UNITED STATES DEPARTMENT OF COM
Economics and Statistics Administration
U.S. Census Bureau
Washington, DC 20233-0001

Final
DSMD
Signed

April 1, 2013

MEMORANDUM FOR Francis Grailand Hall
Chief, Administrative and Customer Services Division

From: Ruth Ann Killion*(Signed)*
Chief, Demographic Statistical Methods Division

Prepared by: Evan Wong and Stephen Mack
Survey of Income and Program Participation,
Demographic Statistical Methods Division

Subject: Survey of Income and Program Participation (SIPP) 2008 Panel: Source and Accuracy Statement for Wave 1 to Wave 11 (core) Public Use Files (S&A-16)[1]

Attached is the Source and Accuracy Statement for the 2008 Panel Wave 1 to Wave 11 (core) Public Use Files of the Survey of Income and Program Participation (SIPP).

Attachment

cc:

J. Fields	(ADDP)	A. Gottschalck	
D. Doyle		K. Bauman	
M. Marlay		T. Renwick	
C. Bennett	(ACSD)	B. O'Hara	
D. Smallwood		R. Kreider	
N. McKee	(DSD)	J. Hisnanick	
Z. McBride		T. Mattingly	(DSMD)
L. Liebrecht			
D. Johnson	(SEHSD)		
J. Day			
J. Eargle			
R. Kominski			

1 This source and accuracy statement can also be accessed through the U.S. Census Bureau website at http://www.census.gov/sipp/source.html .

SOURCE AND ACCURACY STATEMENT FOR THE SURVEY OF INCOME AND PROGRAM PARTICIPATION 2008 WAVE 1 TO WAVE 11 PUBLIC USE FILES

TABLE OF CONTENTS

SOURCE OF DATA 4
- Source of Data 4
- Table A. Sample Loss and Response Rates for SIPP 2008 6
- Table B. Percent of Type as by Nonresponse Status for SIPP 2008 6
- Estimation. 7
- Population Controls. 7
- Use of Weights. 8
- Adjusting Estimates Which Use Less than the Full Sample. 9

ACCURACY OF ESTIMATES 9
- Nonsampling Error 9
- Table C. SIPP Average Coverage Ratios for April 2011 for Age by Race and Sex 10
- Comparability with Other Estimates. 10
- Sampling Variability. 10

USES AND COMPUTATION OF STANDARD ERRORS 10
- Confidence Intervals. 10
- Hypothesis Testing. 11
- Note Concerning Small Estimates and Small Differences. 11
- Calculating Standard Errors for SIPP Estimates. 12
- Standard Error Parameters and Tables and Their Use 12
- Adjusting Standard Error Parameters for Estimates with Less Than Full Sample 12
- Standard Errors of Estimated Numbers. 13
- Standard Error of a Mean. 14
- Standard Error of an Aggregate. 16
- Standard Errors of Estimated Percentages. 16
- Standard Error of a Difference. 18
- Standard Error of a Median. 19
- Standard Errors of Ratios of Means and Medians. 21
- Standard Errors Using SAS or SPSS. 21

REFERENCES 21

TABLES 22
- Table 1. SIPP Panel 2008 Topical Modules 22
- Table 2. SIPP Panel 2008 Reference Months for Each Interview Month 23
- Table 3. Factors to be Used When Using Less Than Full Sample 24
- Table 4. SIPP Generalized Variance Parameters for the 2008 Panel, Wave 1 File 25
  - Wave 2 to Wave 3 File 26
  - Wave 4 to Wave 6 File 27
  - Wave 7 to Wave 9 File 28
  - Wave 10 to Wave 11 File 29
- Table 5. SIPP Topical Module Generalized Variance Parameters for the 2008 Panel 30
- Table 6. Base Standard Errors of Estimated Numbers of Household or Families 31
- Table 7. Base Standard Errors of Estimated Numbers of Persons 32
- Table 8. Base Standard Errors for Percentages of Households or Families 33
- Table 9. Base Standard Errors for Percentages of Persons 34
- Table 10. Distribution of Monthly Cash Income 35

SOURCE AND ACCURACY STATEMENT FOR THE SURVEY OF INCOME AND PROGRAM PARTICIPATION 2008 WAVE 1 TO WAVE 11 PUBLIC USE FILES[2]

SOURCE OF DATA

Source of Data. The data were collected in the 2008 Panel of the Survey of Income and Program Participation (SIPP). The population represented in the 2008 SIPP (the population universe) is the civilian noninstitutionalized population living in the United States. The institutionalized population, which is excluded from the population universe, is composed primarily of the population in correctional institutions and nursing homes (91 percent of the 4.1 million institutionalized people in Census 2000).

The 2008 Panel of the SIPP sample is located in 351 Primary Sampling Units (PSUs), each consisting of a county or a group of contiguous counties. Of these 351 PSUs, 123 are self-representing (SR) and 228 are non-self-representing (NSR). SR PSUs have a probability of selection of one. NSR PSUs have a probability of selection of less than one. Within PSUs, housing units (HUs) were systematically selected from the master address file used for the 2000 decennial census. To account for HUs built within each of the sample areas after the 2000 census, a sample containing clusters of four HUs was drawn from permits issued for construction of residential HUs up until shortly before the beginning of the panel. In jurisdictions that don't issue building permits or have incomplete addresses, we systematically sampled expected clusters of four HUs which were then listed by field personnel.

Households were classified into two strata, such that one strata had a higher concentration of low income households than the other. We oversampled the low income stratum by 44 percent to increase the accuracy of estimates for statistics of low income households and program participation. Analysts are strongly encouraged to use the SIPP weights when creating estimates since households are not selected with equal probability.

Sample households within a given panel are divided into four random subsamples of nearly equal size. These subsamples are called rotation groups and one rotation group is interviewed each month. Each household in the sample was scheduled to be interviewed at four-month intervals over a period of roughly five years beginning in September 2008. The reference period for the questions is the four-month period preceding the interview month. The most recent month is designated reference month 4, the earliest month is reference month 1. In general, one cycle of four interview months covering the entire sample, using the same questionnaire, is called a wave. For example, Wave 1 rotation group 1 of the 2008 Panel was interviewed in September 2008 and data for the reference months May 2008 through August 2008 were collected.

---

2 For questions or further assistance with the information provided in this document contact: Tracy Mattingly of the Demographic Statistical Methods Division at (301) 763-6445 or via the e-mail at Tracy.L.Mattingly@census.gov.

In Wave 1, the 2008 SIPP began with a sample of about 65,500 HUs. About 13,500 of these HUs were found to be vacant, demolished, converted to nonresidential use, or otherwise ineligible for the survey. Field Representatives (FRs) were able to obtain interviews for about 42,000 of the eligible HUs. FRs were unable to interview approximately 10,000 eligible HUs in the panel because the occupants: (1) refused to be interviewed; (2) could not be found at home; (3) were temporarily absent; or (4) were otherwise unavailable. Thus, occupants of about 81 percent of all eligible HUs participated in the first interview of the panel.

For subsequent interviews, only original sample people (those in Wave 1 sample households and interviewed in Wave 1) and people living with them are eligible to be interviewed. The SIPP sample includes original sample people if they move to a new address, unless the new address was more than 100 miles from a SIPP sample area. In this case, FRs attempt telephone interviews.

Since SIPP follows all original sample members, those members that form new households are also included in the SIPP sample. This expansion of original households can be estimated within the interviewed sample, but is impossible to determine within the non-interviewed sample. Therefore, a growth factor based on the growth in the known sample is used to estimate the unknown expansion of the non-interviewed households.

Growth factors account for the additional nonresponse stemming from the expansion of non-interviewed households. They are used to get a more accurate estimate of the weighted number of non-interviewed HUs at each wave, called sample loss. To calculate sample loss we use Formula (1):

$$Sample\ Loss = \frac{(A_1 \times GF) + A_C + D_C}{I_C + (A_1 \times GF) + A_C + D_C} \tag{1}$$

where A_1 is the weighted number of Type A non-interviewed households in Wave 1, A_C is the weighted number of Type A non-interviewed households in the Current Wave, D_C is the weighted number of Type D non-interviewed households in the current wave, I_C is the weighted number of interviewed households in the current wave, and GF is the growth factor associated with the current wave.

