THE SURVEY OF INCOME AND
PROGRAM PARTICIPATION

AN ANALYSIS OF ATTRITION IN THE PSID AND SIPP WITH AN APPLICATION TO A MODEL OF LABOR MARKET BEHAVIOR

No. 191

J. E. Zabel

U. S. Department of Commerce BUREAU OF THE CENSUS

An Analysis of Attrition in the PSID and SIPP with an Application to a Model of Labor Market Behavior

Jeffrey E. Zabel

Department of Economics
Tufts University

and

HHES Division
Bureau of the Census

This paper was written while the author was a research fellow at the U.S. Bureau of the Census. I would like to thank Enrique Lamas, Lee Lillard, Amy Ellen Schwartz, and Dan Weinberg for useful comments and participants at the Center for Economic Studies, U.S. Bureau of the Census Seminar Series and the Conference on Attrition in Longitudinal Surveys. The views expressed here do not necessarily reflect those of the U.S. Census Bureau or the Department of Commerce.

Abstract

This paper analyzes attrition behavior in two major longitudinal surveys; the Panel Study of Income Dynamics (PSID) and the Survey of Income and Program Participation (SIPP). Significant indicators in a model of attrition include measures of mobility and variables that correspond to the interviewer and the interview process. There is evidence that surveys of longer duration and with higher frequency interviews experience higher attrition rates. The estimation results for a model of attrition and labor market behavior show little indication of bias due to attrition but there is some evidence that the labor market behavior of attritors and non-attritors is different.

I. Introduction

The increased availability of longitudinal surveys is a major advance in facilitating the empirical analysis of individual behavior over time. One aspect of panel data sets that has received relatively little attention is the exit behavior, or attrition, of individuals from these longitudinal surveys. This analysis is important since non-random attrition results in a data set that is no longer representative of the population from which the survey is sampled. Once the attrition process is understood, weights can be developed for use in obtaining sample statistics that are representative of the population (Lepkowski 1989, Lepkowski, Kalton, and Kasprzyk 1989).

In the survey methodology literature, studies of longitudinal data have focused on the effect of re-interviews on response accuracy (see, for example, Kasprzyk, Duncan, Kalton, and Singh 1989) rather than on the attrition process itself. A number of important issues need to be resolved in order to get a better understanding of attrition. For example, it is very likely that the interviewer and the interview process can influence the respondent's attachment to the survey but there is little evidence to support this assertion. Knowledge of this relationship can lead to changes in the structure of the survey that will lower attrition rates. This paper analyzes the determinants of attrition to address this and other issues. Attrition equations are estimated that include demographic, location, and mobility indicators and variables that correspond to the interviewer and the interview process.

Two major longitudinal surveys will be analyzed; the Panel Study of Income Dynamics (PSID) and the Survey of Income and Program Participation (SIPP). The PSID is an ongoing panel for which the first annual interview was conducted in 1968. The attrition rate for the full sample in the first year of the panel was 11.9% and by the twenty-first wave, only 48.8% of the original sample was present for all interviews. The SIPP is a series of two-year eight-month panels. Interviews are conducted every four months and data are recorded on a monthly basis. Both the 1984 and 1990 panels (SIPP84 and SIPP90) will be used so that comparisons can be made across panels. In SIPP84, the attrition rate in the first wave was 5.9% while 71.4% of the original sample was present for all eight waves. The comparable values for SIPP90 are 7.6% and 73.4%. This suggests that the attrition process has been fairly constant across panels. The different nature of the PSID and SIPP surveys provides for an interesting comparison of the attrition processes of two popular longitudinal surveys that will result in evidence about how the frequency of interviews and the duration of surveys affect attrition.

An important reason to account for attrition is that ignoring it can result in biased parameter estimates. This can occur when the reason for leaving the survey is related to the behavior that is being modeled. Given the high exit rates for many panel data sets, this problem can be quite troublesome. Until recently, only a few formal analyses of the effect of attrition on economic models of individual behavior have been conducted (Hausman and Wise 1979, Becketti, Gould, Lillard and Welch 1988, and Ridder 1990, 1992).[1] In this study, a model of attrition and labor market behavior is developed and estimated using the PSID and SIPP.

The theoretical analysis and review of previous studies of attrition are included in Section 2, and a two-step estimator (in the spirit of Heckman 1978) of the attrition and labor market model is developed. Also addressed in Section 2 is the issue of duration dependence; the effect of previous interviews on the probability of attrition. The designs of the PSID and SIPP data sets are compared and the attrition processes for male heads of households are estimated in Section 3. Section 4 includes the empirical analysis of the labor market model. Concluding remarks are given in Section 5.

The empirical analysis provides evidence that the characteristics of the interviewer and the interview process have significant impacts on attrition in the PSID and the SIPP. Maintaining the same interviewer and reducing the interview length will tend to reduce attrition. Demographic and location variables appear to have the same effect on attrition for all three data sets but there are significant differences in other aspects of the attrition processes. The results indicate that surveys of longer duration and those with higher frequency interviews will experience higher attrition rates.

The results for the model of attrition and labor market behavior show that the attrition bias correction term is significant in the wage equations for the SIPP90 and PSID samples and the correlation between the error terms in the attrition and labor force participation equations is significant for SIPP84 yet there is little impact on the parameter estimates when attrition is not accounted for. Still, the estimates for the labor market behavior model are different when it is estimated separately for attritors and non-attritors. Thus, care should be taken when merging the two groups to increase sample size.

2. Theoretical analysis and previous studies

Consider a longitudinal survey that includes T waves. Attrition occurs for individual i if he/she leaves the survey in period $T_i \leq T$. If information for individual i is not observed for any wave greater than T_i, attrition is an absorbing state. Attrition is non-absorbing if an individual can return to the sample after exiting. Attrition is considered to be absorbing for this analysis. Thus the attrition process is specified as

$$a^*_{it} = Z_{1it}\alpha_1 + D_{it}\alpha_2 + v_{1it} \qquad i=1,\ldots,N, \; t=1,\ldots,T_i \tag{1}$$

where

individual i leaves in period t (a_{it}=1) if $a_{it}^{*}>0$,

individual i remains in period t (a_{it}=0) if $a_{it}^{*}\leq 0$,

and

$$v_{1it} = \epsilon_{1i} + u_{1it}.$$

Z_{1it} is a 1xM vector of regressors, D_{it} is a 1xT_i vector of wave dummies, and ϵ_{1i} is an unobservable individual effect. It is assumed that u_{1it} is an independently and identically distributed (i.i.d.) normal random variable with mean zero and variance σ_1^2 that is independent of ϵ_{1i} and Z_{1it}.

Since attrition is absorbing in this analysis, duration dependence can be modeled by including wave dummies, D_{it}. The presence of duration dependence is revealed by a monotonic change in the coefficients for the D_{it}'s. Negative duration dependence occurs (monotonically increasing coefficients) if the participant becomes tired of being interviewed and hence is more likely to attrit. Positive duration dependence can arise (monotonically decreasing coefficients) if the interview process becomes less time consuming as the respondent becomes more familiar with the interview process, the individual develops a sense of loyalty to the survey and hence feels more inclined to continue through time, or the interviewer develops a rapport with the respondent (Waterton and Lievesley 1989). Thus, it is unclear, a priori, whether positive or negative duration dependence will arise.

Note the importance of controlling for the individual effect, ϵ_{1i}, when testing for duration dependence. All things equal, individuals with relatively large values of ϵ_{1i} will be more likely to exit from the survey while those with smaller values remain and are less likely to leave the survey in future waves.

This higher probability of attrition in the sample can be mistaken as positive duration dependence (see Heckman (1981) for a general discussion of heterogeneity and state dependence).

An important reason to account for attrition is that it can bias parameter estimates if it is related to the behavior that is being modeled. One means for determining if such a bias is present is to incorporate attrition into the behavioral model. Here, a model of attrition and labor market behavior is presented. The labor market behavior model is chosen since it is a popular model to estimate using the PSID and SIPP.

The first formal model of attrition and labor market behavior was developed by Hausman and Wise (1979). They were concerned that the high dropout rate in the Gary Income Maintenance Experiment might bias the estimate of the experimental effect since it is likely that attrition was linked to income. Hausman and Wise model the attrition process along with the pre- and post-experiment income equations. They find some evidence of attrition bias but it has little impact on the estimate of the experimental effect.

Ridder (1990, 1992) extends the Hausman-Wise model to more than two periods and allows for both state dependence and a more general error structure. He estimates a model of attrition and transportation demand, using data from the Dutch Transportation Panel, that is better able to account for the non-random attrition from the sample than the Hausman-Wise model.

An analysis of attrition in the first fourteen waves of the PSID was carried out by Becketti, Gould, Lillard, and Welch (1988). They find no significant differences in the parameter estimates for a 1968 labor income equation for samples of workers who remained in 1968, 1975, and 1981. However, by looking only at cross-section estimates for 1968, they do not take full advantage of the longitudinal nature of the PSID. Also, by restricting their analysis to workers, a form of selection bias is induced that is similar to the bias that can arise if attrition is disregarded.

