

Case Study IV, Interim Report II

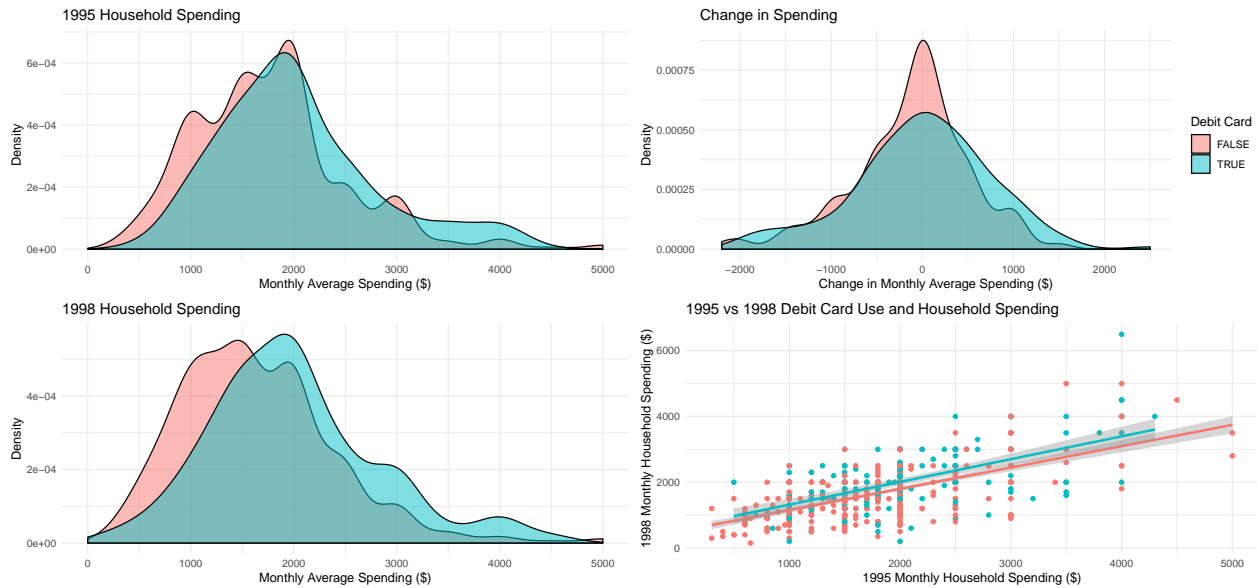
Jake Epstein, Daniel Spottiswood, Man-Lin Hsiao, Michael Tan, Sahil Patel

11/01/2019

Introduction

The goal of this case study is to evaluate the causal impact of debit card ownership on household spending. The data come from the Italy Survey on Household Income and Wealth (SHIW), a 1995-1998 survey of 584 Italian households. The dataset includes 1995 and 1998 monthly household spending, whether the household had exactly one debit card in 1998 and demographic information including family size, geographic region and average age. In this report, we will create a model to estimate the causal impact of debit card ownership on household spending, utilizing propensity score methods to ensure model balance.

Exploratory Data Analysis



We begin our exploratory data analysis by looking at spending. In 1995 and 1998, households with debit cards tended to spend more than households without. The distribution of difference in household spending is centered at around 0, indicating most households spent about the same amount in 1998 as they did in 1995. The distribution of changes for households with debit cards has slightly more weight on the positive side, indicating that these households may have increased their spending slightly relative to non-debit card households. We also looked into spending as a percentage of income or of wealth, and the results were consistent with those above.

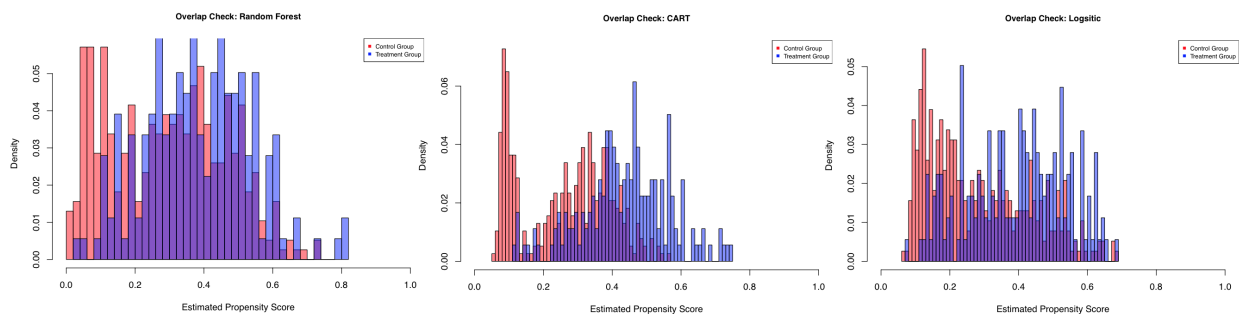


We then examined relationships between demographic characteristics and 1998 spending. Some selected plots are shown above. Our initial analysis indicates that family size is positively associated with spending, which is intuitive given the cost of raising children. Additionally, families with household heads who have higher educational status tend to spend more than those headed by less educated individuals. This may be a function of income or wealth, as higher educated individuals tend to earn more; regardless, it is worth exploring further. Finally, both income and wealth are positively associated with spending, and households with debit cards tend to spend more at all levels of income and wealth.

