Report Cross Validation of Bad Error Predictions

Dirk Cotton

5/9/2018

The consumption data predicted by our models based on HRS wealth and income data showed large errors relative to reported consumption from the CAMS data base for a large number of households. We cross-validated our models to determine if removing those hoouseholds with poor predictions would improve prediction results.

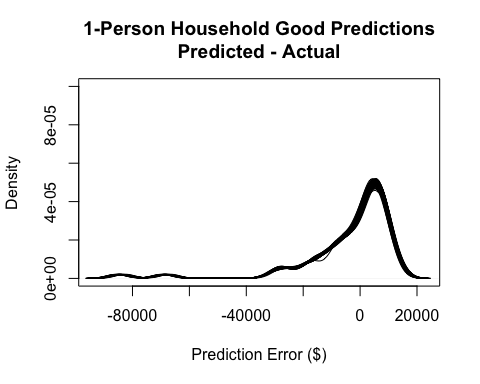
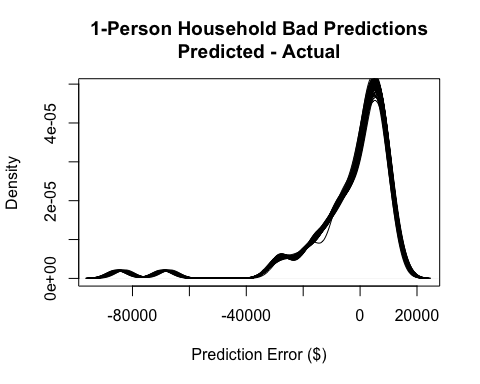
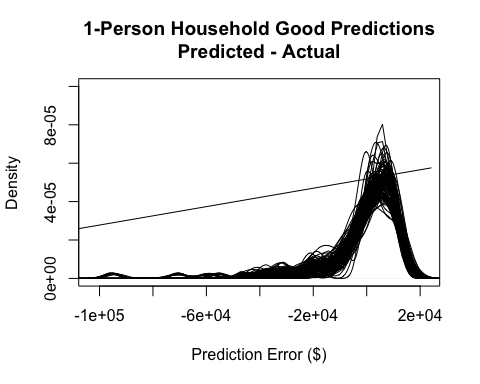
We us a model for one-person households. The regression model for one-person households follows:

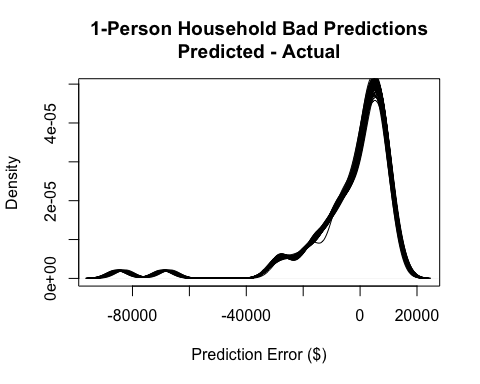
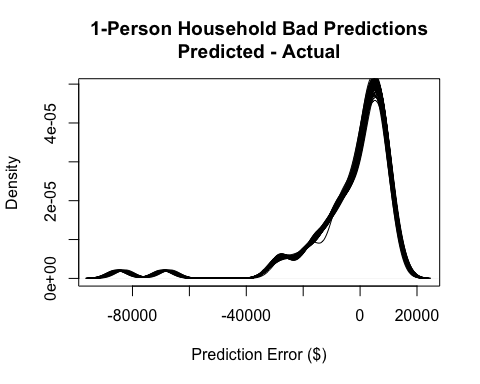
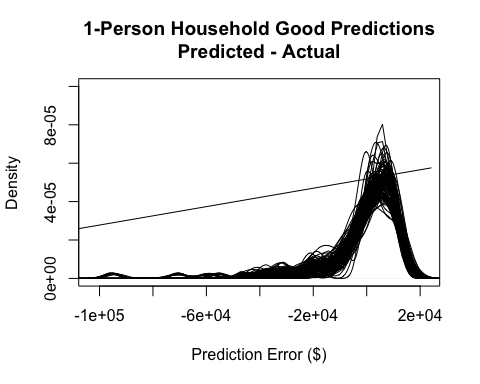
##   
## Call:  
## lm(formula = cams1Pgtw$cndurWOR ~ cams1Pgtw$d5 + cams1Pgtw$d6 +   
## cams1Pgtw$d7 + cams1Pgtw$d8 + cams1Pgtw$d9 + cams1Pgtw$d10 +   
## cams1Pgtw$d11 + cams1Pgtw$d12 + cams1Pgtw$male + cams1Pgtw$ssIncAtRetire1 +   
## cams1Pgtw$portfolioAssetsTotal + cams1Pgtw$pensionIncome +   
## cams1Pgtw$ageAtRetirement1 + cams1Pgtw$male - 1)  
##   
## Coefficients:  
## cams1Pgtw$d5 cams1Pgtw$d6   
## 15441.94 17641.86   
## cams1Pgtw$d7 cams1Pgtw$d8   
## 17742.84 18464.81   
## cams1Pgtw$d9 cams1Pgtw$d10   
## 19295.40 16276.30   
## cams1Pgtw$d11 cams1Pgtw$d12   
## 18519.89 15570.72   
## cams1Pgtw$male cams1Pgtw$ssIncAtRetire1   
## -26.33 254.14   
## cams1Pgtw$portfolioAssetsTotal cams1Pgtw$pensionIncome   
## 237.90 1183.24   
## cams1Pgtw$ageAtRetirement1   
## -27.67

We predicted consumption using the “fit” model of predictions with acceptable errors (< 100%), randomly selecting 55 households with acceptable errors as the prediction set from a data set of 478 CAMS households.

For the second cross-validation, we used 55 households with high errors as the predictions set.

We plotted the density of the eroor results (predicted consumption minus reported consumption) and generated the following two plots:





Our conclusion from these results is that removing households with poor predictions from the model will not improve the model’s predictions.

NOTES:

[1] This output is created by running the following scripts in order:

1. source(‘~/R Projects/HRS/Validate Predicted CAMS Data.R’)
2. source(‘~/R Projects/HRS/CrossValidate 1P HH Good Predictions.R’)
3. source(‘~/R Projects/HRS/CrossValidate 1P HH Bad Predictions.R’)

and then knitting the Word document with this script.

[2] The errant line on the “Good Predictions” graph is caused by a household with an extreme negative prediction that doesn’t fit within the graph’s scale.