Data Appendix

Integrated Household Surveys: An Assessment of U.S. Methods and Innovation

Jason Premo (Federal Reserve Bank of Boston)

Giri Subramaniam (Boston College)[1]

May 2017

[1] Special thanks to Stacy Carlson, Mi Luo, David Zhang, and Taylor Wilson (Bureau of Labor Statistics), whose help and assistance were indispensable to the final results.

Table of Contents

1. Introduction

Comparing estimates from the various different U.S. surveys is fraught with challenges. Different surveys cover different assets, liabilities, and types of income or expenditures. Even when surveys do cover the same topics, questions are rarely (if ever) worded the same, and even slight differences in measurement can confound attempts to compare estimates.

The estimates in the tables of *Integrated Household Surveys: An Assessment of U.S. Methods and Innovation* represent our best attempt to compare the various U.S. surveys. The details of these calculations are described, in qualitative terms, in this document. Those who would like more detail as to the creation of these figures should consult the Stata programs, which are made available on the Federal Reserve Bank of Boston website.

Note that estimates from the Townsend Thai Monthly Survey (TTMS) were provided wholesale by the TTMS team. We therefore do not provide any details of those calculations here. Those who are interested in the construction of the TTMS estimates are encouraged to read the TTMS Users' Manual. A link to this document is provided in the "Codebooks and Data Dictionaries" section below.

2. Notes on the 2012 SCPC and DCPC

2.1 Income imputation in the SCPC

The SCPC does not directly ask for respondent household income, instead asking respondents to report their income range. The question is reproduced below:

- (DE010) Which category represents the total combined **income** of all members of your family living here during the **past 12 months**?

  This includes money from jobs, net income from business, farm or rent, pensions, dividends, interest, social security payments and any other money income received by members of your family who are 15 years of age or older.
  - 1 Less than $5,000
  - 2 $5,000–$7,499
  - 3 $7,500–$9,999
  - 4 $10,000–$12,499
  - 5 $12,500–$14,999
  - 6 $15,000–$19,999
  - 7 $20,000–$24,999
  - 8 $25,000–$29,999
  - 9 $30,000–$34,999
  - 10 $35,000–$39,999
  - 11 $40,000–$49,999
  - 12 $50,000–$59,999
  - 13 $60,000–$74,999
  - o 14 $75,000–$99,999

- 15 $100,000–$124,999
- 16 $125,000–$199,999
- 17 $200,000 or more

In order to produce specific point estimates from which to calculate a mean, a simple imputation process was used. The process uses OLS and the SCF data to predict (logarithmic) income based on sex, age, marital status, race, education, and the above-mentioned income buckets.

$$\log(income)_i = sex_i + age_i + married_i + race_i + education_i + inc\_bucket_i + \varepsilon$$

These specific demographic categories were chosen because they exist in compatible forms in both surveys. This imputation results in a left-skewed income distribution relative to the SCF; some of this is certainly due to the imperfect nature of the imputation, and some is due to differences in sampling between the two surveys. To show that the ultimate results are not unduly influenced by outliers, we report both median and censored incomes (average income less than $200,000) along with regular income.

2.2 Household Adjustment

The CPC samples U.S. consumers, while the other surveys sample groups of consumers (households, family units, or primary economic units). It is therefore necessary to convert per consumer estimates to per household estimates in order to make estimates from the SCPC and DCPC comparable to estimates from other surveys. To this end, per consumer estimates are adjusted by a conversion factor: the total Q4:2012 18+ non-institutional U.S. population divided by the Q4:2012 number of U.S. households.

$$\frac{235036000}{114960000} \approx 2.0445$$

The household estimate is from the Housing Vacancy Survey and the population estimate is from the Current Population Survey, both via Haver Analytics. Note that, for some questions—most notably the household and liability questions described above—this conversion is not necessary as the question already asks for a household-level response.

