6. Nonsampling Errors

This chapter summarizes information about nonsampling errors in the Survey of Income and Program Participation (SIPP) that may affect the results of certain types of analyses. All surveys are subject to various sources of nonsampling errors, and the SIPP is no exception. Nonsampling errors in SIPP include those that are found in most surveys as well as errors that arise because of SIPP's panel (longitudinal) nature. The chapter focuses on the extent of nonsampling errors in SIPP and the impact of those errors on some survey estimates. The following topics are discussed:

- Undercoverage;

- Nonresponse;

- Measurement errors;

   and,

- Effects of nonsampling errors on some survey estimates.

Undercoverage

One source of error in SIPP, as in other household surveys, is differential undercoverage of demographic subgroups. Black males over 15 years of age are most affected by undercoverage. The coverage ratio for this group was on average about 0.82 for the interview months of wave 1 in the 1990 and 1991 SIPP Panels. (Coverage ratio is computed as the survey estimate of the number in the subgroup before post-stratification, divided by a population estimate for the subgroup from population projections based on the most recent census. Such population estimate is generally referred to as a control or a benchmark estimate.) For black males in their mid to late 20s, the coverage ratio was lower, about 0.65 in the same panels (*SIPP Quality Profile*, 3rd Ed. [U.S. Census Bureau, 1998a, Chapter 3]; hereinafter in this chapter, *SIPP Quality Profile*, 3rd Ed). These coverage ratios may understate the magnitude of the coverage problems because census undercounts are not reflected in the coverage ratios before1992. Undercoverage in household surveys is attributed mainly to within-household omissions; the omission of entire households is less frequent. Shapiro et al. (1993) estimated that about 70 percent of the undercoverage for young black males consists of within-household omissions; the corresponding percentage for the white population is about 60 percent.

For the 2004 SIPP Panel (the most recent panel as of March 2008), the coverage ratios for black males over 15 years of age were as a whole about 0.80 and 0.78 for the fourth reference months of wave 1 and wave 5, respectively. Hispanic males over 15 years of age were most affected by undercoverage for the 2004 SIPP Panel. The coverage ratios for this group were as a whole about 0.69 and 0.64, respectively. The coverage ratios for all people were as a whole about 0.89

for the fourth reference months of wave 1 and wave 5 of the 2004 SIPP Panel (i.e., 0.8932 for wave 1 and 0.8897 for wave 5).

To compensate for undercoverage, the Census Bureau uses population controls (independent benchmark estimates) to make post-stratification adjustments on the non-interview adjusted weights to derive the final SIPP weights. Little is known about the effectiveness of the post-stratification adjustments in reducing biases, particularly, when an estimate of interest is not strongly correlated to the characteristics of the population controls. One of the reasons for this is unavailability of appropriate administrative records to be used for assessing bias levels in estimates produced by SIPP for various key demographic and socioeconomic characteristics and events.

Nonresponse and Attrition

Nonresponse and attrition are a major concern in SIPP because of the need to follow the same people over time. In SIPP, nonresponse can occur at several levels: household nonresponse at the first wave (Wave 1) and thereafter (Wave 2+); person nonresponse in interviewed households (Type Z nonresponse); and item nonresponse, including complete nonresponse to topical modules. The interview refusal nonresponse households (Type A nonresponse) at wave 1 were not followed for interview, and thus regarded as a permanent sample loss. An unlocated mover is an original (Wave 1) sample person who moved to an unknown address at Wave 2+, and his/her unlocated household is referred to as a Type D nonresponse household or is simply referred to as Type D nonresponse, at that wave. Prior to the 2001 SIPP Panel (with the exception of Type A nonresponse in Wave 1), sample households with Type A nonresponse for two consecutive waves or Type D nonresponse for three consecutive waves were not followed for interview, and thus regarded as a permanent sample loss. In the 2001 and 2004 SIPP Panels (with the exception of Type A nonresponse household in Wave 1), all Type A and Type D nonresponse households were followed for interview at all waves. Like other longitudinal surveys, one potentially significant issue for SIPP is attrition. Attrition in a longitudinal survey is a phenomenon brought about by some of its eligible sample people ceasing to respond to or participate in the survey at some point in a time period under consideration and never reenter the survey again by the end of the time period for any of the following reasons: Type A, Type D, and Type Z nonresponse; and they are thus referred as the attriters. Correspondingly, the eligible sample people continuing to respond to or participate in the survey for the same time period under consideration are referred to as the continuers. Emphatically, if the characteristics of interest are significantly different between the continuers and attriters then the estimates for these characteristics based on the continuers' reported/imputed data and non-interview and post-stratification adjusted weights may still be significantly biased.