Table A. Sample Loss and Response Rate for SIPP 2008

Wave	Eligible HUs	Interviewed HUs	Type As		Type Ds		Growth Factor	Weighted Sample Loss
			Total	Weighted Rate	Total	Weighted Rate		
1	52,031	42,032	9,999	19.2%				19.2%
2	42,481	39,000	2,921	6.9%	560	1.3%	1.01	26.1%
3	42,779	37,651	4,159	9.7%	969	2.3%	1.02	28.9%
4	43,176	36,195	5,693	13.2%	1,288	2.9%	1.03	32.4%
5	43,422	35,873	6,060	14.0%	1,489	3.3%	1.04	33.2%
6	43,544	34,891	6,894	15.9%	1,759	4.0%	1.04	35.2%
7	43,619	33,827	7,901	18.2%	1,891	4.2%	1.05	37.5%
8	43,609	33,417	8,231	19.0%	1,961	4.3%	1.05	38.2%
9	43,621	32,567	8,880	20.4%	2,174	4.7%	1.04	39.6%
10	43,690	31,445	9,877	22.7%	2,368	5.1%	1.05	41.9%
11	43,720	31,007	10,25	23.5%	2,457	5.3%	1.05	42.7%

Table B. Percent of Type As by Nonresponse Status for SIPP 2008

Wave	Language Problem	Unable to Locate	No One Home	Temporarily Absent	Household Refused	Other
1	1.2%	0.8%	16.6%	3.4%	67.2%	10.9%
2	0.8%		19.2%	5.2%	61.3%	13.4%
3	0.5%		18.6%	5.7%	60.7%	14.5%
4	0.4%		18.4%	3.9%	62.5%	14.7%
5	0.3%		16.6%	3.4%	64.7%	15.1%
6	0.4%		14.8%	3.7%	67.8%	13.3%
7	0.4%		15.3%	2.9%	62.8%	18.7%
8	0.2%		13.7%	2.4%	62.7%	20.9%
9	0.3%		13.8%	2.7%	62.7%	20.5%
10	0.3%		12.0%	2.2%	65.7%	19.9%
11	0.3%		10.8%	1.8%	71.4%	15.8%

Note that in Table A the Wave 1 weighted sample loss rate is the same as the weighted Type A rate since growth factors and Type D (movers) are not applicable until Wave 2.

The public use files include core and supplemental (topical module) data. Core questions are repeated at each interview over the life of the panel. Topical modules include questions which are asked only in certain waves. The 2008 panel topical modules are given in Table 1.

Table 2 indicates the reference months and interview months for the collection of data from each rotation group for the 2008 panel. For example, Wave 1 rotation group 1 of the 2008 panel was interviewed in September 2008 and data for the reference months May 2008 through August 2008 were collected.

Estimation. The SIPP estimation procedure involves several stages of weight adjustments to derive the cross-sectional person level weights. First, each person is given a base weight (BW) equal to the inverse of the probability of selection of a person's household. Next, a Duplication Control Factor (DCF) is used to adjust for subsampling done in the field when the number of sample units is much larger than expected. Then a noninterview adjustment factor is applied to account for households which were eligible for the sample but which FRs could not interview in Wave 1(F_{N1}). Similarly for subsequent waves *i*, the noninterview adjustment factor is (F_{Ni}). A Mover's Weight (MW) is applied in Waves 2+ to adjust for persons in the SIPP universe who move into sample households after Wave 1. The last adjustment is the Second Stage Adjustment Factor (F_{2S}). This adjusts estimates to population controls and equalizes husbands' and wives' weights. The 2008 Panel adjusts weights to both national and state level controls.

The final cross-sectional weight is $FW_c = BW * DCF * FN_1 * F_{2S}$ for Wave 1 and is $FW_c = IW * FN_2 * F_{2S}$ for Waves 2+, where IW is either $BW * DCF * F_{N1}$ or MW. Additional details of the weighting process are in *SIPP 2008: Cross-Sectional Weighting Specifications for Wave 1 and Wave 2+.*

Population Controls. The 2008 SIPP estimation procedure adjusts weighted sample results to agree with independently derived population estimates of the civilian noninstitutional population. National family type controls are obtained by taking the Current Population Survey (CPS) weights and doing a "March type" family equalization. That is, wives' weights are assigned to husbands and then proportionally adjusted to the weights of persons by month, rotation group, race, sex, age, and by the marital and family status of householders. This attempts to correct for undercoverage and thereby reduces the mean square error of the estimates. The national and state level population controls are obtained directly from the Population Division and are prepared each month to agree with the most current set of population estimates released by the U.S. Census Bureau's population estimates and projections program.

The national level controls are distributed by demographic characteristics as follows:

- Age, Sex, and Race (White Alone, Black Alone, and all other groups combined)
- Age, Sex, and Hispanic Origin

The state level controls are distributed by demographic characteristics as follows:

- State by Age and Sex
- State by Hispanic origin
- State by Race (Black Alone, all other groups combined)

The estimates begin with the latest decennial census as the base and incorporate the latest available information on births and deaths along with the latest estimates of net international migration.

The net international migration component in the population estimates includes a combination of:

- Legal migration to the U.S.,
- Emigration of foreign born and native people from the U.S.,
- Net movement between the U.S. and Puerto Rico,
- Estimates of temporary migration, and
- Estimates of net residual foreign-born population, which include unauthorized migration.

Because the latest available information on these components lags the survey date, to develop the estimate for the survey date, it is necessary to make short-term projections of these components.

Use of Weights. There are three primary weights for the analysis of SIPP data. The person month weight (one for each reference month) is for analyzing data at the person level. Everyone in the sample in a given reference month has a person month weight. The person month weight of the household reference person is used to analyze data at the household level (a household may consist of related and unrelated persons). The person month weight of the family reference person is the family weight. Use this weight to analyze family level questions. Weights are also available in the public use files for related subfamilies. Chapter 8 of the *SIPP Users' Guide* provides additional information on how to use these weights.

By selecting the appropriate reference month weight an analyst can obtain the average of an item such as income across several calendar months.

> **Example.** Using the proper weights, one can estimate the monthly average number of households in a specified income range over August 2008 to September 2008. To estimate monthly averages of a given measure, e.g., total, mean, over a number of consecutive months, sum the monthly estimates and divide by the number of months. To form an estimate for a particular month, use the <u>reference month</u> weight for the month of interest, summing over all persons or households with the characteristic of interest whose reference period includes the month of interest.

The core wave file does not contain weights for characteristics that involve a person's or household's status over two or more months (such as, number of households with a 50 percent increase in income between December 2008 and January 2009).

Adjusting Estimates Which Use Less than the Full Sample. When estimates for months with less than four rotations worth of data are constructed from a wave file, factors greater than 1 must be applied. Multiply the sum by a factor to account for the number of rotations contributing data for the month. This factor equals 4 divided by the number of rotations contributing data for the month. For example, July 2008 data are only available from rotations 1-3 for Wave 1 of the 2008 Panel, so a factor of 4/3.1.3333 must be applied. A list of appropriate factors is in Table 3.

ACCURACY OF ESTIMATES

SIPP estimates are based on a sample; they may differ somewhat from the figures that would have been obtained if a complete census had been taken using the same questionnaire, instructions, and enumerators. There are two types of errors possible in an estimate based on a sample survey: sampling and nonsampling. For a given estimator, the difference between an estimate based on a sample and the estimate that would result if the sample were to include the entire population is known as sampling error. For a given estimator, the difference between the estimate that would result if the sample were to include the entire population and the true population value being estimated is known as nonsampling error. We are able to provide estimates of the magnitude of SIPP sampling error, but this is not true of nonsampling error.

Nonsampling Error. Nonsampling errors can be attributed to many sources:

- inability to obtain information about all cases in the sample
- definitional difficulties
- differences in the interpretation of questions
- inability or unwillingness on the part of the respondents to provide correct information
- errors made in the following: collection such as in recording or coding the data, processing the data, estimating values for missing data
- biases resulting from the differing recall periods caused by the interviewing pattern used and undercoverage.

Quality control and edit procedures were used to reduce errors made by respondents, coders and interviewers. More detailed discussions of the existence and control of nonsampling errors in the SIPP can be found in the *SIPP Quality Profile, 1998 SIPP Working Paper Number 230*, issued May 1999.

Undercoverage in SIPP results from missed HUs and missed persons within sample HUs. It is known that undercoverage varies with age, race, and sex. Generally, undercoverage is larger for males than for females and larger for Blacks than for non-Blacks. Ratio estimation to independent age-race-sex population controls partially corrects for the bias due to survey undercoverage. However, biases exist in the estimates to the extent that persons in missed households or missed persons in interviewed households have characteristics different from those of interviewed persons in the same age-race-sex group.

A common measure of survey coverage is the coverage ratio, the estimated population before ratio adjustment divided by the independent population control. Table C below shows SIPP coverage ratios for age-sex-race groups for one month, December 2011, prior to the ratio adjustment. The SIPP coverage ratios exhibit some variability from month to month, but these are a typical set of coverage ratios. Other Census Bureau household surveys [like the CPS] experience similar coverage.