3. Notes on Other Surveys

3.1 Panel Study of Income Dynamics (PSID)

- For many categories, the PSID allows respondents to report a value for the head of household and a spouse. The "household" values reported in the balance sheet are the sum of these two variables, when available. For instance, the wage income in Table 3 is the head of household's wage income plus the spouse's wage income. Note that not all variables have both a head of household and spouse version—for instance, bonuses and overtime are only reported for the head of household.

3.2 Consumer Expenditures Survey (CES)

- The CES data are divided into several different datasets, depending on which calculations we wish to make. This means that, in general, the CES calculations are more "cobbled together" than the calculations for the other surveys. The FMLI files are used for most calculations, but in some cases the OPI, FNA, OPH, HEL, OVB, or MOR files are used. Many of these asset and liability data do not contain any imputed values. See *Consumer Expenditure Interview Survey Public Use Microdata: 2012 Users' Documentation* (2013) for more information about the distinctions between these files.
- Median assets and liabilities are not reported for the CES. This is because various files are necessary to produce the estimates of the individual line items, and the total estimates (that is, total assets) are produced by summing these smaller categories. No individual family is included in every file, meaning there are no total estimates for any single family, and so no median can be calculated.
- The CES data are organized in a large number of different files containing different families and different data. This particular organization method sometimes makes it difficult to determine the meaning of a missing data value. In some files, a missing value indicates that the respondent refused to answer a given question. In other files, a missing value can sometimes mean that the respondent was never asked the question. For the purposes of generating means, it is important that respondents without a given balance-sheet item are given a value of zero, while those who were never asked are excluded. The CES estimates in the table are the best estimates we could construct after consulting with members of the CES team at the Bureau of Labor Statistics to determine the meaning of missing values in different situations.

3.3 Survey of Consumer Finances (SCF)

- The SCF sampling unit is the Primary Economic Unit (PEU), which is slightly less inclusive than a household; according to the SCF 2013 documentation, "[t]he great majority of the time, the PEU and the household are identical".
- For some asset holdings, respondents are asked both market value and face value. We use market value in all cases where it is available.
- For many questions, respondents are asked to report multiple asset/liability accounts, and in these cases all reported accounts are simply added together. Some kinds of property holdings are adjusted by the respondent's self-reported ownership share.

3.4 Health and Retirement Study (HRS)

- As mentioned elsewhere, the HRS is a sample of older Americans. As such, we should generally expect the HRS to have a notably different composition of assets, liabilities, and income than the other surveys, as the sample is composed of individuals who are in a later part of their life cycle.
- Asset variables in the HRS are often net rather than gross values. Where possible, net values are made gross by adding back in the associated liabilities.

3.5 Survey of Income and Program Participation (SIPP)

- The SIPP samples by household. The head and "spouse" of the household are then asked to report both individual and joint assets, liabilities, and income. Individual assets, liabilities, and income are summed within the household to create a household estimate. Joint assets, liabilities, and income are averaged within the household to create a household estimate. Household weights are provided and used.
- The SIPP is administered in panels within which the entire sample is interviewed at 4-month intervals commonly referred to as *waves*. In order to obtain estimates for a year, we multiply these reported amounts by three to get annual figures.
- The SIPP is divided into *core* questions that are asked every time and *topical* questions that vary across waves. Specifically, we use Wave 10 from the core and the topical module, both of which were administered in September–December 2011.

4. Notes on Tables and Figures

4.1 Table 2-a: Balance Sheets – Assets

Currency

SCF: The currency estimate comes from option 63 (Cash, not elsewhere categorized) in questions X4020, X4024, and X4028, which allows respondents to choose any additional assets that they did not mention elsewhere. Given the structure of the survey, it is unlikely that people with very small cash holdings chose to report them. In the public dataset, this coding usually reserved for just currency also includes PayPal and stored value cards.

Bank accounts

PSID: Includes checking or savings accounts, money market funds, certificates of deposit (CD), government bonds, and treasury bills. CDs, bonds, and T-bills do not belong in this category, but there is no way to disaggregate them from the data.

HRS: Includes checking, savings, and money market accounts.

SIPP: Includes only non-interest-earning checking accounts.

