

A RAND NOTE

Sample Dynamics: Some Behavioral Issues

Lee A. Lillard

August 1990

RAND

The research described in this report was supported by The RAND Corporation as part of its program of public service.

This Note contains an offprint of RAND research originally published in a journal or book. The text is reproduced here, with permission of the original publisher.

The RAND Publication Series: The Report is the principal publication documenting and transmitting RAND's major research findings and final research results. The RAND Note reports other outputs of sponsored research for general distribution. Publications of The RAND Corporation do not necessarily reflect the opinions or policies of the sponsors of RAND research.

A RAND NOTE

N-3142-RC

Sample Dynamics: Some Behavioral Issues

Lee A. Lillard

August 1990

RAND

Sample Dynamics: Some Behavioral Issues*

Lee A. Lillard

1. INTRODUCTION

Panel surveys are most useful for the analysis of models of dynamic behavior, change over time, latent variable models, and measurement error—the kinds of models discussed in the paper by Solon. Heckman discusses circumstances when independent cross sections may do as well or better. Panel data are usually treated as observations of a given economic unit for a fixed number of periods. The purpose of this paper is to point out some of the ways that panel samples themselves are dynamic in ways that may influence or be influenced by the behaviors under study.

Almost every empirical study, whatever the behavioral issue, is based on a sample containing relevant information from some target population. In panel surveys, the dynamics of sample membership can play a potentially important role in the content of any particular sample and thus in the consistency and precision of empirical results. While this is true of all longitudinal surveys, it is particularly true of family- or household-oriented surveys like the Panel Study of Income Dynamics (PSID) and the Survey of Income and Program Participation (SIPP). The central focus of this discussion is the PSID because of its unique position as a long-running (20 uninterrupted years), family-oriented survey that attempts to follow all members of the original sample households, and which has collected valuable supplemental information at various points along the way. The purpose of this paper is to point out some aspects of sample dynamics that may have implications for the estimation of behavioral models and that may not be widely recognized. This discussion relies heavily on my own experience with the PSID¹ and other panel data, and in particular on a study of the representativeness of the PSID after 14 years. See Beckett et al. (1983, 1987). The PSID data are described more fully in the appendix to this paper.

Section 2 introduces the issues of sample dynamics that arise in family-oriented panel surveys. These are primarily related to patterns of entry into and exit from the panel. While the implications of sample dynamics should be considered systematically for each type of behavior under study, some general empirical results may be of value. First,

Reprinted from *Panel Surveys*, Wiley Series in Probability and Mathematical Statistics, Daniel Kasprzyk et al. (eds.), pp. 497–511, © 1989 by John Wiley & Sons, Inc. Reprinted by permission.

Section 3 discusses the potential influence of entry and exit of sample members on behavioral outcomes. The approach is applied to earnings and labor supply models. Second, Section 4 considers the influence of recent behavior on the likelihood of leaving the panel, both for original sample members and new entrants. Section 5 summarizes our conclusions.

2. IN WHAT SENSE IS A LONGITUDINAL SAMPLE DYNAMIC?

First, consider the initial sampling frame, which interacts with the other dimensions of sample chance. Are households or individuals selected at random or on the basis of a probability sample, or are certain groups oversampled? The PSID, for example, initially oversampled low-income households. In 1968, the PSID interviewed 4,802 households/families, of whom 2,930 were selected from the Survey Research Center's master sampling frame. These families are called the SRC sample. Another 1,872 families were drawn from the Bureau of the Census's Survey of Economic Opportunity and are called the SEO sample. Thus one portion (about three-fifths) of the resulting sample (SRC) was a self-weighting nationally representative sample and the other portion (SEO) was a low-income sample based on family income prior to the initial 1968 interview in relation to twice the poverty line. Weights are provided that made the initial full sample nationally representative. Similarly, the SIPP provides weights that made that sample representative. There is currently a controversy about whether these weights should be used in the estimation of behavioral models, as is evidenced by discussion at this conference. Without judging whether weighting is appropriate or not, I will make some comments about the implications of sample dynamics for weighting if it is used.

