

# PSID USER GUIDE

A complete user guide was written by Martha Hill and published by Sage in 1991. This guide, *The Panel Study of Income Dynamics: A User's Guide, volume 2. 1991*, is available for purchase from Sage [here](#) .

Three chapters are available below. Additional resources are available within the [documentation](#) section of the PSID website.

## Contents

[Introduction](#) \*

[Study Design](#) \*

Field Procedures

Data Preparation

[Data Quality](#) \*

Content

Data Files

Data Analysis

Getting Started

# INTRODUCTION

## Monograph purpose and content

This monograph describes the origins, design, procedures, and broad analytical potential of one of the major data bases in the social sciences--the Panel Study of Income Dynamics (PSID). After a brief review of the PSID's background, the monograph discusses the major design parameters, field procedures, and data preparation activities. These chapters describe the data collection process and serve as important background for understanding how PSID data come into being and what they represent. Issues of the quality of PSID data are addressed next, reporting evidence about a number of different quality dimensions. The remainder of the monograph delves into the data themselves--what topics are covered; what data files are available; and crucial information regarding analysis issues, key variables, and choice of data files. The "Data Analysis" chapter provides details of several analysis examples, so that the reader can see the assembly of parts needed

to create estimates of earnings regressions, long-run poverty status, changes in women's income following divorce, and correlations between parents' income and a child's adult income.

The monograph takes the reader from the drawing board to a completed product with a minimum of detail. Where most applicable, it notes other PSID documents, such as the PSID's documentation books and User Guide, that can provide further details about particular aspects of the study. In the final chapter, information is provided about obtaining these documents and the data files themselves. Throughout the monograph italics are used to distinguish terms with special meaning in the context of the PSID, and full capital letters are used to designate PSID variable names.

## Overview of the PSID

The Panel Study of Income Dynamics (PSID) is a longitudinal survey of a representative sample of U.S. individuals (men, women, and children) and the families in which they reside. It has been ongoing since 1968. Data are collected annually, and the data files contain the full span of information collected over the course of the study. PSID data can be used for cross-sectional, longitudinal and intergenerational analyses, and for studying both individuals and families. The study emphasizes the dynamic aspects of economic and demographic behavior, but it contains a wide range of measures, including sociological and psychological ones. Between 1968 and 1988, the PSID collected information regarding approximately 37,500 individuals and spanning as much as 21 years of their lives.

The general design and core content of the study have remained largely unchanged, and considerable effort has been expended cleaning the data. These two features greatly enhance the PSID's potential for longitudinal analysis. Preparation and distribution of comprehensive documentation and a User Guide also facilitate use of the PSID data.

The study has been conducted at the Survey Research Center, University of Michigan since its beginning in 1968, with the Inter-University Consortium for Political and Social Research (ICPSR) data archive handling the public distribution of the data files, documentation, and User Guide. PSID data files have been disseminated widely throughout the United States and to numerous

foreign countries.

Starting with a national sample of approximately 4,800 U.S. households in 1968, the PSID has traced individuals from those households since that time, whether or not they are living in the same dwelling or with the same people. Adults have been followed as they have grown older, and children have been observed as they advance through childhood and into adulthood, forming families of their own. Each year information is collected about the PSID's sample members (members of the PSID's 1968 sample families or their offspring) and their current co-residents (spouses,cohabitators, children, and others living with them), even if those co-residents were not part of original-sample families.

Because the original focus of the study was on the dynamics of poverty, the 1968 sample included a disproportionately large number of low-income households. The oversampling of families poor in the late 1960s resulted in a sizable sub-sample of blacks. Probability-of-selection weights enable analysts to make estimates from the sample that are representative of the U.S. population. In the absence of nonresponse bias, the PSID's rules for tracking individuals and families over time lead to accurate representation of the nonimmigrant U.S. population both cross-sectionally each year, and in terms of change, since 1968. To help correct for omissions in representing post-1968 immigrants, a representative sample of 2,043 Latino (Mexican, Cuban, and Puerto Rican) households was added in 1990.

