative distribution function that describes how the probabilities of complete income reporting and incomplete income reporting respectively, are related to the socioeconomic variables. Pi is the probability that the i-th consumer unit is a complete income reporter, Xi is the vector of characteristics of the i-th consumer unit, and β is the vector of unknown parameters. The binomial logit model assumes a cumulative logistic probability distribution for the underlying function. The probability of a complete income response is defined mathematically as

$$Pi = F(Xi\beta) = \frac{1}{1+e^{-(X_i\beta)}} \qquad (3)$$

The Xi's are considered to be oberservations on nonstochastic variables which are independent of each other. The error terms implicit in the model are assumed to follow the Weibull or extreme value distribution and are assumed to be independent. The probability that the i-th consumer unit is not a complete income reporter is mathematically defined as

$$1 - Pi = \frac{e^{-(X_i\beta)}}{1+e^{-(Xi\beta)}} \qquad (4)$$

The logit is obtained by transforming the zero to one interval probability to a log odds ratio which ranges from $-\infty$ to $+\infty$ as $X_i\beta$ goes from $-\infty$ to $+\infty$. The logit is derived below:

$$\text{logit} = \log \frac{P_i}{1-Pi} = \log P_i - \log (1-P_i)$$

$$= -\log(1+e^{-(X_i\beta)}) - [\log (e^{-(X_i\beta)}) - \log(1+e^{-(X_i\beta)})] \qquad (5)$$

$$= X_i\beta$$

Thus, the logit, or logarithm of the odds that a particular alternative will result, is $X_i\beta$. The logit is a linear function of the explanatory variables, but the probabilities are not.

Maximum likelihood estimation is frequently the preferred statistical approach used to estimate the logit coefficients. If we let N_0 be the subset of observations for which the discrete choice Y_i = 1 (complete income reporter) and N_1 be the subset of observations for which Y_i = 0 (not a complete income reporter), the likelihood function L, of the sample can be written as

$$L = \prod_{i\varepsilon N_0} P_i \cdot \prod_{i\varepsilon N_1} (1-P_i)$$

$$= \prod_{Y_i=1} F(X_i\beta) \cdot \prod_{Y_i=0} [1-F(X_i\beta)] \tag{6}$$

$$= \prod_{Y_i=1} \frac{1}{1+e^{-(X_i\beta)}} \cdot \prod_{Y_i=0} \frac{e^{-(X_i\beta)}}{1+e^{-(X_i\beta)}}$$

The logarithmic likelihood is

$$\log L = \sum_{i\epsilon N_0} \log F(X_i\beta) + \sum_{i\epsilon N_1} \log[1-F(X_i\beta)] \tag{7}$$

The expected value of Y_i can be written as

$$E(Y_i) = 1 \cdot F(X_i\beta) + 0 \cdot [1-F(X_i\beta)] \tag{8}$$

$$= F(X_i\beta)$$

Therefore, the expected value of Y_i is the value of the logistic distribution which necessarily falls in the unit interval. The maximum likelihood estimates of the β's are found by setting the derivaties of the likelihood function with respect to the β's equal to zero.

Maximum likelihood parameter estimates for the probability of income reporting model were obtained by using the interative Newton-Raphson optimization procedure. The computer software package (LOGIT) used for the analysis was developed by Antos (1983).

RESULTS

Results of the logic analysis of income response completeness are presented in this section. The sample used for the analysis is described, followed by a description of the results of the logit estimation and tests of the model specification. The estimated parameters are then used to calculate income response probabilities for different consumer units and changes is response probabilities.

Distribution of Variables

In 1983, 4,611 consumer units participated is a second interview of the CE Survey. The scan values and percent distribution of independent variables used in the logit analysis are presented in Table 2. The average age of reference persons in the sample consumer units vas 46 years. The majority of reference persons were white includes a small number of non-white non-blacks) and were hale. More than 50 percent were high school graduates and approximately one-fourth were salaried professionals or managers. The majority of consumer units in the sample owned their own hoses and lived outside a central city. Consumer units were fairly equally distributed among regions.