What then happens to sample composition in a household- or family-oriented survey that runs for many years? The original families or household units do not remain as they were. Couples split up, individuals going their own ways to form new family units; children grow up, leave home, and form their own families; individuals die and the survivors form new household units; some individuals or families decide to leave the survey; and some new individuals join the households of members of the original survey, and perhaps later leave those household units. The implication is that there is entry into and exit from sample households, that these households split to form many households, and that some households exit altogether.

2.1. Attrition

A well-known major source of sample dynamics is attrition from the sample. Attrition of individuals is drastically reduced when all individuals of the original households are followed. This is extremely im-

portant for studies of living arrangements, the emergence of children into adulthood, migration, and poverty among female-headed households, for example. However, not all individuals can be found or will agree to participate in the survey repeatedly as time goes on.

The attrition rate experienced by the PSID has been quite low. After an initial loss of about 14% between the first and second waves, the annual full sample attrition rate has been about 2% among the original sample members. The rates are slightly higher for male heads and slightly lower for wives and female heads (including childless women). See Table 1 for the distribution by years in the sample. The effect of attrition accumulates, leaving only about 60% of the original sample members present after 14 years for the 1981 survey. These attrition rates are similar to and compare favorably with the experience of other annual panel surveys such as the National Longitudinal Surveys (NLS). The NLS surveys are individual- rather than family-oriented, so that the comparison is with male heads or wives and female heads. See Rhoton (1986).

There is some evidence appearing at this conference and elsewhere that response rates are "per survey requested of respondents." That is, for example, the attrition rates are not dramatically lower after two- or three-year intervals in the NLS than after the single-year intervals, and the attrition rates are not much better between the much more frequent surveys in the SIPP. More will be presented later concerning attrition in the PSID.

These rates of attrition, resulting in a loss of 40% of the original sample over 14 years, introduce substantial potential for the remaining sample to become less and less representative over time. Since the issue of the effects of attrition must be addressed separately for each behavioral study, it is very important that the data for those individuals who leave the sample over time be retained and made available for analysis.² These observations may be valid for some analyses, thus almost doubling sample size, and may be used to assess the impact of attrition for others.

2.2. Entry of Nonsample Individuals

A second major source of sample dynamics is much less widely recognized. This is the introduction of "nonsample" individuals who joined the households of original sample members, but who were not a part of the survey in 1968. While they are not a part of the official "sample," the full set of survey information is collected for these individuals. The information gain from using the additional data can be substantial. However, current procedures do not include following these individuals once they leave the household of a sample member. Therefore, the attrition rate for them is higher.

TABLE 1. DURATION IN THE PSID PANEL FOR THE ORIGINAL SAMPLE

Duration (Years)	Full Sample		Male Heads	Wives and Female Heads
	SRC	SEO		
1	14.5	10.1	14.3	11.8
2	2.7	3.4	3.9	2.9
3	2.2	1.9	2.6	2.1
4	1.8	2.6	3.1	1.7
5	2.2	2.5	2.7	2.1
6	2.1	2.1	3.0	1.9
7	1.8	2.3	2.2	1.7
8	2.1	3.0	2.6	2.4
9	1.6	2.5	2.1	2.1
10	1.9	1.6	2.0	1.9
11	1.6	2.4	2.0	1.6
12	1.6	2.7	1.7	1.8
13	2.3	2.7	2.2	1.9
14	61.6	60.4	55.7	65.1

Several issues arise immediately. First, how do these individuals affect the weighting scheme developed to adjust for the initial sampling frame? This an open subject for study. The current procedure in the PSID is to give these individuals a weight of zero so that they are implicitly excluded from any behavioral analysis using the individual weights.³ On the other hand, an unweighted analysis includes these individuals with full weight.