The study's tracking rules, along with its Latino subsample addition in 1990, have meant substantial increases in the number of individuals in the study as it has progressed through time. In 1968 the PSID gathered information about approximately 18,000 individuals; by 1988 this number had grown to a cumulative total of about 37,500. Similarly, the number of family units has increased from just under 5,000 at the beginning of the study to about 7,000 currently, not including Latino households.

The PSID provides a wide variety of information at the family and individual level, as well as some information about the locations in which sample households reside. The central focus of the data is economic and demographic, with substantial detail on income sources and amounts, employment, family composition changes, and residential location. Content of a more sociological or psychological nature is also included in some waves of the study.

Information gathered in the survey applies to the circumstances of the family unit as a whole (e.g., type of housing) and to particular persons in the family unit (e.g., age, earnings). Some data are collected about all individuals in the family unit, but the most extensive data are gathered for the family head (who is male in married-couple families, but female or male otherwise) and wife. Information about the study's core topics (e.g., income, employment, family composition) is gathered annually, and this is supplemented with data on additional topics (e.g., health, wealth, retirement plans, flows of time and money help among families and their friends, and motivation and efficacy) gathered intermittently. The amount and variety of data are substantial; over 300 pages are required to list, by topic and wave, the variables on the study's main, cross-year data file.

The PSID staff merges each new wave of data with prior waves to provide comprehensive coverage of information collected for individuals and families over the entire course of the study. These multi-wave data files become publicly available upon completion of the merging, numerous data-quality checks, and generation of variables. This usually occurs 18-24 months following the completion of interviewing.

### Origin of the PSID

As part of Lyndon Johnson's War on Poverty, the Office of Economic Opportunity (OEO) directed the U.S. Bureau of the Census to conduct a nationwide assessment of the extent to which the War on Poverty was affecting people's economic well-being. This Census study, called the Survey of Economic Opportunity (SEO), completed interviews with about 30,000 households, first in 1966, and again in 1967.

Interest in continuing this national study of economic well-being led OEO to approach the Survey Research Center (SRC) at the University of Michigan about interviewing a subsample of approximately 2,000 low-income SEO households. Professor James N. Morgan, who became the new study's director at SRC, argued successfully for adding a fresh cross-section of households from the SRC national sampling frame so that the new study would be representative of the entire population of the United States, including non-poor as well as poor households. It was also decided to follow, and keep as part of the sample, members of the families

who moved away from their original households to set up new households, such as children who came of age during the study. In this way, the sample could remain representative of the nation's families and individuals over time. The study came to be known as the Panel Study of Income Dynamics. It began interviewing in 1968, successfully completing interviews with 4,802 households across 40 states--1,872 low-income households from the SEO plus 2,930 households drawn from the SRC national sampling frame. The year 1991 marked the study's 24<sup>th</sup> annual wave of interviewing, with its family units having substantially increased in number and having spread to cover all 50 states as well as some other countries.

### Administration and funding

The PSID has been funded principally by a collection of federal agencies, including the Office of Economic Opportunity; the Assistant Secretary for Planning and Evaluation of the Department of Health, Education and Welfare (now Health and Human Services); the Departments of Labor and Agriculture; the National Science Foundation; the National Institute of Child Health and Human Development (NICHD); and the National Institute on Aging (NIA). The Ford, Sloan, and Rockefeller foundations have provided important supplementary grants to the PSID. Since 1983, the National Science Foundation (NSF) has been the principal sponsor of the study, with substantial continuing support from the Office of the Assistant Secretary for Planning and Evaluation (ASPE) of the Department of Health and Human Services. Since 1982, the study has had an advisory Board of Overseers, created by the National Science Foundation to foster input from the national community of scholars, researchers and policy makers.

Throughout its history, the PSID has been conducted at the Survey Research Center, which is located within the Institute for Social Research at the University of Michigan. Professor James N. Morgan, now emeritus, directed the study from 1968-1989. Beginning in 1982, Professor Greg J. Duncan became co-director, and subsequently director when Professor Morgan retired in 1989.

