

codebook for replicating/extending SCF cross sections

emily case

July 15, 2022

This file explains variable definitions, file organization, etc. for replicating and extending the “interactive chart” on the Federal Reserve’s website. The interactive chart does not include cross-sections, and that is the main contribution of this code, but it also demonstrates how to calculate the means and medians and their standard errors. These calculations are not trivial, because the Survey of Consumer Finances (SCF) data includes (1) imputations and (2) weights.

1 main variables for analysis

These variables are all in 2019 dollars, and are the main variables of interest. Each Excel file is titled with the main variable name, and the sheets within an Excel file are the cross-sectional variables.

| | |
|-----------------|---|
| retqliq | Total value of quasi-liquid held by household. Includes IRAs, Keoghs, thrift-type accounts, and future and current account-type pensions. |
| houses | Total value of primary residence of household. Excludes the part of a farm or ranch used in a farming or ranching business. |
| asset | Total value of assets held by household; the sum of financial and nonfinancial assets. |
| debt | Total value of debt held by the household. Includes principal residence debt (mortgages and HELOCs), other lines of credit, debt for other residential property, credit card debt, installment loans, and other debt. |
| networth | Total net worth of household; the difference between assets and debt. See definitions of assets and debt for further clarification. |
| bus | Total value of business(es) in which the household has either an active or non-active interest. Businesses include both actively and nonactively-managed business(es). Value of active business(es) calculated as net equity if business(es) were sold today, plus loans from the household to the business(es), minus loans from the business(es) to the household not previously reported, plus value of personal assets used as collateral for business(es) loans that were reported earlier. Value of nonactive business(es) is calculated as the market value of the business(es). |

2 cross-sectional variables

These variables were used to create the cross-sectional statistics, and are categorical variables. Please note that the last three variables (which start with ‘h’) are used in conjunction with the corresponding main variable. The interactive chart on the Federal Reserve’s website conditions results on whether or not families have the variable in question. For example, rather than taking the average home value of all respondents, they take the average home value of respondents who own a home.

| | |
|-----------------------|--|
| <code>agegroup</code> | Age group of the reference person: 18-34, 35- 54, and 55 or older. |
| <code>racecl4</code> | Class of race of respondent: White, non-mixed race; Black/African-American, non-mixed race; Hispanic/Latino, non-mixed race; Other. |
| <code>racecl</code> | Binary class of race of respondent: White, non-Hispanic; nonwhite or Hispanic. |
| <code>occat1</code> | Occupation categories for reference person: Work for someone else; self-employed/partnership; retired/disabled + (student/homemaker/misc. not working and age 65 or older); other groups not working (mainly those under 65 and out of the labor force). |
| <code>edugroup</code> | Education level categories for reference person: Less than a high school degree; high school degree; some college or associates degree; bachelors degree or graduate school. |
| <code>hbus</code> | Respondent has (or does not have) active or nonactively managed business(es). |
| <code>hretqliq</code> | Respondent has (or does not have) any type of account that would be included in the <code>retqliq</code> variable. |
| <code>hhouses</code> | Respondent owns (or does not own) a house. |

3 files and code organization

Each Excel file corresponds to one of the main variables, and the sheets within the file correspond to different cross-sections of the data. Each Excel file is created using 600 iterations of the Bootstrapping method, as is recommended by the SCF codebook in order to have accurate standard deviations. The means and their standard errors, the medians and their standard errors, and observation counts are reported. *Note that for some cross-sections, there were not enough observations. This is represented by a value of -999 in the Excel file.*

3.1 code and data files

All code was run in Stata.

Stata do files:

- `RUN.do` is the “home base” file, and should run the entire program. In this file, users can change (1) the main variables list, which tells Stata which main variables (that are in dollar amounts) to run analyses on, and (2) the iteration amount. Those just wishing to experiment with the code should choose a very small iteration amount, such as 10. This code takes a very long time to run!!
- `0_load.do` loads the data, cleans it, and creates variables and their value labels.
- `output.do` acts as a giant function. It receives the main variable from the `RUN` file, calculates all of the results, and automatically outputs the results to a formatted Excel file.
- `mdn.do` is a smaller function which calculates means, medians, and standard errors for a particular subset of data. For example, it calculates the statistics for respondents with `racecl4 = 1` and `agegroup = 2`. This is the file that takes so long. Statistics are calculated using the Stata command `scfcombo`,¹ which needs to be installed.
- `excelfmt.do` formats the excel files to look more readable.

Data files downloaded from SCF:

- `rscfp2019.dta` contains all the variables for analysis.
- `p19_rw1.dta` contains the weights for the bootstrap method. *Note: this file is needed in order to use the Stata command `scfcombo`!!*

4 Helpful links/citations

- The SCF data can be downloaded [here](#).
- The SCF interactive chart used to check calculations is [here](#).
- The SCF variable descriptions can be found [here](#).
- The Stata package for `scfcombo` can be found [here](#)
- This [website](#) was helpful to learn about sampling weights.

¹Karen Pence, 2015. “SCFCOMBO: Stata module to estimate errors using the Survey of Consumer Finances,” Statistical Software Components S458017, Boston College Department of Economics, revised 04 Oct 2015.