Other deposit accounts

SIPP: Includes interest-earning accounts.

Certificates of deposit

HRS: Includes government bonds and T-bills.

Bonds

HRS: Includes *all* bonds, meaning that there is likely some double counting with the variable used for Certificates of Deposit.

Mutual funds/hedge funds

SIPP: Includes mutual funds and stocks held in corporations excluding in one's own corporation(s).

Publicly traded equity

PSID: Current market value of all non-IRA stock holdings.

CES: The variable itself, SECESTX, refers to all types of securities, including stocks, mutual funds, private bonds, government bonds, and treasury notes. The variable USBNDX, which contains U.S. Savings Bonds, is netted out to form the "Savings bond" line item.

HRS: Net value of stocks, mutual funds, and investment trusts. This variable cannot be made gross.

Life insurance

PSID: Also includes valuable collections and rights in a trust or estate. It is a net, rather than gross, value.

Retirement accounts

PSID: A combination of head of household and spouse pensions, along with current market value of any IRA accounts or annuities.

HRS: Net value of IRA and Keogh accounts.

SIPP: Includes IRAs, Keogh accounts, 401(k), 403(b), and or thrift plans.

Other important assets

SIPP: This includes equity in other financial investments less any debts against it. This is therefore a net, rather than gross, value.

Business

PSID: Net, rather than gross, value of farm or business.

Primary residence

CES: Unlike other assets, home values are pulled from the OPI file. Identifying the meaning of missing values in this file is difficult. After consultation with the BLS, we used CUTENURE and PROP_ALX to identify which missing values are truly missing and which are missing because the family does not own a home. Additionally, unlike other housing estimates for other surveys, this may not include furnishings.

Other real estate

PSID: Market value of second home only. This variable does not include any other real estate.

HRS: This value is a combination of the *net* value of all investment real estate and the gross value of a second residence. There is unfortunately no way of turning the net value into a gross value.

Vehicles

PSID: The net value of all vehicles is reported in the survey is made gross by adding back in all reported vehicle loans. Unlike vehicles, the aggregate net value of which is reported in one variable, vehicle loans are reported individually. Only three vehicle loans can be reported, meaning that there may be slight attenuation bias if some respondents own more than three cars with car loans.

Unknown assets

SCF: Respondents are allowed to report up to three "other" assets from a list of 42. This row represents any asset reported from that list, with the exception of cash (currency), which is reported in its own row. See the note on "currency" above.

HRS: Respondents are asked to report anything not previously reported that might reasonably be called an asset.

4.2 Table 2-b: Balance Sheets – Liabilities

Revolving debt

SIPP: This includes amount owed for credit cards as well as store bills. There is no way to disintegrate the two in the dataset.

Credit cards / charge cards

PSID: All credit card debt, excluding debt from convenience use.

CES: Data are pulled from the FNA file. The ultimate values are a little less than double the values in the other surveys. So far, we have not been able to confirm for certain that we are treating missing values properly.

Revolving store accounts

CES: Calculated from the FNA file.

Mortgages for primary residence

PSID: Includes only remaining principal.

CES: Calculated from MOR file.

SCF: The SCF does not have information that can directly link loans to primary vs. secondary residences or investment properties. It does, however, allow respondents to report a multitude of mortgages, loans, and land contracts associated with any property they own. We use the first mortgage/land contract reported in the survey to calculate this number.

Mortgages for investment real estate or second home

PSID: Includes only remaining principal.

CES: Calculated from MOR file.

SCF: The SCF does not have information that can directly link loans to primary vs. secondary residences or investment properties. It does, however, allow respondents to report a multitude of mortgages, loans, and land contracts associated with any property they own. We calculate this number by adding all mortgages, loans, and land contracts that can be associated with property, excluding the mortgage/land contract reported in *Mortgages for primary residence.*

SIPP: This includes mortgages on a rental mobile home, principal owed on rental property, and principal owed on mortgage. It is unclear why this estimate is so low relative to the estimates of other surveys.

