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 *for these retired households* is available for only 4182 of those. 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.

CAMS household data is available for 5188 of the households identified as retired in HRS data.

We are interested in household spending for the wave in which each household 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 from HRS data to have retired prior to wave 5 or not to have retired. 2160 households are removed from our CAMS data set on this basis, leaving 2321households.

469 of the remaining CAMS households provide no spending data for the wave of retirement. These observations are rejected leaving 1852 households in our CAMS data set.

We then reject 239 CAMS households that have neither data for current income or any savings in HRS, despite providing consumption data, as we cannot identify the source of this consumption. This action leaves 1613 observations in the CAMS data set.

These actions leave 476 one-person households and 1137 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 modeled 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 dollar-denominated x-variables for that household.

The results of the linear 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   
## -15282 -7028 -2963 2577 144434   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## cams1Pgtw$d5 19514.10 5576.13 3.500 0.000511 \*\*\*  
## cams1Pgtw$d6 20552.94 5833.35 3.523 0.000468 \*\*\*  
## cams1Pgtw$d7 21044.81 5428.28 3.877 0.000121 \*\*\*  
## cams1Pgtw$d8 24308.91 5590.15 4.349 1.69e-05 \*\*\*  
## cams1Pgtw$d9 22271.85 5689.33 3.915 0.000104 \*\*\*  
## cams1Pgtw$d10 22565.72 5296.88 4.260 2.48e-05 \*\*\*  
## cams1Pgtw$d11 21482.78 5244.39 4.096 4.95e-05 \*\*\*  
## cams1Pgtw$d12 20413.80 5269.07 3.874 0.000122 \*\*\*  
## cams1Pgtw$male -616.32 1481.08 -0.416 0.677506   
## cams1Pgtw$ssIncAtRetire1 -3701.16 1882.19 -1.966 0.049849 \*   
## cams1Pgtw$portfolioAssetsTotal -2235.75 1137.02 -1.966 0.049858 \*   
## cams1Pgtw$pensionIncome -1213.93 2524.52 -0.481 0.630846   
## cams1Pgtw$ageAtRetirement1 -54.35 77.67 -0.700 0.484403   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 13260 on 463 degrees of freedom  
## Multiple R-squared: 0.5794, Adjusted R-squared: 0.5676   
## F-statistic: 49.06 on 13 and 463 DF, p-value: < 2.2e-16

A few modifications are made to the regression model for two-person households:

* All HRS two-person households consist of one male and one female, so gender is not a factor
* Social Security retirement income for the second household member must be added to the model
* The age of the second household member must also be added to the model

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

##   
## Call:  
## lm(formula = cams2Pgtw$cndurWOR ~ cams2Pgtw$d5 + cams2Pgtw$d6 +   
## cams2Pgtw$d7 + cams2Pgtw$d8 + cams2Pgtw$d9 + cams2Pgtw$d10 +   
## cams2Pgtw$d11 + cams2Pgtw$d12 + cams2Pgtw$ssIncAtRetire1 +   
## cams2Pgtw$ssIncAtRetire2 + cams2Pgtw$portfolioAssetsTotal +   
## cams2Pgtw$pensionIncome + cams2Pgtw$ageAtRetirement1 + cams2Pgtw$ageAtRetirement2 -   
## 1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -34576 -12786 -4975 6314 192232   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## cams2Pgtw$d5 33324.1 6198.7 5.376 9.26e-08 \*\*\*  
## cams2Pgtw$d6 38374.8 6208.8 6.181 8.92e-10 \*\*\*  
## cams2Pgtw$d7 36234.1 6080.5 5.959 3.39e-09 \*\*\*  
## cams2Pgtw$d8 33635.2 6213.5 5.413 7.56e-08 \*\*\*  
## cams2Pgtw$d9 32589.5 6434.1 5.065 4.77e-07 \*\*\*  
## cams2Pgtw$d10 33116.9 5909.9 5.604 2.64e-08 \*\*\*  
## cams2Pgtw$d11 34375.4 5987.8 5.741 1.21e-08 \*\*\*  
## cams2Pgtw$d12 41174.4 6007.2 6.854 1.18e-11 \*\*\*  
## cams2Pgtw$ssIncAtRetire1 -3510.3 2818.6 -1.245 0.2133   
## cams2Pgtw$ssIncAtRetire2 -3128.4 2626.0 -1.191 0.2338   
## cams2Pgtw$portfolioAssetsTotal 8067.3 1821.6 4.429 1.04e-05 \*\*\*  
## cams2Pgtw$pensionIncome -1812.2 5439.2 -0.333 0.7391   
## cams2Pgtw$ageAtRetirement1 158.8 110.6 1.436 0.1514   
## cams2Pgtw$ageAtRetirement2 -310.6 130.8 -2.375 0.0177 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 21930 on 1123 degrees of freedom  
## Multiple R-squared: 0.6474, Adjusted R-squared: 0.643   
## F-statistic: 147.3 on 14 and 1123 DF, p-value: < 2.2e-16

Coefficients determined by the respective linear regression results are used to impute a “predicted” amount of consumer non-durable spending for retired households from the HRS data base when no spending is available for the household’s wave of retirement.