

Appendix D. Description of the SIPP 1987 Panel File and Data Quality

DESCRIPTION OF SIPP 1987 PANEL FILE

The estimates presented in this report are based on the fourth SIPP panel file. This file contains monthly data for persons over a 28-month period. The staggered SIPP design (described in appendix A) means that the actual reference periods are January 1987 to April 1989, October 1986 to January 1989, November 1986 to February 1989, and December 1986 to March 1989. The period covered by the 1987 longitudinal panel file consists of 28 interview months (seven interviews) for rotations 1, 2, 3, and 4. Data from all four rotation groups are available only for the reference period January 1987 through January 1989.

Each person in the panel file has been assigned three weights: a weight for calendar year 1987, a weight for calendar year 1988, and a weight for the 28-month reference period. In order to receive a non-zero weight, a person must have an observation for each month of the relevant reference period (in this report, 1987 and 1988) or have a complete set of observations up until the time he or she died or became institutionalized. The data shown in this report are affected if characteristics of persons with an incomplete set of observations differed from those with a complete set.

Table D-1 shows three categories of sample persons by sex, age, and program participation status. The numbers in the table are unit counts; they are not weighted. The category "Complete set of interviews obtained" includes 24,429 persons. The next category, "Interviewed in first wave, left sample for reasons other than death or institutionalization" includes 6,403 persons. The final category includes 4,896 persons who were not a member of a SIPP household during the first wave of interviews, but who subsequently became a member of a sample household.

A comparison of the first two columns shows the characteristics of those who completed the full set of interviews are reasonably close to the characteristics of those who dropped out of the sample. The major differences in the age distribution are for young adults and for the elderly. Young adults are underrepresented and the elderly are overrepresented in the group of persons who completed the full set of interviews. The data in table D-1 are, as noted, unweighted, and any potential problem caused by unrepresentative age distributions are minimized when the file is weighted to independent controls.

TIME-IN-SAMPLE BIAS

The use of the panel file to obtain estimates for 1987 and 1988 raises the issue of time-in-sample bias. There is ample evidence that certain measures vary according to

Table D-1. Percent Distribution: Three Categories of Sample Persons

Characteristic	Complete set of interviews obtained ¹	Interviewed in first wave, left sample for reasons other than death or institutionalization	Not a member of sample household during first wave, interview obtained in second or later waves
Total	24,429 (100.0)	6,403 (100.0)	4,896 (100.0)
SEX			
Male	47.4	49.8	51.0
Female	52.6	50.2	49.0
AGE AT FIRST INTERVIEW			
Under 18 years	28.2	26.3	36.6
Under 6 years	10.4	8.8	24.0
18 to 24 years	9.1	16.7	21.4
25 to 44 years	30.7	31.9	33.8
45 to 64 years	19.3	16.5	10.7
65 years and over	12.7	8.6	3.6
75 years and over	5.0	3.2	1.6
PROGRAM PARTICIPATION, FIRST MONTH IN SAMPLE:			
Persons 18 years and over	17,537 (100.0)	4,717 (100.0)	3,406 (100.0)
Participated in major assistance program	8.6	9.7	9.7
AFDC or general assistance ..	2.1	2.8	2.3
Food stamps	4.7	4.9	4.2
Medicaid	5.0	5.5	5.8
Public/subsidized housing	3.0	3.5	2.5
SSI	2.1	1.5	2.4
Did not participate	91.3	90.2	81.5

¹Includes 614 persons who died or were institutionalized during the 28-month period.

the number of times the respondent has been visited. In the CPS, for example, the measured unemployment rate is always higher for the group of households being interviewed for the first time than for the groups being interviewed for the second or later times.

Time-in-sample bias arises when a person's response to a survey question (or the interviewer's method of asking a question) is influenced by what occurred in a previous visit. The overlapping SIPP sample design provides the data that allows for an examination of the presence of time-in-sample bias in SIPP estimates. That is, it is possible in SIPP to obtain estimates for a given time period from two or more separate panels and the amount of time respondents will have spent in the SIPP panel will differ for each

of the panels. For example, estimates for each of the four quarters of 1987 can be obtained from both the 1986 and 1987 panels (respondents in the 1986 will have had more visits).