At the household level, the rate of sample loss for the 1991 Panel rose from about 8 percent at Wave 1 to more than 21 percent by Wave 8 (last wave). For the same panel, 23 percent of the original sample persons who participated in Wave 1 missed one or more interviews for which they were eligible in later waves. At the item level, the nonresponse rate is typically around 10 percent or less for items on income amounts for the 1984, 1985, 1986, 1992, and 1993 SIPP Panels. However, the nonresponse rates for items on asset amounts vary from about 13 to 42

percent for the 1984 and 1986 SIPP Panels. These sample loss rates (overall cumulative household level nonresponse rates) are excerpted from Chapter 5 of *SIPP Quality Profile,* 3rd Ed.

Prior to the 2001 SIPP Panel, the rates of household sample loss at Wave 1 varied from about 8 to 9 percent; however, the rates of household sample loss at Wave 1 for the 2001 and 2004 SIPP Panels increased to about 13 and 15 percent, respectively. By Wave 8 (last wave for the 1991 SIPP Panel), the rates of sample loss for the 1992, 1993, 1996, 2001, and 2004 SIPP panels rode about 25, 26, 31, 30, and 34 percent, respectively. By Wave 12 (last wave) of the 1996 and 2004 SIPP panels (the longest panels), the rates of household sample loss rose to about 36 and 37 percent, respectively. These sample loss rates are excerpted from Benton (2008).

For Wave 1 of the 2004 SIPP Panel, the nonresponse rates for items on asset amounts vary from about 13 to 25 percent (versus a variation from about 13 to 42 percent for the 1984 and 1986 SIPP Panels as pointed out earlier). As described by Bruun and Moore (2005), the large improvement for the item nonresponse rates on asset amounts for Wave 1 of the 2004 SIPP Panels over those of the earlier SIPP Panels is attributable to the implementation of new and expanded follow-up questions (in Wave 1 of the 2004 SIPP Panel) for the respondents who initially responded "don't know" or refused to respond to questions concerning the amount of income produced by their assets. These follow-up questions provided the initial respondents with a multiple-choice range of income amounts from which to select.

Nonresponse reduces the effective sample size (and, therefore, increases sampling error) and introduces bias in the survey estimates. The Census Bureau uses a combination of weighting and imputation methods to reduce the biasing effects of nonresponse at all three levels (household, person, and item nonresponse levels) in SIPP. The effectiveness of those procedures remains a matter of ongoing review and research (SIPP Quality Profile, 3rd Ed., Chapters 4, 5, and 8).

Measurement Errors

Measurement errors are associated with the data collection phase of the survey. They may vary across SIPP panels because of changes in data collection procedures over the years. Most core survey items in SIPP are used consistently at every panel, although there have been occasional changes to improve the clarity of some items. The data collection method, which was face-to-face (in-person) interviewing for the early panels, was changed to a maximum use of telephone interviewing in February 1992. Telephone interviewing was used as the primary mode of data collection between February 1992 and January 1996 for all waves except Waves 1, 2, and 6, for which face-to-face interviewing was used. The switch to telephone interviewing has had no known adverse effects on data quality.

Computer-assisted interviewing (CAI) was introduced with the 1996 SIPP Panel. The effects of CAI on survey responses have yet to be determined (*SIPP Quality Profile*, 3rd Ed., Section 11.3). For the 1996 Panel, computer-assisted personal interviewing (CAPI) was used for Waves1 and 2. After Wave 2, the field representatives used the CAI instrument in face-to-face interviews with approximately one-third of the respondents; for the remaining interviews, the

field representatives used the CAI instrument but conducted telephone interviews from their homes.

The combination of face-to-face interviews and telephone interviews used across waves is prespecified and varies for different subgroups of the sample according to the following scheme (Waite, 1996). Sample members are assigned to one of three interviewing mode subgroups. For each subgroup, a pattern of interviewing modes is designated and repeated every three waves. Thus, for Waves 3, 4, and 5, subgroup 1 is assigned the sequence face-to-face, telephone, telephone; subgroup 2, the sequence telephone, face-to-face, telephone; and subgroup 3, the sequence telephone, telephone, face-to-face. Under this scheme, which is applied with each rotation group, one-third of the sample is interviewed in person each wave and each month, and every household is interviewed in person once a year. The same sequence is repeated for Waves 6 and beyond, with a cycle of three waves (*SIPP Quality Profile*, 3rd Ed.).

As mentioned earlier, the switch from in-person to telephone interviewing has had no known adverse effects on data quality. Therefore, for the 2001 and 2004 SIPP Panels, in-person interviewing was generally required for Wave 1, and maximum effort was imposed on telephone interviewing for Wave 2 and beyond in order to reduce cost. As it turned out, about 77 to 80 percent of the household interviews were accomplished by telephone interviewing for Wave 2 and beyond in the 2004 SIPP Panel.