Table C. SIPP Average Coverage Ratios for December 2011 for Age by Race and Sex						
Age	White Only		Black Only		Residual	
	Male	Female	Male	Female	Male	Female
<15	0.83	0.83	0.73	0.72	0.77	0.86
15	0.92	0.88	0.81	0.69	0.98	0.98
16-17	0.87	0.86	0.81	0.70	0.99	0.97
18-19	0.83	0.84	0.80	0.72	0.98	0.99
20-21	0.74	0.75	0.65	0.68	1.00	0.93
22-24	0.65	0.66	0.65	0.69	0.89	0.88
25-29	0.64	0.70	0.44	0.58	0.78	0.78
30-34	0.75	0.81	0.51	0.71	0.76	0.77
35-39	0.83	0.87	0.63	0.77	0.73	0.84
40-44	0.82	0.88	0.66	0.75	0.80	0.90
45-49	0.83	0.87	0.81	0.70	0.98	1.01
50-54	0.84	0.89	0.79	0.86	0.99	1.01
55-59	0.91	0.97	0.83	1.04	0.98	1.05
60-61	0.95	1.01	0.89	1.02	1.02	1.04
62-64	1.02	1.04	0.89	1.01	1.03	1.06
65-69	0.93	0.93	1.07	1.00	0.99	0.96
70-74	0.96	0.95	1.06	1.08	1.00	0.97
75-79	0.91	0.97	1.10	1.07	0.99	1.00
80-84	0.98	1.02	1.02	1.02	0.99	0.95
85+	0.94	0.93	1.08	1.02	0.95	1.04

Comparability with Other Estimates. Caution should be exercised when comparing this data with data from other SIPP products or with data from other surveys. The comparability problems are caused by such sources as the seasonal patterns for many characteristics, different nonsampling errors, and different concepts and procedures. Refer to the *SIPP Quality Profile* for known differences with data from other sources and further discussions.

Sampling Variability. Standard errors indicate the magnitude of the sampling error. They also partially measure the effect of some nonsampling errors in response and enumeration, but do not measure any systematic biases in the data. The standard errors for the most part measure the variations that occurred by chance because a sample rather than the entire population was surveyed.

USES AND COMPUTATION OF STANDARD ERRORS

Confidence Intervals. The sample estimate and its standard error enable one to construct a confidence interval. A confidence interval is a range about a given estimate that has a known probability of including the result of a complete enumeration. For example, if all possible samples were selected, each of these being surveyed under essentially the same conditions and using the same sample design, and if an estimate and its standard error were calculated from each sample, then:

1. Approximately 68 percent of the intervals from one standard error below the estimate to one standard error above the estimate would include the average result of all possible samples.

2. Approximately 90 percent of the intervals from 1.645 standard errors below the estimate to 1.645 standard errors above the estimate would include the average result of all possible samples.

3. Approximately 95 percent of the intervals from two standard errors below the estimate to two standard errors above the estimate would include the average result of all possible samples.

The average estimate derived from all possible samples is or is not contained in any particular computed interval. However, for a particular sample, one can say with a specified confidence that the average estimate derived from all possible samples is included in the confidence interval.

Hypothesis Testing. Standard errors may also be used for hypothesis testing, a procedure for distinguishing between population characteristics using sample estimates. The most common types of hypotheses tested are 1) the population characteristics are identical versus 2) they are different. Tests may be performed at various levels of significance, where a level of significance is the probability of concluding that the characteristics are different when, in fact, they are identical.

To perform the most common test, compute the difference $X_A - X_B$, where X_A and X_B are sample estimates of the characteristics of interest. A later section explains how to derive an estimate of the standard error of the difference $X_A - X_B$. Let that standard error be S_{DIFF}. If $X_A - X_B$ is between $(-1.645 \times S_{DIFF})$ and $(+1.645 \times S_{DIFF})$, no conclusion about the characteristics is justified at the 10 percent significance level. If, on the other hand $X_A - X_B$, is smaller than $(-1.645 \times S_{DIFF})$ or larger than $(+1.645 \times S_{DIFF})$, the observed difference is significant at the 10 percent level. In this event, it is commonly accepted practice to say that the characteristics are different. We recommend that users report only those differences that are significant at the 10 percent level or better. Of course, sometimes this conclusion will be wrong. When the characteristics are the same, there is a 10 percent chance of concluding that they are different.

Note that as more tests are performed, more erroneous significant differences will occur. For example, at the 10 percent significance level, if 100 independent hypothesis tests are performed in which there are no real differences, it is likely that about 10 erroneous differences will occur. Therefore, the significance of any single test should be interpreted cautiously. A Bonferroni correction can be done to account for this potential problem that consists of dividing your stated level of significance by the number of tests you are performing. This correction results in a conservative test of significance.

Note Concerning Small Estimates and Small Differences. Because of the large standard errors involved, there is little chance that estimates will reveal useful information when computed on a base smaller than 75,000. Also, nonsampling error in one or more of the small number of cases providing the estimation can cause large relative error in that particular estimate. Care must be taken in the interpretation of small differences since even a small amount of nonsampling error can

cause a borderline difference to appear significant or not, thus distorting a seemingly valid hypothesis test.

Calculating Standard Errors for SIPP Estimates. There are three main ways we calculate the Standard Errors (SEs) for SIPP Estimates. They are as follows:

- Direct estimates using replicate weighting methods;
- Generalized variance function parameters (denoted as a and b); and
- Simplified tables of SEs based on the a and b parameters.

While the replicate weight methods provide the most accurate variance estimates, this approach requires more computing resources and more expertise on the part of the user. The Generalized Variance Function (GVF) parameters provide a method of balancing accuracy with resource usage as well as smoothing effect on SE estimates across time. SIPP uses the Replicate Weighting Method to produce GVF parameters (see K. Wolter, *Introduction to Variance Estimation,* for more information). The GVF parameters are used to create the simplified tables of SEs.

Standard Error Parameters and Tables and Their Use. Most SIPP estimates have greater standard errors than those obtained through a simple random sample because of its two-stage cluster sample design. To derive standard errors that would be applicable to a wide variety of estimates and could be prepared at a moderate cost, a number of approximations were required.

Estimates with similar standard error behavior were grouped together and two parameters (denoted as a and b) were developed to approximate the standard error behavior of each group of estimates. Because the actual standard error behavior was not identical for all estimates within a group, the standard errors computed from these parameters provide an indication of the order of magnitude of the standard error for any specific estimate. These a and b parameters vary by characteristic and by demographic subgroup to which the estimate applies. Table 4 provides a and b parameters for the core domains to be used for the 2008 Panel Wave 1 to Wave 11 estimates. The base a and b parameters for the topical modules for Wave 1 to Wave 11 are found in Table 5.
For those users who wish further simplification, we have also provided base standard errors for estimates of totals and percentages in Tables 6 through 9. Note that these base standard errors only apply when data from all four rotations are used and must be adjusted by an f factor provided in Table 4. The standard errors resulting from this simplified approach are less accurate. Methods for using these parameters and tables for computation of standard errors are given in the following sections

Adjusting Standard Error Parameters for Estimates Which Use Less Than the Full Sample
If some rotation groups are unavailable to contribute data to a given estimate, then the estimate and its standard error need to be adjusted. The adjustment of the estimate is described in the previous section. The standard error is adjusted by multiplying the appropriate a and b parameters by a factor equal to 4 divided by the number of rotation groups contributing data to the estimate or it can be taken from Table 3 where the factor is given for each single reference month, May 2008 to August 2008.

For monthly and quarterly estimates, use Table 3 to select the adjustment factor appropriate to the number of rotation months. Multiply this factor by the a and b base parameters of Table 4 to produce a and b parameters for the variance estimate for a specific subgroup and reference period.

Illustration 1.
Using Table 4 for Wave 1 of the 2008 panel, the base a and b parameters for total number of households are -0.00002703 and 3,179, respectively. Using Table 3 for Wave 1, the factor for June 2008 is 2 *since only two rotation months of data are available*. So the a and b parameters for the variance estimate of a white household characteristic in June 2008 based on Wave 1 are:

$$-0.00002703 \times 2 = -0.00005406 \; and \; 3{,}179 \times 2 = 6{,}358, respectively.$$

Similarly, the factor from Table 3 for the third quarter of 2008 is 1.0370, since the only data available are the eleven rotation months from Wave 1. (Rotation 1 provides three rotation months, rotation 2 provides three rotation months, rotation 3 provides three rotation months, and rotation 4 provides two rotation months of data.) Thus, the a and b parameters for the variance estimate of a white household characteristic in the third quarter of 2008 are:

$$-0.00002703 \times 1.0370 = -0.00002803 \; and \; 3{,}179 \times 1.0370 = 3{,}297, respectively.$$

Standard Errors of Estimated Numbers. The approximate standard error, s_x, of an estimated number of persons, households, families, unrelated individuals and so forth, can be obtained in two ways. Both apply when data from all four rotations are used to make the estimate. However, only Formula (2) should be used when less than four rotations of data are available for the estimate. Note that neither method should be applied to dollar values.

The standard error may be obtained by the use of Formula (2):

$$s_x = f \times s, \tag{2}$$

where f is the appropriate f factor from Table 4, and s is the base standard error on the estimate obtained by interpolation from Tables 6 or 7.