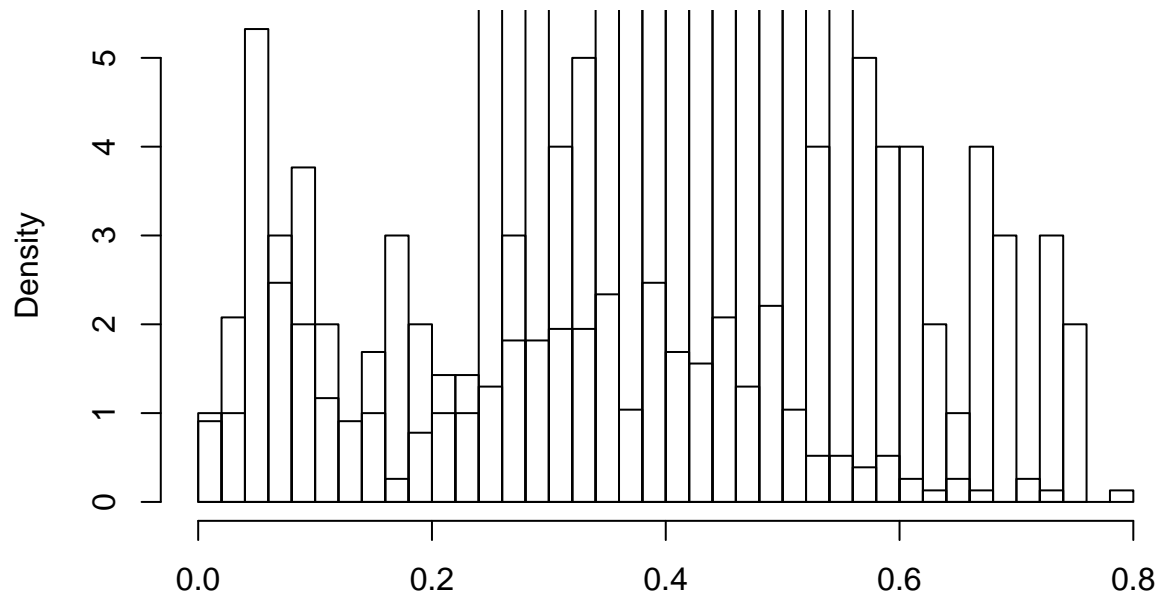
We move forward with attempts to balance the covariates using matching, weighting and propensity scores as these plots illustrate significant differences in the covariates between the treatment and control groups. $y =$ “1998 Spending (% of Income)”, $x =$ “Geography”)

Data Balancing

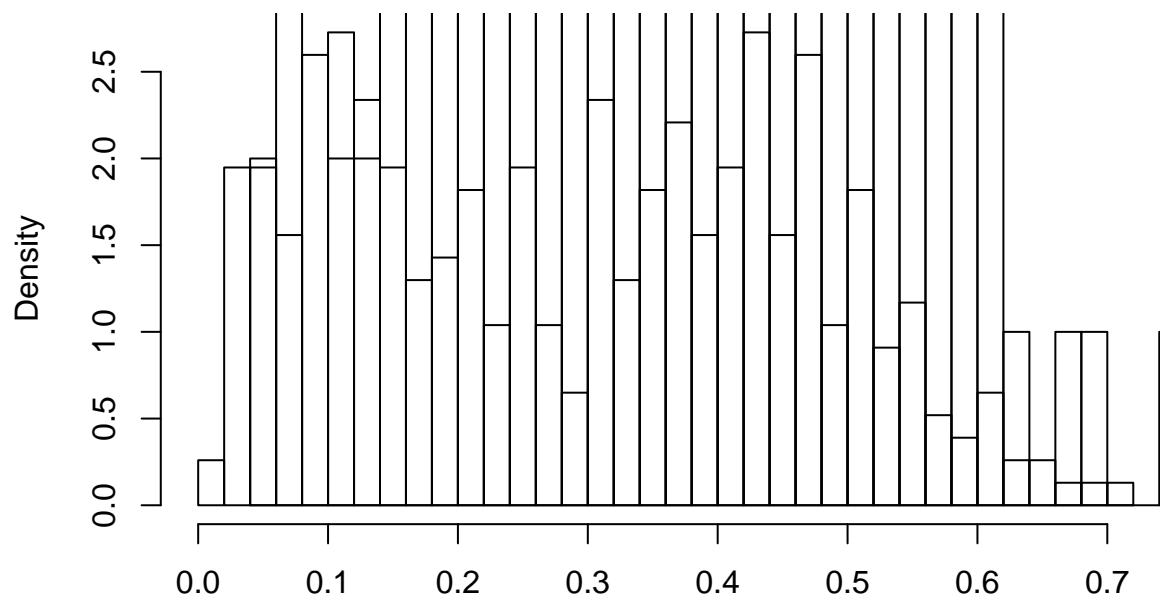
To examine data balance between control and treatment groups, we looked at three different propensity score models: logistic regression, random forest, and CART, a decision-tree based machine learning model. The overlap in propensity scores for each model is shown below. Logistic regression and random forest show good overlap, whereas CART does not. Given that logistic and random forest perform similarly, and logistic regression is a simpler and more interpretable model, we will use the logistic regression model for propensity scores.