### Illustrative uses of the PSID

Two key features give the PSID its unique analytic power: (i)

individuals are followed over very long time periods and in the context of their family setting; and (ii) families are tracked across generations, with interviews often conducted simultaneously with multiple generations of the same families. The type of information the study collects, in conjunction with these unique qualities, builds a number of strengths, including the following:

- Continuous representation of families and of individuals of all ages.
- Long annual--and, in some cases, monthly--time series of employment, income, and demographic information, reported through annual interviews.
- Extensive intergenerational information, with a long time-series of adulthood information obtained from each generation directly, information after individuals have become adults as well as during their childhoods, and comparable detail for all children from the same families.
- Coverage of diverse supplemental topics (e.g., health, wealth, saving, kinship).
- Recent additions of information, accessible to data users under special circumstances, about neighborhoods (e.g., Census tract) and about health and mortality (e.g., from Medicare records and the National Death Index).
- Extensive longitudinal, as well as cross-sectional, checking of the data and comprehensive documentation of the full data set since its start in 1968.

These features make the PSID one of the most widely used and influential data sets in the social science research community. Some 200 institutions have requested copies of the PSID data. And over 700 publications using PSID data have appeared in economic, demographic and sociological journals and books. The data are also extensively used for dissertations, reports, conference presentations, and working papers. A comprehensive bibliography is available from the PSID staff upon request (see final section of this monograph).

Areas of basic economic research addressed with the data include: labor supply, consumption, life-cycle earnings, unions, compensating wage differentials, dynamic aspects of income distribution and various methodological studies. PSID topics of

interest to researchers in several disciplines--demographers, sociologists, psychologists, and economists--include poverty and welfare experiences during adulthood or childhood; motivation and economic mobility; changes in family structure (e.g., births, divorce, remarriage); child support; out-of-wedlock births; teenage childbearing; and the intergenerational transmission of economic status. This diversity of topics reflects the philosophy of the PSID to ask limited sets of questions about a wide variety of topics rather than extensive questions about only a few topics. The study's multi-faceted information is couched in the context of substantial detail about income, employment, and family composition.

## **STUDY DESIGN**

### Overall design

The PSID gathers information about families and all individuals in those families through its annual interviews.<sup>1</sup> A single primary adult--usually the male adult head,<sup>2</sup> if there is one--serves as the sole respondent. Sometimes the wife (or cohabitor, referred to as "wife") of the head agrees to grant an interview when the head does not. The single household respondent provides information about him/herself and about all other family members.<sup>3</sup>

The study's original households constitute a national probability sample of U.S. households as of 1967. Its rules for following household members were designed to maintain a representative sample of families at any point in time as well as across time. To accomplish this, the PSID tracks members of its wave-1 (1968) families, including all those leaving to establish separate family units. Children born to a member of an original-sample member are classified as sample members and are eligible for tracking as separate family units when they set up their own households. Ex-spouses and other adult sample members who move out of PSID family units are tracked to their new family units. This procedure replicates the population's family-building activity and produces a dynamic sample of families each year. New PSID families form when children grow up and establish separate households or when marriage partners go separate ways. This results in growth over time in both the number of family units and

the number of people residing with a sample member at some time during the study.

Information is gathered about all persons residing in the family unit, but in most waves there is only one respondent per family unit (usually the head). The most detailed information is collected each year about the heads of family units. Since the late 1970s, however, the PSID has sought to collect the same detail for wives/"wives" (by "wives" we mean cohabitators) as for heads. For special supplements gathering retrospective history information in 1976 and 1985, the study conducted separate interviews with all wives/"wives" of heads as well as their husbands. Except for the very early years of the study, cohabitators have been treated in a similar manner to husbands and wives.

The general design of the study has remained largely unchanged over time; however, the mode of interviewing has changed. From 1968 through 1972, the PSID conducted in-person interviews. In 1973, to reduce costs, the study began taking the majority of interviews by telephone. Since that time, in-person interviews have been conducted only with respondents who do not have telephones (roughly 500 each year), or who have special circumstances which make a telephone interview unfeasible. To further reduce costs, and because long interviews are difficult by telephone, interview length was also reduced in 1973. The interview averaged about one hour when it was conducted in person; since the change to telephone interviewing the length has averaged 20 to 30 minutes.

