# Replicate Weights for the 1996 SIPP Panel are now available through the Ferret system on the Internet!

Users can now access replicate weights for estimating variances of survey estimates for the 1996 SIPP Panel.  This site provides details on how users can download the replicate weights and estimate variances using these weights.

## SIPP SPECIFICS

The SIPP sample is interviewed in waves.  Waves are four-month periods during which the full sample is interviewed.  There are currently twelve waves in the 1996 panel.  Approximately one quarter of the full sample is interviewed in each month within a wave (the SIPP sample is split this way to ease interviewer burden).   These quarter-samples, called rotation groups,  are interviewed every four months, once per wave, for all twelve waves.

SIPP interviews gather data covering the four months immediately preceding the interview month. These preceding months are called reference months.  For example, the four reference months corresponding to July interviews are March, April, May, and June.  Table 1 shows the wave, rotation group, interview month, and reference month associations for all waves of the 1996 panel (Note: the time period covered by an interview is different for each rotation group):

**Table 1.  Wave, Rotation Group, Interview Month, and Reference Month for the 1996 SIPP Panel**

| Wave | Rotation Group | Interview Month | Reference Months and Corresponding Number | | | |
|------|------|------|------|------|------|------|
| | | | 1 | 2 | 3 | 4 |
| 1 | 1 | April '96 | December '95 | January '96 | February '96 | **March '96** |
| | 2 | May '96 | January '96 | February '96 | **March '96** | **April '96** |
| | 3 | June '96 | February '96 | **March '96** | **April '96** | **May '96** |
| | 4 | July '96 | **March '96** | **April '96** | **May '96** | **June '96** |
| 2 | 1 | August '96 | **April '96** | **May '96** | **June '96** | **July '96** |
| | 2 | September '96 | **May '96** | **June '96** | **July '96** | **August '96** |
| | 3 | October '96 | **June '96** | **July '96** | **August '96** | **September '96** |
| | 4 | November '96 | **July '96** | **August '96** | **September '96** | **October '96** |
| 3 | 1 | December '96 | **August '96** | **September '96** | **October '96** | **November '96** |
| | 2 | January '97 | **September '96** | **October '96** | **November '96** | **December '96** |
| | 3 | February '97 | **October '96** | **November '96** | **December '96** | **January '97** |
| | 4 | March '97 | **November '96** | **December '96** | **January '97** | **February '97** |

| | | | | | |
|---|---|---|---|---|---|
| 4 | 1 | April '97 | **December '96** | **January '97** | **February '97** | **March '97** |
| | 2 | May '97 | **January '97** | **February '97** | **March '97** | **April '97** |
| | 3 | June '97 | **February '97** | **March '97** | **April '97** | **May '97** |
| | 4 | July '97 | **March '97** | **April '97** | **May '97** | **June '97** |
| 5 | 1 | August '97 | **April '97** | **May '97** | **June '97** | **July '97** |
| | 2 | September '97 | **May '97** | **June '97** | **July '97** | **August '97** |
| | 3 | October '97 | **June '97** | **July '97** | **August '97** | **September '97** |
| | 4 | November '97 | **July '97** | **August '97** | **September '97** | **October '97** |
| 6 | 1 | December '97 | **August '97** | **September '97** | **October '97** | **November '97** |
| | 2 | January '98 | **September '97** | **October '97** | **November '97** | **December '97** |
| | 3 | February '98 | **October '97** | **November '97** | **December '97** | **January '98** |
| | 4 | March '98 | **November '97** | **December '97** | **January '98** | **February '98** |
| 7 | 1 | April '98 | **December '97** | **January '98** | **February '98** | **March '98** |
| | 2 | May '98 | **January '98** | **February '98** | **March '98** | **April '98** |
| | 3 | June '98 | **February '98** | **March '98** | **April '98** | **May '98** |
| | 4 | July '98 | **March '98** | **April '98** | **May '98** | **June '98** |
| 8 | 1 | August '98 | **April '98** | **May '98** | **June '98** | **July '98** |
| | 2 | September '98 | **May '98** | **June '98** | **July '98** | **August '98** |
| | 3 | October '98 | **June '98** | **July '98** | **August '98** | **September '98** |
| | 4 | November '98 | **July '98** | **August '98** | **September '98** | **October '98** |
| 9 | 1 | December '99 | **August '98** | **September '98** | **October '98** | **November '98** |
| | 2 | January '99 | **September '98** | **October '98** | **November '98** | **December '98** |
| | 3 | February '99 | **October '98** | **November '98** | **December '98** | **January '99** |
| | 4 | March '99 | **November '98** | **December '98** | **January '99** | **February '99** |
| 10 | 1 | April '99 | **December '98** | **January '99** | **February '99** | **March '99** |
| | 2 | May '99 | **January '99** | **February '99** | **March '99** | **April '99** |
| | 3 | June '99 | **February '99** | **March '99** | **April '99** | **May '99** |
| | 4 | July '99 | **March '99** | **April '99** | **May '99** | **June '99** |

| | | | | | | |
|---|---|---|---|---|---|---|
| 11 | 1 | August '99 | **April '99** | **May '99** | **June '99** | **July '99** |
| | 2 | September '99 | **May '99** | **June '99** | **July '99** | **August '99** |
| | 3 | October '99 | **June '99** | **July '99** | **August '99** | **September '99** |
| | 4 | November '99 | **July '99** | **August '99** | **September '99** | **October '99** |
| 12 | 1 | December '99 | **August '99** | **September '99** | **October '99** | **November '99** |
| | 2 | January 2000 | **September '99** | **October '99** | **November '99** | December '99 |
| | 3 | February 2000 | **October '99** | **November '99** | December '99 | January 2000 |
| | 4 | March 2000 | **November '99** | December '99 | January 2000 | February 2000 |