The model of attrition and labor market behavior presented here is similar in nature to those developed by Hausman-Wise and Ridder. In this case non-workers are included, the labor force participation (LFP) equation is modeled along with the attrition and labor market activity equations, and a two-step estimator is developed. Consider the following panel data model of labor market behavior

$$LFP^*_{it} = R_{1it}\gamma_1 + \epsilon_{2i} + u_{2it} \qquad i=1,\ldots,N,\ t=1,\ldots,T^*_i \tag{2}$$

$$Y_{it} = X_{1it}\beta_1 + \epsilon_{3i} + u_{3it} \qquad i=1,\ldots,N,\ t\epsilon\tau_i \tag{3}$$

where

i works in period t ($LFP_{it}=1$) if $LFP^*_{it}>0$

i does not work in period t ($LFP_{it}=0$) if $LFP^*_{it}\leq 0$,

X_{1it} and R_{1it} are 1xK and 1xP vectors of regressors, ϵ_{2i} and ϵ_{3i} are individual effects, u_{2it} is an i.i.d. normal random variable with mean zero and variance σ^2_2 that is independent of ϵ_{2i} and R_{1it}, u_{3it} is an i.i.d. stochastic disturbance term with mean zero and variance σ^2_3 that is independent of ϵ_{3i}, X_{1it}, and R_{1it}, $\tau_i=\{t\leq T^*_i | LFP^*_{it}>0\}$, and $T^*_i=T_i-1$ for $T_i<T$ and $T^*_i=T$ for $T_i=T$. For this analysis, Y_{it} represents a measure of labor market activity such as labor income, wages, or hours of work. The labor market behavior model is estimated along with the attrition equation (1). Note that the LFP equation (2) is defined for periods $1,\ldots,T^*_i$ to account for attrition from the sample.

If (2) and (3) derive from a dynamic utility maximization problem, the individual effects are related to the marginal utility of initial wealth which is a function of all information available at t=1 (Heckman and MaCurdy 1980). This implies that ϵ_{2i} and ϵ_{3i} are correlated with R_{1it} and X_{1it}, respectively, and any consistent estimator of γ_1 and β_1 must account for this correlation. One approach is to model the individual effects as functions of the means of the time-varying variables. This method has advantages over others such as fixed effects which is inconsistent in the presence of selectivity (Zabel 1994a).

Consider the following specifications for ϵ_{2it} and ϵ_{3it}

$$\epsilon_{2i} = \bar{R}_{1i}\gamma_2 + R_{2i}\gamma_3 + w_{2i} \qquad w_{2i} \sim N(0,\sigma^2_{w2}) \tag{4}$$

and

$$\epsilon_{3i} = \bar{X}_{1i}\beta_2 + X_{2i}\beta_3 + w_{3i}. \tag{5}$$

Using (4) and (5) to substitute out ϵ_{2i} and ϵ_{3i} in (2) and (3) gives

$$LFP^*_{it} = R_{1it}\gamma_1 + \bar{R}_{1i}\gamma_2 + R_{2i}\gamma_3 + v_{2it} \tag{6}$$

$$Y_{it} = X_{1it}\beta_1 + \bar{X}_{1i}\beta_2 + X_{2i}\beta_3 + v_{3it}, \tag{7}$$

where

$$v_{jit} = w_{ji} + u_{jit}, \quad j=2,3.$$

It is also possible to specify ϵ_{1i} as a function of the means of the time-varying regressors in the attrition equation. But this results in endogeneity bias since some of the means are dependent on a_{it}. For example, the mean of age, $\overline{AGE}$, increases with each subsequent wave in the survey since age increases over time. This will cause a positive bias in the estimated coefficient for $\overline{AGE}$. To avoid this, the first period values, rather than the means, of the time-varying variables are included in the attrition equation in Section 3. Note that this is a way of controlling for ϵ_{1i} so that a test of duration dependence can be undertaken.

The attrition and behavioral equations can be estimated jointly using maximum likelihood given assumptions on the distribution of the error terms. But this estimator is quite complicated since two of the dependent variables are binary and the third is censored. Instead, a simpler, consistent, two-step estimator is developed. Though less efficient, this estimator requires fewer assumptions about the distribution of the error terms since it is not necessary to specify the joint distribution of all the errors. The second step is to estimate

$$Y_{it} = X_{1it}\beta_1 + \bar{X}_{1i}\beta_2 + X_{2i}\beta_3 + \beta_4\lambda_{1it} + \beta_5\lambda_{2it} + v^*_{3it} \qquad t\epsilon t_i \tag{8}$$

where

$$v^*_{3it} = v_{3it} - \beta_4\lambda_{1it} - \beta_5\lambda_{2it},$$

and λ_{1it} and λ_{2it} are sample selection bias correction terms. λ_{1it} accounts for the non-random selection of individuals who participated in the survey and λ_{2it} accounts for the participation in the labor force. The derivation of these terms is given in Appendix 1.

The first step of the two-step procedure is to estimate the attrition and participation equations using maximum likelihood. It is assumed that v_{1it} and v_{2it} follow a bivariate normal distribution with zero means, unit variances, and correlation ρ. The log-likelihood function to be maximized is

$$\text{LogL} = \sum_{a_{it}=1} \log[\phi(-Z_{it}\alpha)] + \sum_{\substack{a_{it}=0,\\ LFP_{it}=0}} \log[B(Z_{it}\alpha, -X_{it}\beta, \rho)] + \sum_{\substack{a_{it}=0,\\ LFP_{it}=1}} \log[B(Z_{it}\alpha, X_{it}\beta, -\rho)] \tag{9}$$

where ϕ and B are the standard normal univariate and bivariate distribution functions. If ρ is significant, this is an indication of attrition bias in the participation equation, while a significant λ_{1it} in (8) is evidence of attrition bias in the behavioral equation.[2]

3. Empirical analysis of attrition in the PSID and SIPP

The data sets used in this study are Waves I-XXI (1968-1988) of the PSID and the 1984 and 1990 panels of the SIPP (hereafter SIPP84 and SIPP90). 4,802 families were first interviewed for the PSID in 1968 and were re-interviewed on an annual basis. This includes 1,872 low-income families from the Survey of Economic Opportunity (SEO). The remaining 2,930 families were drawn from the Survey Research Center (SRC) master sampling frame.

The design of the SIPP is different than that of the PSID in a number of ways. First, interviews are conducted three times a year and information for each of the four months prior to the interview is recorded. Second, only eight interviews are conducted so the length of the panel for a given individual is two years and eight months.[3] New panels are started each year resulting in a series of overlapping panels. This allows for a comparison of results across panels. Third, the sample size for the SIPP is much larger. Given these important differences between the PSID and the SIPP, a comparison of the results should provide useful information about the relationship between attrition and the frequency of interviews and the duration of surveys (see Appendix 2 for a further discussion of differences between the PSID and SIPP).

Attrition rates are given in Table 1. Individuals are included if they were interviewed (self or proxy) in the first wave.[4] The attrition rate was 11.9% in the first wave of the PSID and only 48.8% of the original sample members was present for all twenty-one waves. The attrition rated dropped to

3.3% in the second wave and it was always less than or equal to 2.6% after that. There is somewhat of a trend in the unconditional attrition rate (the percentage of attritors with respect to the number in sample in wave 1) but this is because the sample shrinks over time so the number of potential attritors also falls. There is no indication of a trend in the conditional attrition rate (the percent of attritors with respect to the number in sample in the same wave).

- Table 1 here -

5.9% of the original sample members left SIPP84 in the first wave and the annual exit rate declined slowly for each wave. 71.4% of the respondents were present for all eight waves. The comparable rates for SIPP90 are 7.6% and 73.4%. Thus while the exit rate for the first wave of SIPP90 was higher than that for SIPP84, it was lower for each subsequent wave of SIPP90 versus SIPP84.

Though the initial exit rate was higher for the PSID, the attrition rate was generally greater for the SIPP after the first wave. One explanation for this deference is that there is more time to interview individuals in the PSID for whom the first attempt at an interview was unsuccessful. See Zabel (1994b) for other reasons for the differential attrition rates for the two surveys.

Since the SRC sample is representative of the U.S. population it is used for this analysis. The sample is restricted to male heads of households who were aged 25 to 64 in the first wave of the panel (results in Zabel 1994b show similar attrition behavior for female heads and wives). Definitions of the variables used in this analysis are given in Table 2.[5] Generally, the demographic and location variables are comparable across samples. Means, by attrition status, for the first wave values for these variables are given in Table 3. Individuals who left the survey are more likely to live in urban areas, the South or West, be nonwhite, not married, have fewer children, and not own homes.[6]

- Tables 2 and 3 here -

While attention has been paid to the effect of the interviewer and interview process on response accuracy in the survey literature (Groves and Kahn 1979, Groves 1989, Kasprzyk, Duncan, Kalton, and Singh 1989), there appears to be little discussion of the effect on attrition. Respondents who completed the 1985 panel of the SIPP were asked to give the main reason why they continued to participate in the survey (Meier 1988). The most often cited reason was that they liked the interviewer. Estimating the attrition equation (1) allows for a number of variables that measure the characteristics of the interviewer and the interview process to be added as regressors. These include the length of the interview, the number of contacts between the respondent and the interviewer, whether there was a change in the interviewer, whether the interview was given to a proxy, whether the interview was conducted by phone, the number of adjustments that were made to certain variables, and the length of time the interviewer spent editing the response forms. These variables are expected to be positively correlated with attrition.