HEL/HELOC

CES: Calculated by combining data from HEL and OPH files. We have been helpfully assured by people at the BLS that these files contain an exhaustive list of respondents who have HELs or HELOCs.

Loans for improvement

HRS: Gross value of all other home loans, aside from those on a primary and secondary residence. This may include loans on investment properties and HELOCs, but is not exclusively either of these.

Loans on vehicles

PSID: Only includes remaining principal.

CES: Calculated from the OVB file (Owned Vehicles – Detailed Questions).

Other loans

PSID: Includes only loans from relatives.

SIPP: All other loans, including medical bills not covered by insurance, money owed to private individuals, educational loans, and other debt not covered. This should, theoretically, exclude mortgages, home equity loans, and car loans. However, given the relative size of this estimate, the situation may be more complicated than it seems. Once again, there is no way to disentangle this amount any further.

4.3 Table 3: Income Statements

Total Income

SIPP: The SIPP includes a "reaggregated" measure of income. However, it is unclear what was aggregated to create it. For total income, we report the mean "reaggregated" income. Therefore, total income on the first line of the table is not the sum total of the subcategories. Total income as implied by summing the subcategories is approximately $6,500 greater.

Wages and salaries

PSID: Includes wages for head of household and spouse, as well as bonus, overtime, tips, and commissions for head of household.

Professional practice or trade

PSID: Only for head of household.

Other labor earnings

PSID: Only for head of household. Value is net rather than gross.

Business Income

PSID: Value is net rather than gross.

SCF: Includes business and farm income.

Rent

PSID: Value is net rather than gross.

Interest, dividends, etc.

PSID: Includes interest, dividends, trusts, and annuities.

SCF: Includes interest, dividends, and capital gains.

HRS: The HRS includes a broad "capital income" variable which is "the sum of household business or farm income, self-employment earnings, business income, gross rent, dividend and interest income, trust funds or royalties, and other asset income."

Other transfer receipts, from business

PSID: Includes only worker's compensation benefits.

HRS: Includes only worker's compensation benefits

All other income

PSID: Includes pension, IRA, child support, alimony, help from relatives, help from others, and "other" income.

HRS: Includes pension and "other" income.

SIPP: We once again use a reaggregated measure provided in the dataset. The codebook does not mention what variables are included in the construction of this variable.

Employment Taxes

CES: Calculated using employment taxes contained in the MEMI file.

Other Taxes

CES: Includes only property taxes. Note that expenditures are reported for a calendar quarter. This estimate is produced by taking the quarter's estimate and multiplying by 4.

SCF: Includes only real estate taxes. Yearly taxes are computed by adjusting by the frequency of the tax bill.

SIPP: Includes only property taxes. Once again, we use midpoints as explained above.

4.4 Table 4: Statement of Cash Flows

Net income

For all surveys, this figure corresponds to the "Net Income" line of the Income Statements for each respective survey (Table 3).

Consumption Expenditures

In general, consumption expenditures are all expenditures identifiable from a survey. When possible, taxes are excluded.

PSID: Consumption includes estimates for expenditures on food, housing, transportation, education, health, clothing, entertainment, and communication, and excludes property and employment taxes.

CES: All expenditures, excluding property and employment taxes.

SCF: Includes expenditures for food, housing, transportation, education, financial services and insurance, charitable giving, and miscellaneous spending, and excludes property taxes.
HRS: All expenditures.
SIPP: Includes expenses related to rent, utilities and housing, work, commuting, parking and tolls, child care and child support expenses. Taxes are midpoints of reported bins of income and property taxes.

Transfers

PSID: Pension contributions.
SCF: Withdrawals from IRA.

Lending

Lending is calculated as the change in the dollar value of mutual funds, CDs, bonds, and loans to others from the previous period to the present period, converted to an annual rate. For instance, for the SCF, we take the sum of the above-mentioned categories for 2013, subtract out the sum for 2010, and divide by 3. Prior year balance sheets are available in the excel file used to create the tables.