This document applies to estimating variances using replicate weights for survey estimates for the months highlighted in bold in Table 1. These are the months for which full-sample data is currently available, or will be coming available as more data is collected and released.

## SIPP VARIANCE ESTIMATION METHOD

Users should use Fay's method for estimating variances for the 1996 SIPP Panel. Fay's method is a modified form of the basic balanced repeated replication method (BRR) of variance estimation. An advantage Fay's method has over the BRR is the inclusion of both half-samples in each replicate whereas the BRR includes only one-half of the sample. For 0 # $k$ < 1, the modification produces replicate estimates by weighting observations by the factor $k$ or 2-$k$ in place of the 0 or 2 used in the original version of half-sample replication. With $k$ > 0, all replicate totals will be positive for totals that are positive in the full sample, unlike half-sample replication. This feature has practical advantages when working with nonlinear estimators, such as weighted totals from the SIPP which are nonlinear as a result of complex ratio estimation included in the SIPP weighting process.

The variance formulas for Fay's method and the BRR differ in that the formula for Fay's method includes a perturbation factor (1-k). The perturbation factor takes into account the use of both half-samples in each replicate. The variance formula for Fay's method is:

$$\text{Var}(\textbf{2}_0) = (1 / (G (1 - k)^2)) \; E_{i=1}^{G} (\textbf{2}_i - \textbf{2}_0)^2 \qquad \textbf{(1)}$$

where  G  =  number of replicates
1-k  =  perturbation factor
i  =  replicate i,  i = 1 to G
$\textbf{2}_i$  =  ith estimate of the parameter **2** based on the observations included in the ith replicate
$\textbf{2}_0$  =  survey estimate of the parameter **2** based on the full sample

The 1996 SIPP Panel uses 108 replicate weights which are calculated based on a perturbation factor of 0.5 (1-k = 0.5, where k = 0.5). Inserting these values into (1) results in the SIPP variance formula of:

$$\mathrm{Var}(\theta_0) = (1 / (108 * (0.5)^2 )) \; \mathbf{E}_{i=1}^{108} (\theta_i - \theta_0)^2 \tag{2}$$

For more details on the BRR and Fay's method, see "User's Guide to WestvarPC", 1997 (this document can be downloaded from the internet at http://www.westat.com/wesvarpc/files /wesvar-2.12/weszip2.exe.  The applicable file is appena.pdf).

NOTE:        The Census Bureau used VPLX software (see below) to compute the replicate weights that are available on the internet.

## ACCESSING THE REPLICATE WEIGHTS

To access the 1996 replicate weights and SIPP data files, follow the procedure outlined below:

1.    Go to http://ferret.bls.census.gov/cgi-bin/ferret.  This address accesses the Federal Electronic Research and Review Extraction Tool (FERRET) web site.  Enter your E-mail address in the box and click on "Continue"

2.    The next screen lists Current Population Survey and SIPP data files followed by options for the type of data desired.  Click on the circle next to "Survey of Income and Program Participation -- 1996 Panel."  Then click on "Continue."

3.    Data users are prompted for data type options on the next screen.  Click on the circle next to the statement "I want to create my own tables or extract my own data files, i.e. microdata."  Then click on "Continue."

      Note: users accessing FERRET for the first time will get a one-time screen asking them if they're sure they want to create their own tables.  Click on "Continue.").

4.    On the fourth screen, users can begin the process of extracting data from the 1996 SIPP data file.

      a.    In the upper left-hand corner of the screen, select the wave or waves that contain the data of interest.  These data can come from any month in which full sample data was collected.  Note that data for some months fall in two separate waves, so you may need to get replicate weights from 2 waves in order to obtain a full month's data.  Table 1 will help you determine which months fall into which waves.

            For users interested in downloading data from multiple waves, multiple waves can be accessed at the same time by following steps i and ii:

            i.    click on one wave;
            ii.   click on the other waves while simultaneously pressing the CTRL key.

      b.    Go to the next box on the screen.  This box prompts data users to enter words that are used to search against variable names and descriptions.  Enter in the box:

            i.    the variables for which you would like to calculate variances.

        ii.      the following variables (individual words need to be typed once, only): wave, rotation group, reference month, calendar month, calendar year, suid (unit ID), person number, person final weight.

        iii.     the word "replicate."

    c.     The next prompt asks you to choose what kind of search you want. Both options are okay, so click on either option.

    d.     Next, a box shows variable groupings. Users have the option to subset on these variable groupings. To subset on the variable groupings, click on the variable grouping(s) of interest. To subset on two or more variable groupings, select the additional groupings by clicking on the grouping while simultaneously pressing on the CTRL key.

    e.     Go to the last section on the screen. This section prompts data users to select the types of data they want. Click on Edited, Recodes, Weights, Topcodes, Sample Control, Geography, and Replicate Weights. Unedited and Flags are optional.

    f.     Click on "Continue."