How many nonsample individuals are we talking about? Table 2 gives this information for male heads and for wives and female heads in the PSID. Compared to the 4,888 sample member male heads ever present in the PSID, the sample size can be enhanced by 32% to 6,452 by the inclusion of nonsample male heads. Similarly, the sample of wives and female heads can be increased from 6,292 by 26% to 7,903. Note, however, that this is virtually all wives. Since nonsample individuals are not followed when a split occurs, there are no nonsample female heads. Therefore, the sample composition is clearly and directly affected by the behavior of the individuals involved, namely marriage and separation. Wives are clearly overrepresented in an unweighted analysis of adult women, and the nonsample wives of sample members are completely excluded in a weighted analysis.

This introduces the second issue related to nonsample individuals. The inclusion of these individuals is *prima facie* not unrelated to certain behavior. They have moved into the household of a sample individual and thus have experienced events related to migration, living arrangements, and possibly marriage. The exit of these individuals is clearly

TABLE 2. SAMPLE SIZES, SAMPLE MEMBERSHIP, AND ATTRITION

Subsample	Total Sample	Male Heads	Wives and Female ^a Heads
Total ever present	30,903	6,452	7,903
Percentage sample members	76.5	75.8	79.6
Percentage remaining in 1981	66.5	66.5	71.7
Percentage sample members and remaining in 1981		49.4	57.2

^aIncludes single persons.

more directly related in an inverse way to these same events.

2.3. Implications for Cross-Sectional Samples

Cross-sectional sample sizes, as they would appear on each year's distribution tape, are presented in Table 3. The sample sizes are remarkably stable across the years for both male heads and women who are wives or female heads, as well as for the separate wife and female head subgroups. If the cross sections were independent, they approximately replicate a CPS-type survey. However, observations are not independent over time and the stability of sample size masks a great deal of compositional change in them over time. In fact, there is a tendency for the cross-sectional sample sizes to grow over time. Increasing sample sizes will not occur in individual-oriented samples like the NLS.

The proportion who are sample members declines substantially over the 14-year period. This is due to the marriage of children who split off from the parental household as they grow up, and to the dissolution and reformation of family and household relationships that involve new non-sample persons. Analyses based on weighted data would implicitly include only the sample individuals, and thus the increasing sample size is turned into a decline.

Those present in 1981 (and thus present in the file) are a declining proportion of the earlier years' potential full samples. The resulting sample sizes for earlier years increasingly underrepresent the corresponding years' samples. These represent the samples for earlier years that would be obtained from the 1981 distribution tape if the non-response records (as of 1981) were not incorporated.

The fraction representing the SRC sample is relatively stable, with some decline due to the split off of male children from the SEO sample (who had larger families) and due to greater marital instability among low-income households (resulting in more nonsample additions).

TABLE 3. CROSS-SECTIONAL SAMPLE SIZES AND COMPOSITIONS

Year	Male Head			Wives and Female Heads		
	Sample Size	Sample Members	Pres. 1981	Sample Size	Sample Members	Pres. 1981
1968	3253	100.0	57.3	4286	100.0	65.3
1969	2955	97.2	66.0	4007	98.0	73.1
1970	3054	93.8	68.6	4136	95.7	75.3
1971	3175	90.7	70.8	4294	93.6	76.7
1972	3326	87.7	73.2	4480	90.9	77.9
1973	3479	85.0	75.6	4655	88.8	79.6
1974	3620	83.2	78.3	4814	87.3	81.6
1975	3754	82.0	81.1	4950	86.3	83.5
1976	3831	80.2	84.9	5067	84.9	85.9
1977	3950	78.6	87.4	5141	83.8	88.4
1978	4069	77.1	90.2	5249	82.6	90.7
1979	4214	75.5	93.4	5450	80.2	96.2
1980	4351	74.1	97.2	5585	80.2	96.2
1981	4129	75.4	100.0	5372	81.1	100.0
Person Years	51,160	83.4	81.7	67,486	87.5	84.3