The attrition equation (1) is estimated by probit and the results are given in Table 4. The coefficient estimates are transformed to measure changes in the probability of attrition by multiplying by the probability density function evaluated at sample means. The design of the SIPP is such that information for the current period is not available for attritors so lagged variables are used in the attrition equation and the initial period of analysis is the second wave.

- Table 4 here -

Recall that the first-period values of the time-varying variables are included in the equation (but not in Table 4 to save space). Thus the results are conditional on the first period values. The coefficient estimates and their significance levels are very similar for SIPP84 and SIPP90. The differences between the PSID and SIPP samples apply less to the parameter estimates than to the significance levels which are largely due to the smaller sample size for the PSID. T-tests are carried out to determine if the differences in the individual parameters for the three surveys are zero for the common variables. The only differences that are significant at the 1% level are those for AGE in SIPP84 and the PSID and SIPP90 and the PSID (see (Zabel 1994b) for an explanation of these differences). Next, the likelihood ratio (LR) test is used to determine if the SIPP84 and SIPP90 samples can be merged. The LR statistic is significant at the 1% level so the attrition processes are different for the SIPP84 and SIPP90 samples even though there are no significant differences in the individual coefficients.

Generally, changes in the demographic variables tend to have relatively little impact on attrition. This result is in contrast to the numerous significant differences in the means of these variables for attritors and non-attritors. Blacks and other nonwhites are more likely to attrit than whites for SIPP84 and SIPP90 but not for the PSID. Home-ownership is a significant (negative) indicator of attrition in all three samples, movers are more likely to exit in both SIPP84 and SIPP90, and PSID sample members whose residence in 1968 was in the same state in which they were born are less likely to attrit. Surprisingly, participation in the labor force in the previous period is only significant for SIPP90, though it has a negative effect on attrition for all three samples. See Zabel (1994b) for more discussion of the effects of the demographic variables. The coefficient estimates for the PSID sample are generally similar to those in Lillard and Panis (1994) even though they use a somewhat different sample and a different set of regressors.

Many of the variables that correspond to the characteristics of the interviewer and interview process have significant impacts on attrition. For all three samples the number of imputations has a positive effect on attrition. For the PSID, the number of contacts has a strong positive effect on attrition while the length of time the interviewer spent editing the forms is positively correlated (though not significant) with attrition in SIPP84 and SIPP90. These are signs of respondents who may spend less time preparing for the interviews and who are less interested in the surveys.

Surprisingly, interview length has a negative effect on attrition. This may arise because interviewers spend more time with respondents who enjoy the survey more and hence are less likely to attrit. In the PSID, there was an explicit attempt to reduce the length of the interview in 1973 (Hill 1992, though for some unknown reason the reduction in length began in 1972). This allows for a test of whether an exogenous change in interview length affects attrition. A variable is included that equals the 1968 interview length for waves prior to 1972 and the 1972 interview length for all waves after, and including, 1972. The 1968, rather than 1971, interview length is used because not all individuals were present in 1971. The estimated coefficient is positive and the t-statistic is 2.80. This is evidence that the reduction in interview length in 1972 had a negative affect on attrition.

Since some interviews are given to a proxy, the binary variable PROXY is included to determine if this affected the probability of attrition for SIPP84 and SIPP90. PROXY is also interacted with interview length since the length should not matter if the individual is not present. Both of these variables are not significant for either SIPP sample. Thus whether the interview was given to the individual or to a proxy does not seem to affect attrition.

One important positive indicator of attrition in the SIPP is a change in the interviewer. But since the interview change might arise because there was a problem with the interview, the causality of the relationship between interview change and attrition is not clear. To minimize the probability that the change in interviewer was due to a problem interview, INTCH only includes cases where there was a change to another field interviewer and not to a supervisor since the latter is often called in when a problem arises. The estimated coefficients for this variable are positive in both cases and significant for SIPP84. This result supports the hypothesis that the interviewer builds up a positive rapport with the respondent. A new interviewer will not have established this rapport and hence the possibility of attrition is greater.

Variables relating to telephone interviews are included to test whether the mode of interview affects attrition. It is hypothesized that the rapport between the interviewer and the respondent is stronger when interviews are conducted in person rather than by telephone. One case where the change to phone interview was random was the seventh wave of the 1990 panel.[7] Over half the sample members that received personal interviews for the first six waves were contacted by phone in the seventh wave. The variable PHONE7 is a dummy variable that is one in the final wave for individuals whose first six interviews were conducted in person and for whom the seventh was a phone interview. The estimated coefficient is actually negative but is insignificant. This indicates that the mode of interview does not affect attrition (for individuals whose first six interviews were in person).

Almost all interviews were conducted in person for the first five waves of the PSID but starting with the sixth wave the vast majority of interviews were conducted by telephone. The dummy variable PHONE73 indicates whether an interview was conducted by phone for individuals who had personal interviews for the first five waves but who had a phone interview in the sixth wave. The estimated coefficient for this variable is positive but insignificant. Again, this is evidence that the mode of interview does not affect attrition.[8]

Wave dummies are included to account for possible duration dependence and they are highly significant in all three samples (using the LR test). Given that the first wave values of the time-varying variables and the time-invariant variables account for the unobserved heterogeneity (ϵ_{1i} in equation 1), the wave dummies will measure duration dependence.[9] Any unobserved heterogeneity that remains can induce spurious duration dependence. Thus the results on duration dependence are conditional on the ability of the first wave values of the time-varying variables and the time-invariant variables to control for the unobservable heterogeneity.[10]

There is strong evidence of positive duration dependence in the SIPP as indicated by the monotonically decreasing coefficient estimates. The reduction in the probability of attrition between the first and last waves is estimated to be 4.3% for SIPP84 and 5.2% for SIPP90. The larger decrease for SIPP90 is consistent with the differential attrition rates in Table 1. The PSID exhibits less evidence of positive duration dependence. The large coefficient estimate in the first year is consistent with the relatively high exit rate in the first wave of the survey, but there is not a consistent pattern of declining coefficients for the later waves. This result is in agreement with Becketti et al. (1988) who find no evidence of duration dependence using a Weibull hazard specification. The positive duration dependence in the SIPP may be due to the shorter length of the panel and to the known ending date (individuals may be more likely to continue if they know the end is near).

As a test for linear duration dependence, a linear trend variable is included instead of the wave dummies. This variable is negative and significant for all three samples. The restrictions that are imposed on the coefficients for the wave dummies to obtain the trend variable are not rejected for either SIPP84 or SIPP90 but they are significantly rejected for the PSID. This supports the hypothesis of positive duration dependence in SIPP84 and SIPP90. The estimates of the trend coefficients for the SIPP84 and SIPP90 samples show that the probability of attrition decreases by 0.63% and 0.72% per wave for SIPP84 and SIPP90. Note that one can reach the conclusion that there is positive duration dependence in the PSID based solely on the results for the linear trend variable but the rejection of the linear trend in favor of the more general wave dummies does not lead to the same deduction.

The results in this section provide some evidence about the effects of survey duration and interview frequency on attrition. Surveys of shorter duration are likely to exhibit a higher degree of positive duration dependence which will decrease attrition rates. More frequent interviews are likely to lead to higher exit rates since there is less time to get interviews when the first attempt is unsuccessful. Also, the results suggest that maintaining the same interviewer and decreasing the interview length will lead to lower attrition rates.

4. Empirical analysis of the impact of attrition on labor supply behavior

The results from the last section indicate that there are significant differences in the underlying behavior of the attritors and non-attritors. These differences can bias parameter estimates if they are related to the behavior that is being modeled. In this section, the effect of attrition on a model of labor supply is analyzed. It should be noted that attrition bias is model specific. Evidence of bias in the labor supply model does not imply that bias will exist in any other model (or vice versa).

The impact of attrition is calculated by determining the significance of parameters that indicate the presence of attrition bias and by comparing the parameter estimates for models that do and do not account for attrition and for samples of attritors and non-attritors. The behavioral model that is estimated is the participation equation (6) and the hours of work equation

$$\ln H_{it} = \beta_1 \ln W_{it} + \beta_2 \overline{\ln W}_{it} + X_{1it}\beta_3 + \bar{X}_{1i}\beta_4 + X_{2i}\beta_5 + v_{3it} \tag{11}$$

where $\ln H_{it}$ and $\ln W_{it}$ are the logs of hours worked and wages. Note that $H_{it}=0$ and $\ln W_{it}$ is not observed if $LFP_{it}=0$. Wages are likely to be correlated with v_{3it} so instruments for wages are used to obtain consistent estimates of the β's. H_{it} is defined to be the number of hours usually worked per week times the number of weeks worked per period.[11] W_{it} is equal to the ratio of labor income to H_{it}. Nominal variables are put in real terms using the Consumer Price Index.

The parameters in the hours equation that are of the most interest to labor economists are β_1 and β_2. β_1 is the intertemporal substitution elasticity and $\beta_1+\beta_2$ is the elasticity due to a parallel shift in the wage profile. The latter is useful for measuring permanent changes in wages that arise from changes in taxes or other policy instruments. Thus this analysis will focus on these two parameters.