Incomplete income reporter consumer units accounted for approximately 13 percent of the sample. An examination of only the lean values and percent distribution of variables reveals that incomplete income respondents in the sample were different from complete income respondents. Incomplete income respondent reference persona were, on average, *lightly older and were sore likely to be hale than were complete income respondent reference persons. They were also sore likely to be college graduates and to be among the self-employed, when considering the educational and occupational categories as defined in this study. Incomplete income respondent consumer units were sore likely to be homeowners and were sore likely to live outside a central city. Consumer unite living-in the

Table 2. Mean Values and Percent Distribution of Variables in Logit Analysis

Independent Variable	Full Sample (n=4611)	Complete Income Reporter (n=4018)	Incomplete Income Reporter (n=593)
Age of Reference Person (continuous)			
Age	45.91	45.59	48.03
Age squared	2428.23	2404.79	2587.07
Race of Reference Person			
Black	11.08	11.03	11.47
White and other	88.92	88.97	88.53
Sex of Reference Person			
Male	67.73	67.10	72.01
Female	32.27	32.90	27.99
Education of Reference Person			
No school	0.54	0.47	1.01
Elementary	25.53	25.91	22.93
High School	52.11	52.27	51.10
College	11.00	10.55	14.00
Postgraduate	10.82	10.80	10.96
Principal Occupation of Reference Person			
Salaried Professional or Manager	21.66	21.73	21.24
Sales	17.76	18.09	15.51
Services	8.72	9.23	5.23
Laborer	11.88	12.10	10.46
Craft	6.68	6.72	6.41
Self-employed	6.12	5.00	13.66
Retired	15.25	15.43	14.00
Not working and other	11.93	11.70	13.49
Housing Tenure			
Owned	59.57	58.19	68.97
Rented	40.43	41.81	31.03
Degree Urban			
Inside a central city	34.85	35.09	33.22
Outside a central city	65.15	64.91	66.78
Region			
Northeast	22.53	20.81	34.23
South	27.83	28.70	21.92
Northcentral	26.15	25.73	29.01
West	23.49	24.76	14.84

Northeastern and Northcentral regions of the country were more likely to be incomplete income respondents than were consumer units living in the South or West. In this study reference persons in incomplete and complete income respondent consumer units did not differ in terms of their race.

Logit Results

Results of the logit analysis are displayed in Table 3. All variables included in the model, their coefficients, and asymptotic standard errors are presented. Results of the statistical tests used to describe the explanatory power of the model are also presented.

To test the overall significance of the set of variables included in the model, the likelihood ratio statistic was used.[5] The resulting Chi square value was significant at the $\alpha = 0.01$ level. The

[5] The test statistic is $\chi^2 = -2(\log \text{Likelihood}_R - \log \text{Likelihood}_U)$. The statistic is asymptotically Chi-square distributed with the degrees of freedom equal to the number of coefficients set equal to zero. The log likelihood function for the restricted model, represented by R, is obtained when the function is maximized with respect to the intercept only. The log likelihood of the unrestricted model, U, is obtained when the function is maximized with respect to all the coefficient estimates corresponding to the intercept and all explanatory variables.

null hypothesis that all of the coefficients (except the intercept) are equal to zero was rejected. This means that at least one of the independent variables was important in explaining the probability of complete income response.

Table 3. Estimated Model Parameters and Standard Errors

Independent Variable	Estimated Parameter	Asymptotic Standard Error
Age of Reference Person		
Age	-0.0463*	0.0174
Age squared	0.0004**	0.0002
Race of Reference Person (White and other)		
Black	-0.2709***	0.1538
Sex of Reference Person (Female)		
Male	-0.1257	0.1092
Education of Reference Person (High school)		
No school	-0.7433	0.4910
Elementary	0.1479	0.1220
College	-0.3684**	0.1457
Postgraduate	0.0185	0.1636
Principal Occupation of Reference Person (Salaried Professional or Manager)		
Sales	0.0180	0.1568
Services	0.3513	0.2226
Laborer	0.0644	0.1833
Craft	-0.0749	0.2123
Self-employed	-1.0839*	0.1737
Retired	0.0810	0.2048
Not working and other	-0.2467	0.1826
Housing Tenure (Rented)		
Owned	-0.2601**	0.1109
Degree Urban (Outside a central city)		
Inside a central city	-0.0831	0.1016
Region (South)		
Northeast	-0.8287*	0.1265
Northcentral	-0.3702*	0.1276
West	0.2166	0.1491
Constant	3.7568*	0.4143
Likelihood Ratio Statistic	182.253> 37.57	
Likelihood Ratio Index	0.052	

*Statistically significant at the 0.01 level.
**Statistically significant at the 0.05 level.
***Statistically significant at the 0.10 level.