Borrowing

Change in liabilities from the previous period to the present period, converted to an annual rate. The total liabilities refers to the first line of Table 2-b. Prior year balance sheets are available in the excel file used to create the tables.

4.5 Table 5: Balance Sheets (TTMS and DCPC/SCPC)

In the 2012 SCPC, households were asked five questions with regard to assets and liabilities, reproduced below.[2]

- (pu010) **Last month,** about how much was the unpaid balance on **all** your credit cards that you carried over from the previous month?

- (DE014) What is the approximate market value of your **primary home**?
  Please enter your answer below in **thousands** of dollars.

- (DE015) About how much do you owe on loans for your **primary home**, including mortgages, home equity loans, and home equity lines of credit?
  Please enter your answer below in **thousands** of dollars.

- (DE016) Excluding the market value of your primary home, what is the approximate value of your household's **other assets**?
  Include real estate other than your primary home.
  Please enter your answer below in **thousands** of dollars.

[2] Note that the wordings of these questions were adjusted depending on whether the respondent reported owning a home.

- (DE019) Excluding the debt on the mortgage of your primary home, what is the approximate dollar amount of your remaining **debts**?
  Examples of other debts include credit card debt, student loan debt, and car loan debt.
  Please enter your answer below in **thousands** of dollars.

The responses to these questions were used to generate estimates for primary home value, unknown assets, revolving debt, mortgage for primary residence, and other loans. The currency holdings estimate is from the DCPC, which asks respondents to report their currency holdings at the end of each day.

Note that four of the five questions above ask respondents to report their assets/liabilities in thousands of dollars. Unfortunately, many respondents seem to have missed this instruction, resulting in implausibly high values for these questions. Upon reviewing the data on a case-by-case basis, we decided to set a cutoff of 3,000 for these questions, which would correspond to $3,000,000 in assets/liabilities. Thus, when creating the estimate, all responses less than or equal to 3,000 were adjusted by $1,000 to create nominal values, while those above 3,000 used as is.

TTMS estimates are reported without adjustment from data provided by the TTMS team. Those who are interested in more detail about their methods of calculation should refer to the TTMS documentation linked in Section 5 below.

4.7 Table 7: Statements of Cash flows (TTMS and DCPC), and Table 9: DCPC Statement of Account Flows

The details of the calculations for the Statement of Cash Flows and the Statement of Account Flows are largely the same, although the latter has greater detail. Therefore, we will focus on the Statement of Account Flows. Again, this document will deal only with the DCPC estimates; those interested in the details of the TTMS calculations should consult the reference material in Section 5 below.

The Statement of Account Flows attempts to track all movements of liquidity between different accounts that can be gleaned from the DCPC. Oftentimes these movements can be tracked by way of the payment instrument used to make the payment; for instance, liquidity transferred using a debit card presumably came from a Demand Deposit Account (DDA). In other cases, the respondent reports directly where the funds came from; for instance, when reporting cash withdrawals, respondents are asked the source of the funds.

While it is generally easy to link a payment instrument to a type of liquidity account, two exceptions stand out:

- It is not possible to know from the DCPC how a money order was originally purchased. As a compromise, we split money order consumption payments half-and-half between currency and DDA consumption.

- Text message payments are placed in Revolving Debt, as they are often added to one's cell phone bill. This process is, in some respects, a way of making a payment with a revolving debt account.

Italicized figures mean that the figure in the cell is known to suffer from attenuation bias because it is missing certain types of flows. For instance, Financing flows from DDA to Other Debt ($643) is italicized because other debt in this case is only mortgage payments. Since mortgage payments are not the totality of transfers to "other debt," we can be certain that there is significant measurement error in this figure. Cash deposits to DDA, on the other hand, should represent a complete estimate.

Production

The three estimates in the Production line are, in fact, only two estimates: net income and currency income. Currency income is measured in the DCPC's "cash withdrawals" module, which includes cash received as income. Net income is taken from the top line of Table 6, which is the imputed SCPC yearly household income divided by 12. It is assumed that all non-currency income is received in the form of direct deposit, and therefore it is placed in the DDA category.