Alternatively, s_x may be approximated by Formula (3):

$$s_x = \sqrt{ax^2 + bx} \tag{3}$$

This formula was used to calculate the base standard errors in Tables 6 and 7. Here x is the size of the estimate and a and b are the parameters from Table 4 which are associated with the characteristic being estimated (and the wave which applies). Use of Formula (3) will generally provide more accurate results than the use of Formula (2).

Illustration 2.
Suppose SIPP estimates based on Wave 1 of the 2008 panel show that there were 2,000,000 females aged 25 to 44 with a monthly income of greater than $6,000 in September 2008. The appropriate parameters and factor from Table 4 and the appropriate general standard error from Table 7 are:

$$a = -0.00002917 \qquad b = 3{,}584 \qquad f = 0.989 \qquad s = 85{,}282$$

Using Formula (2), the approximate standard error is:

$$s_x = 0.989 \times 85{,}282 = 84{,}344.$$

Using Formula (3), the approximate standard error is:

$$s_x = \sqrt{(-0.00002917 \times 2{,}000{,}000^2) + (3{,}584 + 2{,}000{,}000)} = 83{,}972\ females.$$

Using the standard error based on Formula (3), the approximate 90-percent confidence interval as shown by the data is from 1,861,866 to 2,138,134 females $(i.e., 2{,}000{,}000 \pm 1.645 \times 83{,}972)$. Therefore, a conclusion that the average estimate derived from all possible samples lies within a range computed in this way would be correct for roughly 90% of all samples.

Standard Error of a Mean. A mean is defined here to be the average quantity of some item (other than persons, families, or households) per person, family or household. For example, it could be the average monthly household income of females age 25 to 34. The standard error of a mean can be approximated by Formula (4) below. Because of the approximations used in developing Formula (4), an estimate of the standard error of the mean obtained from this formula will generally underestimate the true standard error. The formula used to estimate the standard error of a mean $\bar{x}$ is:

$$s_{\bar{x}} = \sqrt{\left(\frac{b}{y}\right) s^2}, \tag{4}$$

where y is the size of the base, s^2 is the estimated population variance of the item and b is the parameter associated with the particular type of item.

The population variance s^2 may be estimated by one of two methods. In both methods, we assume x_i is the value of the item for i^{th} unit. (A unit may be person, family, or household). To use the first method, the range of values for the item is divided into c intervals. The lower and upper boundaries of interval j are Z_{j-1} and Z_j, respectively. Each unit, x_i, is placed into one of c intervals such that $Z_{j-1} < x_i \leq Z_j$. The estimated population mean, $\bar{x}$, and variance, s^2, are given by the formulas:

$$\bar{x} = \sum_{j=1}^{c} p_j m_j$$

$$s^2 = \sum_{j=1}^{c} p_j m_j^2 - \bar{x}^2 \tag{5}$$

where $m_j = (Z_{j-1} + Z_j)/2$, and p_j is the estimated proportion of units in the interval j. The most representative value of the item in the interval j is assumed to be m_j. If the interval c is open-ended, or no upper interval boundary exists, then an approximate value for m_c is

$$m_c = \frac{3}{2} Z_{c-1}.$$

In the second method, the estimated population mean, $\bar{x}$, and variance, s^2 are given by:

$$\bar{x} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}$$

$$s^2 = \frac{\sum_{i=1}^{n} w_i x_i^2}{\sum_{i=1}^{n} w_i} - \bar{x}^2 \tag{6}$$

where there are n units with the item of interest and w_i is the final weight for i^{th} unit. (Note that $\sum w_i = y$.)

<u>Illustration 3.</u>
Suppose that based on Wave 1 data, the distribution of monthly cash income for persons age 25 to 34 during the month of September 2008 is given in Table 10. Using these data, the mean monthly cash income for persons aged 25 to 34 is \$2,530. Applying Formula (5), the approximate population variance, s^2, is:

$$s^2 = \left(\frac{1{,}371}{39{,}851}\right)(150)^2 + \left(\frac{1{,}651}{39{,}851}\right)(450)^2 + \cdots + \left(\frac{1{,}493}{39{,}851}\right)(9{,}000)^2 - (2{,}530)^2 = 3{,}159{,}887.$$

Using Formula (4) and a base b parameter of 3,584, the estimated standard error of a mean $\bar{x}$ is:

$$s_{\bar{x}} = \sqrt{\frac{3{,}584}{39{,}851{,}000} \times 3{,}159{,}887} = \$16.86$$

Thus, the approximate 90-percent confidence interval as shown by the data ranges from \$2,502.27 to \$2,557.73.

Standard Error of an Aggregate. An aggregate is defined to be the total quantity of an item summed over all the units in a group. The standard error of an aggregate can be approximated using Formula (7). As with the estimate of the standard error of a mean, the estimate of the standard error of an aggregate will generally underestimate the true standard error. Let y be the size of the base, s^2 be the estimated population variance of the item obtained using Formula (5) or Formula (6) and b be the parameter associated with the particular type of item. The standard error of an aggregate is:

$$s_x = \sqrt{b \times y \times s^2}. \tag{7}$$

Standard Errors of Estimated Percentages. The reliability of an estimated percentage, computed using sample data for both numerator and denominator, depends upon both the size of the percentage and the size of the total upon which the percentage is based. Estimated percentages are relatively more reliable than the corresponding estimates of the numerators of the percentages, particularly if the percentages are 50 percent or more, e.g., the percent of people employed is more reliable than the estimated number of people employed. When the numerator and denominator of the percentage have different parameters, use the parameter (and appropriate factor) of the numerator. If proportions are presented instead of percentages, note that the standard error of a proportion is equal to the standard error of the corresponding percentage divided by 100.

There are two types of percentages commonly estimated. The first is the percentage of people sharing a particular characteristic such as the percent of people owning their own home. The second type is the percentage of money or some similar concept held by a particular group of people or held in a particular form. Examples are the percent of total wealth held by people with high income and the percent of total income received by people on welfare.

For the percentage of people, the approximate standard error, $s_{(x,p)}$, of the estimated percentage p can be obtained by the formula:

$$s_{(x,p)} = f \times s, \tag{8}$$

when data from all four rotations are used to estimate p. In this formula, f is the appropriate f factor from Table 4 (for the appropriate wave) and s is the base standard error of the estimate from Tables 8 or 9.

Alternatively, it may be approximated by the formula:

$$s_{(x,p)} = \sqrt{\frac{b}{x}(p)(100-p)}, \tag{9}$$

from which the standard errors in Tables 8 and 9 were calculated. Here x is the size of the subclass of social units which is the base of the percentage, p is the percentage $(0 < p < 100)$, and b is the parameter associated with the characteristic in the numerator. Use of Formula (9) will give more accurate results than use of Formula (8) above and should be used when data from less than four rotations are used to estimate p.

Illustration 4.
Suppose that in September 2008, 6.7 percent of the 16,812,000 persons in nonfarm households with a mean monthly household cash income of \$4,000 to \$4,999 were black. Using Formula (9), a b parameter of 3,534, and a factor of 1 from Table 3 since all four rotations are used, the approximate standard error is:

$$s_{(x,p)} = \sqrt{\frac{3{,}534}{16{,}812{,}000} \times 6.7 \times (100 - 6.7)} = 0.36\ percent$$

Consequently, the 90 percent confidence interval as shown by these data is from 6.11 to 7.29 percent.

For percentages of money, a more complicated formula is required. A percentage of money will usually be estimated in one of two ways. It may be the ratio of two aggregates:

$$p_I = 100\left(\frac{x_A}{x_N}\right),$$

or it may be the ratio of two means with an adjustment for different bases:

$$p_I = 100\left(\hat{p}_A\left(\frac{\bar{x}_A}{\bar{x}_N}\right)\right),$$

where x_A and x_N are aggregate money figures, $\bar{x}_A$ and $\bar{x}_N$ are mean money figures, and $\hat{p}_A$ is the estimated number in group A divided by the estimated number in group N. In either case, we estimate the standard error as

$$s_I = \sqrt{\left(\frac{\hat{p}_A\bar{x}_A}{\bar{x}_N}\right)^2\left[\left(\frac{s_p}{\hat{p}_A}\right)^2 + \left(\frac{s_A}{\bar{x}_A}\right)^2 + \left(\frac{s_B}{\bar{x}_N}\right)^2\right]}, \tag{10}$$

where s_p is the standard error of $\hat{p}_A, s_A$ is the standard error of $\bar{x}_A$ and s_B is the standard error of $\bar{x}_N$. To calculate s_p, use Formula (9). The standard errors of $\bar{x}_N$ and $\bar{x}_A$ may be calculated using Formula (4).

It should be noted that there is frequently some correlation between $\hat{p}_A, \bar{x}_N$, and $\bar{x}_A$. Depending on the magnitude and sign of the correlations, the standard error will be over or underestimated.