Four sets of estimates for each data set are given in Table 5. Model 1 is the attrition model, Model 2 is the labor supply model that ignores attrition, and Models 3 and 4 are the labor supply models for non-attritors and attritors, respectively. Estimating these models will provide information about the consequences of ignoring the attrition decision.

For the sake of brevity, estimates are only given for the parameters in the hours equation, the selection bias correction terms in the wage equation and the correlation between the disturbance terms in the attrition and LFP equations.[12]

The regressors in the wage equation are terms in the cubic expansion of the age and education variables, and also location indicators (region and Metropolitan Statistical Area), and the number of family members. The participation equation includes all exogenous regressors in the wage and hours equations.[13]

All models are estimated for white heads of households who were continuously married and not in school for the length of the survey.[14] There are 3,232, 3,787, and 329 such individuals in the SIPP84, SIPP90, and PSID samples, respectively. These restrictions are imposed so that the results can be compared to previous analyses of life cycle labor supply models. The percentage of attritors is 23.8%, 18.8%, and 50.8% for the three samples.

An issue that arises in sample selection models is the identification of the parameters in the behavioral equations. One means of identifying the parameters in the hours and wage equations is to include unique variables in the attrition equation (1) and the LFP equation (6). Many exclusion restrictions arise for the SIPP samples since all the variables in the attrition equation are lagged variables. For the PSID sample, lagged nonlabor income is included in the attrition equation but not in the hours or wage equations. Also, the number of imputations, the length of the interview, whether there was a change in the interviewer, and whether there are family members other than the husband, wife, and children are unique to the attrition equation. It is assumed that these variables affect attrition but do not affect nor are affected by labor market behavior.

In the standard labor supply model, the identification of the parameters in the hours and wage equations arises through the nonlinearity of the selection bias correction term since the variables in the LFP equation are the same as those in the hours and wage equations. When the attrition equation is included in the model, two sample selection bias correction terms are added to the behavioral equations. These two terms are functions of the regressors in both the attrition and LFP equations as long as the correlation between their disturbance terms is not zero (see Appendix 1). Thus the exclusion restrictions in the attrition equation alone can identify the parameters in the hours and wage equations.

- Table 5 here -

The estimates/(t-values) for the correlation between the error terms in the attrition and LFP equations for the PSID, SIPP84, and SIPP90 samples are 0.8921/(0.85), -0.4110/(3.36), and -0.2013/(1.72), respectively. Thus there is evidence that attrition bias exists in the SIPP84 sample but not in the PSID or SIPP90 samples.

The estimates of the intertemporal substitution elasticity for Model 1 for the PSID, SIPP84, and SIPP90 samples are 0.1130, -0.0164, and 0.0415, respectively and none are significantly different from zero.[15] The values for the PSID and SIPP90 are comparable to earlier estimates in the literature (see MaCurdy 1981, Altonji 1986, and Zabel 1994a). The negative estimate for SIPP84 is the opposite of what is expected, though it is insignificant and very small in magnitude.

The estimates of the elasticity due to a parallel shift in the wage profile are -0.0625, -0.0821, and -0.0565 for the PSID, SIPP84, and SIPP90 samples and all three are significant. MaCurdy's (1981) estimate of -0.07 is close to all three estimates. It appears that the sample of married men react to a permanent change in wages but not to a temporary one.

The coefficients for the sample selection bias correction terms, λ_1 and λ_2, are proportional to the correlations between the disturbance terms in the behavioral equations and the attrition and LFP equations, respectively (see

Appendix 1). Thus significant estimates are an indication of selection bias. The coefficient estimates for λ_1 are insignificant in the hours equation for all three samples but are significant in the wage equation for the PSID and SIPP90 samples. Hence there is some evidence of attrition bias in the wage equation in the PSID and SIPP90 but no evidence of attrition bias in the hours equation. The estimate of the coefficient for λ_2 is significant in the hours equation for SIPP84 and SIPP90 and in the wage equation for the PSID and SIPP90. Thus there is evidence of bias due to the non-random selection into the labor force.

Another way of checking for attrition bias is to compare the parameter estimates for the different models. First consider Models 1 and 2. Model 2 does not correct for potential attrition bias. The parameter estimates change very little when Model 2 is used. Thus, correcting for attrition bias has little impact on the estimates of the wage elasticities. This is true even though the attrition bias indicator is significant in the wage equation for the SIPP90 and PSID samples and the correlation between the error terms in the attrition and LFP equations is significant in SIPP84. This result is similar to Ziliak and Kniesner (1994) who find little difference in the parameter estimates for a labor supply equation using the SRC subsample of the PSID when they correct for possible attrition bias using the probit two-step estimator. The same is true for Lamas, Tin, and Eargle (1994) who estimate earnings equations using SIPP90.

Another useful comparison is the models for non-attritors and attritors; Models 3 and 4, respectively. The across model differences of the estimates of the wage elasticities are generally larger than for Models 1 and 2 (though not significant). The significance levels for the LR test that the coefficients are equal (except the constant) in the LFP equation for Models 3 and 4 are less than 0.001 for all three samples. The significance levels are also less than 0.001

for the same test for the wage equation (assuming that the coefficients are the same for the LFP equation so that λ_{2it} will be the same for the full sample and the two sub-samples). These rejections may not be too surprising given the large sample sizes.

The comparison of Models 3 and 4 provides some evidence that the labor supply behavior of the attritors and non-attritors is different. But there is little difference is the estimates for Models 1 and 2 even though λ_{1it} is significant for the SIPP90 and PSID samples. Similar results are obtained by Ziliak and Kniesner (1994) using the PSID and Rendtel and Buechel (1994) using the German Socio-Economic Panel.

These contrary results may arise because there are relatively few observations where attrition occurs (atr_{it}=0). In this circumstance, the attrition bias can be small, even if the two groups are different, since a large majority of the observations are selected (atr_{it}=1). The small sample of attritors also explains why the estimates for the non-attritors (Model 3) show little change when the attritors are included (Model 2).

One might be tempted to add the attritors to the non-attritors to increase the sample size. But there are a number of reasons why this might not be a good strategy. First, the percentage increase in the sample size is not large, particularly for the panels that are shorter in duration. As a result, there is no evidence of an increase in efficiency when the full sample is used (compare the standard errors for Models 2 and 3). Second, these results give some indication that attritors and non-attritors are different so care must be taken when merging the two groups. Thus, the small, if any, gains in efficiency may not outweigh the parameter bias that arises when estimating a labor supply model using the full sample of attritors and non-attritors.

5. Conclusion

In this paper, attrition models were estimated using samples of adult men from the PSID and the 1984 and 1990 panels of the SIPP. A number of variables that relate to the interviewer and the interview process are significant. Incorporating these characteristics into the weighting scheme that corrects for non-random attrition should lead to more accurate estimates of population parameters. Both the length of the interview and whether there was a change in the interviewer are positively correlated with attrition. This indicates that shorter interviews and maintaining the same interviewer can help reduce attrition.

Generally, the coefficient estimates in the attrition equation are similar for the three samples for the common variables. One important difference between the SIPP and PSID is the strong evidence of positive duration dependence in the SIPP and the weaker evidence of duration dependence in the PSID. Also, while the PSID suffers from a high exit rate in the first wave, the attrition rate for the second wave on is generally less than that for the SIPP. These results provide some evidence that surveys of longer duration and those with higher frequency interviews are likely to exhibit higher attrition rates.

A model of attrition and labor supply behavior was estimated using a two-step procedure. Generally, it appears that accounting for attrition bias has little impact on the parameter estimates. This result is supported by other studies of the impact of attrition on labor market activity (Hausman and Wise 1979, Becketti et al. 1988, Lamas, Tin, and Eargle 1994, Rendtel and Buechel 1994, and Ziliak and Kniesner 1994). However, care must be taken when merging the samples of attritors and non-attritors since the parameter estimates for the two groups are significantly different for the three samples.

Endnotes

1. A parallel literature investigates the exit behavior of firms. See, for example, Hall (1987) or Olley and Pakes (1992) for models that incorporate the exits of firms.

2. It is possible to account for the autocorrelation induced by the presence of the individual effects, ϵ_{1i} and ϵ_{2i}, but this requires two-fold numerical integration. Guilkey and Murphy (1993) provide Monte carlo evidence that shows that accounting for the autocorrelation in the random effects probit model does not lead to more accurate parameter estimates or more precise standard errors.

3. Information from a ninth interview that was conducted for half of the sample for SIPP84 is not used so that the results for SIPP84 and SIPP90 would be more comparable.

4. It is possible to include respondents who joined the survey at a later date, but they are not considered in this study. Lillard (1989) calculates that the initial sample of male heads of households in the PSID can be increased by 32% by including "non-sample" individuals.

5. For the analysis in this section, random samples of 75% of SIPP84 and 65% of SIPP90 were drawn to make the size of the data sets more manageable.

6. Zabel (1994b) shows that similar results entail when the correlations between these variables are controlled for by regressing a variable that is one for individuals who ever left the survey and zero for those who remained on the

demographic and location variables. Comparable results using the same technique are reported in Fitzgerald, Gottschalk, and Moffitt (1994) for the PSID and in Lamas, Tin, and Eargle (1994) for SIPP90.

7. Even this experiment was not totally random since the interviewers had some discretion over whether or not the interview was conducted by phone and not all respondents had phones.