Histogram of `debit.ps$ps_basicLog[which(debit.ps$debit_card1998 ==`

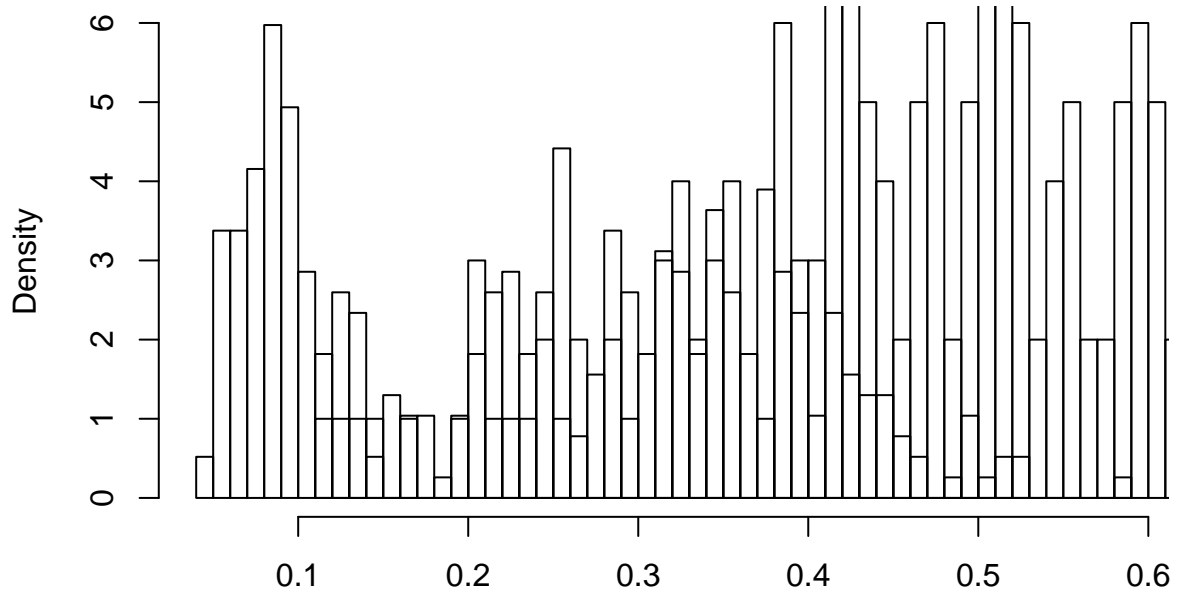


stogram of `debit.ps$ps_randomForest[which(debit.ps$debit_card1998`



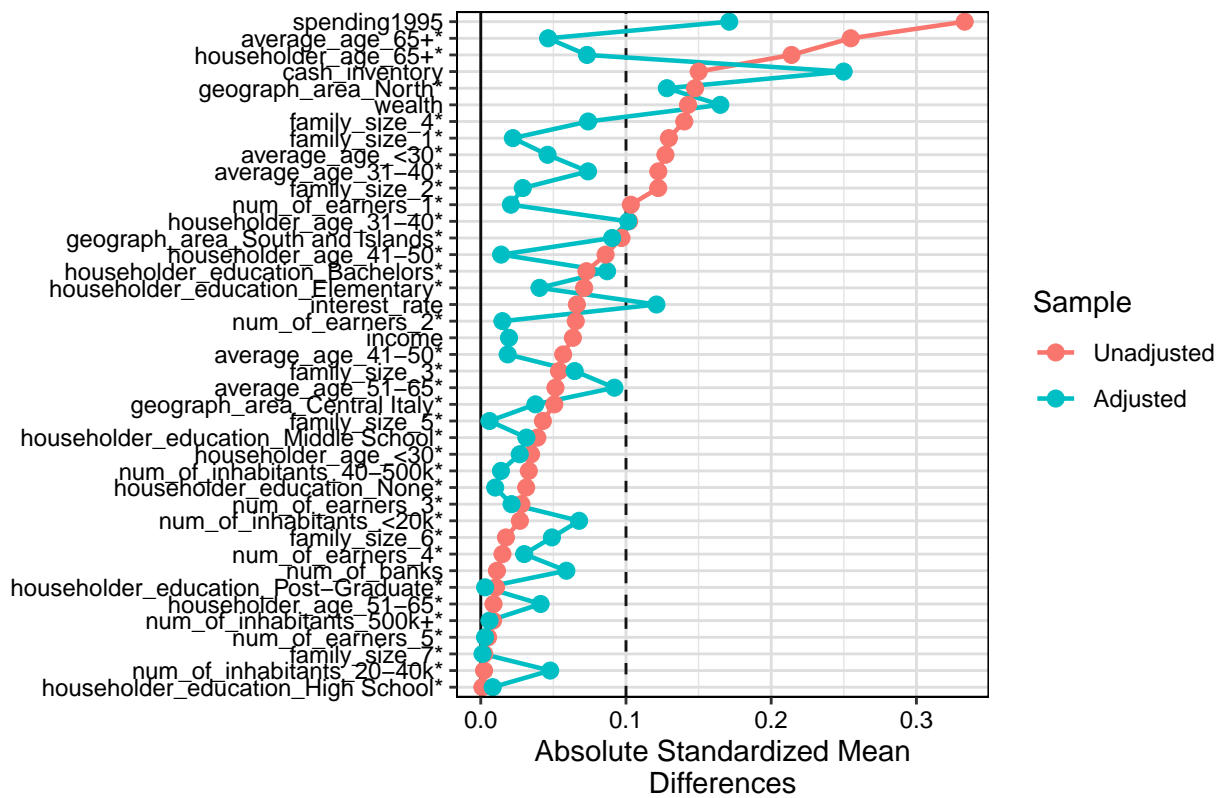
`debit.ps$ps_randomForest[which(debit.ps$debit_card1998 == 0)]`