As discussed in the "Content" chapter, the PSID has maintained a core of questions addressing issues relevant to income dynamics and demographic change. In addition to the central core, there have been a number of supplements to the core, adding questions on a wide variety of other topics. These supplements have led to the creation and release of a number of special files that complement the main PSID data files.

## **SAMPLE DESIGN**

### Sample frame



The initial sample for the PSID actually consisted of two independent samples: a cross-sectional, national sample (based on stratified multistage selection of the civilian noninstitutional population of the U.S.) and a national sample of low-income families.<sup>4</sup> The cross-section sample was drawn by the Survey Research Center (SRC). Commonly called the SRC sample, it was an equal probability sample of households in the 48 coterminous states designed to yield about 3,000 completed interviews. (In fact 2,930 interviews were taken in 1968 from this sample).

The second sample of responding PSID families, known as the SEO sample, came from the Survey of Economic Opportunity (SEO), conducted by the Bureau of the Census for the Office of Economic Opportunity. The PSID selected from the SEO's sample, the goal being to obtain about 2,000 low-income families with heads under 60 years old. In fact, 1,872 families were successfully interviewed. The SEO sample was confined to Standard Metropolitan Statistical Areas (SMSAs) and to non-SMSAs in the Southern region, and it involves unequal selection probabilities.

Both the SRC and SEO were subject to nonresponse in the first wave (1968). Three factors played a special role in preventing successful interviews with the SEO sample:

1. There was nonresponse in the original Census survey from which the SEO sample was selected.
2. Sampled Census respondents were asked by the Bureau of the Census to sign a release to allow their names to be passed on to SRC. Approximately one quarter of the households failed to sign the release.
3. OEO failed to transmit some sampled addresses to SRC.

The PSID sample combines the SRC and SEO samples. Both samples are probability samples (i.e., samples for which every element in the population has a known non-zero chance of selection). Their combination is also a probability sample. The combination, however, is a sample with unequal selection probabilities, and as a result compensatory weighting is needed in estimation, at least for descriptive statistics. (The various disciplines disagree about the need for weighting in model-based estimation.) Weight adjustments are also needed to attempt to compensate for differential nonresponse in 1968 and subsequent waves. As explained in the

"Data Analysis" chapter, and detailed in the PSID's technical documentation, weights supplied on PSID data files are designed to compensate for both unequal selection probabilities and differential attrition.

### Latino supplemental sample

The original PSID sample contains too few Latino households to provide reliable estimates either for Latinos as a group or for major subgroups of Latinos. In addition, Latinos entering the U.S. since 1968 are not represented in the basic PSID sample unless they co-reside with persons in the U.S. in 1968. To help reduce these shortfalls a sample of 2,043 Latino households was interviewed and added to the PSID sample beginning in the 1990 wave. Funding for this supplemental sample came from the Ford Foundation, the Rockefeller Foundation, the ASPE in the Department of Health and Human Services, the Employment Training Administration in the Department of Labor, and NSF.

The 1990 addition of a Latino sample is designed to provide precisely the kind of representative information about Latinos that is now available for blacks and non-Latino whites in the original PSID. The sample was originally selected for the Latino National Political Survey (LNPS), a 1989 study of the political participation of Latino households cosponsored by the InterUniversity Program for Latino Research and administered by the Center for Mexican American Studies of the University of Texas at Austin. The LNPS did not attempt to cover the entire Latino population in the U.S. However, it covered at least 89 percent of the three largest Latino subgroups--the Mexican, Puerto Rican and Cuban origin populations.

Latino sample members were asked extensive background information in their initial interview in 1990, including marital and fertility histories. In addition, questions were added to the 1990 interview to enable calculation of selection-probability weights so that the Latino sample can be combined with the existing PSID sample for analysis purposes. The PSID plans to continue gathering comparable information from the Latino and original PSID samples and to combine the two samples in PSID data files.