Illustration 5.
Suppose that in September 2008, 9.8% of the households own rental property, the mean value of rental property is \$72,121, the mean value of assets is \$78,734, and the corresponding standard errors are 0.18%, \$5,468, and \$2,703, respectively. In total there are 86,790,000 households. Then, the percent of all household assets held in rental property is:

$$100\left(0.098 \times \frac{72{,}121}{78{,}734}\right) = 9.0\%$$

Using Formula (10), the appropriate standard error is:

$$s_I = \sqrt{\left(\frac{0.098 \times 72{,}121}{78{,}734}\right)^2 \left[\left(\frac{0.0018}{0.098}\right)^2 + \left(\frac{5{,}468}{72{,}121}\right)^2 + \left(\frac{2{,}703}{78{,}734}\right)^2\right]} = 0.7\%.$$

Standard Error of a Difference. The standard error of a difference between two sample estimates is approximately equal to

$$s_{(x-y)} = \sqrt{s_x^2 + s_y^2}, \tag{11}$$

where s_x and s_y are the standard errors of the estimates x and y. The estimates can be numbers, percents, ratios, etc. The above formula assumes that the correlation coefficient between the characteristics estimated by x and y is zero. If the correlation is really positive (negative), then this assumption will tend to cause overestimates (underestimates) of the true standard error.

Illustration 6.
Suppose that for September 2008 SIPP estimates show the number of persons age 35-44 years with monthly cash income of \$4,000 to \$4,999 was 4,880,200 and the number of persons age 25-34 years with monthly cash income of \$4,000 to \$4,999 in the same time period was 4,810,800. Then, using the parameters $a = -0.00001504$ *and* $b = 3{,}584$ from Table 4 and Formula (3), the standard errors of these numbers are approximately 130,891 and 129,976, respectively. The difference in sample estimates is 69,400 and using Formula (11), the approximate standard error of the difference is:

$$\sqrt{130{,}891^2 + 129{,}976^2} = 184{,}462.$$

Suppose that it is desired to test at the 10 percent significance level whether the number of persons with monthly cash income of \$4,000 to \$4,999 was different for people age 35-44 years than for people age 25-34 years. To perform the test, compare the difference of 69,400 to the product $1.645 \times 184{,}462 = 303{,}440$. Since the difference is not greater than 1.645 times the standard error of the difference, the data show that the two age groups are not significantly different at the 10 percent significance level.

Standard Error of a Median. The median quantity of some items such as income for a given group of people is that quantity such that at least half the group have as much or more and at least half the group have as much or less. The sampling variability of an estimated median depends upon the form of the distribution of the item as well as the size of the group. To calculate standard errors on medians, the procedure described below may be used.

The median, like the mean, can be estimated using either data which have been grouped into intervals or ungrouped data. If grouped data are used, the median is estimated using Formulas (12) or (13) with $p = 0.5$. If ungrouped data are used, the data records are ordered based on the value of the characteristic, then the estimated median is the value of the characteristic such that the weighted estimate of 50 percent of the subpopulation falls at or below that value and 50 percent is at or above that value. Note that the method of standard error computation which is presented here requires the use of grouped data. Therefore, it should be easier to compute the median by grouping the data and using Formulas (12) or (13).

An approximate method for measuring the reliability of an estimated median is to determine a confidence interval about it. (See the section on sampling variability for a general discussion of confidence intervals.) The following procedure may be used to estimate the 68-percent confidence limits and hence the standard error of a median based on sample data.

1. Determine, using either Formula (8) or Formula (9), the standard error of an estimate of 50 percent of the group.

2. Add to and subtract from 50 percent the standard error determined in step 1.

3. Using the distribution of the item within the group, calculate the quantity of the item such that the percent of the group with more of the item is equal to the smaller percentage found in step 2. This quantity will be the upper limit for the 68-percent confidence interval. In a similar fashion, calculate the quantity of the item such that the percent of the group with more of the item is equal to the larger percentage found in step 2. This quantity will be the lower limit for the 68-percent confidence interval.

4. Divide the difference between the two quantities determined in step 3 by two to obtain the standard error of the median.

To perform step 3, it will be necessary to interpolate. Different methods of interpolation may be used. The most common are simple linear interpolation and Pareto interpolation. The appropriateness of the method depends on the form of the distribution around the median. If density is declining in the area, then we recommend Pareto interpolation. If density is fairly constant in the area, then we recommend linear interpolation. Note, however, that Pareto interpolation can never be used if the interval contains zero or negative measures of the item of interest. Interpolation is used as follows. The quantity of the item such that p percent have more of the item is:

$$X_{pN} = A_1 \times \exp\left[\left(\frac{\ln\left(\frac{pN}{N_1}\right)}{\ln\left(\frac{N_2}{N_1}\right)}\right)\ln\left(\frac{A_2}{A_1}\right)\right] \tag{12}$$

if Pareto Interpolation is indicated and:

$$X_{pN} = \left[A_1 + \left(\frac{PN - N_1}{N_2 - N_1}\right)(A_2 - A_1)\right], \tag{13}$$

if linear interpolation is indicated, where:

N	is the size of the group,
A_1 *and* A_2	are the lower and upper bounds, respectively, of the interval in which X_{pN} falls
N_1 *and* N_2	are the estimated number of group members owning more than A_1 and A_2, respectively
exp	refers to the exponential function and
ln	refers to the natural logarithm function

Illustration 7.
To illustrate the calculations for the sampling error on a median, we return to Table 10. The median monthly income for this group is \$2,158. The size of the group is 39,851,000.

1. Using Formula (9), the standard error of 50 percent on a base of 39,851,000 is about 0.5 percentage points.

2. Following step 2, the two percentages of interest are 49.5 and 50.5.

3. By examining Table 10, we see that the percentage 49.5 falls in the income interval from \$2,000 to \$2,499. (Since 55.5% receive more than \$2,000 per month, the dollar value corresponding to 49.5 must be between \$2,000 and \$2,500.) Thus, $A_1 = \$2{,}000, A_2 = \$2{,}500, N_1 = 22{,}106{,}000$ *and* $N_2 = 16{,}307{,}000$.

In this case, we decided to use Pareto interpolation. Therefore, using Formula (12), the upper bound of a 68% confidence interval for the median is

$$\$2{,}000 \times \exp\left[\left(\frac{\ln\left(\frac{0.495 \times 39{,}851{,}000}{22{,}106{,}000}\right)}{\ln\left(\frac{16{,}307{,}000}{22{,}106{,}000}\right)}\right) \times \ln\left(\frac{2{,}500}{2{,}000}\right)\right] = \$2{,}174.$$

Also by examining Table 10, we see that 50.5 falls in the same income interval. Thus, A_1, A_2, N_1 *and* N_2 are the same. We also use Pareto interpolation for this case. So the lower bound of a 68% confidence interval for the median is

$$\$2{,}000 \times \exp\left[\left(\frac{\ln\left(\frac{0.505 \times 39{,}851{,}000}{22{,}106{,}000}\right)}{\ln\left(\frac{16{,}307{,}000}{22{,}106{,}000}\right)}\right) \times \ln\left(\frac{2{,}500}{2{,}000}\right)\right] = \$2{,}142.$$

Thus, the 68-percent confidence interval on the estimated median is from $2,142 to $2,174.

4. Then the approximate standard error of the median is

$$\frac{\$2{,}174 - \$2{,}142}{2} = \$16$$

Standard Errors of Ratios of Means and Medians. The standard error for a ratio of means or medians is approximated by:

$$s_{\frac{x}{y}} = \sqrt{\left(\frac{x}{y}\right)^2 \left[\left(\frac{s_y}{y}\right)^2 + \left(\frac{s_x}{x}\right)^2\right]}, \tag{14}$$

where x and y are the means or medians, and s_x and s_y are their associated standard errors. Formula (14) assumes that the means are not correlated. If the correlation between the population means estimated by x and y are actually positive (negative), then this procedure will tend to produce overestimates (underestimates) of the true standard error for the ratio of means.

Standard Errors Using SAS or SPSS. Standard errors and their associated variance, calculated by SAS or SPSS statistical software package, do not accurately reflect the SIPP's complex sample design. Erroneous conclusions will result if these standard errors are used directly. We provide adjustment factors by characteristics that should be used to correctly compensate for likely under-estimates. The design effect (DEFF) factors that are available in Table 4, must be applied to SAS or SPSS generated variances. The square root of DEFF can be directly applied to similarly generated standard errors. These factors approximate design effects which adjust statistical measures for sample designs more complex than a simple random sample.

REFERENCES

U.S. Census Bureau (1999). *SIPP Quality Profile, 1998,* SIPP Working Paper No. 230. Washington, DC: U.S. Census Bureau, May 1999.