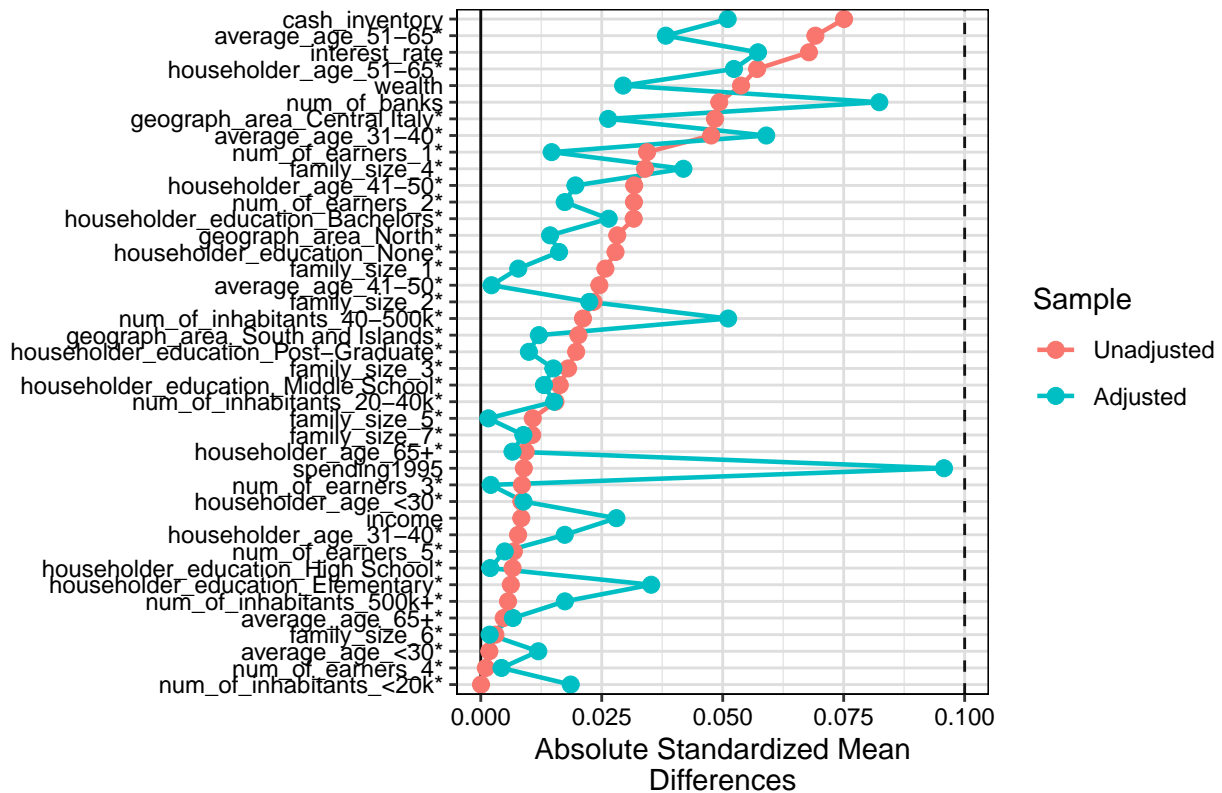
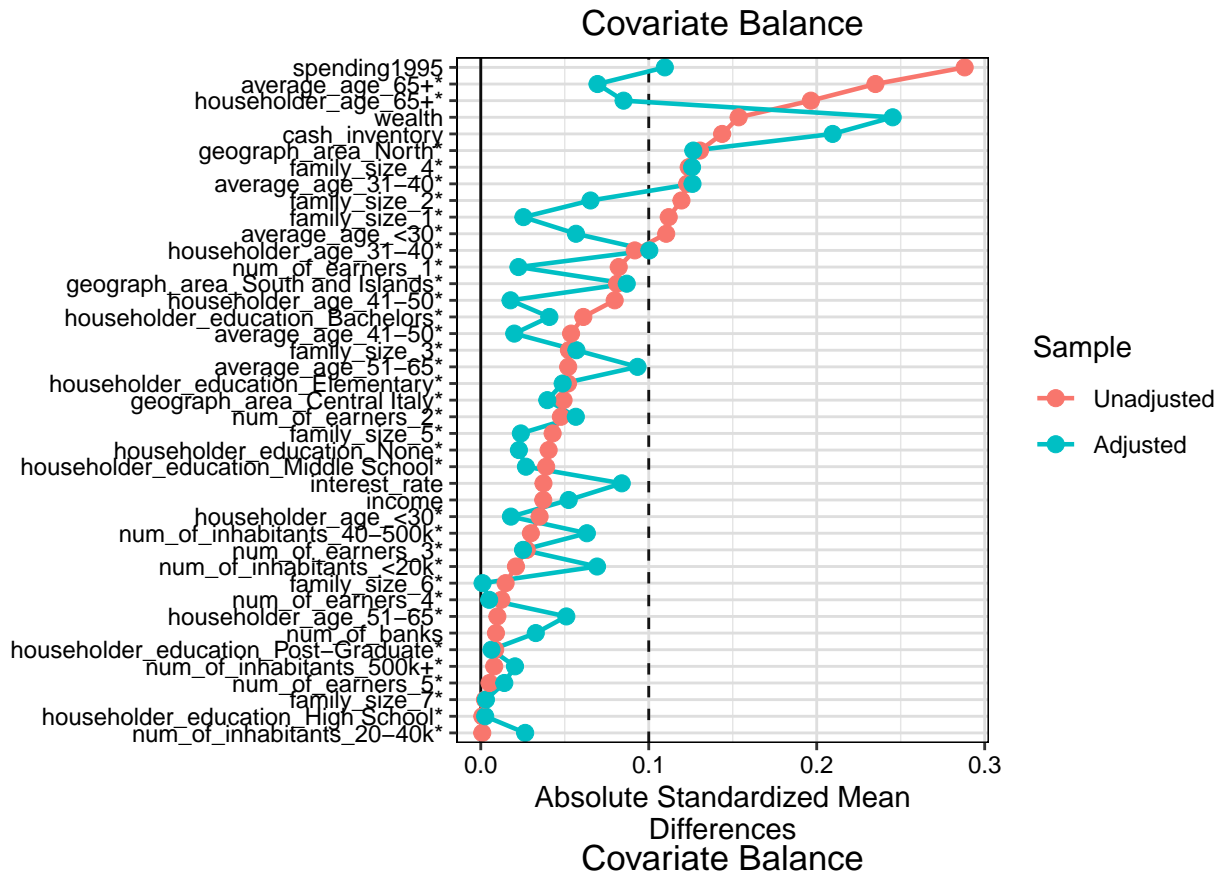
Histogram of debit.ps\$ps_CART[which(debit.ps\$debit_card1998 == (