## Tracking rules

The PSID's tracking rules call for following members of the original family units and their adult offspring to whatever living arrangements they experience. Information is gathered about these sample members and their co-residents if they are living in a household (i.e., non-institutional) situation. A family member who moves out of a PSID family is eligible for interviewing as a separate family unit if he or she is a sample member and he or she is 18 years old or older and living in a different, independent household.<sup>5</sup> If a sample member 18 or older moves to an institution such as a prison, a college dormitory, or the military the PSID records this fact and attaches an institutional status data record to the family he or she left. The PSID keeps track of the location of sample members living in institutional housing. Interviews are attempted with them if and when they leave the institution to set up their own households.

# DATA QUALITY

## Response rates

As noted at the end of the "Field Procedures" chapter, the PSID takes a number of special measures to try to ensure high follow-up response rates. Annual response rates have been exceedingly high in every year except the first. In 1968, the PSID's first year, 76 percent of sampled families were successfully interviewed. In 1969, interviews were attempted with the heads of family units containing adults who were members of 1968 interviewed families. The response rate in 1969 was 88.5 percent. Since 1969, annual response rates have ranged between 96.9 and 98.5 percent. With a minor exception in 1990, no attempt has been made to recontact attriters from previous years. Even small attrition from wave to wave cumulates over time. As of 1988, the response rate for individuals who lived in 1968 households was 56.1 percent. The level of cumulative response is sufficiently low to raise concerns, and this has prompted direct investigation of possible attrition biases.

## Unicon study

In 1982, the Unicon Research Corporation was commissioned by the National Science Foundation to conduct comparisons of the descriptive characteristics of individuals who had attrited and those still remaining in the panel, and to estimate a series of models of earnings, labor supply and migration using data from early panel waves to see if subsequent attriters differed from respondents in behavioral terms. We quote directly from their results (Beckett et al., 1988: 490-491):

"In this article we examined the dynamics of participation in the PSID and considered whether attrition has affected the representativeness of the PSID. We found some observable variables that are correlated with attrition, but these variables explain only a negligible portion of the attrition in the PSID. We found no compelling evidence that attrition (or entry) has any effect on estimates of the parameters of the earnings equations we studied.

The 1968 PSID is quite unlike the population of the United States if we use the CPS as a benchmark..1 Weighting the PSID with the weights supplied by ISR goes a long way toward making the PSID sample resemble the CPS sample. While there are statistically significant differences in the empirical distributions of observable characteristics, most of these differences are of no practical significance or can be explained by known differences in coding of answers across the two surveys. For some variables, particularly income and education, there is some reason to believe that the reports in the PSID may be more accurate than those in the CPS. At any rate, the PSID participants behave almost identically, conditional on their observed characteristics, to participants in the CPS."

## Lillard-Waite study of marital histories

As part of a larger study of marriage and divorce, Lee Lillard and Linda Waite conducted an analysis of the quality of panel and retrospective marital histories in the PSID. Again, we quote directly from their report (Lillard and Waite, 1989: 252-253):

"Our comparison of panel and retrospective histories produced a detailed picture of the agreements and disagreements between the two. To summarize briefly, we found substantial levels of agreement on marital status as of the first survey interview, and substantial agreement on the occurrence of the first marriage. We found that the dates of first marriage matched best for those who were either married as of the initial interview or who married during the survey in the most typical pattern--living at home until marriage and then moving out. For these people dates from the panel and retrospective histories matched very well indeed.

Disruptions also appeared to be captured well by both types of histories, although we do observe disruptions in the panel that are not reported in the retrospective history and respondents report a substantial number of disruptions that the panel history misses. For those disrupted by both histories, dates of disruption match within a year for three-quarters; we suspect the other quarter are reporting on two different events.

...[O]n balance this data set is among the very best for studying the beginnings and ends of marriages. The large sample at all ages, the long panel period, the wealth of other information, and the multiple measures of the events in question all make the PSID an excellent source of information on marriage and divorce."