The likelihood ratio[6] index was calculated as a measure of goodness-of-fit of the binomial logit model and is analogous to the R-squared in ordinary least squares regression. The index is a measure of how well the model approximates the observed data. An index of 0.052 was obtained for the response model. Although this value seems low, it may be reasonable since values of the index between 0.2 and 0.4 are considered extremely good fits (Hensher and Johnson 1981).

[6] /The index is

$$\rho^2 = 1 - \frac{\log \text{Likelihood}_U}{\log \text{Likelihood}_R}.$$

Generally the likelihood ratio index has an upper bound of about 0.3; it is unlikely that an index value would approach one because that could only happen if all individuals predicted probabilities were exactly zero or one (Kinsey and Lane 1978; Pindyck and Rubinfeld 1981; Tardiff 1976).

When examining individual coefficients, there are several points to keep in mind. Unlike a regression coefficient, a logit parameter estimate itself does not directly provide information concerning the quantitative effect of a change in an independent variable on the probability of belonging to one group as opposed to another (e.g., complete versus incomplete income reporters). Instead, a logit coefficient measures the change in the log of the odds (e.g., Pi/(1-Pi)) associated with a unit change in the independent variable. Relative values of the coefficients, however, are meaningful within a variable grouping and thus they can be ranked relative to one another. Thus, it is possible to rank coefficients from the lowest to the highest value and to interpret a coefficient as having a greater effect on the dependent variable than those ranked below it. For example, if parameter A has a coefficient of 0.2 and parameter B a coefficient of 0.4, then parameter B may be interpreted as having a larger effect on the choice probability than parameter A. One cannot, however, assert that 8 has twice the effect on the dependent variable as A. Like regression, the relationship between the independent variables and the probability of belonging to a group can be investigated by examining the signs of the coefficients.

To determine the impact of individual variables on the log (Pi/(1-Pi)), the relative values and signs of the coefficients were examined. In addition, asymptotic t-tests were used to determine the significance of individual coefficients at the 0.01, 0.05, and 0.10 percent levels of significance.

All independent variables except age entered the binomial logit equation as dummy variables. It vas necessary, therefore, to omit one of the dummy categories for each variable before estimating the final equation. In Table 3 each omitted category is identified in parentheses next to the variable name. The result for each dummy variable is interpreted relative to the omitted group. For example, the positive coefficient for cleat implies that consumer units in the West have a greater probability of being complete income reporters than do consumer units living in the South. In contrast, the negative coefficient on Northeast indicates that consumer units in the Northeast are less likely than consumer units in the South to be complete income respondents.

Among the socioeconomic variables included in the probability model, age, age squared, race, one of the education dummy variables (college), one of the occupation variables (self-employed), housing tenure, and two of the region variables (Northeast and Northcentral) were significant. For the most part, these results were consistent with the hypotheses and with the findings of previous researchers.

The negative coefficient for the age variable indicates that as age increased, the probability of complete income response decreased. If consumer units with older reference persons have high incomes and/or receive income from numerous sources, they may be less likely to divulge their incomes or to note dollar amounts received from each income source. Or, they slay be reacting to the questionnaire design or interview procedure. However, the coefficient for age squared was positive, although small. This indicates that for most of the sample, as age increased, the effect of age on the probability of complete income reporting (while negative) diminished.

Consumer units with black reference persons were less likely than those with reference persons of other races to be complete income reporters. There was no a priori reason why this might be the case.

Of all the education variables, only college was significant in the probability model. Consumer units with college educated reference persons were significantly less likely than those in the omitted category (high school degree or ease college) to be complete income reporters. This we consistent, in general, with the hypothesis and findings of previous researchers. Although the other education coefficients were not significant, the bimodal result for the signs relative to the probability of complete income response may indicate that the effects of education on the probability are nonlinear or that there is an interaction of education and other variables in the model.

Self-employment of the reference person vas an important variable in the probability of complete reporting model. Consumer units with self-employed reference persons were significantly less likely to be complete income reporters than were consumer units with salaried professional or manager reference persons. Consumer unit with reference persona in the other occupations did not differ significantly from those represented by the omitted group (salaried professional or manager) in their completeness of income response. Coder and Feldman (1984) reported a similar finding for SIPP: nonresponse rates for individuals with self-employment income exceeded the rates for individuals with income from wages and salaries. A related finding was also reported by Lillard, Smith, and Welch (1986) in their examination of the proportion of nonreporting white sales by type of earnings.