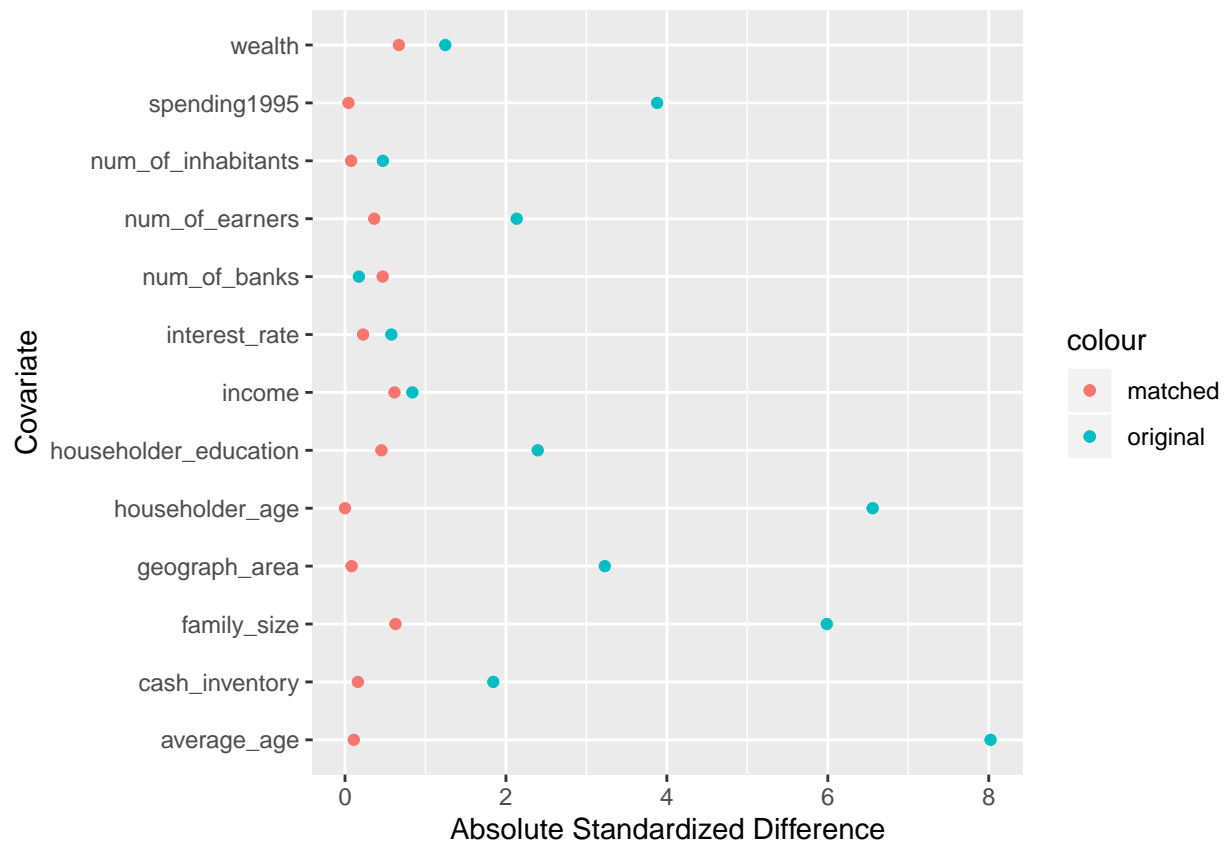


debit.ps\$ps_CART[which(debit.ps\$debit_card1998 == 0)]