U.S. Census Bureau (2008). "Chapter 8: Using Sampling Weights on SIPP Files," *Survey of Income and Program Participation Users' Guide,* 3rd Ed. Washington, DC: U.S. Census Bureau.

Wolter, Kirk M. (2007). "Chapter 7: Generalized Variance Functions," *Introduction to Variance Estimation,* 2nd Ed. New York: Springer, pp. 272-297.

TABLES

Table 1. 2008 Panel Topical Modules			
W1	• Recipiency History • Employment History • Tax Rebates	W7	• Assets and Liabilities • Real Estate, Dependent Care, and Vehicles • Int Acct, Stocks, Mortg, Rental, Val of Bus, Other • Medical Expenses/Utilization of Health Care Services • Poverty (Work-related Expenses/Child Support Paid)
W2	• Work Disability • Education & Training History • Marital History • Migration History • Fertility History • Household Relationships • Tax Rebates	W8	• Annual Income and Retirement Accounts • Taxes • Child Care • Work Schedule
W3	• Welfare Reform • Retirement and Pension Plan Coverage	W9	• Informal Care-giving • Adult Well-being
W4	• Assets and Liabilities • Real Estate, Dependent Care, and Vehicles • Int Accts, Stocks, Mortg.,Val of Bus, Rental, Other • Medical Expenses/Utilization of Health Care Services • Poverty (Work-related Expenses/Child Support Paid) • Child Well-Being	W10	• Assets and Liabilities • Real Estate, Dependent Care, and Vehicles • Int Acct, Stocks, Mortg, Rental, Val of Bus, Other • Medical Expenses/Utilization of Health Care Services • Poverty (Work-related Expenses/Child Support Paid) • Child Well-Being
W5	• Annual Income and Retirement Accounts • Taxes • Child Care • Work Schedule	W11	• Retirement and Pension Plan Coverage
W6	• Adult Well-being • Child Support Agreements • Support for Non-household Memebers • Functional Limitations and Disability-Adults • Functional Limitations and Disability-Children • Employer-Provided Health Benefits	W12 - W16	• There are no topical modules planned for Waves 12 – 16.

Table 2. SIPP Panel 2008 Reference Months (horizontal) for Each Interview Month (vertical)[3]

Month of Interview	Wave / Rotation	2008								2009												2010												2011												2012							
		2nd Quarter		3rd Quarter			4th Quarter			1st Quarter			2nd Quarter			3rd Quarter			4th Quarter			1st Quarter			2nd Quarter			3rd Quarter			4th Quarter			1st Quarter			2nd Quarter			3rd Quarter			4th Quarter			1st Quarter			2nd Quarter			3rd Quarter			4th Quar.	
		May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov
Sep 08	1/1	1	2	3	4																																																			
Oct	1/2		1	2	3	4																																																		
Nov	1/3			1	2	3	4																																																	
Dec	1/4				1	2	3	4																																																
Jan 09	2/1					1	2	3	4																																															
Feb	2/2						1	2	3	4																																														
Mar	2/3							1	2	3	4																																													
Apr	2/4								1	2	3	4																																												
May	3/1									1	2	3	4																																											
Jun	3/2										1	2	3	4																																										
July	3/3											1	2	3	4																																									
Aug	3/4												1	2	3	4																																									
Sep	4/1													1	2	3	4																																								
Oct	4/2														1	2	3	4																																							
Nov	4/3															1	2	3	4																																						
Dec	4/4																1	2	3	4																																					
Jan 10	5/1																	1	2	3	4																																				
Feb	5/2																		1	2	3	4																																			
Mar	5/3																			1	2	3	4																																		
Apr	5/4																				1	2	3	4																																	
May	6/1																					1	2	3	4																																
Jun	6/2																						1	2	3	4																															
July	6/3																							1	2	3	4																														
Aug	6/4																								1	2	3	4																														
Sep	7/1																									1	2	3	4																													
Oct	7/2																										1	2	3	4																												
Nov	7/3																											1	2	3	4																											
Dec	7/4																												1	2	3	4																										
Jan 11	8/1																													1	2	3	4																									
Feb	8/2																														1	2	3	4																							
Mar	8/3																															1	2	3	4																						
Apr	8/4																																1	2	3	4																					
May	9/1																																	1	2	3	4																				
Jun	9/2																																		1	2	3	4																			
July	9/3																																			1	2	3	4																		
Aug	9/4																																				1	2	3	4																	
Sep	10/1																																					1	2	3	4																
Oct	10/2																																						1	2	3	4															
Nov	10/3																																							1	2	3	4														
Dec	10/4																																								1	2	3	4													
Jan 12	11/1																																									1	2	3	4												
Feb	11/2																																										1	2	3	4										
Mar	11/3																																											1	2	3	4									
Apr	11/4																																												1	2	3	4								
May	12/1																																													1	2	3	4							
Jun	12/2																																														1	2	3	4						
July	12/3																																															1	2	3	4					
Aug	12/4																																																1	2	3	4				
Sep	13/1																																																	1	2	3	4			
Oct	13/2																																																		1	2	3	4		
Nov	13/3																																																			1	2	3	4	
Dec	13/4																																																				1	2	3	4

3 The SIPP 2008 panel has been extended to go through Wave 16.

Table 3. Factors to be Used When Using Less Than Full Sample	
Number of Available Rotation Months[4]	**Factor**
Monthly Estimate[5]	
1	4.0000
2	2.0000
3	1.3333
4	1.0000
Quarterly Estimate[6]	
6	1.8519
8	1.4074
9	1.2222
10	1.0494
11	1.0370
12	1.0000

4 The number of available rotation months for a given estimate is the sum of the number of rotations available for each month of the estimates.

5 Adjustment factors for monthly estimates are equal to 4 divided by the number of rotation groups contributing data to the estimate

6 Adjustment factors for quarterly estimates are calculated as follows:
Assume:

1. No change within rotation (i.e., no change in value for a variable across months).
2. Rotations are independent.
3. All sigmas are equal.

The monthly factor for each month are equal to 4 divided by the number of rotation groups contributing data to the estimate. Therefore, the variance of the estimate for the full sample is: $\sum_{Rotation} Var(X_{Jan} + X_{Feb} + X_{March}) = 36\sigma^2$. The variance of the estimate for less than a full sample is: the sum of the squared monthly factors for each rotation month $* \sigma^2$. The adjustment factor for the quarterly estimate is: (the sum of the squared monthly factors for each rotation month $* \sigma^2$) $/(36\sigma^2)$.

Table 4. SIPP Generalized Variance Parameters for the 2008 Panel, Wave 1

Domain	Parameters		DEFF[6]	*f*
	a	*b*		
Poverty and Program Participation, Persons 15+				
Total	-0.00001532	3,651	1.84	1.000
Male	-0.00003163	3,651		
Female	-0.00002971	3,651		
Income and Labor Force Participation,				
Total	-0.00001504	3,584	1.80	0.989
Male	-0.00003105	3,584		
Female	-0.00002917	3,584		
Other, Persons 0+				
Total (or White)	-0.00001223	3,661	1.84	1.000
Male	-0.00002496	3,661		
Female	-0.00002397	3,661		
Black, Persons 0+	-0.00009339	3,534	1.78	0.983
Male	-0.00020096	3,534		
Female	-0.00017447	3,534		
Hispanic, Persons 0+	-0.00009852	4,588	2.31	1.119
Male	-0.00019194	4,588		
Female	-0.00020241	4,588		
Households				
Total (or White)	-0.00002703	3,179	1.60	1.000
Black	-0.00021922	3,179		
Hispanic	-0.00023147	3,179		

Notes on Domain Usage for Table 4:

Poverty and Program Participation	Use these parameters for estimates concerning poverty rates, welfare program participation (e.g., foodstamp, SSI, TANF), and other programs for adults with low incomes.
Income and Labor Force	These parameters are for estimates concerning income, sources of income, labor force participation, economic well being other than poverty, employment related estimates (e.g., occupation, hours worked a week), and other income, job, or employment related estimates.
Other Persons	Use the "Other Persons" parameters for estimates of total (or white) persons aged 0+ in the labor force, and all other characteristics not specified in this table, for the total or white population.
Black/Hispanic Persons	Use these parameters for estimates of Black and Hispanic persons 0+.
Households	Use these parameters for all household level estimates.