The quarterly estimates in table D-2 are shown for the four quarters of 1987. Estimates from each panel file are shown separately for comparison. The estimates shown are of median income of nonfarm households, number of households receiving Social Security or Railroad Retirement, number of households receiving food stamps, and number of households with low monthly income.

The figures in table D-2 provide very little evidence regarding the existence of time-in-sample bias for several reasons. First, most of the observed differences are smaller than the differences that could be explained by sampling error. Second, a single observation is not sufficient to identify a pattern of bias. Third, differences may be attributable to attrition bias rather than time-in-sample bias. In spite of these qualifications, however, the observed relationships offer some reason to be cautious in interpreting the differences that have been presented earlier in this report—both the differences between CPS and SIPP estimates and the differences between the 1986 and 1987 estimates that were obtained from the SIPP.

Table D-2. Selected Monthly Averages by Quarter From 1986 and 1987 SIPP Cross-Sectional Files

(Numbers in thousands)

Selected characteristics	Source of estimate				1986 panel to 1987 panel
	1986 panel	Standard error	1987 panel	Standard error	
Median Income of Nonfarm Households					
1987, Quarter 1	\$1,958	\$28	\$2,004	\$23	0.98
1987, Quarter 2	2,034	19	2,051	19	0.99
1987, Quarter 3	2,054	20	2,070	20	0.99
1987, Quarter 4	2,079	21	2,094	22	0.99
Number of Nonfarm Households Receiving Social Security or Railroad Retirement					
1987, Quarter 1	24,805	457	24,837	438	1.00
1987, Quarter 2	24,839	457	24,996	439	0.99
1987, Quarter 3	24,773	457	25,200	440	0.98
1987, Quarter 4	24,994	458	25,423	441	0.98
Number of Nonfarm Households Receiving Food Stamps					
1987, Quarter 1	6,340	261	6,230	248	1.02
1987, Quarter 2	6,218	259	6,142	247	1.01
1987, Quarter 3	5,900	253	5,881	242	1.00
1987, Quarter 4	5,789	251	5,833	241	0.99
Number of Nonfarm Households with Low Monthly Income					
1987, Quarter 1	11,471	341	11,291	324	1.02
1987, Quarter 2	11,017	335	10,691	317	1.03
1987, Quarter 3	10,840	333	10,648	316	1.02
1987, Quarter 4	10,644	330	10,558	315	1.01

OTHER ISSUES OF DATA QUALITY

Two major determinants of the quality of income data collected in household surveys are the magnitude of missing responses and the accuracy of the responses that are provided. This appendix has been included to supply information concerning nonresponse rates for selected income questions, the average amounts of income reported in the survey or assigned in the imputation of missing responses, and the extent to which the survey figures underestimate numbers of income recipients and amounts of income received.

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 examination of nonresponse rates. Adjustments to account for noninterviews 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 a very important factor in assessing the quality of survey data. Nonresponses to income questions cannot be considered random since experience has shown that persons with the highest nonresponse rates have reported characteristics such as education levels and occupations that, in general, differ from population averages. The most frequent causes of nonresponse are the inability of the respondent to answer the question because of either a 1) lack of knowledge or 2) refusal to answer. The first reason is especially important in situations of proxy response when one household member answers questions for another household member not present at the time of the interview. The practice of accepting proxy interviews from household members deemed "qualified" to answer is a standard procedure in the CPS and most other surveys conducted by the Bureau. During the seven interviews of the SIPP 1987 panel, an average of 36 percent of the interviews were taken from proxy respondents.