Year	Wives			Female Heads		
	Sample Size	Sample Members	Pres. 1981	Sample Size	Sample Members	Pres. 1981
1968	3012	100.0	67.3	1346	100.0	60.9
1969	2811	97.2	75.2	1269	99.8	68.2
1970	2870	94.1	76.9	1309	99.5	71.6
1971	2932	90.9	78.3	1362	99.5	73.3
1972	3038	87.1	79.4	1416	99.3	74.7
1973	3134	83.8	80.8	1468	99.4	76.8
1974	3206	81.2	82.9	1521	99.7	78.4
1975	3248	79.5	84.8	1568	99.6	80.8
1976	3287	77.1	87.1	1596	99.6	84.0
1977	3305	75.1	89.0	1627	99.8	87.5
1978	3371	73.2	90.7	1630	99.9	91.0
1979	3477	70.5	92.9	1693	100.0	93.0
1980	3552	69.1	95.7	1701	100.0	97.6
1981	3353	69.9	100.0	1743	100.0	100.0
Person Years	44,596	81.4	84.9	21,249	87.5	82.3

2.4. Implications for Longitudinal Samples

A particular advantage of panel data is the ability to study behavior in a longitudinal context. This includes components of variance models for panel data, dynamic models for individual time series, event dating and duration models, and many others. Frequently, the particular period of

time for which a behavioral unit is observed is only indirectly relevant (e.g., as an explanatory factor). All relevant intervals of observation may be used whether or not the beginning and ending dates of the time interval are the same. Similarly, unequal numbers of replication, or different interval lengths, may be a computational nuisance, but may present no conceptual difficulty.

There are some basic choices to be made in choosing a longitudinal sample, each of which may affect the appropriate choice of behavioral model and estimation procedure (see Table 4). First, are current (1981) nonrespondents to be included? Doing so enhances sample sizes substantially, by 50% for men and by 40% for women (assuming non-sample individuals are included). One implication of including the nonrespondents for behavioral modeling is that individuals are not observed for the same length of time (duration), and some are observed for non-overlapping time periods. This can introduce substantial complexity into the development of statistical software for dynamic models. Even this use of the data assumed that these individuals are missing at random.

The second choice is whether to include nonsample individuals for whom longitudinal data are available. This is related to the choice of whether to do weighted or unweighted analysis. There are potentially 6,452 male heads and 7,903 female heads and wives. If weighted analysis is used, then the longitudinal data for 1,564 nonsample men and 1,611 nonsample women receives zero weight. This is a substantial loss of information for these important groups. Unweighted analysis using the SRC (self-weighting) sample results in an even larger loss of sample size.

2.5. Individual Status Changes over Time

A third sense in which a panel sample is dynamic is that the individuals or households in a sample may change their "status" over time. For example, children become adults, heads of households, wives, and parents over the course of time. Single individuals marry, and couples divorce so that wives become female heads of households, and so forth. Therefore, the composition of a sample of individuals of a particular type can vary substantially over time and samples can be expanded to include individuals "ever as" the type. This is particularly important in the analysis of transitions and other dynamic behavior. To illustrate the substantial amount of change in status, Table 5 reports marital status and transitions between wife and female head (including childless) for women over the period that they are either.

3. ARE BEHAVIOR AND SAMPLE DYNAMICS RELATED?

This question must be answered in the context of each particular be-

TABLE 4. LONGITUDINAL SAMPLE SIZES

Subsample	Male Heads ^a		Wives and Female Heads ^a	
	Total	SRC	Total	SRC
<i>Total ever present</i>	6452	3928	7903	4554
Sample members	4888	3107	6292	3618
Nonsample	1564	821	1611	936
<i>Present in 1981^b</i>	4291	2638	5670	3250
Sample members	3189	2029	4522	2546
Nonsample heads	1102	609	1148	704
<i>Nonrespondents in 1981^c</i>	2161	1290	2233	1304
Sample members	1699	1078	1770	1182
Nonsample heads	462	212	463	232

^aIncludes single persons.

^bPresent on the 1981 Wave 14 public use tape.

