Missing Spending Data Report

Dirk Cotton

4/17/2018

We identify 18010 retired households from the HRS database. However, spending data in the CAMS database is available for only 1161 of these households. To replace the missing spending data in our data set of retired households identified from HRS data, we impute retired household spending when the data is missing from CAMS.

We reject 1908 CAMS households that are not determined to have retired.

We are interested in household spending for the wave in which each houshold was determined to have first retired (wave of retirement, or WOR). CAMS data is available beginning with wave 5 so we first remove all CAMS households determined to have retired prior to wave 5 or not to have retired (2792 households removed).

We then reject CAMS 325 households that have neither data for current income or any savings.

These actions leave 560 one-person households and 1348 two-person households with which we can model missing non-durable consumption expenditures by household when that data is missing from CAMS.

We impute spending separately for one-person and two-person households using multiple linear regression models.

For one-person households, the y-variable for our model is consumer non-durable expenditures for the household for the wave of retirement. The model’s x-variables are: + Social Security income for the wave of retirement + Total portfolio assets for the wave of retirement + Pension and annuity income for the wave of retirement + age at retirement + gender (equals 1 for male, else zero), and + Wave number of retirement (5 through 12)

Wave number is modelled as a series of dummy variables, d5 through d12. If a household retired in wave 11, for example, d11 for that household will equal 1 and dummy variables for all other waves will equal zero for that household. (Only one of the eight dummy variables for a household will equal 1.)

The multiple progression is performed suppressing calculation of a y-intercept value. All x-variables expressed in dollars are normalized by dividing each x-variable for a household by the sum of all x-variables for that household.

The results of the multiple regression analysis for one-person households are summarized:

##   
## Call:  
## lm(formula = cams1Pgtw$cndurWOR ~ cams1Pgtw$d5 + cams1Pgtw$d6 +   
## cams1Pgtw$d7 + cams1Pgtw$d8 + cams1Pgtw$d9 + cams1Pgtw$d10 +   
## cams1Pgtw$d11 + cams1Pgtw$d12 + cams1Pgtw$male + cams1Pgtw$ssIncAtRetire1 +   
## cams1Pgtw$portfolioAssetsTotal + cams1Pgtw$pensionIncome +   
## cams1Pgtw$ageAtRetirement1 - 1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -18046 -6381 -2839 2759 139208   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## cams1Pgtw$d5 12896.59 5126.52 2.516 0.012166 \*   
## cams1Pgtw$d6 15479.36 5278.33 2.933 0.003502 \*\*   
## cams1Pgtw$d7 14746.85 4999.08 2.950 0.003315 \*\*   
## cams1Pgtw$d8 15505.54 5193.04 2.986 0.002955 \*\*   
## cams1Pgtw$d9 15412.34 5259.47 2.930 0.003527 \*\*   
## cams1Pgtw$d10 16124.81 4873.28 3.309 0.000999 \*\*\*  
## cams1Pgtw$d11 15663.99 4796.43 3.266 0.001160 \*\*   
## cams1Pgtw$d12 14668.09 4793.26 3.060 0.002321 \*\*   
## cams1Pgtw$male -467.35 1322.51 -0.353 0.723941   
## cams1Pgtw$ssIncAtRetire1 -1712.16 2305.26 -0.743 0.457972   
## cams1Pgtw$portfolioAssetsTotal 8962.08 2246.84 3.989 7.55e-05 \*\*\*  
## cams1Pgtw$pensionIncome 6387.21 3492.11 1.829 0.067938 .   
## cams1Pgtw$ageAtRetirement1 -48.29 70.26 -0.687 0.492126   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 12900 on 547 degrees of freedom  
## Multiple R-squared: 0.6165, Adjusted R-squared: 0.6074   
## F-statistic: 67.64 on 13 and 547 DF, p-value: < 2.2e-16