Homeownership was negatively related to the probability of complete intone response. This finding was consistent with the hypothesised relationship. However, it is unlikely that homeownership exerts an independent effect on the propensity to respond, Whether the relationship is exerted rather through income is a subject for future research.

Coefficents for two of the region variables, Northeast and Northcentral, were statistically significant. Consumer units living in the Northeast and those living in the Northcentral regions of the country were significantly less likely than those living in the South to be complete income reporters. A ranking of the coefficients indicates that consumer units living in the Northeast were less likely to be complete income reporters than were those living in the Northcentral region. These differences may be related to regional patterns of cooperation or to differences in the value of income received or types of income received.

Tests for the combined contribution of variables represented by more than two dummy variables (u in the case of education, occupation, sad region) or by more than one continuous variable (as in the case of age) were performed using the likelihood ratio statistic. This procedure was followed since the variables in each set were expected to be interrelated, and because the combined influence of the set of variables was of interest. The results are supported as indicated by the Chi square statistic values in Table 4.

Table 4. Chi Square Tests for Contribution of Sets of Variables

Independent Variable Set	Chi Square Statistic	Degrees of Freedom
Age of Reference Person	9.20**	2
Education of Reference Person	11.52**	4
Principal Occupation of Reference Person	59.18*	7
Region	72.74*	3

*Statistically significant at the 0.01 level.
**Statistically significant at the 0.05 level.

The likelihood ratio statistic was computed for four separate sets of restrictions on the model, each testing the joint significance of a set of relevant parameters. All of the variable sets contributed significantly to the income response model as explanatory variables. The set of parameters for age (age and age squared) was statistically significant as expected from the t-teats. The combined influence of the education parameters was significant in determining the probability of response completeness, although only the college parameter was significant when individual coefficients were examined. A similar finding resulted for occupation: the occupation parameter set was significant although only the parameter for the self-employed was significant according to the t-tests. As expected, region as a set of variables was important in the income response model.

Probability of Income Response Completeness and Impact of Changing Characteristics

Since logit coefficients cannot directly provide a quantitative assessment of an independent variable's impact on the probability of an event occurring, an alternative procedure was followed. First, the estimated coefficients and consumer unit characteristics (at the sample means) were used to compute the probability of complete income reporting according to the logistic cumulative distribution function:

$$P_i = \frac{1}{1 + e^{-(X_i \beta)}} .$$

The probability of being an incomplete income reporter was calculated as (1-Pi). Next the values of selected variables were altered to determine the change in the probabilities. This procedure was used to predict the effect of a change in an independent variable on the probability of being a complete income reporter.

On average, consumer units participating in a second interview of the CE Survey in 1983 had a 0.8849 probability of being complete income reporters (Table 5). This value represents an "average" response probability for the full sample.

The impact of changes in the independent variables on the probabilities vas examined by calculating the probabilities at particular values of the independent variables and then changing one value while holding all others constant. This procedure was employed since the majority of the independent variables are categorical sad thus only non-marginal changes are meaningful. Only the variables found to be significant in individual t-tests or those belonging to nets of variables that were significant were changed in order to observe the effect on the probabilities.

To assess the impact of changes, a representative consumer unit vas identified for comparison. Characteristics of the representative consumer unit were selected as the mean values of the continuous variables and as the categorical variables with the highest frequency of occurrence for the sampled (see Table 2). The reference person in the representative consumer unit vas 45.9 years of age (the mean of age squared was 2428.28), white (or other nonblack), male, a high school graduate (with perhaps some college), and a salaried professional or manager. The representative consumer unit vas a homeowner, lived outside a central city, and lived in the South. The values representing these characteristics were multiplied by the estimated logit parameters to obtain the representative consumer unit's income response probabilities. Next, the values of the previously identified significant variables or variable sets were changed, one at a time, and the probabilities were recalculated. All changes in the probability values are discussed in reference to the respresentative consumer unit.

Results of the sample probability calculations for complete and incomplete income response are presented in Table 5. Changes in the probabilities were in the same direction as the signs of the estimated parameters. The representative consumer unit had a 0.9018 probabilty of being a complete income respondent.