6 DEFF=b/sample interval, where sample interval=1,989

Table 4.(Continued) SIPP Generalized Variance Parameters for the 2008 Panel, Wave 2-3

Domain	Parameters		DEFF[6]	*f*
	a	*b*		
Poverty and Program Participation, Persons 15+				
Total	-0.00001786	4,295	2.16	1.083
Male	-0.00003687	4,295		
Female	-0.00003465	4,295		
Income and Labor Force				
Total	-0.00001721	4,137	2.08	1.063
Male	-0.00003552	4,137		
Female	-0.00003338	4,137		
Other, Persons 0+				
Total (or White)	-0.00001434	4,327	2.18	1.087
Male	-0.00002926	4,327		
Female	-0.00002811	4,327		
Black, Persons 0+	-0.00011484	4,376	2.20	1.093
Male	-0.00024713	4,376		
Female	-0.00021452	4,376		
Hispanic, Persons 0+	-0.00011685	5,561	2.80	1.232
Male	-0.00022778	5,561		
Female	-0.00023994	5,561		
Households				
Total (or White)	-0.00003137	3,722	1.87	1.082
Black	-0.00025251	3,722		
Hispanic	-0.00026735	3,722		

Notes on Domain Usage for Table 4:

Poverty and Program Participation	Use these parameters for estimates concerning poverty rates, welfare program participation (e.g., foodstamp, SSI, TANF), and other programs for adults with low incomes.
Income and Labor Force	These parameters are for estimates concerning income, sources of income, labor force participation, economic well being other than poverty, employment related estimates (e.g., occupation, hours worked a week), and other income, job, or employment related estimates.
Other Persons	Use the "Other Persons" parameters for estimates of total (or white) persons aged 0+ in the labor force, and all other characteristics not specified in this table, for the total or white population.
Black/Hispanic Persons	Use these parameters for estimates of Black and Hispanic persons 0+.
Households	Use these parameters for all household level estimates.

6 DEFF=b/sample interval, where sample interval=1,989

Table 4.(Continued) SIPP Generalized Variance Parameters for the 2008 Panel, Wave 4-6

Domain	Parameters		DEFF[6]	*f*
	a	*b*		
Poverty and Program Participation, Persons 15+				
Total	-0.00001993	4,834	2.43	1.149
Male	-0.00004111	4,834		
Female	-0.00003867	4,834		
Income and Labor Force Participation,				
Total	-0.00001855	4,500	2.26	1.109
Male	-0.00003827	4,500		
Female	-0.00003600	4,500		
Other, Persons 0+				
Total (or White)	-0.00001592	4,851	2.44	1.151
Male	-0.00003248	4,851		
Female	-0.00003122	4,851		
Black, Persons 0+	-0.00012441	4,818	2.42	1.147
Male	-0.00026711	4,818		
Female	-0.00023288	4,818		
Hispanic, Persons 0+	-0.00012848	6,302	3.17	1.312
Male	-0.00025001	6,302		
Female	-0.00026432	6,302		
Households				
Total (or White)	-0.00003401	4,037	2.03	1.127
Black	-0.00026961	4,037		
Hispanic	-0.00029139	4,037		

Notes on Domain Usage for Table 4:

Poverty and Program Participation	Use these parameters for estimates concerning poverty rates, welfare program participation (e.g., foodstamp, SSI, TANF), and other programs for adults with low incomes.
Income and Labor Force	These parameters are for estimates concerning income, sources of income, labor force participation, economic well being other than poverty, employment related estimates (e.g., occupation, hours worked a week), and other income, job, or employment related estimates.
Other Persons	Use the "Other Persons" parameters for estimates of total (or white) persons aged 0+ in the labor force, and all other characteristics not specified in this table, for the total or white population.
Black/Hispanic Persons	Use these parameters for estimates of Black and Hispanic persons 0+.
Households	Use these parameters for all household level estimates.

6 DEFF=b/sample interval, where sample interval=1,989

Table 4.(Continued) SIPP Generalized Variance Parameters for the 2008 Panel, Wave 7-9

Domain	Parameters		DEFF[6]	*f*
	a	*b*		
Poverty and Program Participation, Persons 15+				
Total	-0.00002221	5,426	2.73	1.217
Male	-0.00004571	5,426		
Female	-0.00004319	5,426		
Income and Labor Force Participation,				
Total	-0.00002011	4,913	2.47	1.158
Male	-0.00004139	4,913		
Female	-0.00003911	4,913		
Other, Persons 0+				
Total (or White)	-0.00001765	5,409	2.72	1.216
Male	-0.00003594	5,409		
Female	-0.00003467	5,409		
Black, Persons 0+	-0.00014401	5,635	2.83	1.241
Male	-0.00030883	5,635		
Female	-0.00026984	5,635		
Hispanic, Persons 0+	-0.00013176	6,604	3.32	1.343
Male	-0.00025629	6,604		
Female	-0.00027116	6,604		
Households				
Total (or White)	-0.00003687	4,425	2.22	1.180
Black	-0.00028880	4,425		
Hispanic	-0.00031165	4,425		

Notes on Domain Usage for Table 4:

Poverty and Program Participation	Use these parameters for estimates concerning poverty rates, welfare program participation (e.g., foodstamp, SSI, TANF), and other programs for adults with low incomes.
Income and Labor Force	These parameters are for estimates concerning income, sources of income, labor force participation, economic well being other than poverty, employment related estimates (e.g., occupation, hours worked a week), and other income, job, or employment related estimates.
Other Persons	Use the "Other Persons" parameters for estimates of total (or white) persons aged 0+ in the labor force, and all other characteristics not specified in this table, for the total or white population.
Black/Hispanic Persons	Use these parameters for estimates of Black and Hispanic persons 0+.
Households	Use these parameters for all household level estimates.

6 DEFF=b/sample interval, where sample interval=1,989

Table 4.(Continued) SIPP Generalized Variance Parameters for the 2008 Panel, Wave 10-11

Domain	Parameters		DEFF[6]	*f*
	a	*b*		
Poverty and Program Participation, Persons 15+				
Total	-0.00002316	5,688	2.86	1.247
Male	-0.00004766	5,688		
Female	-0.00004507	5,688		
Income and Labor Force Participation,				
Total	-0.00002171	5,331	2.68	1.207
Male	-0.00004467	5,331		
Female	-0.00004224	5,331		
Other, Persons 0+				
Total (or White)	-0.00001851	5,701	2.87	1.250
Male	-0.00003769	5,701		
Female	-0.00003638	5,701		
Black, Persons 0+	-0.00015183	5,978	3.01	1.279
Male	-0.00032574	5,978		
Female	-0.00028438	5,978		
Hispanic, Persons 0+	-0.00013671	6,966	3.50	1.379
Male	-0.00026565	6,966		
Female	-0.00028165	6,966		
Households				
Total (or White)	-0.00003865	4,637	2.33	1.125
Black	-0.00030277	4,637		
Hispanic	-0.00032246	4,637		

Notes on Domain Usage for Table 4:

Poverty and Program Participation	Use these parameters for estimates concerning poverty rates, welfare program participation (e.g., foodstamp, SSI, TANF), and other programs for adults with low incomes.
Income and Labor Force	These parameters are for estimates concerning income, sources of income, labor force participation, economic well being other than poverty, employment related estimates (e.g., occupation, hours worked a week), and other income, job, or employment related estimates.
Other Persons	Use the "Other Persons" parameters for estimates of total (or white) persons aged 0+ in the labor force, and all other characteristics not specified in this table, for the total or white population.
Black/Hispanic Persons	Use these parameters for estimates of Black and Hispanic persons 0+.
Households	Use these parameters for all household level estimates.

6 DEFF=b/sample interval, where sample interval=1,989

Table 5. SIPP Topical Module Generalized Variance Parameters for the 2008 Panel

Characteristics	Parameters	
	a	*b*
Employment History, Wave 1		
Both Sexes, Age 18+	-0.00001504	3,584
Male, Age 18+	-0.00003105	3,584
Female, Age 18+	-0.00002917	3,584
Recipiency History, Wave 1		
Both Sexes, Age 18+	-0.00001532	3,651
Male, Age 18+	-0.00003163	3,651
Female, Age 18+	-0.00002971	3,651
Fertility History, Wave 2		
Women	-0.00002596	3,240
Births	-0.00004735	5,907
Education History, Wave 2	-0.00001836	4,412
Marital History, Wave 2		
Some Household Members	-0.00002780	6,677
All Household Members	-0.00002566	8,113
Migration History, Wave 2	-0.00002060	4,939
Household Relationship, Wave 2	-0.00001359	4,093
Welfare Reform, Wave 3	-0.00005229	12,135
Assets and Liabilities		
Wave 4	-0.00001905	4,671
Wave 7	-0.00002124	5,178
Wave 10	-0.00002321	5,696
Child Well-Being (Under 18),		
Wave 4	-0.00005835	4,508
Wave 10	-0.00006757	5,292
Child Care (Age 0 to 15), Wave 5	-0.00006277	4,821
Wave 8	-0.00006694	5,216
Work Schedule (15+), Wave 5	-0.00001826	4,423
Child Support, Wave 6	-0.00004807	6,062
Support for Non-Household Members, Wave 6	-0.00002493	6,062
Health and Disability - Adults, Wave 6	-0.00002375	7,585

Table 6. Base Standard Errors of Estimated Numbers of Households or Families			
Size of Estimate	**Standard Error**	**Size of Estimate**	**Standard Error**
200,000	25,194	30,000,000	266,539
300,000	30,843	40,000,000	289,676
500,000	39,784	50,000,000	302,283
750,000	48,673	60,000,000	305,666
1,000,000	56,142	70,000,000	300,138
2,000,000	79,056	80,000,000	285,181
3,000,000	96,404	90,000,000	259,166
5,000,000	123,366	95,000,000	240,955
7,500,000	149,406	99,500,000	220,696
10,000,000	170,549	105,000,000	189,180
15,000,000	203,969	110,000,000	150,423
25,000,000	250,162	117,610,000	447

Note: These estimates are calculations using the Household Total (or White) a and b parameters from Table 4.