^cMissing from the 1981 Wave 14 public use tape.

havioral model and application of panel data. In this section we take two simplistic approaches, in the context of the PSID panel. The first is to determine if the parameters of a behavioral model are different for those individuals entering or exiting. We consider earnings and labor supply equations here. The second is to determine whether attrition from the sample is related to demographic characteristics and recent behavior. These results are only meant to be suggestive of potential areas for further research. More rigorous approaches should be undertaken in the context of particular applications.⁴

If an analysis is to be based on individual time series for the behavioral variables or on event dates over the interval, then it is important for the duration of the interval of observation to be unrelated to the behavioral process being studied. In the case of event histories, censoring of event durations should be random. In the case of individual time series, or pooled replications, the outcomes for each period should be unrelated to attrition.

3.1. Is Behavior Related to Entry and Exit from a Panel?

One approach to considering the relationship between behavior and sample dynamics is to compare the behavior of individuals who have different patterns of entry and exit.⁵

Separate comparisons were made for male heads, for female heads,

TABLE 5. MARITAL STATUS AND TRANSITIONS FOR ADULT WOMEN

Duration as Wife or Head (Years)	No Change		One Change	Two Changes	Three or More
	Always a Wife	Always a Head			
1	67.6	32.4			
2	46.4	44.7	8.7		
3	41.3	40.0	17.0	1.6	
4	40.3	38.5	18.8	2.2	0.0
5	40.5	33.1	19.8	4.3	2.1
6	43.5	23.0	22.6	7.2	3.4
7	44.2	28.1	21.2	4.1	2.3
8	40.0	25.0	24.2	7.9	2.8
9	37.1	27.4	24.4	7.5	3.4
10	39.6	20.1	29.9	7.7	2.6
11	36.8	19.4	24.8	13.9	5.0
12	34.3	22.5	26.4	12.2	4.4
13	36.0	17.1	29.5	10.6	5.8
14	51.4	20.7	19.0	6.4	2.3

and for wives in the PSID. The behavior is represented by a standard log annual labor earnings equation and a simple log annual hours of work equation for those who worked in each group. Standard regression procedures are used and no attempt is made to correct for self-selection of workers. Both weighted and unweighted comparisons were made. The regression equations include race (indicators for black and for other nonwhite), years of schooling (linear spline below and above high school), potential work experience and its square, and census regions. The entry and exit patterns to be compared here are represented by a comparison of (1) those who subsequently leave the panel versus those who do not, and (2) sample versus nonsample individuals.

To compare those who leave the panel with those who do not, we compare the estimated equations for 1967 earnings and hours (from the original 1968 survey) for three distinct subsets of the original 1968 sample members: (1) those who remain throughout the panel to 1981, (2) those who leave after 1975 but before 1981, and (3) those who leave at or before 1975. Various degrees of difference are allowed and tested (using a standard *F*-test), including a simple intercept difference and a fully different set of coefficients. For the most part, these results suggest that labor income and hours of work are related to these variables in the same way for those who stay and for those who leave. This was true for male heads, female heads, and wives.

To consider entry into a panel, we compare the estimated equations for 1980 earnings and hours (from the 1981 survey) for sample and for

nonsample individuals. Again, various degrees of difference are allowed and tested (using a standard F -test), including a simple intercept difference and a fully different set of coefficients. However, only unweighted comparisons were made, since nonsample individuals are assigned zero weight. And again, no significant differences were found.

3.2. Is the Attrition Rate Related to Recent Behavior?

Another approach to considering the relationship between behavior and sample dynamics is to compare the rate of attrition among individuals with different behavior. This introduces a wider range of behavior than the earnings and hours of work equations considered above.

In this section we consider a simple descriptive model of attrition from the sample and thus of duration in the panel. We use a simple (Weibull) proportional hazards model. In each period t , an individual has a probability of leaving the survey before the next period $t + 1$ given by a Weibull hazard function. The Weibull specification allows for an increasing, decreasing, or constant rate of exit as the duration in the sample increases. The proportional hazard formulation allows for the introduction of time-varying covariates; a log-linear regression specification is used here.