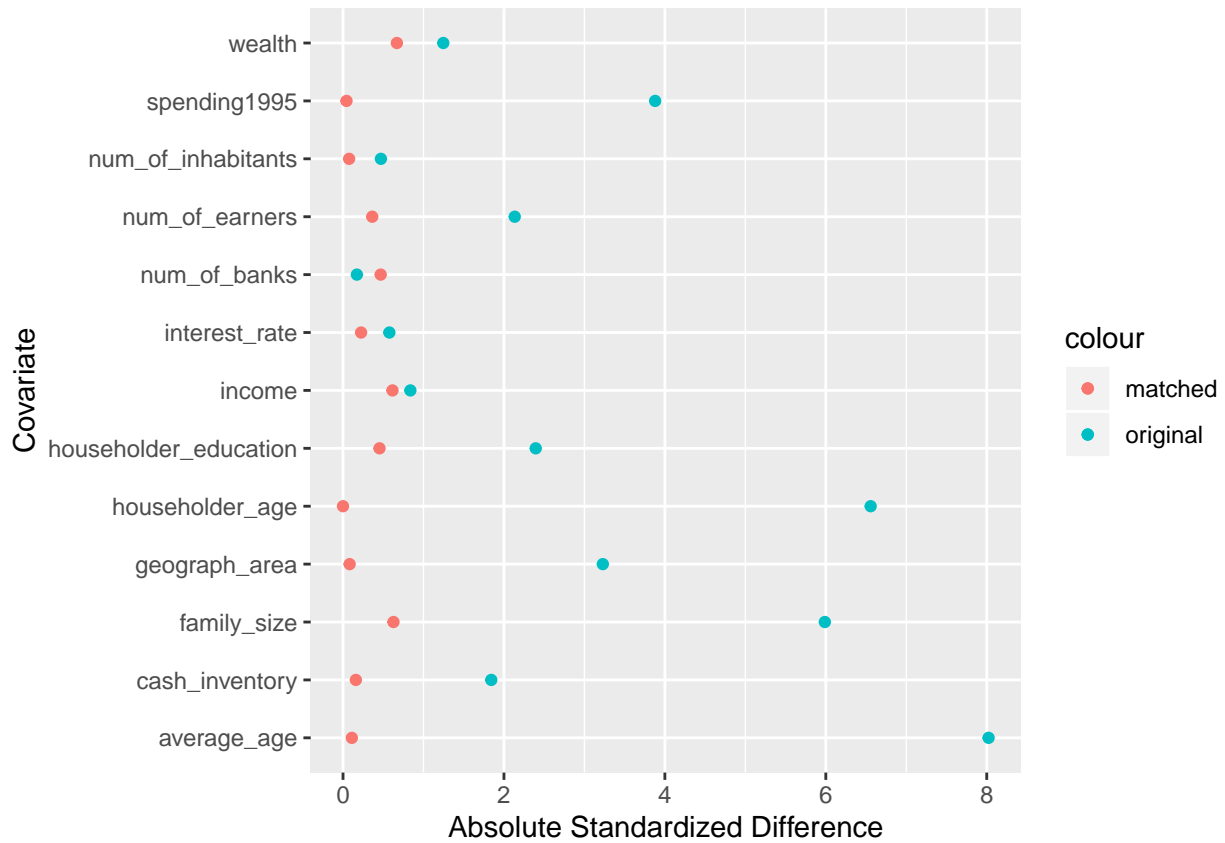
Covariate Balance



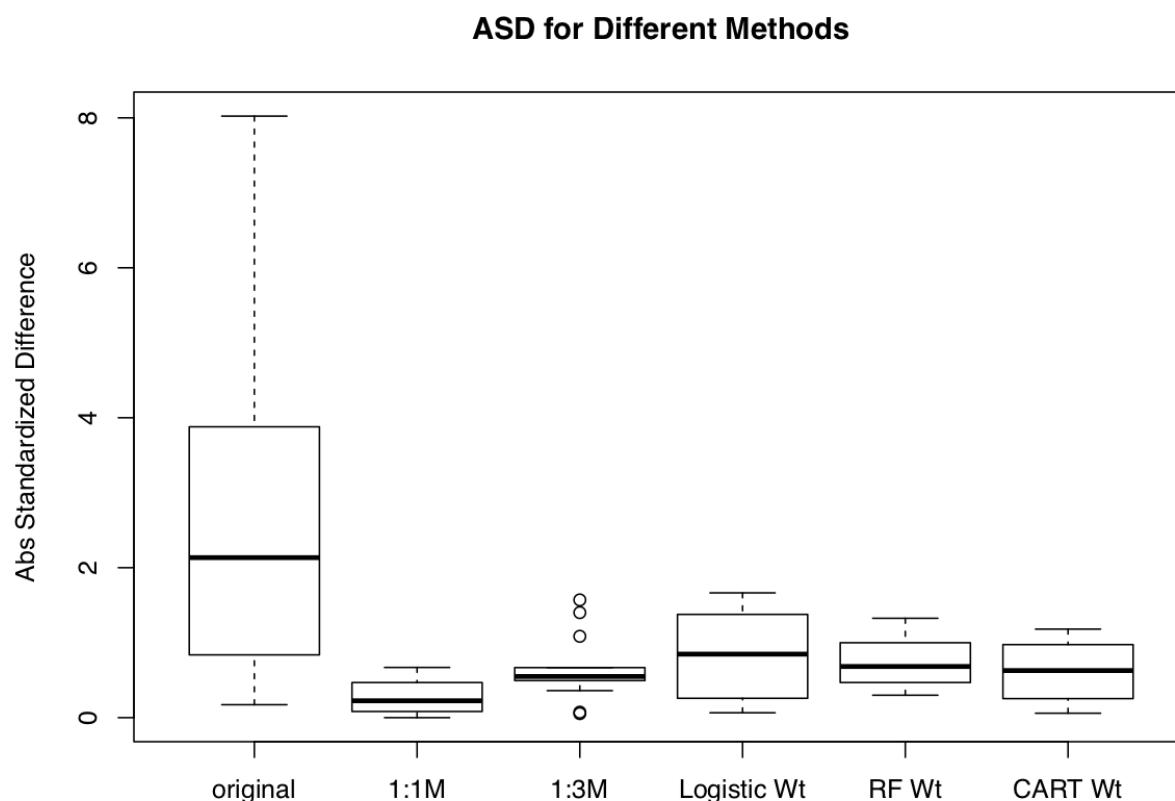




pdf
2



We calculate propensity scores using general logistic regression. We then exclude the samples in the nonoverlapping region. One can see from the histogram that there is significant overlap between the propensity scores of the control and debit_card1998 groups; however, we expect this can be improved upon.



