

When working with SCF data, standard error calculations can overestimate the reliability of regressions and other statistics unless two other kinds of error are accounted for: imputation error and sample variability error. As an example of what effect the failure to account for these types of errors can have, take a look at the p values for the following three regressions:

Regression estimates of correlation between income and homeownership among families with head aged 25-40

	(1)	(2)	(3)
I(Homeowner)	0.144 (0.037) p=0.000	0.144 (0.084) p=0.088	0.144 (0.089) p=0.106
N	6,940	1,388	1,388

Note: OLS regression estimates of log of family income on left-hand side and homeowner indicator variable on right hand side. Includes controls for education, age, and mortgage indicator variable. Families with head of household aged 25-40 only. Estimated standard errors are in parentheses. Column (1) shows standard error without correction for multiple imputation or sampling variance. Column (2) shows an estimated standard error with a correction for multiple imputation [from Stata ado *micombine*], but still no correction for sampling variance, and column (3) shows an estimated standard error with a correction for multiple imputation and the sampling variance [from Stata ado *scfcombo*].

Adjusting for these both types of error caused the p value to increase from 0 to over 0.1, which substantially alters the statistical significance of the result. Note that adjusting for imputation error will always increase the standard error, adjusting for sampling error may increase or decrease the standard error.

To account for these errors, the downloadable Replicate Weight Files are required.

1) Replicate Weights

- a. We construct replicate weights to account for sampling error.
- b. Following bootstrapping theory, we take 999 samples from the SCF data in a way that mirrors the original sampling process.
 - i. In any given bootstrap replicate sample, an SCF family could be sampled zero times or multiple times.
- c. Each time the data is resampled, a new version of the data is created, providing a sample that *could* have been drawn originally.
 - i. Households may get a different weight each time a sample is drawn.
 - ii. The weights sum to the total US population, similar to the base SCF weights.
- d. Because each bootstrap replicate sample will have slightly different composition, taken together they will provide a distribution for each variable.
 - i. We use this distribution to estimate sampling variability.
- e. We estimate sampling variability only on the first implicate.

2) Defining the two types of error in greater detail

- a. Imputation Error
 - i. Missing data in the SCF are imputed 5 times, meaning that each SCF family has 5 separate observations (called “implicates”) in the final data.
 - ii. The estimated standard error of a statistic is smaller than the true standard error when estimated without correcting for multiple imputation.
 - iii. We estimate imputation variance through the MISECOMP macro in SAS or the MICOMBINE .ado file in Stata.
 - 1. A description and the SAS program for MISECOMP can be found in the [SCF Codebook](#) under MACRO MISECOMP.
 - 2. The user-written .ado file, MICOMBINE for Stata is available [here](#). An example and additional detail are found in the [SCF Codebook](#).
 - 3. Both find the variance or estimates across the 5 implicates, corrected by $(1+[1/5])$.
 - 4. Both are appropriate for simple statistics (such as means) or more complicated estimates (such as regression coefficient estimates).
 - b. Sample Variability / Error
 - i. Results from using a sample, rather than a population.
 - ii. Estimated standard errors can be further corrected to include an estimate of sampling variability. Correcting for sampling variability may increase or decrease the standard errors.
 - iii. We estimate sampling variability by a set of bootstrap replicate weights described below.
- 3) Programs for calculating Standard Errors using the Replicate Weight files
- a. [Stata Module to Estimate Errors using the Survey of Consumer Finances](#)
 - i. The user-written .ado file, SCFCOMBO, for Stata “calculates and combines the imputation uncertainty and bootstrapped standard errors for estimation commands run on the Survey of Consumer Finances (SCF).”
 - b. SAS Macro from the [SCF Codebook](#)
 - i. The macro can be found in the SAMPLING ERROR section, and is named MEANIT.
 - c. Either program will incorporate both imputation and sampling variability estimates, and require the download of the Replicate Weight files.

File*	Last Updated	Documentation
Replicate weights--X42001: CPORT version (35 MB ZIP)	9/4/2014	
Replicate weights--X42001: COPY/EXPORT version (35 MB ZIP)	9/4/2014	Codebook (2.6 MB ASCII)
Replicate weights--X42001: Stata version (35 MB ZIP)	9/4/2014	