Table 5. Sample Probability Calculations for Complete and Incomplete Income Reporters

Case Description	Probabilities Complete Income Reporter	 Incomplete Income Reporter
Sample (at the mean)	0.8849	0.1151
Representative Consumer Unit[a]	0.9018	0.0982
Changes to the Representative Consumer Unit		
Age (increase of one standard deviation)	0.8918	0.1082
Black	0.8750	0.1250
No school	0.8136	0.1864
Elementary	0.9141	0.0859
College	0.8640	0.1360
Postgraduate	0.9034	0.0966
Sales	0.9034	0.0966
Services	0.9288	0.0712
Laborer	0.9073	0.0927
Craft	0.8949	0.1051
Self-employed	0.7564	0.2436
Retired	0.9087	0.0913
Not working and other	0.8777	0.1223
Rented	0.9225	0.0775
Northeast	0.8003	0.1997
Northcentral	0.8638	0.1362
West	0.9194	0.0806

[a] In all calculations, the "representative consumer unit" was characterized as follows: age = 45.9 years; (mean of age squared = 2428.23); race = white or other; sex = male; education = high school; occupation = salaried professional or manager; housing tenure = owned; degree urban = outside a central city; region = South.

[7]/Due to the procedure used to identify characteristics, the representative consumer unit for this analysis would be expected to differ from a representative consumer unit for another analysis if definitions for the categorical variables differed. For example, the percentage of consumer units with a reference person in the high school group would decrease if reference persons with some college, but not a four-year degree, were reclassified into the college category which currently refers to four-year college graduates only.

An increase in age (and consequently in age squared) decreased the probability of complete income response only slightly. Holding all other values constant, an increase in age of one standard deviation, equivalent to 17.9 years, and a one standard deviation increase in age squared (1801.84 years) decreased the probability to 0.8918. This was expected given the signs and values of the age coefficients.

With a change in race, the income completeness probabilities changed only slightly. Compared to a consumer unit with a white (or other non-black) reference person and the same representative characteristics, one with a black reference person had a smaller probability of being a complete income reporter (0.875).

A nonlinear pattern emerged in the probabilities with respect to increases in educational attainment. For example, the probability of complete income response decreased to 0.864 with education equal to four years of college, while an increase in education beyond four years of college resulted in a slight increase in the complete income response probability to 0.9034. Consumer units with reference persons reporting no schooling were the least likely to be complete income respondents with a probability of 0.8136.

The greatest impact on the response probabilities was recorded for the self-employed, as vas expected. A change in only the reference person occupation from salaried professional or manager to self-employed resulted in a probability of complete income response of 0.7564. Consumer units with reference persons in service occupations but with the same other representative characteristics were most likely (Pc = 0.9288) to be complete income respondents. If the reference person was not working, the probability decreased to 0.8777. Changes in probabilities were small when the reference person was employed in a sales or craft occupation, as a laborer, or the reference person was retired. Thus, it appears that consumer units with reference persons in these latter occupational groups, holding all else constant, behaved similarly in their response to income, as defined in this study.

A change in housing tenure only marginally affected the probabilities. A renter with the same other characteristics as a homeowner had a probability of being a complete income respondent of 0.9225.

As expected from the logic results, consumer units residing in different regions of the country varied in the probabilities of complete income response. Consumer units residing in the Northeast were the least likely of all consumer units in the regional categories to be complete income respondents; their probability of response was 0.8003. Consumer units residing in the Northcentral region were more likely to be complete income respondents than were consumer units residing in the Northeast. Consumer unite in the West and those in the South were equally likely to be complete income respondents.

SUMMARY

In this section results from the analysis of consumer unit income reporting are reviewed, along with implications for future research. A binomial logit model was tested to identify socioeconomic factors which are related to the completeness of income response in the CE Interview Survey. Completeness of income response, as defined by the Division of Consumer Expenditure Surveys in the BLS, was used to define the dependent variable. Data collected during the second interview in 1983 were analyzed. The sample was composed of 4,018 complete income reporters and 593 incomplete income reporters.

Results of the logic analysis indicated that age, race, education, and occupation of the reference person, housing tenure, and region in which the consumer unit resided were significant variables in determining the probability of being a complete income reporter as opposed to being an incomplete income reporter. With increases in the reference person's age, the probability that the consumer unit was a complete income respondent decreased. However, with increases in age, the effect of age on the probability (while negative) diminished. Consumer units with black reference persons were less likely to be complete income reporters than were consumer units with a nonblack reference person. If the reference person was a college graduate, the probability that the consumer unit would be a complete income reporter decreased. The greatest influence on the probabilities was exerted by the self-employment occupation variable for the reference person. Consumer units in which reference persons were self-employed were the least likely to be complete income reporters, compared to consumer units in which the reference persons were in other occupations. Homeowners were less likely than were renters to be complete income reporters. Residing in the Northeast or in the Northcentral regions also negatively affected the probability that a consumer unit would provide complete responses to the income questions. Neither sex of the reference person nor degree urban contributed significantly to the probability model.