8. For the SIPP samples, the variable PHONE indicates whether an interview was conducted by phone for individuals who chose this mode of interview starting with the second wave. Rather than provide evidence that phone interviews are more likely to lead to attrition, the significantly positive estimates indicate that the type of person who chooses to be interviewed by phone is more likely to attrit. This may be a signal of individuals who are less interested in the survey and hence are more likely to leave before it ends.

9. An alternative is the fixed effect probit estimator that conditions on predicted ϵ_{1i} (Heckman 1981). But individuals who leave in the first period and non-attritors must be excluded. Also, it is necessary to estimate a fixed effect for each individual. This estimator is nonlinear and it can be difficult to implement for a large number of individuals.

10. The attrition equations were estimated using the procedure developed by Butler and Moffitt (1982) to determine the proportion of the variance of the unobservables that is attributable to the individual effect (ρ). When only the time varying regressors are included, ρ is estimated to be 0.55, 0.45, and 0.50 for SIPP84, SIPP90, and the PSID, respectively. When the first period values of

the time-varying variables and the time-invariant variables are included, these estimates are 0.35, 0.44, and 0.56. Finally, when the wave dummies are included, all three estimates of ρ go to zero. Lillard and Panis (1994) also find that the variance of the individual effect in the attrition equation is 0.

11. Since the number of weeks in a period can be either 17 or 18 for the SIPP, hours worked and income variables are standardized so that each period has 17.333 weeks.

12. The means of the time-varying variables, rather than the first-period values, are included in the attrition equation in Model 1. If they are not included, estimates of the correlation between the disturbance terms in the attrition and LFP equations approach one and the MLE program does not converge.

13. Wave dummies were included in the LFP, hours and wage equations but they are not significant in any of the equations for all three samples and were excluded.

14. To be included in the sample, individuals cannot have missing values for any of the variables used in the analysis or major imputations made to the hours worked or labor income variables (see Horrowitz and McFadden (1994) for an analysis of missing values and attrition). Also, for individuals working in consecutive periods, the absolute value of the change in wages or hours worked cannot exceed 200%. The maximum number of hours worked per period is 4,680 for the PSID and 1,560 for the SIPP.

15. As a check, the models were estimated using the deviations from means (fixed effect) estimator. The results are very similar.

Appendix 1
Two-step Estimator for the Bivariate Selection Model

A simplified version of the model that is estimated in this study is (subscripts have been suppressed to ease notation)

$$a^* = Z\alpha + u_1 \tag{A1}$$

$$LFP^* = R\gamma + u_2 \tag{A2}$$

$$Y = X\beta_1 + u_3 \tag{A3}$$

where

$a=1$ if $a^*>0$, $a=0$ if $a^*\leq 0$,

$LFP=1$ if $LFP^*>0$, $LFP=0$ if $LFP^*\leq 0$.

LFP is only observed if a=1 and Y is only observed if a=LFP=1 (note that in (1) $a_{it}=1$ if $a^*_{it}\leq 0$ but the reverse is true here. This is done to simplify the exposition). Assume that u_1 and u_2 are bivariate normal with zero means, unit variances, and correlation ρ_{12}. Assume u_3 has mean zero, variance σ_{33} and covariances σ_{13}, σ_{23} with u_1 and u_2, respectively. Then

$$E[u_3|u_1,u_2] = (\sigma_{13}, \sigma_{23})\cdot\begin{pmatrix}1 & \rho_{12}\\ \rho_{12} & 1\end{pmatrix}^{-1}\cdot\begin{pmatrix}u_1\\ u_2\end{pmatrix}$$

$$= \frac{\sigma_3\cdot(\rho_{13}-\rho_{12}\rho_{23})\cdot u_1}{1-\rho_{12}^2} + \frac{\sigma_3\cdot(\rho_{23}-\rho_{12}\rho_{13})\cdot u_2}{1-\rho_{12}^2}$$

$$= a_1u_1 + a_2u_2. \tag{A4}$$

It follows that

$$E[u_3|u_1>-Z\alpha,u_2>-R\gamma] = a_1E[u_1|u_1>-Z\alpha,u_2>-R\gamma] + a_2E[u_2|u_1>-Z\alpha,u_2>-R\gamma] \tag{A5}$$

where

$$E[u_1|u_1>-Z\alpha,\ u_2>-R\gamma] = \left[\phi(Z\alpha)\cdot\Phi\left(\frac{R\gamma+\rho_{12}Z\alpha}{(1-\rho_{12}^2)^{1/2}}\right) + \rho_{12}\phi(R\gamma)\cdot\Phi\left(\frac{Z\alpha+\rho_{12}R\gamma}{(1-\rho_{12}^2)^{1/2}}\right)\right]/B$$

$$= \lambda_1 + \rho_{12}\lambda_2 \tag{A6}$$

$$E[u_2|u_1>-Z\alpha,\ u_2>-R\gamma] = \left[\phi(R\gamma)\cdot\Phi\left(\frac{Z\alpha+\rho_{12}R\gamma}{(1-\rho_{12}^2)^{1/2}}\right) + \rho_{12}\phi(Z\alpha)\cdot\Phi\left(\frac{R\gamma+\rho_{12}Z\alpha}{(1-\rho_{12}^2)^{1/2}}\right)\right]/B$$

$$= \lambda_2 + \rho_{12}\lambda_1 \tag{A7}$$

and

$$B = \mathrm{Prob}(u_1>-Z\alpha,\ u_2>-R\gamma) = B(Z\alpha, R\gamma, \rho_{12}).$$

Substituting into (A4) gives

$$E[u_3|u_1>-Z\alpha,\ u_2>-R\gamma] = a_1(\lambda_1 + \rho_{12}\lambda_2) + a_2(\lambda_2 + \rho_{12}\lambda_1)$$

$$= (a_1+\rho_{12}a_2)\cdot\lambda_1 + (a_2+\rho_{12}a_1)\cdot\lambda_2$$

$$= \sigma_3\rho_{13}\lambda_1 + \sigma_3\rho_{23}\lambda_2. \tag{A8}$$

Thus

$$E[Y|u_1>-Z\alpha,\ u_2>-R\gamma] = X\beta + E[u_3|u_1>-Z\alpha,\ u_2>-R\gamma]$$

$$= X\beta_1 + \beta_2\lambda_1 + \beta_3\lambda_2. \tag{A9}$$

Consistent estimates of β_1 can be obtained by estimating (A3) with λ_1 and λ_2 included. Since α, γ, and ρ are not observed, they need to be estimated to obtain values for λ_1 and λ_2. These can be obtained by estimating (A1) and (A2) using maximum likelihood.

Appendix 2
Further Differences in the design of the PSID and SIPP

Three differences in the design of the PSID and SIPP were discussed in Section 3. Two relate to the frequency of interviews and the duration of the surveys. The third is the large difference in the sample sizes for the SIPP and the PSID. As previously mentioned, only 2,930 families are used from the PSID. The 1984 panel of the SIPP began with 19,878 families though this number was randomly reduced by 3,400 in wave 5 because of budgetary reasons. The 1990 panel began with 21,900 families. This includes some carryovers from the 1989 panel. These households are not included in the sample from the 1990 panel that is used in this analysis.

A fourth way in which the surveys diverge is that the data correspond to different reference periods. Most respondents in the PSID are interviewed in the spring of each year. Information on variables such as marital status, family members, and location variables apply to the time of interview, but information on income and hours worked is for the previous calendar year. Thus for studies that use the income and hours of work variables the period of analysis is usually a calendar year. In this case, a person is considered to have left in a period (say 1970) if he/she was present for the 1970 interview but not for the 1971 interview. Thus the information that is available for 1970 is included in the attrition equation for 1970. It is only possible to use lagged values of the income variables, say for 1969, since the information for 1970 is not available until the 1971 interview.

Since all information in the SIPP corresponds to the four months prior to the interview, no information on the current values of variables that can change unsystematically over time are available for the attrition equation. In this case, information from the last month of the previous interview is used so the initial period in the sample data set is the second wave.

A fifth difference between the surveys is that respondents in the SIPP are not paid for their participation while those in the PSID were paid ten dollars starting with the second interview and also received five dollars for mailing in their address verification form each year. This may have caused the attrition rates in the PSID to be lower than if the respondents received no compensation.

Hill (1992) lists eight strategies that are pursued to lower attrition rates in the PSID. Besides the payment strategy, the other one that is not followed for the SIPP is the use of certified mail as a means for obtaining forwarding addresses. This strategy is not used because of the shorter time between waves in the SIPP.

References

Ahn, H. and J.L. Powell. 1993. "Semiparametric Estimation of Censored Selection Models with a Nonparametric Selection Mechanism." **Journal of Econometrics** 58: 3-29.

Altonji, J.G. 1986. "Intertemporal Substitution in Labor Supply: Evidence from Micro Data." **Journal of Political Economy** 94: S176-S215.

Becketti, S., W. Gould, L. Lillard and F. Welch. 1988. "The Panel Study of Income Dynamics after Fourteen Years: An Evaluation." **Journal of Labor Economics** 4: 472-92.

Butler, J.S. and R. Moffitt. 1982. "A Computationally Efficient Quadrature Procedure for the One-Factor Multinomial Probit Model." **Econometrica** 50: 761-764.