### Curtin, Juster, and Morgan study of wealth

As part of a general assessment of the quality of wealth data from surveys presented at the 1988 NBER Conference on the Measurement of Saving, Investment and Wealth, Curtin, Juster and Morgan (1988) evaluated wealth data gathered in the 1984 wave of the PSID, the 1983 Survey of Consumer Finances, and the 1984 Wealth Supplement to the Survey of Income and Program Participation (SIPP). A number of quality dimensions were investigated: sample and questionnaire design, response rates and nonresponse bias, ability to represent the upper tail of the income and wealth distribution, the size of measurement error, the importance of item nonresponse and imputations, and the degree to which the household survey adequately represents national wealth. They conclude (p. 544) that:

1. Measured against the standards set by previous household wealth surveys, all three of these data sets stand up quite well. They do not differ substantially among themselves when it comes to measuring total wealth and the distribution of wealth in the great bulk of the U.S. population.
2. The unique design characteristics of the SCF [Survey of Consumer Finances] give it the highest overall potential for wealth analysis of the three data sets examined....Comparing PSID to SIPP, one gets mixed picture, but, in general, PSID had the advantage. Although its basic sample design is less well suited to measuring wealth than SIPP (because it oversamples low-income families, for whom wealth holdings are relatively unimportant), its general descriptive characteristics, taking SCF as the benchmark, look to be closer to actual population characteristics than are those of SIPP. Although PSID is not able to describe the details of wealth holding nearly as well as SIPP because of its highly aggregated nature, its measurement error characteristics look to be consistently better than are those of SIPP. The PSID has a lower item nonresponse rate than SIPP and thus less need to construct imputed values, and it appears to be a somewhat closer match to external control totals."

### Other evidence on representativeness

Research papers periodically provide additional data on the representativeness of the PSID sample. In an article on PSID data quality Duncan and Hill (1989) compared 1980 official program totals and PSID reports of aggregate transfer income of various types. They found that the PSID accounted for 92 percent of income from the Aid to Families With Dependent Children (AFDC) program, 84 percent of Supplemental Security Income and 85 percent of Social Security income. As a frame of comparison, Current Population Survey reports for calendar year 1979 show that the CPS accounts for about 77 percent of AFDC, 69 percent of Supplemental Security Income and 91 percent of Social Security (U.S. Bureau of the Census, 1983, Table A-2, p. 216). The Census Bureau's Survey of Income and Program Participation does considerably better than the CPS in matching up with program aggregates than the CPS, accounting for about 79 percent of Aid to Families With Dependent Children, 94 percent of Supplemental

Security Income and 101 percent of Social Security (U.S. Bureau of the Census, 1985, Table D-3, p. 47).

As part of an analysis of the consequences of teenage childbearing, Duncan and Hoffman (forthcoming) compared high school graduation and marriage rates of black and white women in the PSID (at age 25) and Current Population Survey (at ages 25-29). Although there is some tendency for modest but persistent differences in some of these rates (e.g., black marriage rates are higher in the CPS than PSID; white schooling rates are somewhat higher in the PSID than CPS), the trends for both racial groups track fairly closely over the two decades.

As part of a research proposal submitted to the National Institute on Aging, Ken R. Smith compared the mortality experience of the PSID sample from 1968 to 1984 with life tables for the U.S. taken from 1980 Vital Statistics sources. He found close agreement in the five-year survival rates calculated from the two sources.

### Validation study

A crucial component of the quality of data from any survey such as the PSID is the validity of responses to the questions posed. To investigate this, the National Science Foundation, at the urging of the Board of Overseers, funded a two-wave validation study of the PSID instrument. Attempting to validate responses from actual PSID respondents was judged too costly, so the strategy adopted was to secure the cooperation of a large firm, interview a sample of workers (about 500) from that firm using the PSID instrument and then, whenever possible, to check carefully the responses recorded in the interviews against actual company records.

Evidence from the validation study sample (detailed in [Bound et al., 1989](#)) shows that the amount of measurement error in cross-sectional reports of annual earnings is rather low, with the ratio of error-to-total variance ranging from .15 to .30, depending on the year of measurement. Error in reports of annual work hours is higher (.28 to .37), while error in reports of hourly earnings, obtained by dividing annual earnings by annual hours, is disturbingly high (.67 to .69).