This analysis has allowed us to identify various socioeconomic variables related to income response completeness. This is a first step in determining which factors are important for us to focus upon as we develop ways to increase income response completeness. For this analysis, we did not attempt to test whether the socioeconomic variables influenced income completeness through their effect on income or whether the variables independently influenced income completeness. Our basic assumption was that the completeness of income reporting can be statistically linked to the socioeconomic characteristics of the consumer unit. This is in contrast to income/earnings studies which link the probabilities to characteristics of individuals. Based an the results from our study we can report that complete income reporting consumer units and incomplete income reporting consumer units are different in terms of the socioeconomic variables found to be significant in our model. However, it would be presumptuous for us to say at this time that the pattern of incomplete income reporting is related to the missing income itself. Future research on this issue seems desirable.

Results from this study have important implications for researchers within and outside of the Bureau of Labor Statistics. Researchers interested in using income from the CE Survey need to be aware that complete and incomplete reporters of income are different; this difference may lead to biased estimation results. We at the BLS are interested in selecting or developing an income imputation procedure for the CE Survey. Although we cannot state that the incompleteness of income is related to income itself, we need to be ready to consider model-based procedures which can recognize the response mechanism (for references see David, et al. 1983; Fay 1986). Focusing on factors related to income report completeness is important when revising data collection procedures to improve data quality.

For whatever reason, consumer units which are identified as incomplete income reporters are unwilling or reluctant to provide the requested income information. Reasons for incomplete income response include the inability to understand a question, the mistaken belief that a question does not apply, or the unavailability of the information either from a lapse of memory or lack of knowledge (Fay 1986). There may be a general fear of government or other uses of the data. Incomplete income reporters may have an income-elastic demand for privacy or they may simply place a higher price on their tine for completing the survey Millard, Smith, Welch 1986) compared to complete reporters. Interview length may also be a factor in whether consumer units complete the income questions. Income questions are placed at the end of the Interview questionnaire. The Interview takes, on average, two to three hours to complete. Some survey respondents may become fatigued or frustrated before the income questions are asked. As a result they say refuse to answer the questions entirely, they may provide only basic income information, or they may lie and distort their incomes in order to evade reporting. The presence or absence of others during the interview process may also affect the responses of individuals who regard the divulging of income information as inappropriate (Little and Samuhel 1983). Failure to respond to one question in a survey say induce further nonresponse to subsequent items, thus, nonresponse may lead to further nonresponse (Fay 19861. Fay (1986) rtes that nonresponse may arise from causes which are not restricted to the behavior of the respondents. For example, interviewers may skip questions due to misunderstanding of the questionnaire or avoid questions that embarrass them. Since income of all consumer unit members is to be recorded, incomplete reports may result if other members are not available to be interviewed. Thus, the circumstances of incomplete reporting are varied and complex. Future research is needed to more specifically identify factors which are related to incomplete income response.

Future research could include the testing of various specifications of the probability of income response completeness model. For example, the dependent variable could be defined to represent the three types of complete income reporting situations plus incomplete reporting categories for combinations of refusals and don't knows. Or, the dependent variable could be defined in terms of the reporting of income by source (e.g.. wages and salaries, self-employment income, retirement, and other). Or, the dependent variable could be defined in terms of the reporting of individuals within consumer units. In lieu of the reference person's characteristics, those of the survey respondent could be included as explanatory variables in the model. Additional socioeconomic variables which might be related to income response completeness include marital status of the survey respondent or reference person, number of children, number of persons within the consumer unit with an income source, work status of consumer unit members in terms of fulltime and parttime, and interaction terms far age, education, and occupation. In order to teat hypotheses concerning the data collection features of the survey, administrative variables could be added to the model. These might include the total number of minutes for the interview, month in which the interview was conducted, general survey nonresponse relative to specific income nonresponse, and whether records were used in answering an interviewer's questions.

In conclusion, we repeat, this study must be considered as one of exploration. However, we think that our results are sufficiently promising to warrant future research on the incompleteness of income reporting and the missingness of income in the Consumer Expenditure Survey.