Nonresponses are assigned values prior to producing estimates from the survey data. The procedure used to assign or impute 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 from whom the value is taken is termed the "donor." Values from donors are stored in a matrix defined by demographic and economic data available for both donors and nonrespondents. Each cell of the matrix defines a unique combination of demographic and economic characteristics. For example, the imputation of an amount for monthly wage and salary income is based on eight different variables. These were 1) occupation, 2) sex, 3) age, 4) race, 5) educational attainment, 6) weeks worked, 7) usual hours worked per week, and 8) place of residence.

The second important determinant of data quality and probably the one examined most closely by users of the income data collected in household surveys is the accuracy of reported (and imputed) amounts. In general, household surveys have a tendency to underestimate the number of persons receiving income and the average amount received. These problems result for a variety of reasons including random response error, misreporting of sources of income, failure to report the receipt of income from a specified source, and failure to report the full amount received. The net effect of these kinds of problems is, for most income types, underestimation or underreporting of income amounts. The extent of underreporting is measured by comparing survey estimates with independently derived estimates, usually based on administrative data that are, generally, more reliable than the estimates derived from the survey. It should be noted that the independent estimates are subject to errors themselves. In addition, independent estimates do not reflect income attributable to the "underground" economy, some of which may be reported in the survey.

COMPUTATION OF POVERTY STATUS IN SIPP AND COMPARISON WITH CPS ESTIMATES

Official poverty data in the CPS are based on questions on income received in the preceding calendar year which are asked in the March supplement. Family composition is fixed as of the survey date and assumed to have been constant over the previous year (in the case of 1987 poverty status, the data were collected in March 1988). In this report using the 1987 SIPP panel, income information was collected for each month. Family composition data was updated on a monthly basis also. A person's annual poverty status was determined by comparing the sum of the person's monthly income (family income or unrelated individual income as appropriate) against the sum of the appropriate monthly poverty thresholds. If the sum of the monthly incomes was below the sum of the monthly poverty thresholds, the person was classified as below the poverty level for the year.

Poverty estimates vary considerably between the CPS and SIPP. An earlier study showed that an approach that adjusts poverty for changes in household composition results in a poverty estimate that is about 5 percent lower than an estimate based on an approach that does not adjust for changes in household composition.¹

Other than treatment of changes in household composition, there are several other differences between CPS and SIPP that should be noted in comparing results from the two surveys. First, the shorter recall period in SIPP

results in more accurate data on the receipt of transfer income. This difference would tend to result in SIPP poverty estimates that are lower than CPS poverty estimates. A second difference concerns the way in which self-employment income is recorded. It is possible to record negative amounts in CPS, but not in SIPP. This difference would also tend to result in SIPP estimates of poverty that are lower than CPS estimates. It is also possible because of its more frequent interviews that SIPP has better reporting of intermittent income than does the CPS. If such income tends to cluster at the lower end of the earnings distribution, this would tend to lower the number of poor.

Paradoxically, wage and salary income estimates tend to be lower in SIPP than in the CPS. It is possible that persons tend to report net rather than gross wage and salary income in the SIPP. This would tend to result in SIPP estimates of poverty that are higher than the CPS estimates for those persons/families with wage and salary income.

The data file used in this report is based on persons for whom a complete set of observations was obtained over a 28-month period. These persons were weighted to reflect the total population as of March 1, 1987. The total weighted number of persons in 1987 will fall short of the independent estimates of the total population because some persons with positive weights are excluded from the analysis, namely, those who died or were institutionalized. The total estimate for 1988 will fall short of independent estimates for the same reasons and because of natural increase and net migration between 1987 and 1988.

Tables D-3 and D-4 compare selected poverty rates and year-to-year changes in these rates from the March 1988 and 1989 CPS (in which 1987 and 1988 poverty statistics were collected) with the 1987 SIPP panel file figures. In general, SIPP estimates are considerably lower, with the CPS poverty rate for 1988 being 3.0 (± 0.07) percentage points greater than the comparable SIPP estimate.