Moving forward with 1:1 matching, we see that the absolute mean differences are improved significantly across almost all covariates.

Regression Adjustment

We attempt to fit models on top of the 1:1 matching with replacement. We first use a basic linear model before fitting more flexible models, random forest and extreme gradient boosting.

```
## ATT_mix ATE_mix ATT_mix_boost ATE_mix_boost ATT_mix_forest
## 1 218.6245 127.1942 286.7505 169.6116 249.509
## ATE_mix_forest
## 1 137.8689

## ERROR: factor householder_education has new levels None
## ERROR: factor family_size has new levels 6
## ERROR: factor family_size has new levels 6
## ERROR: factor num_of_earners has new levels 4
## ERROR: factor num_of_earners has new levels 4
## ERROR: factor family_size has new levels 7
## ERROR: factor family_size has new levels 6
## ERROR: factor num_of_earners has new levels 4
## ERROR: factor num_of_earners has new levels 5
## ERROR: factor family_size has new levels 6
## [1] "== 11 =="
```



```

## [1] "== 12 =="
## [1] "== 13 =="
## ERROR: factor num_of_earners has new levels 4
## ERROR: factor family_size has new levels 7
## ERROR: factor family_size has new levels 6, 7
## ERROR: factor family_size has new levels 7
## ERROR: factor householder_education has new levels None
## ERROR: factor family_size has new levels 7
## ERROR: factor num_of_earners has new levels 5
## ERROR: factor family_size has new levels 7
## ERROR: factor family_size has new levels 7
## ERROR: factor family_size has new levels 7
## [1] "== 24 =="
## ERROR: factor family_size has new levels 7
## ERROR: factor family_size has new levels 6, 7
## ERROR: factor num_of_earners has new levels 5
## ERROR: factor num_of_earners has new levels 4
## ERROR: factor householder_age has new levels <30
## ERROR: factor householder_education has new levels None, Post-Graduate
## ERROR: factor family_size has new levels 7
## ERROR: factor num_of_earners has new levels 5
## [1] "== 33 =="
## ERROR: factor num_of_earners has new levels 5
## ERROR: factor family_size has new levels 7
## ERROR: factor family_size has new levels 6, 7
## ERROR: factor num_of_earners has new levels 5
## ERROR: factor num_of_earners has new levels 4
## ERROR: factor num_of_earners has new levels 4
## ERROR: factor num_of_earners has new levels 4, 5
## ERROR: factor num_of_earners has new levels 4
## ERROR: factor num_of_earners has new levels 4
## ERROR: factor num_of_earners has new levels 5
## ERROR: factor num_of_earners has new levels 5
## ERROR: factor family_size has new levels 6, 7
## [1] "== 46 =="
## ERROR: factor family_size has new levels 6, 7
## ERROR: factor family_size has new levels 6, 7
## ERROR: factor family_size has new levels 6, 7
## ERROR: factor num_of_earners has new levels 4
## ERROR: factor num_of_earners has new levels 4
## [1] "== 52 =="
## ERROR: factor family_size has new levels 6, 7
## [1] "== 54 =="
## ERROR: factor family_size has new levels 6, 7
## ERROR: factor family_size has new levels 7
## ERROR: factor family_size has new levels 7
## ERROR: factor family_size has new levels 7
## ERROR: factor family_size has new levels 7
## ERROR: factor num_of_earners has new levels 4, 5
## [1] "== 61 =="
## ERROR: factor num_of_earners has new levels 4
## ERROR: factor num_of_earners has new levels 5
## ERROR: factor num_of_earners has new levels 5
## ERROR: factor family_size has new levels 7

```

```

## ERROR: factor family_size has new levels 6, 7
## ERROR: factor family_size has new levels 6, 7
## ERROR: factor family_size has new levels 6, 7
## ERROR: factor householder_education has new levels None
## ERROR: factor num_of_earners has new levels 4
## ERROR: factor family_size has new levels 7
## ERROR: factor num_of_earners has new levels 4
## ERROR: factor family_size has new levels 6
## ERROR: factor family_size has new levels 7
## ERROR: factor householder_education has new levels High School
## ERROR: factor num_of_earners has new levels 4
## ERROR: factor family_size has new levels 6
## ERROR: factor family_size has new levels 6
## ERROR: factor family_size has new levels 7
## ERROR: factor num_of_earners has new levels 4
## [1] "== 81 =="
## ERROR: factor num_of_earners has new levels 4
## ERROR: factor num_of_earners has new levels 4
## ERROR: factor family_size has new levels 7
## ERROR: factor num_of_earners has new levels 5
## ERROR: factor family_size has new levels 7
## ERROR: factor num_of_earners has new levels 4
## ERROR: factor num_of_earners has new levels 4, 5
## ERROR: factor family_size has new levels 6, 7
## ERROR: factor family_size has new levels 6
## ERROR: factor num_of_earners has new levels 4
## ERROR: factor family_size has new levels 6, 7
## ERROR: factor family_size has new levels 7
## ERROR: factor family_size has new levels 7
## [1] "== 95 =="
## ERROR: factor num_of_earners has new levels 4
## ERROR: factor family_size has new levels 7
## ERROR: factor family_size has new levels 7
## [1] "== 99 =="
## [1] "== 100 =="
## ERROR: factor family_size has new levels 6, 7
## ERROR: factor family_size has new levels 7
## ERROR: factor family_size has new levels 6
## ERROR: factor num_of_earners has new levels 4
## ERROR: factor family_size has new levels 6, 7
## ERROR: factor num_of_earners has new levels 4, 5
## [1] "== 107 =="
## ERROR: factor num_of_earners has new levels 4
## ERROR: factor family_size has new levels 6
## ERROR: factor num_of_earners has new levels 4
## [1] "== 111 =="
## ERROR: factor family_size has new levels 7
## ERROR: factor householder_education has new levels None, High School
## [1] "== 114 =="
## ERROR: factor householder_education has new levels None
## [1] "== 116 =="
## ERROR: factor num_of_earners has new levels 4
## [1] "== 118 =="
## ERROR: factor num_of_earners has new levels 4