The estimates of the Weibull parameters and the regression coefficients are presented in Table 6 for both sample and nonsample individuals and for both male heads and for female heads and wives. These estimates are based on the longest period of continuous participation in the survey as a male head or as a wife or female head. Children, other household members, and nonsample persons enter into these categories after 1968.

First, the estimates indicate the significantly greater exit rate after the first interview. This is a widely recognized phenomenon in most panel surveys. The Weibull parameters (gamma 2 in particular) indicate further that the hazard rate is constant thereafter, indicating that an exponential model is sufficient. The level of the hazard rate (gamma 1) across groups indicates that women have a lower hazard rate than men—the rate is $\exp(\text{gamma } 1)$ —and that sample individuals have a lower hazard rate than nonsample individuals.

Some general patterns emerge for both men and women. Two clear results are that nonwhites are more likely to leave the panel, and that leaving the sample is related to mobility, having moved, or planning to move in the near future. Studies focusing on migration and related behavior may be affected by attrition.

The rate of exit increases with age and is especially great for men who have recently retired or become disabled. Therefore, behaviors related to aging and retirement may be affected.

The exit rate is greater for low-income individuals as measured by

TABLE 6. PROPORTIONAL HAZARD MODEL FOR SAMPLE ATTRITION

Variable	Male Heads		Female Heads and Wives	
	Sample	Nonsample	Sample	Nonsample
First interview	1.5006*** (0.1479)	-0.2127 (0.2147)	1.5092*** (0.1351)	-1.1929*** (0.2481)
SEO subsample	-0.0177 (0.0674)	0.0301 (0.1093)	-0.1211* (0.0660)	0.2611** (0.1269)
Age spline < 25	-0.0141 (0.0243)	-0.0402 (0.0329)	0.0207 (0.0199)	0.0346 (0.0262)
Age spline 25 < A < 45	-0.0092* (0.0047)	-0.0022 (0.0093)	-0.0116*** (0.0043)	0.0322*** (0.0107)
Age spline > 45	0.0265*** (0.0036)	-0.0072 (0.0117)	0.0367*** (0.0029)	0.0147 (0.0154)
Nonwhite	0.3277*** (0.0650)	0.5704*** (0.1085)	0.3021*** (0.0645)	0.4726*** (0.1282)
Taxable income (\$1,000)	-0.0182*** (0.0034)	0.0026 (0.0045)	-0.0427*** (0.0055)	-0.0120*** (0.0033)
Might move	0.2645*** (0.0598)	0.1886** (0.0925)	0.2532*** (0.0560)	0.4970*** (0.1036)
Just moved	0.3185*** (0.0717)	0.1199 (0.0971)	0.4037*** (0.0672)	1.1744*** (0.1062)
Welfare received	0.0316 (0.1121)	0.4150** (0.1666)	-0.0508 (0.0791)	-0.5949** (0.2548)
Retired or disabled	0.3539*** (0.0905)	0.8493*** (0.2151)		
Unemployed	0.5899*** (0.1130)	0.5517*** (0.1763)		
<i>Hazard Parameters</i>				
Gamma 1	-3.2989*** (0.5655)	-2.1600** (0.8956)	-4.5286*** (0.5056)	-4.1105*** (0.6875)
Gamma 2	0.0164 (0.0491)	-0.1483 (0.0970)	-0.0557 (0.0466)	0.0202 (0.0958)
Log l	-6347.3	-1878.1	-7287.6	-1542.8
Sample size	4,888	1,564	6,292	1,611
Mean number periods	8.1	4.8	8.7	4.6

Notes: Standard errors in parentheses; *, **, and *** denote significance levels of .10, .05, and .01, respectively, from zero.

the level of taxable family income. However, receipt of welfare is not associated with attrition except for nonsample men who are more likely to leave and nonsample women who are less likely to leave. This may be related to eligibility requirements for welfare. While the SEO subsample was a lower-income group in 1967, there is no difference in the

exit rate of SRC and SEO men. SRC and SEO women are different, but in opposite directions for sample and nonsample women. The signs are in a reasonable direction.