COMPARISON OF SIPP AND CPS INCOME ESTIMATES

Table 1 of this report examined the distribution of family (or individual) income in 1987 and 1988 for each fully-interviewed SIPP respondent. Table D-5 compares some of these estimates with those derived from the March 1989 CPS. The reference period for both sets of estimates is calendar year 1988.

The SIPP median family (or individual) income for all persons in 1988 was \$30,689, not significantly different from the comparable CPS figure (\$30,287). There were no statistically significant differences between SIPP and CPS estimates by race.

SIPP estimates of income for those between the ages of 18 and 24 (\$30,523) were higher than the comparable CPS estimates.

¹See John F. Coder, et. al., Preliminary Data from the SIPP 1983-84 Longitudinal Research File. SIPP Working Paper No. 8702.

Table D-3. Comparison of CPS and SIPP Poverty Rates, by Selected Characteristics: 1987 and 1988

(Estimates from SIPP based on 1987 panel file)

Characteristics	Percent below the poverty level		Percentage point difference CPS-SIPP	CPS/ SIPP
	CPS ¹	SIPP		
1987				
AGE				
Total	13.4	10.8	2.6	1.24
Under 18 years	20.3	17.7	2.6	1.15
18 to 64 years	10.6	8.3	2.3	1.28
65 years and over	12.5	8.9	3.6	1.40
SEX				
Male	11.8	9.5	2.3	1.24
Female	14.9	11.9	3.0	1.25
RACE AND HISPANIC ORIGIN				
White	10.4	7.9	2.5	1.32
Black	32.4	30.2	2.2	1.07
Hispanic origin ²	28.0	22.8	5.2	1.23
1988				
AGE				
Total	13.0	10.0	3.0	1.30
Under 18 years	19.5	16.4	3.1	1.19
18 to 64 years	10.5	7.6	2.9	1.38
65 years and over	12.0	8.7	3.3	1.38
SEX				
Male	11.5	8.8	2.7	1.31
Female	14.5	11.2	3.3	1.29
RACE AND HISPANIC ORIGIN				
White	10.1	7.3	2.8	1.37
Black	31.3	28.8	2.5	1.09
Hispanic origin ²	26.7	21.9	4.8	1.22

¹Standard errors for the 1987 and 1988 Current Population Survey poverty rates can be derived from Appendix B of Current Population Reports, Series P-60, No. 171, "Poverty in the United States: 1988-89".

²Persons of Hispanic origin may be of any race.

Table D-4. Comparison of 1987-88 Year-to-Year Percentage Point Change in Poverty Rate Between CPS and SIPP

Characteristic	CPS/1	SIPP
AGE		
Total	-0.4	-0.8
Under 18 years	-0.8	-1.3
18 to 64 years	-0.1	-0.7
65 years and over	-0.5	-0.2
SEX		
Male	-0.3	-0.7
Female	-0.4	-0.7
RACE AND HISPANIC ORIGIN		
White	-0.3	-0.6
Black	-1.1	-1.4
Hispanic origin ²	-1.3	-0.9

¹Standard errors for the Current Population Survey figures can be obtained from Appendix B of the Current Population Reports, Series P-60, No. 171, "Poverty in the United States: 1988-89."

²Persons of Hispanic origin may be of any race.

Table D-5. Comparison of CPS and SIPP Estimates of Median Family or Individual Income, All Persons: 1988

Selected characteristics	SIPP		CPS	
	Median income	Standard error	Median income	Standard error
All persons	\$30,689	\$290	\$30,287	\$80
RACE AND HISPANIC ORIGIN				
White	32,037	340	31,723	92
Black	18,359	636	17,896	310
Hispanic origin ¹	22,203	966	20,839	244
AGE				
Under 18 years	31,119	540	30,403	145
18 to 24 years	30,523	1,059	28,205	410
25 to 44 years	33,269	538	32,964	147
45 to 64 years	33,616	846	34,556	250
65 years and over	16,947	815	16,213	193

¹Persons of Hispanic origin may be of any race.