Response errors in SIPP include errors of recall, errors in proxy respondents' reports, and other errors associated with the panel nature of SIPP. The SIPP uses a 4-month recall period to reduce memory error, and respondents are encouraged to use financial records and an event calendar to facilitate recall. Although the level of accuracy for self-response is generally believed to be higher than for proxy response (see Moore, 1988, for a contrary view), achieving a higher proportion of self-response would increase data collection costs and might lead to some increase in person nonresponse rates (*SIPP Quality Profile*, 3rd Ed., Section 4.5.3).

A potential source of response error that arises from the panel nature of SIPP is the time-in-sample effect (or panel conditioning). This effect occurs when the responses given at later waves are affected by the respondents' experiences of being interviewed in previous waves. The extent of this error is difficult to evaluate because it is often confounded with other sources of error, particularly attrition. Thus far, studies have found little evidence of systematic biases resulting from time-in-sample effects (Pennell and Lepkowski, 1992; McCormick et al., 1992).

Measurement errors can also occur when respondents misinterpret questions. For example, when asked about earnings, some respondents may have reported take-home pay instead of gross earnings. There is also some evidence of confusion in regard to welfare programs, such as the old Aid to Families with Dependent Children and general assistance programs.

Another response error identified through the panel nature of SIPP is the seam phenomenon. Research has consistently indicated that respondents tend to report the same status (e.g., employment or program participation) and the same amounts (e.g., Social Security income) for all 4 months within a wave, with most reported changes occurring between the last month of one wave and the first month of the subsequent wave. This phenomenon results in an overstatement of changes at the on-seam months (the boundary between interviews in successive waves of a panel) and an understatement of changes at the off-seam months. The seam phenomenon affects most variables for which monthly data are collected. As a result of the rotation group pattern, the phenomenon has relatively small effects on cross-sectional estimates based on all four rotation groups. That is because there is only one rotation group (or one-fourth of the sample) that is on seam and three rotation groups off seam for any given pair of calendar months. The effects of the seam phenomenon on longitudinal estimates are not well known (*SIPP Quality Profile*, 3rd Ed., Chapter 6).

Effects of Nonsampling Errors on Survey Estimates

A considerable amount of research has been conducted to investigate the various sources of nonsampling error in SIPP. The results of the research are summarized in the *SIPP Quality Profile*, 3rd Ed. The research includes, for example, the SIPP Record Check Studies (Marquis and Moore, 1989a,b, 1990; Marquis et al., 1990) that compared SIPP responses on program participation with administrative records. Despite the volume of this methodological research, it remains difficult to quantify the combined effects of nonsampling errors on SIPP estimates. The problem is made more complex because the effects of nonsampling error of different types on survey estimates vary, depending on the estimate under consideration. There are, however, some findings about nonsampling error that SIPP users should bear in mind when conducting their analyses and examining their results. Those findings include the following:

- Some demographic subgroups are underrepresented in SIPP because of undercoverage and nonresponse. They include young black males, Hispanic males, metropolitan residents, renters, people who changed addresses during a panel (movers), and people who were divorced, separated, or widowed. The Census Bureau uses weighting adjustments and imputation to correct the underrepresentation. Those procedures, however, may not be fully correct for all potential biases (*SIPP Quality Profile*, 3rd Ed., Chapter 8).

- The SIPP estimates of income from Social Security, Railroad Retirement, and Supplemental Security programs represent more than 95 percent of the amounts reported by administrative sources. The SIPP estimates of unemployment income, workers' compensation income, veteran's income, and public assistance income, however, are low relative to the amounts reported by administrative sources (Coder and Scoon-Rogers, 1996).

- Evaluation studies typically find that SIPP estimates (as well as other survey estimates) of property income are generally poor. Among the different types of property income,

reports of interest and dividend income are most prone to error. Respondents are often confused about those two sources of income, and both sources tend to be underreported (Coder and Scoon-Rogers, 1996).

● The SIPP estimates of assets, liabilities, and wealth are low relative to estimates from the Federal Reserve Board (Eargle, 1990).

● For SIPP panels before 1996, the estimates of the percentages of people in poverty were lower than those found in the Current Population Survey (CPS) (Shea, 1995a).

● The SIPP estimates of the working population differ from those produced from CPS. The differences may be explained largely by substantial conceptual and operational differences in the collection of labor force data in the two surveys (*SIPP Quality Profile*, 3rd Ed., Chapter 10).

● The SIPP estimates of people without any health insurance coverage are much lower than the CPS estimates. There are reasons to believe that the SIPP estimates are more accurate (McNeil, 1988).