Although annual earnings were reported fairly reliably, it was also discovered that workers with lower-than-average earnings

tended to overreport and high-wage workers to underreport their earnings—a covariance almost always assumed to be zero in measurement error models. This covariance reduced from 18 to 24 percent the biasing effects due to errors in measuring earnings when earnings is a right-hand independent variable. Mean-reverting error also produced biases to right-hand side variable coefficients when annual earnings is a dependent variable that ranged from 10 to 17 percent. The restricted variability of true earnings from the single-company sample probably leads to an overstatement of these biases.

Furthermore, the validation data set also showed a surprisingly small decrement to reliability when going from cross-sectional measures of earnings level to panel measures of annual earnings change--there as more "news" than "noise" when earnings were differenced over either one- or four-year intervals. [\(Note\)](#). Reliability was also fairly high in panel reports of change in annual work hours. Indeed, apparently turbulent employment conditions produced cross-sectional reports of earnings and hours in one of the survey waves that were less reliable than the corresponding change measures.

The company sample also provided validation for retrospective reports over a two-and-a-half year period of spells of nonemployment with the firm. It showed that only one-third of the spells of nonemployment appearing in company records were reported in the interviews. Shorter and more distant spells were less likely to be reported, although the fraction of presumably salient longer and more recent spells unreported still exceeded one-third. Furthermore, the incidence of reporting error appeared to be correlated with typical right-hand measures such as age and schooling. Thus, all of the ingredients for coefficient bias due to measurement errors would appear to be present in unemployment event-history data.

### "SEAM" Transitions

In 1984 the PSID began coding information on labor-force status and program participation on a monthly basis. As has been found in other studies where the measurement period (e.g. month) is less than the length of the reference period (e.g. year), observed transitions tend to concentrate at the beginnings and ends of the



reference period. Hill (1987) compared the PSID with the Survey of Income and Program Participation (SIPP) in terms of the disproportionate concentration of transitions at the 'seam'. Perhaps because of the PSID's longer reference period, he found the extent of seam problems appreciably greater in the PSID than in SIPP data--especially for Food Stamp reciprocity.

Taking advantage of overlap in the 1984 and 1985 PSID reference periods, Hill used dual reports of employment status for the same month to examine individual characteristics associated with 'seam amplifying' and 'seam attenuating' inconsistencies. Age and race were found to be very strong predictors of seam amplifying inconsistencies--with Blacks and older individuals having significantly higher rates of concentration of transitions at seams. Gender and income, on the other hand, were the sole significant predictors of 'seam attenuating' inconsistencies--with high income females exhibiting a greater propensity for this type of response error.

The extent to which these types of response errors affect the estimated parameters of event-history models has not been fully worked out. Hill and Hill (1986), however, have demonstrated that with SIPP data whether the week in question is a 'seam week' is by far the most important predictor of transitions from unemployment and the existence of excessive seam (or insufficient within wave) transitions has profound impacts on the estimated survival functions. Interestingly, despite a smaller sample size, the chi-square goodness-of-fit statistic of the proportional-hazards model was more than twice as high for the PSID as for the SIPP. This suggests that the original over-sampling of low-income and minority households in the PSID has notable analytic advantages.

Taken as a whole, these different studies examine a variety of aspects of data quality, and the general results are supportive of the PSID data being valid and not subject to major nonresponse bias. Still, an analyst of any data set should be sensitive to possibilities of low validity or nonresponse bias for his or her particular analysis.

Author's notes

1. This refers to the deliberate oversampling of low-income families in the PSID's initial wave.
2. These design characteristics include a long questionnaire focused on components of wealth and an oversample of high-income households.
3. Bound and Krueger (1989) report a similar finding for the March 1977 and 1978 Current Population Survey, using Social Security earnings records for those same individuals for validation.
4. There is no evidence, however, that the measurement errors in measures such as earnings are higher in the PSID than in other surveys. Indeed, the PSID's substantial editing and across-wave consistency checking should make measurement errors of this type less problematic than in surveys not following such procedures.
5. This result held in a CPS-Social Security validation study as well (Bound and Krueger, 1989).