Men who were unemployed in the last year are significantly more likely to leave the sample before the next interview than those who were not. This raises the possibility that models of unemployment spells and job change will be affected by attrition. If there is heterogeneity in the probability of unemployment, then studies of unemployment based on detailed data collected in special surveys in the eighties may under represent the level of unemployment. These issues deserve further study.

4. CONCLUDING REMARKS

This paper has been concerned with possible interrelationships between the entry and exit of individuals to and from a panel survey and the behavior of the individuals in the panel. First, we described the most important dimensions of entry and exit, namely the entry of nonsample individuals who become members of the households of sample households and the exit of individuals through attrition. It was shown that the inclusion of nonsample individuals or of the earlier observations for those who leave the panel can dramatically affect the size and character of the sample used for the estimation of a particular behavioral model.

These issues are particularly important for a long-running family-oriented panel like the PSID, which we consider in detail, but are also relevant for shorter household-oriented panels like the SIPP and for long-running individual-oriented panels like the NLS.

Even the simple descriptive information about the dynamics of sample size in a panel survey that we presented here suggest some implications concerning weighting and imputation. Remember that the original weights at the beginning of panel are designed to adjust for stratified sampling procedures. Since panel data are most useful for models requiring longitudinal data, let us consider that case. We would also argue that all individuals with relevant panel intervals be included, whether or not an individual remains in the panel in the most recent survey (assuming for now that attrition is random). If one believes that weighting is appropriate for behavioral models, then the original 1968 sample weights would seem most appropriate. There is no need to update the weights over time. This begs the question of how to weight nonsample individuals should they be included. One possibility is to weight them the same as the other (sample) members of the 1968 households from which they are generated. If unweighted analysis is used, then the issue does not arise.

Some analysts have suggested imputation procedures to "fill-in" the

missing values generated by wave nonresponse. While imputation procedures may be useful for generating missing values for random item nonresponse, the idea seems futile in the case of wave nonresponse generated by attrition, or the entry of nonsample individuals. Consider the following implication of full imputation. There are 4,888 male heads who are sample members in the PSID. While there are potentially 68,432 person years of data, only 39,555 are actually observed, or 57.8%. Imputation would almost double the number of person years without adding any real information. While the suggestion is clearly inappropriate in the context of a very long panel like the PSID, it is less obvious in the context of shorter panels like the rotating groups in the SIPP. These issues can only become more important as time progresses.

In addition, we examined the relationship between entry and exit of individuals and their behavior. For simple earnings and hours-of-work regression equations we found no evidence of differences for those who subsequently leave the sample and those who do not and we found no difference for those nonsample individuals who entered the sample from the original sample members. However, we did find that the exit rate from the PSID panel after a given interview was related to certain individual "behavior" reported in that interview. That is, leaving the panel is related to planned and actual mobility, to age, retirement and disability, to income (but only partially to receipt of welfare), and to having experienced unemployment. This suggests that further, more careful, study should be given to these issues.

APPENDIX: FEATURES OF THE PSID DATA

The basic character of the PSID will be important to our discussion. In this section we outline the essential features. While the PSID data are distributed in a number of forms, we will limit our discussion to the individual longitudinal tape.

In 1968, the PSID interviewed 4,802 households/families. Of this group, 2,930 families were selected from SRC's master sampling frame. These families (and/or the members of these families) are called the SRC sample. The other 1,872 families were drawn from the Bureau of the Census's Survey of Economic Opportunity (SEO) and are called the SEO sample.

A major goal of the PSID was to facilitate study of the determinants of poverty. Since a random sample of 5,000 families would include too few poverty and minority families, the SEO sample was added to the PSID. The SEO sample is a subset of the approximately 30,000 families interviewed in 1966 and 1967 for the Survey of Economic Opportunity. One important criterion for inclusion in the PSID was that the household had family income in 1966 less than or equal to twice the

1966 poverty line for the corresponding family size.