REFERENCES

ANDREWS, FRANK M., and HERZOG, A.R. (1986), "The Quality of Survey Data as Related to Age of Respondent," Journal of the American Statistical Association, 81 (394), 403-410.

ANTOS, JOSEPH (1983), LOGIT package version 3/83, available at the Bureau of Labor Statistics.

CHAMPNEY, TIMOTHY F., and FELL, RALPH (1982), "Imputation of Income: A Procedural Comparison," American Statistical Association, 1982 Proceedings of the Section on Survey Research Methods.

CODER, JOHN P., and FELDMAN, ANGELA M. (1984), "Early Indications of Item Nonresponse on the Survey of Income and Program Participation," American Statistical Association, 1984 proceedings of the Section on Survey Research Methods.

DAVID, MARTIN H., LITTLE, RODERICK, SAMUHEL, MICHAEL, and TRIEST, ROBERT (1983), "Imputation Models Based on the Propensity to Respond," American Statistical Association, 1983 Proceedings of the Section on Business and Economics Statistics.

DOMENCICH, T., and MCFADDEN, D. (I975), Urban Travel Demand: A Behavioral Analysis, Amsterdam: North Holland Press.

FAY, ROBERT E. (1986), "Causal Models for Patterns of Nonresponse," Journal of the American Statistical Association, 81 (394), 354-365.

GREENLEES, JOHN S., REECE, WILLIAM S., and ZIESCHANG, KIMBERLY D., (1982), "Imputation of Hissing Values When the Probability of Response Depends on the Variable Being Imputed," Journal of the American Statistical Association, (Applications Section) 77 (378), 251•261.

HENSHER, D.A., and JOHNSON, L.W. (1981), Applied Discrete-Choice Modelling, London: Croon Helm.

JUDGE, G.G., RILL, R.C., GRIFFITHS, W., LUTKEPOHL, A., and LEE, T. (1982), Introduction to the Theory and Practice of Econometrics, New York: John Wiley and Sons.

KINSEY, JEAN, and LANE, SYLVIA (1978), "The Effect of Debt on Perceived Household Welfare," Journal of Consumer Affairs, 12 (I), 48-62.

LILLARD, LEE, SMITH, JAMES P., and WELCH, FINIS (1986), "What Do We Really Know about Wages? The Importance of Nonreporting and Census Imputation," Journal of Political Economy, 94 (3), 489-506.

LITTLE, RODERICK J.A., and RUBIN, DONALD B. (1983) in Wright, Tommy (ad.), Statistical Methods and the Improvement of Data Quality, Orlando: Academic Press, Inc.

LITTLE, RODERICK J.A., and SAMUHEL, MICHAEL E. (1983), "Alternative Models for CPS Income Imputation, "American Statistical Association, 1983 Proceedings of the Section on Survey Research Methods.

LITTLE, RODERICK J.A., and SCHLUCHTER, HARK D. (1985), "Maximum Likelihood Estimation for Hued Continuous and Categorical Data with Hissing Values," Biometrika, 73 (3), 497-512.

MADDALA, G.S. (1977), Econometrics, New York: McGraw Hill Book Company.

MADDALA, G.S. (1983), Limited-dependent and Qualitative Variables in Econometrics, Cambridge: Cambridge University Press.

PINDYCK, R.S., and RUBINFELD, D.L. (1981), Econometric Models and Economic Forecasts (2nd ad.), New York: McGraw-Hill Book Company.

TARDIFF, T. (1976), "A Note on Goodness of Fit Statistics for Probit and Logit Models," Transportation, 5(4), 377-388.

U.S. DEPARTMENT OF LABOR, BUREAU OF LABOR STATISTICS, (1986), Consumer Expenditure Survey: Interview Survey 1984, bulletin 226), Washington, D.C.: U.S. Government Printing Office.

ACKNOWLEDGEMENTS

Thanks are extended to Paul Hsen, Eva Jacobs, Julie Nelson. Diane Woodbury, and Kimberly D. Zieschang for providing comments on an surlier draft of this paper. Thanks are also extended to Herb Cover for computer assistance, to Anna Sanders for expert typing, to Brent Moulton and Adriana Silberstein for providing input into the development of the paper, and to Kirk Kaneer for preparing presentation materials. Any errors which remain in the paper are the sole responsibility of the authors. This paper was originally presented at the Third Annual Census Research Conference, March 29 - April 1, 1987, Baltimore, Maryland.