

Appendix D. Data Quality

Two major determinants of the quality of data collected in household surveys are the magnitude of missing responses and the accuracy of the responses that are provided. This appendix provides information about the imputation process for nonresponse cases and provides a comparison of selected items with independent estimates.

Nonresponse in this discussion refers to missing responses to specific questions or "items" on the questionnaire. Noninterviews or complete failure to obtain cooperation from any household member have not been considered in this discussion of nonresponses. Adjustments to account for noninterview are made by proportionally increasing the survey weights of interviewed households. Missing responses to specific questions are assigned a value in the imputation phase of the data processing operation.

Nonresponse is an important factor in assessing the quality of survey data. Nonresponse occurs when respondents do not answer questions because of a lack of knowledge or a refusal to answer. Nonresponses are assigned values prior to producing estimates from the survey data. The procedure used to assign or impute most responses for missing data for SIPP are of a type commonly referred to as a "hot deck" imputation method. This process assigns values reported in the survey by respondents to nonrespondents. The respondent for whom the value is taken is termed the "donor". Values from donors are assigned by controlling for demographic and economic data available for both donors and nonrespondents. For example, for every asset and liability item, there was a common set of characteristics used in the imputation process. These included the age, race, sex, and years of schooling of the person, and total household income during the four month reference period. For other items, additional characteristics were used in the imputation process. For example, the imputation of the current market value of own home used the common set of characteristics listed above plus the original purchase price of the home.

A second important determinant of data quality is the accuracy of reported and imputed amounts. Response errors are the result of a variety of factors including random response error, misreporting or failure to report asset ownership, misreporting of asset and liability values, and misreporting of the sources of income and the full amount received. In general, household surveys have a tendency to underestimate the number of persons receiving income and the average amount received.

The extent of response error is measured by comparing survey estimates with independently derived estimates. A comparison of SIPP aggregate asset amounts in 1988 with estimates derived from the Flow of Funds data of the Federal Reserve Board (FRB) is shown in Table D-1. The Flow of Funds Balance Sheet data provides estimates as of the end of the year.

[Table D-1](#) [TXT - 2.3K]

There are several conceptual and coverage differences between SIPP and FRB balance sheet data. First, the household sector in the FRB Balance Sheet includes nonprofit institutions and private trusts not covered in SIPP. In order to make the source more comparable, a rough estimate of the financial assets held by the nonprofit sector and personal trusts was obtained from

the Federal Reserve Board. The second difference is that the SIPP universe consists of the noninstitutional resident population. Excluded are Armed Forces personnel living in military barracks, citizens residing abroad, and institutional persons such as correctional facility inmates and nursing home residents. The asset holdings of these groups is included in the FRB Balance Sheet estimates. A third limitation is that the household sector of the FRB Balance Sheet is estimated as a residual after allocations are made to all other sectors. As a result, allocation errors can lead to inaccuracies in the household sector estimates, especially in assets where the amount held by households comprise a small percentage of the total.

Finally Table D-2 shows a comparison of SIPP estimates of median value of own home with similar data from the 1990 Decennial Census. Interviewing for the census took place on April 1990.

[Table D-2](#) [TXT - 1.1K]