Fitzgerald, J., P. Gottschalk, and R. Moffitt. 1994. "A Study of Sample Attrition in the Michigan Panel Study of Income Dynamics." unpublished paper.

Groves, R.M. 1989. **Survey Errors and Survey Costs.** New York: John Wiley and Sons.

Groves, R.M. and R.L. Kahn. 1979. **Surveys by Telephone, a National Comparison with Personal Interviews.** New York: Academic Press.

Guilkey, D.K. and J.L. Murphy. 1993. "Estimation and Testing in the Random Effects Probit Model." **Journal of Econometrics** 59: 301-307.

Hall, B. 1987. "The Relationship between Firm Size and Firm Growth in the U.S. Manufacturing Sector." **Journal of Industrial Economics** 35: 583-606.

Hausman, J.A. and D. Wise. 1979. "Attrition Bias in Experimental and Panel Data: The Gary Income Maintenance Experiment." **Econometrica** 47: 455-73.

Heckman, J.J. 1978. "Dummy Endogenous Variables in a Simultaneous Equations System." **Econometrica** 46: 931-59.

———. 1981. "Statistical Models for Discrete Panel Data." In **Structural Analysis of Discrete Data with Econometric Applications**, ed. C.F. Manski and D. McFadden, 114-178. Cambridge MA: The MIT Press.

Heckman, J.J. and T. MaCurdy. 1980. "A Life-Cycle Model of Female Labor Supply." **Review of Economic Studies** 47: 47-74.

Hill, M.S. 1992. **The Panel Study of Income Dynamics: A User's Guide.** Newbury Park: Sage Publications.

Horowitz, J.L., and C.F. Manski. 1994. "Joint Censoring of Regressions and Outcomes: Survey Nonresponse and Attrition.", unpublished paper.

Kasprzyk, D., G. Duncan, G. Kalton, and M.P. Singh. 1989. **Panel Surveys.** New York: John Wiley and Sons.

Lamas, E., J. Tin, and J. Eargle. 1994. "The Effect of attrition on Income and Poverty Estimates from the Survey of Income and Program Participation." unpublished paper, U.S. Bureau of the Census.

Lepkowski, J.M. 1989. "Treatment of Wave Nonresponse in Panel Surveys." In **Panel Surveys**, ed. D. Kasprzyk, G. Duncan, G. Kalton, and M.P. Singh, 348-374. New York: John Wiley and Sons.

Lepkowski, J.M., G. Kalton, and D. Kasprzyk. 1989. "Weighting Adjustments for Partial Nonresponse in the 1984 SIPP Panel," U.S. Bureau of the Census. **Survey of Income and Program Participation Selected Papers: 1989 Meetings of the American Statistical Association**, Washington, D.C.

Lillard, L.A. 1989. "Sample Dynamics: Some Behavioral Issues." In: **Panel Surveys**, ed. D. Kasprzyk, G. Duncan, G. Kalton, and M.P. Singh, 497-511. New York: John Wiley.

Lillard, L.A. and W.A. Panis. 1994. "Attrition from the PSID: Marital Status, Income and Mortality." unpublished paper, Rand Corporation.

MaCurdy, T.E. 1981. "An empirical Model of Labor Supply in a Life Cycle Setting." **Journal of Political Economy** 89: 1059-1085.

Meier, A.J. 1988. "SIPP 85: Results of Debriefing Respondents." Internal Census Bureau memorandum to R.P. Singh, September 26.

Olley, G.S. and A. Pakes. 1992. "The Dynamics of Productivity in the Telecommunications Equipment Industry." Discussion Paper 92-2, Center for Economic Studies, U.S. Bureau of the Census, Washington, DC.

Rendtel, U. and F. Buechel. 1994. "Tests for Non-Ignorable Panel Attrition and Their Application on Wage Estimates form the German Socio-Economic Panel (GSOEP)." unpublished paper, German Institute for Economic Research.

Ridder, G. 1990. "Attrition in Multi-wave Panel Data." In: **Panel Data and Labor Market Studies**, ed. J. Hartog, G. Ridder and J. Theeuwes, 45-67. Amsterdam: North-Holland.

————. 1992. "An Empirical Evaluation of Some Models for Non-random Attrition in Panel Data." unpublished paper, Rijksuniversiteit Groningen.

Waterton, J. and D. Lievesley. 1989. "Evidence of Conditioning Effects in the British Social Attitudes Panel Survey." In **Panel Surveys**, ed. D. Kasprzyk, G. Duncan, G. Kalton, and M.P. Singh, 319-339. New York: John Wiley and Sons.

Zabel, J.E. 1994a. "Estimating Wage Elasticities for Models of Labor Supply Behavior." unpublished paper, Tufts University.

————. 1994b. "A Comparison of Attrition in the Panel Study of Income Dynamics and the Survey of Income and Program Participation." unpublished paper, U.S. Bureau of the Census.

Ziliak, J. and T. Kniesner. 1994. "The Sensitivity of Life-Cycle labor Supply Estimates to Sample Attrition." unpublished paper, University of Oregon.

Table 1
Number and Percentage of Attritors

WAVE	FULL SAMPLE NUMBER SAMPLE	EXIT	PERCENT MARG	CUM*	COND	MALE HEADS NUMBER SAMPLE	EXIT	PERCENT MARG	CUM*	FEMALE WIVES/HEADS NUMBER SAMPLE	EXIT	PERCENT MARG	CUM*
						PSID							
1	18192	2166	11.9	11.9	11.9	2344	343	14.6	14.6	2694	354	13.1	13.1
2	16026	600	3.3	15.2	3.7	2001	67	2.9	17.5	2340	67	2.5	15.6
3	15426	401	2.2	17.4	2.6	1934	64	2.7	20.2	2273	60	2.2	17.9
4	15025	423	2.3	19.7	2.8	1870	61	2.6	22.8	2213	38	1.4	19.3
5	14602	440	2.4	22.2	3.0	1809	62	2.6	25.5	2175	51	1.9	21.2
6	14162	401	2.2	24.4	2.8	1747	71	3.0	28.5	2124	55	2.0	23.2
7	13761	378	2.1	26.4	2.7	1676	54	2.3	30.8	2069	49	1.8	25.0
8	13383	470	2.6	29.0	3.5	1622	60	2.6	33.4	2020	60	2.2	27.2
9	12913	392	2.2	31.2	3.0	1562	48	2.0	35.4	1960	46	1.7	29.0
10	12521	316	1.7	32.9	2.5	1514	50	2.1	37.5	1914	63	2.3	31.3
11	12205	373	2.1	35.0	3.1	1464	44	1.9	39.4	1851	39	1.4	32.7
12	11832	391	2.1	37.1	3.3	1420	34	1.5	40.9	1812	48	1.8	34.5
13	11441	319	1.8	38.9	2.8	1386	45	1.9	42.8	1764	39	1.4	36.0
14	11122	267	1.5	40.3	2.4	1341	38	1.6	44.4	1725	39	1.4	37.4
15	10855	314	1.7	42.1	2.9	1303	43	1.8	46.2	1686	36	1.3	38.8
16	10541	330	1.8	43.9	3.1	1260	50	2.1	48.4	1650	51	1.9	40.6
17	10211	353	1.9	45.8	3.5	1210	55	2.3	50.7	1599	50	1.9	42.5
18	9858	369	2.0	47.8	3.7	1155	43	1.8	52.6	1549	52	1.9	44.4
19	9489	325	1.8	49.6	3.4	1112	43	1.8	54.4	1497	48	1.8	46.2
20	9164	289	1.6	51.2	3.2	1069	45	1.9	56.3	1449	45	1.7	47.9
21	8875			48.8		1024			43.7	1404			52.1
						SIPP84							
1	43781	2572	5.9	5.9	5.9	11115	646	5.8	5.8	14025	738	5.3	5.3
2	41209	2512	5.7	11.6	6.1	10469	628	5.7	11.5	13287	726	5.2	10.4
3	38697	2100	4.8	16.4	5.4	9841	545	4.9	16.4	12561	658	4.7	15.1
4	36597	1761	4.0	20.4	4.8	9296	465	4.2	20.5	11903	519	3.7	18.8
5	34836	1514	3.5	23.9	4.3	8831	374	3.4	23.9	11384	438	3.1	22.0
6	33322	1240	2.8	26.7	3.7	8457	312	2.8	26.7	10946	356	2.5	24.5
7	32082	828	1.9	28.6	2.6	8145	214	1.9	28.6	10590	235	1.7	26.2
8	31254			71.4		7931			71.4	10355			73.8
						SIPP90							
1	47870	3632	7.6	7.6	7.6	12236	866	7.1	7.1	15375	1009	6.6	6.6
2	44238	2294	4.8	12.4	5.2	11370	512	4.2	11.3	14366	600	3.9	10.5
3	41944	1933	4.0	16.4	4.6	10858	486	4.0	15.2	13766	568	3.7	14.2
4	40011	1747	3.6	20.1	4.4	10372	482	3.9	19.2	13198	513	3.3	17.5
5	38264	1423	3.0	23.0	3.7	9890	349	2.9	22.0	12685	367	2.4	19.9
6	36841	980	2.0	25.1	2.7	9541	250	2.0	24.1	12318	279	1.8	21.7
7	35861	710	1.5	26.6	2.0	9291	195	1.6	25.7	12039	220	1.4	23.1
8	35151			73.4		9096			74.3	11819			76.9

* - last entry in this row is the percent remaining in sample for each data set.