5.    The next screen is titled "Ferret Select Variables." A box containing a list of variables should appear on this screen. These are the variables whose labels contain any of the words that were entered on the previous screen. Select the variables of interest:

    NOTE:     Make sure you select calendar month, calendar year and each of the 108 replicate weights (repwt1-repwt108).

    a.     Click on the first variable. This variable should be highlighted after selection.

    b.     Select each subsequent variable by clicking on the variable while simultaneously pressing the CTRL key.

    c.     Click on "Continue" when selection is complete.

6.    "Select Variable Values" is the title of the sixth screen. Users can subset the data using any of the variables that were selected on the "Ferret Select Variables" screen. Users who are interested in specific months' data can subset on calendar month and calendar year.

    a.     Select the month(s) of interest by subsetting on calendar month:

        i.      Go to calendar month (RHCALMN).
        ii.     Click on a month of interest. This month should be highlighted after selection.
        iii.    Select each additional calendar month by clicking on calendar month while simultaneously pressing the CTRL key.

    b.     Select the year of interest by subsetting on calendar year:

        i.      Go to calendar year (RHCALYR).

        ii.      Enter the year of interest in both boxes.

    c.      Select values on which to subset for any additional variables.

    d.      After making your selections, go to the next section.  This section has the prompt "Choose the output you want."  Select the type of data set (ASCII or SAS) you want by clicking on the appropriate circle.

    e.      Go to the bottom of the screen, under the "Continue" and "Cancel choices" buttons.  Click on the first circle ("Selected values above").

    f.      Click on "Continue."

7.      Next is a screen titled "Verify Your Selections."  A list of the variables (items) and the subsetted ranges you selected are presented at the top of the screen.   Underneath these lists users are prompted to select an output type.  Create a file to download from the internet that contains the selected variables and records meeting the subsetted range constraints:

    a.      Click on the square next to "Download Resulting File."

    b.      Click on a circle for type of record format (ASCII files only) and type of file compression desired.

    c.      Go to the bottom of the screen and click on "Get Results."

8.      Finally, the last screen presents data users with download options.  Click on the option of interest (option titles and the number of options are different between ASCII and SAS data sets).  Click on the last option ("Download custom code book") for descriptions of values for each variable on the file.  Click on "File" and "Save As" from the menu bar to create and download a file containing the value descriptions.

    NOTE:      After clicking on the option to download the file, users may receive a message box with the title "Unknown File Type."  If this happens, click on the "Save File" button to continue downloading the data set.

## SOFTWARE PACKAGES FOR COMPUTING REPLICATE VARIANCES

1.      VPLX

    VPLX is a programming language that is designed to compute variance estimates and related statistics for complex samples using replication methods.  Input files must be in ASCII format.  All input file data must be numeric.  VPLX reads in blank fields as zeros.  VPLX documentation can be found on the INTERNET at http://www.census.gov/sdms/www/vwelcome.html.

2.      SAS

SAS software must be purchased through the SAS Institute. A sample SAS program for computing variance estimates using Fay's method is provided below.

a. Read in the ASCII file or SAS dataset that contains the full person weight, replicate weights, and variables of interest before implementing the following code:

b. The PROC MEANS sums the full person weight and replicate weights for one characteristic of interest. Fullweight represents the survey estimate, which, in this example, is an estimated total.

```
PROC MEANS DATA = <libref>.<filename1> SUM NOPRINT;
  WHERE statement (if applicable);
  VAR fullweight repwt-repwt108;
  OUTPUT OUT = <filename2> SUM=fw rw1-rw108;
```

c. The data step computes the variance using Fay's method for the estimated total computed in the PROC MEANS above. The variables **fw**, **var** and **se** represent the estimated total, variance and standard error, respectively.

```
DATA <filename2>  (KEEP= fw var se);
  SET <filename2> END=eof;
  IF _N_ = 1 THEN ss = 0;
  ARRAY repwts{1-109} fw rw1-rw108;

  DO I = 2 TO 109;
    ss = ss + (repwts{i} - repwts{1})**2;
  END;

  IF eof THEN DO;
    var = (4/108) * ss;          /*  This equals formula (2)    */
    se = (var)**0.5;
    OUTPUT;
  END:
```

3. WestVarPC

WestVarPC is a software package that computes variance estimates using replicate weights. For the 1996 SIPP Panel, users should use Fay's method with 108 replicates and k=0.5.

WestVarPC is located on the INTERNET at http://www.westat.com/wesvar/index.html.