The 1968 SRC sample may be treated as approximately a random sample of U.S. families. The User Guide to the PSID (SRC, 1984) suggests unweighted analysis if only the SRC sample is included and weighted analysis if the SRC and SEO are combined. In 1968, the PSID calculated weights for each family that represent the ex-ante probability that a family appears in the PSID. Each individual in the sample in 1968 is assigned the weight corresponding to the family. Since 1978 the individual weights have been recalculated annually for each individual in the sample to account for differential nonresponse rates in the succeeding waves of the PSID.

A substantial number of individuals in each wave after 1968 are assigned a weight of zero. These are the so-called nonsample individuals, that is, persons who entered the sample after 1968 through marriage or a living arrangement that placed them in the same household with a sample person. The PSID assigns a weight of zero to these individuals to indicate that they are not a part of the original panel design. However, the same survey information is collected for them. A child born into the sample household is given an individual weight that is the average of the parents' weights. Therefore, any weighted analysis implicitly omits them.

Another important feature of nonsample persons is that no attempt is made to continue interviewing them if they stop residing with a sample person. Sample persons, on the other hand, are pursued even if they leave their original family. Such sample persons are called split-offs.

Until recently, behavioral researchers have not had access to the records of those individuals or families who leave the PSID sample—nonrespondents. Any family or individual that does not respond to the most current wave survey is removed from the distribution tape.

The most frequently used analysis samples are (1) the SRC probability sample and (2) all sample members (with positive weights) from the combined SRC and SEO samples. Analysis has almost always been limited to those individuals present on the current distribution tape (1981 in this case). Therefore we consider each of these subgroups. The results reported in this paper are based on a nonrespondent file constructed for an evaluation of the representativeness of the PSID as of 1981 (Wave 14). The data thus refer to the period 1968–1981 rather than 1983 as in the Non-Respondent File recently released by the PSID.

REFERENCES

- Beckett, S., Gould, W., Lillard, L., & Welch, F., (1983), *Attrition from the PSID*, Unicon Report.
- Beckett, S., Gould, W., Lillard, L., & Welch, F., (1987, rev.), "The PSID after Four-

- teen Years: An Evaluation," *Journal of Labor Economics*, forthcoming.
- Lillard, L., & Waite, L. (1987), *Children and Marital Stability* (RAND working paper), RAND Corporation, Santa Monica, CA.
- Lillard, L. (1983), "A Model of Wage Expectations in Labor Supply," in *Panel Data on Incomes* (A. B. Atkinson & F. A. Cowell, eds.), ICERD Volume.
- Lillard, L., & Willis, R., (1978), "Dynamic Aspects of Earning Mobility," *Econometrica*, 46.
- Rhoton, P. (1986), "Attrition and the National Longitudinal Surveys of Labor Market Experience: Avoidance, Control and Correction," *IASSIST Quarterly*, 23-35.
- Survey Research Center (1984), *User Guide to the PSID*, Inter-university Consortium for Political and Social Research, Institute for Social Research, The University of Michigan, Ann Arbor.

NOTES

*Support for this paper was provided by the RAND Corporation, but it builds on earlier work by Beckett, Gould, Lillard, and Welch (1983, 1987) funded by the National Science Foundation.

¹See Lillard & Willis (1978) and Lillard (1983) for examples of models of earnings, poverty, and labor supply dynamics, and see Lillard & Waite (1987) for an application to marital stability and childbearing. Ongoing research includes analysis of the relationship of attrition to marital stability and the relationship between retrospective reports of marital status and contemporaneous reports over the period of a panel.

²The PSID has, for example, recently constructed a file of data for nonrespondents who had left the survey prior to 1983.

³An unsuspecting analyst might even overstate the sample size he has used since nonsample individuals are not otherwise indicated without a careful look at the documentation.

⁴One example, under study by the author, is the relationship between attrition and divorce. Individuals experiencing a divorce may be less willing to respond to a detailed survey about their behavior. If so, then estimated divorce rates will be understated unless the estimation procedure addresses the issue.

⁵These results are taken largely from the analysis of Beckett et al. (1983, 1987). That study also compared these behavioral relationships in the PSID with those in data from the CPS.