Table 7. Base Standard Errors of Estimated Numbers of Persons			
Size of Estimate	**Standard Error**	**Size of Estimate**	**Standard Error**
200,000	27,050	110,000,000	504,705
300,000	33,124	120,000,000	513,038
500,000	42,749	130,000,000	518,886
750,000	52,334	140,000,000	522,333
1,000,000	60,405	150,000,000	523,426
2,000,000	85,282	160,000,000	522,180
3,000,000	104,273	170,000,000	518,578
5,000,000	134,161	180,000,000	512,570
7,500,000	163,614	190,000,000	504,070
10,000,000	188,114	200,000,000	492,950
15,000,000	228,393	210,000,000	479,027
25,000,000	289,623	220,000,000	462,048
30,000,000	314,361	230,000,000	441,659
40,000,000	356,191	240,000,000	417,363
50,000,000	390,480	250,000,000	388,426
60,000,000	419,085	260,000,000	353,712
70,000,000	443,106	270,000,000	311,292
80,000,000	463,258	275,000,000	286,149
90,000,000	480,028	280,000,000	257,387
100,000,000	493,761	299,340,000	4,636

Notes: (1) These estimates are calculations using the Other Persons 0+ a and b parameter from Table 4.

(2) To calculate the standard for another domain multiply the standard error from this table by the appropriate f factor from Table 4.

Table 8. Base Standard Errors for Percentages of Households or Families

Base of Estimated Percentages	Estimated Percentages					
	≤ 1 or ≥ 99	2 or 98	5 or 95	10 or 90	25 or 75	50
200,000	1.25%	1.77%	2.75%	3.78%	5.46%	6.30%
300,000	1.02%	1.44%	2.24%	3.09%	4.46%	5.15%
500,000	0.79%	1.12%	1.74%	2.39%	3.45%	3.99%
750,000	0.65%	0.91%	1.42%	1.95%	2.82%	3.26%
1,000,000	0.56%	0.79%	1.23%	1.69%	2.44%	2.82%
2,000,000	0.40%	0.56%	0.87%	1.20%	1.73%	1.99%
3,000,000	0.32%	0.46%	0.71%	0.98%	1.41%	1.63%
5,000,000	0.25%	0.35%	0.55%	0.76%	1.09%	1.26%
7,500,000	0.20%	0.29%	0.45%	0.62%	0.89%	1.03%
10,000,000	0.18%	0.25%	0.39%	0.53%	0.77%	0.89%
15,000,000	0.14%	0.20%	0.32%	0.44%	0.63%	0.73%
25,000,000	0.11%	0.16%	0.25%	0.34%	0.49%	0.56%
30,000,000	0.10%	0.14%	0.22%	0.31%	0.45%	0.51%
40,000,000	0.09%	0.12%	0.19%	0.27%	0.39%	0.45%
50,000,000	0.08%	0.11%	0.17%	0.24%	0.35%	0.40%
60,000,000	0.07%	0.10%	0.16%	0.22%	0.32%	0.36%
70,000,000	0.07%	0.09%	0.15%	0.20%	0.29%	0.34%
80,000,000	0.06%	0.09%	0.14%	0.19%	0.27%	0.32%
90,000,000	0.06%	0.08%	0.13%	0.18%	0.26%	0.30%
105,000,000	0.05%	0.08%	0.12%	0.17%	0.24%	0.28%
110,000,000	0.05%	0.08%	0.12%	0.16%	0.23%	0.27%
117,610,000	0.05%	0.07%	0.11%	0.16%	0.23%	0.26%

Note: These estimates are calculations using the Households Total (or White) b parameter from Table 4.

Table 9. Base Standard Errors for Percentages of Persons

Base of Estimated Percentages	Estimated Percentages					
	≤ 1 or ≥ 99	2 or 98	5 or 95	10 or 90	25 or 75	50
200,000	1.35%	1.89%	2.95%	4.06%	5.86%	6.76%
300,000	1.10%	1.55%	2.41%	3.31%	4.78%	5.52%
500,000	0.85%	1.20%	1.86%	2.57%	3.71%	4.28%
750,000	0.70%	0.98%	1.52%	2.10%	3.03%	3.49%
1,000,000	0.60%	0.85%	1.32%	1.82%	2.62%	3.03%
2,000,000	0.43%	0.60%	0.93%	1.28%	1.85%	2.14%
3,000,000	0.35%	0.49%	0.76%	1.05%	1.51%	1.75%
5,000,000	0.27%	0.38%	0.59%	0.81%	1.17%	1.35%
7,500,000	0.22%	0.31%	0.48%	0.66%	0.96%	1.10%
10,000,000	0.19%	0.27%	0.42%	0.57%	0.83%	0.96%
15,000,000	0.16%	0.22%	0.34%	0.47%	0.68%	0.78%
25,000,000	0.12%	0.17%	0.26%	0.36%	0.52%	0.61%
30,000,000	0.11%	0.15%	0.24%	0.33%	0.48%	0.55%
40,000,000	0.10%	0.13%	0.21%	0.29%	0.41%	0.48%
50,000,000	0.09%	0.12%	0.19%	0.26%	0.37%	0.43%
60,000,000	0.08%	0.11%	0.17%	0.23%	0.34%	0.39%
70,000,000	0.07%	0.10%	0.16%	0.22%	0.31%	0.36%
100,000,000	0.06%	0.08%	0.13%	0.18%	0.26%	0.30%
110,000,000	0.06%	0.08%	0.13%	0.17%	0.25%	0.29%
120,000,000	0.05%	0.08%	0.12%	0.17%	0.24%	0.28%
130,000,000	0.05%	0.07%	0.12%	0.16%	0.23%	0.27%
140,000,000	0.05%	0.07%	0.11%	0.15%	0.22%	0.26%
150,000,000	0.05%	0.07%	0.11%	0.15%	0.21%	0.25%
160,000,000	0.05%	0.07%	0.10%	0.14%	0.21%	0.24%
170,000,000	0.05%	0.06%	0.10%	0.14%	0.20%	0.23%
180,000,000	0.04%	0.06%	0.10%	0.14%	0.20%	0.23%
190,000,000	0.04%	0.06%	0.10%	0.13%	0.19%	0.22%
200,000,000	0.04%	0.06%	0.09%	0.13%	0.19%	0.21%
210,000,000	0.04%	0.06%	0.09%	0.13%	0.18%	0.21%
220,000,000	0.04%	0.06%	0.09%	0.12%	0.18%	0.20%
230,000,000	0.04%	0.06%	0.09%	0.12%	0.17%	0.20%
240,000,000	0.04%	0.05%	0.09%	0.12%	0.17%	0.20%
250,000,000	0.04%	0.05%	0.08%	0.11%	0.17%	0.19%
280,000,000	0.04%	0.05%	0.08%	0.11%	0.16%	0.18%
299,340,000	0.03%	0.05%	0.08%	0.10%	0.15%	0.17%

Notes: (1) These estimates are calculations using the Other Persons 0+ a and b parameter from Table 4.

(2) To calculate the standard for another domain multiply the standard error from this table by the appropriate f factor from Table 4.

Table 10. Distribution of Monthly Cash Income Among People 25 to 34 Years Old
(Not Actual Data, Only Use for Calculation Illustrations)

	Interval of Monthly Cash Income												
	Under $300	$300 to $599	$600 to $899	$900 to $1,199	$1,200 to $1,499	$1,500 to $1,999	$2,000 to $2,499	$2,500 to $2,999	$3,000 to $3,499	$3,500 to $3,999	$4,000 to $4,999	$5,000 to $5,999	$6,000 and Over
Number of People in Each Interval (in thousands)	1,371	1,651	2,259	2,734	3,452	6,278	5,799	4,730	3,723	2,519	2,619	1,223	1,493
Cumulative Number of People with at Least as Much as Lower Bound of Each Interval (in thousands)	39,851 (Total People)	38,480	36,829	34,570	31,836	28,384	22,106	16,307	11,577	7,854	5,335	2,716	1,493
Percent of People with at Least as Much as Lower Bound of Each Interval	100	96.6	92.4	86.7	79.9	71.2	55.5	40.9	29.1	19.7	13.4	6.8	3.7