Table 2
Definitions of Sample Variables

Variable	Definition
NKIDLT6	Number of child less than six years old living at home.
NKID6T17	Number of child between six and seventeen years old living at home.
AGE	Age in years.
HS	1 if high school degree only, 0 otherwise.
COL	1 if college degree, 0 otherwise.
NORTHC	1 if living in the North-Central region, 0 otherwise.
SOUTH	1 if living in the South, 0 otherwise.
WEST	1 if living in the West, 0 otherwise.
MSA	1 if living in Metropolitan Statistical Area (MSA) or Primary MSA, 0 otherwise.
URBAN1*	1 if living in an urban area, 0 otherwise, for years 1968-1982.
URBAN2	1 if living in an urban area, 0 otherwise, for years 1983-1986.
MAR	1 if married and spouse present, 0 otherwise.
WIDOW	1 if widower, 0 otherwise.
DIVSEP	1 if divorced or separated, 0 otherwise.
BLACK	1 if race listed as black, 0 otherwise.
OTHNW	1 if race listed as other nonwhite, 0 if race listed as black or white.
NFAM	Number of family members.
EXFAM	1 if family members include others than husband, wife, or children, 0 otherwise.
SCHOOL	1 if in school, 0 otherwise.
LFP	1 if worked, 0 otherwise.
OWN	1 if home is owned by household member, 0 otherwise.
MOVE	1 if moved between waves, 0 otherwise.
PMOVE	1 if might move in next couple of years, 0 otherwise.
PROXY	1 if interview given to proxy, 0 otherwise.
ILENGTH	Length of interview in minutes.
ILENG6872	Length of 1968 interview in minutes for years prior to 1972 and length of 1972 interview in minutes for after (and including) 1972.
CLENGTH	Length of time interviewer spent editing forms, in minutes.
INTCH	1 if there was a change in the interviewer where the new interviewer is another field interviewer and not a field supervisor.
NCALLS	Number of contacts made by the interviewer.
PHONE	1 if interview was conducted by phone, 0 otherwise.
PHONE7	1 in Wave 8 of SIPP90 if interview was conducted in person for the first six waves and by phone in the seventh, 0 otherwise.
PHONE73	Same as PHONE for individuals in PSID for whom interviews were conducted in person for 1968-72 and by phone in 1973, 0 otherwise.
NIMP	The number of imputations made to asset variables is SIPP.
NIMPMI	The number of minor imputations made to selected variables.
NIMPMA	The number of major imputations made to selected variables.
HHINC	Total household income in thousands of dollars divided by the Consumer Price Index.
NONLINC	household income minus individual's total labor earnings in thousands of dollars divided by the Consumer Price Index.
lnH	natural log of [(number of hours usually worked per week at job)x (number of weeks worked at job)].
lnW	natural log of [(total labor income divided by hours of work)/ Consumer Price Index].

* - There are two urban variables because the source of the information for

Table 3
Wave 1 Variable Means for Attritors and Non-attritors
(standard deviations for means in parentheses)

Variable	**SIPP84**		**SIPP90**		**PSID**	
	NAT	**AT**	**NAT**	**AT**	**NAT**	**AT**
AGE	42.477	42.004	42.524	41.849*	37.509	40.917**
	(0.168)	(0.280)	(0.159)	(0.286)	(0.515)	(0.533)
HS	0.540	0.545	0.555	0.565	0.516	0.391**
	(0.007)	(0.012)	(0.007)	(0.013)	(0.030)	(0.025)
COL	0.258	0.242	0.301	0.227**	0.209	0.126**
	(0.006)	(0.010)	(0.007)	(0.011)	(0.025)	(0.017)
SCHOOL	0.057	0.057	0.016	0.027*	0.007	0.013
	(0.003)	(0.006)	(0.002)	(0.004)	(0.005)	(0.006)
NKIDLT6	0.320	0.270**	0.315	0.273*	0.906	0.761*
	(0.010)	(0.015)	(0.009)	(0.017)	(0.057)	(0.033)
NKID6T17	0.707	0.586**	0.627	0.560*	1.708	1.440*
	(0.015)	(0.023)	(0.014)	(0.024)	(0.091)	(0.086)
NFAM	3.236	2.989**	3.100	2.913**	4.693	4.048**
	(0.023)	(0.039)	(0.022)	(0.042)	(0.105)	(0.107)
EXFAM	0.094	0.116*	0.108	0.175**	0.079	0.086
	(0.004)	(0.008)	(0.005)	(0.010)	(0.016)	(0.014)
MAR	0.850	0.768**	0.827	0.719**	0.986	0.917**
	(0.005)	(0.010)	(0.006)	(0.012)	(0.007)	(0.014)
WIDOW	0.008	0.012	0.009	0.010	0.000	0.011
	(0.001)	(0.003)	(0.001)	(0.003)	(0.000)	(0.005)
DIVSEP	0.075	0.134**	0.080	0.147**	0.007	0.032**
	(0.004)	(0.008)	(0.004)	(0.009)	(0.005)	(0.009)
BLACK	0.072	0.112**	0.058	0.126**	0.069	0.110
	(0.004)	(0.008)	(0.003)	(0.009)	(0.015)	(0.016)
OTHNW	0.022	0.034*	0.030	0.045*	0.014	0.054**
	(0.002)	(0.004)	(0.003)	(0.005)	(0.007)	(0.012)
HHINC	10.579	10.310	15.071	13.033**	10.901	10.215
	(0.233)	(0.332)	(0.250)	(0.457)	(0.283)	(0.387)
MSA/URBAN1	0.722	0.785**	0.804	0.842**	0.653	0.724
	(0.007)	(0.010)	(0.006)	(0.010)	(0.029)	(0.023)
NORTHC	0.278	0.186**	0.265	0.165**	0.339	0.263*
	(0.007)	(0.009)	(0.006)	(0.010)	(0.028)	(0.023)
SOUTH	0.293	0.350**	0.299	0.401**	0.253	0.303
	(0.007)	(0.012)	(0.007)	(0.013)	(0.026)	(0.024)
WEST	0.162	0.199**	0.192	0.237**	0.166	0.164
	(0.005)	(0.010)	(0.006)	(0.011)	(0.022)	(0.019)
OWN	0.750	0.636**	0.772	0.613**	0.747	0.590**
	(0.006)	(0.012)	(0.006)	(0.013)	(0.026)	(0.026)
number	4587	1688	4633	1448	277	373
percent	0.731	0.269	0.762	0.238	0.426	0.574

*, ** - Difference between means for attritors and nonattritiors is significant at the 5% and 1% levels, respectively.

Table 4

Parameter Estimates for the Attrition Equation*

Variable	SIPP84 Prob	SIPP84 t-stat	SIPP90 Prob	SIPP90 t-stat	PSID Prob	PSID t-stat
AGE	-0.0008	0.9461	0.0012	1.5088	0.0047	3.6347
AGE^2	0.0000	0.1337	-0.0002	2.1583	-0.0004	2.9057
HS	-0.0286	2.1797	-0.0049	0.3787	0.0201	0.9665
COL	-0.0483	2.8213	-0.0207	1.1203	0.0047	0.1837
AGE*HS	0.0007	3.2971	0.0002	0.9598	-0.0004	1.1451
AGE*COL	0.0013	5.1696	0.0005	2.2609	0.0000	0.0072
NKIDLT6	-0.0138	3.0952	-0.0032	0.7891	0.0003	0.0988
NKID6T17	-0.0023	0.6857	-0.0086	2.6413	0.0017	0.5664
NFAM	0.0040	1.6546	0.0004	0.1588	-0.0026	0.8685
EXFAM	0.0060	1.0892	0.0067	1.3640	-0.0026	0.3632
MAR	-0.0060	0.4747	-0.0105	0.6930	0.0114	1.0972
WIDOW	0.0108	1.0694	-0.0237	0.6973		
DIVSEP	0.0021	0.1479	-0.0075	0.4622	0.0194	1.6200
BLACK	0.0096	2.7962	0.0168	5.4878	0.0010	0.1700
OTHNW	0.0104	1.7728	0.0107	2.3270	0.0152	1.7641
SCHOOL	-0.0001	0.0264	-0.0133	1.4619	0.0330	1.2335
HHINC	0.0001	0.5720	-0.0001	0.4389	-0.0002	0.2851
$HHINC^2$	0.0000	0.8076	0.0000	0.8335	0.0001	1.7748
LFP	-0.0052	1.3806	-0.0161	4.2714	-0.0075	0.9455
MSA/URBAN1	0.0180	2.0910	0.0082	0.8053	0.0046	1.1941
URBAN2					0.0060	0.6558
NORTHC	0.0023	0.1794	-0.0043	0.2280	0.0122	1.0464
SOUTH	0.0076	2.9931	0.0142	5.7918	-0.0001	0.0295
WEST	0.0046	1.5042	0.0102	3.7276	-0.0061	1.2472
OWN	-0.0122	2.4708	-0.0196	4.4369	-0.0163	4.0724
MOVE	0.0116	2.5115	0.0133	2.9737	0.0027	0.7310
SAMEST68					-0.0098	2.8494