● The SIPP estimates of the number of births compare favorably with the CPS estimates. Both surveys, however, provide estimates that are low relative to the records from the National Center for Health Statistics (NCHS). The SIPP estimates of the number of marriages are fairly comparable with the NCHS counts, but the SIPP estimates of the number of divorces are consistently lower than the NCHS estimates (*SIPP Quality Profile*, 3rd Ed., Chapter 10).

● In two studies by Vaughn and Scheuren (2002) and Hall and Sae-Ung (2004) on the effects of attrition on the SIPP earnings estimates, they defined an *attriter* as an original sample person who was not interviewed during the final wave of a given panel; and a *continuer*, in contrast, was someone who was interviewed in both the first wave and the final wave of the panel (but may or may have been interviewed during all of the waves in between). Comparing the quartile estimates of the 1996 to 1999 annual earnings of the continuers using the 1996 SIPP Panel data with those using the earnings administrative records from the Social Security Administration (SSA) yielded the following results (Hall and Sae-Ung, 2004): The percent differences for medians were typically within 10% and moderately likely to be statistically significant, so are those of the 75th percentiles; and the percent differences for the 25th percentile were larger (up to 15% or more) and usually statistically significant. In addition, the median annual earnings estimates based on the SSA administrative records for all attriters were 10% to 25% lower than those for all continuers between the years 1992 to 2001 for the 1992, 1993, and 1996 SIPP Panels, and all of the differences were statistically different for all years. This indicates that the difference in earnings between the continuers and attriters may be a significant cause of the bias in the SIPP earnings estimates.

● In spell analyses, Kalton et al. (1992) found that spell durations of multiples of 4 months (e.g., 4 months, 8 months, 12 months) were particularly common, a feature that can be

explained by the seam phenomenon. For the 2004 SIPP Panel, the U.S. Census Bureau (Moore, 2007) used new dependent interviewing (DI) procedures in the 2004 SIPP Panel questionnaire designed to reduce seam bias of a number of characteristics (e.g., government transfer program participation, school enrollment, employment, health insurance coverage, etc.). Analyses showed that the new DI procedures are capable of substantially lowering the seam biases in the 2004 SIPP Panel when compared with those in the 2001 SIPP panel; however, even with the clear improvements, seam bias still afflicts data collected for the 2004 SIPP Panel. Further fine-tuning of the current DI procedures is unlikely to yield substantial additional improvement in seam bias. Therefore, for future questionnaire redesign for the SIPP, new approaches such as event calendar history methods will be considered.

● The latest evaluation of sample loss in the SIPP and CPS Annual Social and Economic Supplement (CPS-ASEC) done by Czajka, Mabli, and Cody (2008) yields the following conclusion. They recommendation to prospective users of SIPP data at the Social Security Administration (SSA) is that they should not hesitate to use the 2001 SIPP Panel any more than they would hesitate to use the 1996 SIPP Panel as a source of information (data) on current and potential beneficiaries served by programs that the SSA administers. Neither attrition bias nor match bias (in linking of SSA administrative records to the survey data) provides any more reason to avoid the 2001 panel than the earlier panels. However, there are two areas of concern stand out. The first is a wave 1 effect that elevates poverty rates during the first wave of each new panel. The second stems from the divergent (inconsistent) trends between the SIPP and CPS-ASEC estimates for the material well-being of the elderly either cross-sectionally or over time (longitudinally). They recommend that the Office of Research, Evaluation, and Statistics (ORES) of the SSA encourages the Census Bureau to undertake an assessment of how these two surveys can present such inconsistent estimates.

● Similar to the finding by Czajka, Mabli, and Cody (2008), the study by Sae-Ung, Sissel, and Mattingly (2007) also found the inconsistency between the SIPP and CPS-ASEC annual estimates of the 2001 health insurance coverage rates and annual low-income rates below 150% and 200% poverty thresholds, even with the following longitudinal enhancements added to the degree of systematic similarity between the 2001 SIPP Panel and the 2002 and 2003 CPS_ASEC Supplement: They created a 2002 and 2003 CPS-ASEC quasi longitudinal file by simulating the movers and the deceased, barracked, expatriated, and institutionalized survey universe leavers between March 2002 and March 2003 among the CPS-ASEC 2002 sample people who no longer belonged their households that remained in sample in the 2003 CPS-ASEC. They then longitudinally weighted the 2002 and 2003 CPS-ASEC quasi longitudinal file using the 2001 SIPP Panel longitudinal weighting procedure, the same March 2002 controls (benchmark population estimates) as those of SIPP for the post stratification weight adjustment, and the same longitudinal interview definition as that of SIPP. In the next phase of their study, they will attempt to determine what causes the inconsistency using modeling approaches.