Consumption

Not all payments measured in the DCPC are consumption.[3] We exclude mortgage payments from consumption measures, as well as 91 percent of payments to "financial services and insurance" merchants. The 91 percent figure was chosen based on preliminary evidence from the 2015 DCPC about the percentage of payments to "financial services and insurance" that are for non-homeowner insurance.

Financing

Each day, DCPC respondents were asked to report all of their withdrawals of cash, deposits of cash, and any time they loaded a prepaid card. They were also asked the source of the funds or, in the case of prepaid loads, the payment instrument used. These three modules represent the majority of transfers. The two exceptions are transfers to other debt, which represent mortgage payments by any means, and foreign currency exchange, which was its own survey question.

4.8 Figure 3: Survey Coverage Ratios

A given survey *k* contains N_k measurements of assets and liabilities (stocks) and M_k measurements of income and expenditure items (flows). We wish to compare the sums of these stocks and flows to outside benchmark to get a sense of how much they cover. For the stocks, we use balance-sheet information from the Federal Reserve's Flow of Funds (FoF) survey as our outside

[3] There is some evidence to suggest that DCPC payments match consumption pretty closely. See Schuh (2017) for detailed evidence.

benchmark, and for the flows we use National Income and Product Accounts (NIPA) estimates of income and personal consumption.

$$\text{Balance sheet coverage ratio} = \frac{\left(\sum_{i=1}^{N_k} stock_i\right)}{FoF\ estimate}$$

$$\text{Income statement coverage ratio} = \frac{\left(\sum_{i=1}^{M_k} flow_i\right)}{NIPA\ estimate}$$

The asset and liability estimates from FoF include assets held and liabilities owed by non-profits in addition to those of households, and these cannot be disentangled. According to one Federal Reserve estimate from 2001, "[s]eparate estimates, also published in the accounts, indicate that nonprofit organizations have in recent years accounted for 5 to 7 percent of the assets and liabilities of the combined sector." We therefore adjust the FoF estimate of combined assets and liabilities by a factor of 0.94.

For the sum of stocks in each survey, we are concerned with the total assets and liabilities reported in Tables 2-a and 2-b. For the sum of flows, we are concerned with total income in Table 3 and Total expenditures, including taxes. Total expenditures excluding taxes are the consumption expenditures reported in Table 4, while taxes are reported in Table 3.

5. Codebooks and Data Dictionaries

The following are links to the codebooks and data users' guides and data dictionaries that we relied upon heavily for the production of these figures.

PSID: http://psidonline.isr.umich.edu/Guide/documents.aspx

CES: https://www.bls.gov/cex/pumd_doc.htm

SCF: https://www.federalreserve.gov/econres/files/codebk2013.txt

HRS: https://hrs.isr.umich.edu/documentation

SIPP: https://www.census.gov/programs-surveys/sipp/tech-documentation/data-dictionaries.html

TTMS: https://dataverse.harvard.edu/file.xhtml?fileId=2994992&version=1.0

DCPC (forthcoming): https://www.bostonfed.org/publications-and-data.aspx#/search?siteTpc=&topic=&jel=&type=&series=&dept=2b802ad4820f45ae86c411c243ee3d86

SCPC: https://www.bostonfed.org/publications/survey-of-consumer-payment-choice/2012-survey.aspx

Stata programs and Excel files (forthcoming): https://www.bostonfed.org/publications-and-data.aspx#/search?siteTpc=&topic=&jel=&type=&series=&dept=2b802ad4820f45ae86c411c243ee3d86

References

Schuh, Scott. 2017. “Measuring Consumer Expenditures with Payment Diaries.” *Economic Inquiry* forthcoming.

Teplin, Albert M. 2001. “The U.S. Flow of Funds Accounts and Their Uses.” *Federal Reserve Bulletin* 432–441 (July). https://www.federalreserve.gov/pubs/bulletin/2001/0701lead.pdf, retrieved on 5/2/2017.