Table 4 - Continued

Variable	SIPP84 Prob	SIPP84 t-stat	SIPP90 Prob	SIPP90 t-stat	PSID Prob	PSID t-stat
ILENGTH	-0.0002	1.7741	-0.0002	1.4144	-0.0004	2.9210
ILEN6872					0.0004	2.7966
PROXY	0.0057	1.1375	0.0050	0.9824		
LENGPR	-0.0001	0.3584	-0.0002	0.9667		
CLENGTH	0.0001	0.4102	0.0003	1.8570		
NCALL					0.0047	5.4552
INTCH	0.0131	5.3654	0.0033	1.2328		
PHONE	0.0230	6.1720	0.0131	4.6505		
PHONE7/PHONE73			-0.0096	1.4588	0.0003	0.0202
NIMP/NIMPI**	0.0111	5.4109	0.0101	5.7879	0.0035	2.7700
NIMPMA**					0.0050	4.3673
T1					-0.0710	2.0847
T2	-0.0694	3.3818	-0.1010	5.3093	-0.1223	3.5584
T3	-0.0751	3.6638	-0.1178	6.1490	-0.1272	3.7045
T4	-0.0808	3.9230	-0.1219	6.3927	-0.1518	4.3492
T5	-0.0851	4.1123	-0.1209	6.3024	-0.1486	4.2437
T6	-0.0923	4.4841	-0.1326	6.8876	-0.1412	4.0154
T7	-0.0969	4.6845	-0.1398	7.2539	-0.1675	4.6065
T8	-0.1123	5.4264	-0.1532	7.7673	-0.1530	4.3277
T9					-0.1611	4.4479
T10					-0.1474	4.1988
T11					-0.1516	4.3073
T12					-0.1654	4.5596
T13					-0.1559	4.5228
T14					-0.1827	5.1108
T15					-0.1676	4.7909
T16					-0.1731	4.8571
T17					-0.1503	4.3548
T18					-0.1884	5.1187
T19					-0.1601	4.5666
T20					-0.1868	5.1844
sample size	37,625		36,855		6,989	
% atr=1/atr=0	0.955/ 0.045		0.955/ 0.045		0.947/ 0.053	

* - The coefficients are estimates of a change in the probability of attrition. A positive sign indicates an increase in the probability of attrition.

** - The imputation variables in the PSID are standardized to have a mean of zero and standard deviation of one. This is done because the number of variables for which imputations were possible changed from year to year.

Table 5
Parameter Estimates for Labor Supply Models
(absolute value for t-statistic in parentheses)**

Variable	SIPP84 TEMP	SIPP84 PERM*	SIPP90 TEMP	SIPP90 PERM*	PSID TEMP	PSID PERM*
Model 1						
Hours Equation						
LNWAGE	-0.0164	-0.0821	0.0415	-0.0565	0.1130	-0.0625
	(0.09)	(3.51)	(0.19)	(3.51)	(1.14)	(2.60)
NONLINC	0.0003	0.0007	0.0002	0.0002	0.0204	0.0303
	(0.43)	(1.36)	(0.59)	(0.60)	(1.73)	(2.18)
NKIDLT6	0.0036	-0.0023	-0.0180	0.0050	-0.0019	-0.0207
	(0.38)	(0.56)	(1.62)	(1.71)	(0.17)	(1.45)
NKID6T17	-0.0057	0.0028	-0.0210	-0.0006	0.0017	0.0077
	(0.70)	(1.34)	(2.27)	(0.32)	(0.34)	(1.08)
AGE	-0.0064	0.0024	-0.0004	0.0031	-0.0047	-0.0030
	(1.70)	(4.39)	(0.13)	(6.76)	(3.49)	(3.98)
HS		0.0259		0.0431		0.0462
		(2.81)		(5.28)		(4.68)
COL		0.0875		0.0943		0.1141
		(5.69)		(7.70)		(6.37)
Selection Bias Correction Terms						
ρ		0.4110		0.2013		-0.8921
		(3.36)		(1.72)		(0.85)
HOURS: $\lambda_1 \backslash \lambda_2$	0.0073	-0.1979	0.0111	-0.2632	0.0538	-0.2508
	(0.23)	(2.09)	(0.44)	(3.76)	(1.34)	(1.56)
WAGE: $\lambda_1 \backslash \lambda_2$	0.0261	0.1605	-0.1394	0.5853	0.1975	0.3368
	(0.49)	(0.98)	(3.66)	(3.15)	(3.32)	(2.49)
NOBS	19,626		23,477		3,893	
Model 2						
Hours Equation						
LNWAGE	-0.0800	-0.0888	0.0593	-0.0561	0.1102	-0.0647
	(0.43)	(3.79)	(0.28)	(3.47)	(1.09)	(2.69)
NONLINC	0.0004	0.0007	0.0002	0.0002	0.0206	0.0291
	(0.60)	(1.50)	(0.53)	(0.58)	(1.75)	(2.10)
NKIDLT6	0.0034	-0.0022	-0.0186	0.0051	-0.0047	-0.0199
	(0.35)	(0.52)	(1.71)	(1.79)	(0.43)	(1.41)
NKID6T17	-0.0057	0.0028	-0.0213	-0.0006	0.0005	0.0081
	(0.69)	(1.34)	(2.31)	(0.30)	(0.10)	(1.15)
AGE	-0.0053	0.0025	-0.0004	0.0032	-0.0049	-0.0030
	(1.34)	(2.55)	(0.14)	(6.77)	(3.46)	(3.97)
HS		0.0267		0.0425		0.0455
		(2.88)		(5.18)		(4.59)
COL		0.0902		0.0933		0.1142
		(5.88)		(7.58)		(6.31)
Selection Bias Correction Terms						
HOURS EQ		-0.2302		-0.2632		-0.2627
		(2.02)		(3.42)		(1.59)
WAGE EQ		0.1363		0.5626		-0.3323
		(0.78)		(3.00)		(2.39)
NOBS	18,858		22,766		3,726	

Table 5 - Continued

Variable	SIPP84 TEMP	SIPP84 PERM	SIPP90 TEMP	SIPP90 PERM	PSID TEMP	PSID PERM
			MODEL 3 - non-attritors			
			Hours Equation			
LNWAGE	-0.2291	-0.1042	0.0779	-0.0548	0.1218	-0.0926
	(1.10)	(4.05)	(0.35)	(3.40)	(1.57)	(3.28)
NONLINC	0.0006	0.0009	0.0002	0.0002	0.0203	0.0106
	(0.93)	(1.77)	(0.43)	(0.54)	(1.54)	(0.63)
NKIDLT6	0.0063	-0.0005	-0.0171	0.0056	-0.0120	-0.0716
	(0.62)	(0.12)	(1.52)	(1.99)	(0.65)	(3.53)
NKID6T17	-0.0052	0.0024	-0.0221	-0.0001	0.0081	0.0767
	(0.57)	(1.02)	(2.39)	(0.03)	(1.25)	(5.60)
AGE	-0.0018	0.0028	-0.0006	0.0032	-0.0061	-0.0120
	(0.35)	(4.41)	(0.17)	(6.79)	(3.66)	(4.23)
HS		0.0244		0.0412		0.1009
		(2.46)		(4.87)		(4.13)
COL		0.0924		0.0945		-1.1721
		(5.74)		(7.58)		(2.67)
			Selection Bias Correction Terms			
HOURS EQ		-0.3421		-0.2638		0.0314
		(2.04)		(3.31)		(1.97)
WAGE EQ		-0.2369		0.4828		-0.0161
		(1.50)		(2.54)		(0.52)
NOBS	17,241		21,462		3,240	
			Model 4 - attritors			
			Hours Equation			
LNWAGE	-0.1147	-0.0168	-0.0650	-0.0780	-0.1349	-0.2011
	(0.77)	(0.18)	(0.35)	(1.37)	(1.53)	(4.13)
NONLINC	-0.0001	-0.0017	0.0023	0.0006	0.0395	0.0701
	(0.05)	(0.41)	(0.38)	(0.32)	(0.81)	(1.37)
NKIDLT6	-0.0366	-0.0239	-0.1160	-0.0075	-0.0063	-0.0280
	(0.49)	(1.17)	(1.23)	(0.57)	(0.24)	(1.01)
NKID6T17	-0.0375	0.0133	0.0405	-0.0134	-0.0082	0.0238
	(0.59)	(1.53)	(0.62)	(1.42)	(0.51)	(1.04)
AGE	-0.0029	0.0002	-0.0020	0.0008	-0.0097	-0.0178
	(0.15)	(0.10)	(0.08)	(0.52)	(2.72)	(1.56)
HS		0.0640		0.0871		-0.0637
		(1.90)		(2.73)		(0.31)
COL		0.1200		0.1021		-0.5170
		(2.01)		(2.20)		(0.56)
			Selection Bias Correction Terms			
HOURS EQ		-0.0272		-0.1279		-0.1889
		(0.10)		(1.37)		(0.55)
WAGE EQ		1.4456		0.5226		-0.0863
		(4.92)		(2.36)		(0.40)
NOBS	1,617		1,304		486	

* - The estimates is this column are the sum of the estimates of the coefficients for the current and mean variables.

** - t-statistics are obtained using heteroskedastic-consistent standard errors.