```

```

## ERROR: factor num_of_earners has new levels 4
## ERROR: factor family_size has new levels 6, 7
## [1] "==" 122 ==
## ERROR: factor family_size has new levels 7
## ERROR: factor family_size has new levels 6, 7
## ERROR: factor family_size has new levels 7
## ERROR: factor num_of_earners has new levels 5
## [1] "==" 127 ==
## ERROR: factor family_size has new levels 6, 7
## ERROR: factor family_size has new levels 6, 7
## ERROR: factor num_of_earners has new levels 5
## [1] "==" 131 ==
## ERROR: factor family_size has new levels 7
## ERROR: factor num_of_earners has new levels 4
## ERROR: factor family_size has new levels 7
## ERROR: factor num_of_earners has new levels 4
## ERROR: factor family_size has new levels 7
## [1] "==" 137 ==
## ERROR: factor family_size has new levels 7
## ERROR: factor num_of_earners has new levels 4
## ERROR: factor family_size has new levels 6
## ERROR: factor family_size has new levels 6, 7
## [1] "==" 142 ==
## ERROR: factor family_size has new levels 7
## ERROR: factor family_size has new levels 7
## ERROR: factor family_size has new levels 7
## ERROR: factor family_size has new levels 7
## ERROR: factor family_size has new levels 6
## [1] "==" 148 ==
## ERROR: factor num_of_earners has new levels 4
## ERROR: factor householder_education has new levels High School
## ERROR: factor householder_education has new levels High School
## [1] "==" 152 ==
## ERROR: factor num_of_earners has new levels 4
## ERROR: factor num_of_earners has new levels 4
## [1] "==" 155 ==
## ERROR: factor family_size has new levels 6
## ERROR: factor num_of_earners has new levels 5
## ERROR: factor num_of_earners has new levels 4
## ERROR: factor num_of_earners has new levels 4
## ERROR: factor num_of_earners has new levels 5
## ERROR: factor family_size has new levels 7
## ERROR: factor householder_education has new levels None
## [1] "==" 163 ==
## ERROR: factor family_size has new levels 7
## ERROR: factor family_size has new levels 7
## ERROR: factor family_size has new levels 6, 7
## [1] "==" 167 ==
## ERROR: factor family_size has new levels 7
## ERROR: factor num_of_earners has new levels 5
## [1] "==" 170 ==
## ERROR: factor family_size has new levels 7
## ERROR: factor family_size has new levels 7
## ERROR: factor family_size has new levels 6, 7

```

```

## ERROR: factor family_size has new levels 7
## ERROR: factor family_size has new levels 6
## ERROR: factor family_size has new levels 7
## [1] "== 177 =="
## ERROR: factor num_of_earners has new levels 4
## ERROR: factor family_size has new levels 7
## ERROR: factor householder_education has new levels High School
## ERROR: factor num_of_earners has new levels 4
## ERROR: factor family_size has new levels 6, 7
## [1] "== 183 =="
## [1] "== 184 =="
## ERROR: factor family_size has new levels 7
## ERROR: factor num_of_earners has new levels 5
## ERROR: factor num_of_earners has new levels 4
## [1] "== 188 =="
## ERROR: factor householder_education has new levels None
## [1] "== 190 =="
## ERROR: factor family_size has new levels 6
## ERROR: factor num_of_earners has new levels 4
## ERROR: factor family_size has new levels 6
## [1] "== 194 =="
## [1] "== 195 =="
## ERROR: factor family_size has new levels 6, 7
## ERROR: factor family_size has new levels 6, 7
## ERROR: factor family_size has new levels 7
## ERROR: factor householder_education has new levels None, High School
## ERROR: factor num_of_earners has new levels 5

## % latex table generated in R 3.5.3 by xtable 1.8-4 package
## % Mon Nov 4 18:30:13 2019
## \begin{table}[ht]
## \centering
## \begin{tabular}{rrrrr}
## \hline
## & mean & V2 & V3 & V4 \\
## \hline
## 1 & 112.0971 & 126.7246 & 127.1942 & 169.6116 \\
## 2 & 57.6081 & 79.1400 & 106.9714 & 61.0165 \\
## 3 & 169.0854 & 258.1435 & 218.6245 & 286.7505 \\
## 4 & 70.9231 & 80.4682 & 67.4613 & 79.9212 \\
## \hline
## \end{tabular}
## \end{table}

```

Conclusions

From our model, we can conclude that having a debit card caused the Italian households with debit cards to increase their spending by about \$460 per month.