

Investigating the Imputation of Assets and Liabilities in the CE Interview Survey

Geoffrey Paulin, Ph.D

Senior Economist

Consumer Expenditure Surveys (CE) Program

CE Microdata Users' Workshop

July 20, 2018

Washington, DC



The Consumer Expenditure Survey (CE) collects information on:

- Expenditures
- Income
- Taxes (income and other)
- Assets and Liabilities



Nonresponse is a problem for each. However, corrections are in place for most of these items:

- **Expenditures: Since the 1980s**
- **Income: Since 2004**
- **Taxes (income only): Since 2013**



Assets and Liabilities are currently under investigation.



Assets and Liabilities

Project Overview

- “The purpose of this team is to initiate and conduct a research project designed to impute missing Interview asset and liability data, leveraging models from income imputation and other relevant procedures.”
- “The goal is to implement this into production with 2017 Quarter 2 data.”

Source: Charter for the Asset and Liability Imputation Team, 9-9-2014



This presentation describes three aspects of the project:

- What asset and liability data are collected?
- What processes have been considered for imputation of missing values?
- What are the next steps in the investigation?



Asset/Liability Data

Assets:

- Retirement accounts
- Stocks, bonds, mutual funds
- Checking, savings, money market, CDs
- Whole life insurance
- Other, including annuities, trusts, royalties

Liabilities:

- Credit cards
- Student loans
- Other loans, including medical and personal

Collection

- Questions are asked in the final survey (4th interview)
- Most are asked in two parts: Did you have _____? If yes, how much?
- For some items, only a total value is collected. In these cases, it is not clear whether \$0 means:
 - ▶ No, I did not have such an account or
 - ▶ Yes, I had an account, but it is empty.

Collection, continued:

- For each asset/liability, the total value/balance/amount owed is collected:
 - ▶ As of today
 - ▶ As of one year ago today
- Bracket questions are asked when the respondent cannot provide a specific value.

The team considered several methods:

- Survey of Consumer Finance method (multiple imputation, iterative process)
- Regression trees
- Hotdeck

...But none is feasible.

Going back to the original motivations (charter):

A system based on income imputation processing is being investigated.

- ▶ Regression-based, multiple imputation of each component asset/liability, from which “total change in” values can be derived.
- ▶ For each component, separate models are run for demographic groups across which large variation in parameter estimates is observed or expected.

For example:

Consider IRAX.

- Amount reported when asked: “What is the total value of all retirement accounts such as 401(k)s, IRAs, and Thrift Savings Plans that you or your household own/owns?”
- Expected to vary considerably by age
- Preliminary tests support use of one model for each age group (group 1: $\text{age} < A$; group 2: $A \leq \text{age} < B$, etc.)

Bracket imputation will also be used:

- Respondent identifies range in which asset/liability falls (e.g., less than \$X; \$X to \$Y; etc.)
- Five values are selected based on current methods used in income imputation; each falls within the specified bracket range.
- Open-ended brackets (\$Z or over) also are treated in income imputation

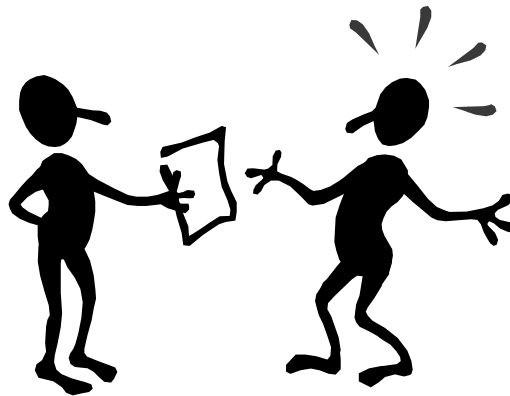
Related Challenges:

- How to distinguish \$0 meaning no balance from \$0 meaning no account.
- How to identify groups upon which to base models. That is:
 - ▶ Grouping variable: Is IRAX (e.g.) based on age, occupation, or something else?
 - ▶ Variable range: If age, where do the breaks occur—under 35, 35 to 64, 65 and older, or under 25, 25 to 34, etc.?

Work in progress:

- Identifying groups, and selecting variables to include within each model.
 - ▶ ANOVA/Chow tests have been used so far to test differences/pooling potential across groups.
 - ▶ Variables used in income imputation are considered the “starter group,” with some to be added, deleted, or redesigned. (Example: Age ranges used in binary variables could be widened or narrowed.)

Comments/Suggestions are welcome!



Contact Information

Geoffrey Paulin, Ph.D.

Senior Economist

Consumer Expenditure Survey Program

www.bls.gov/cex

202-691-5132

paulin.geoffrey@bls.gov

