Case Study IV, Interim Report I

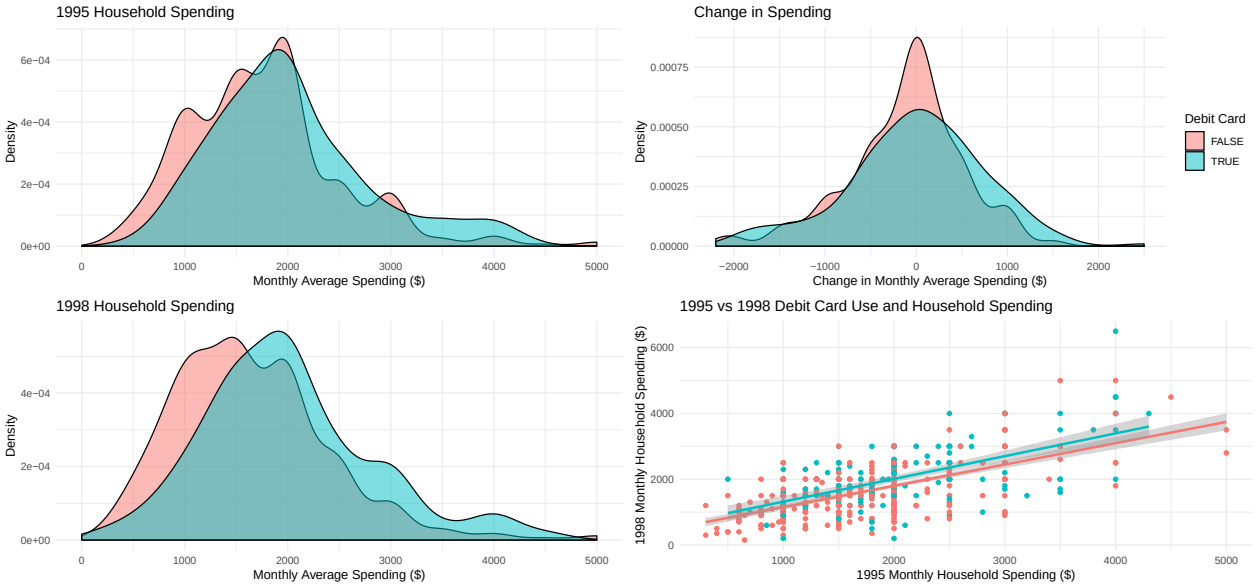
Jake Epstein, Daniel Spottiswood, Man-Lin Hsiao, Michael Tan, Sahil Patel

10/29/2019

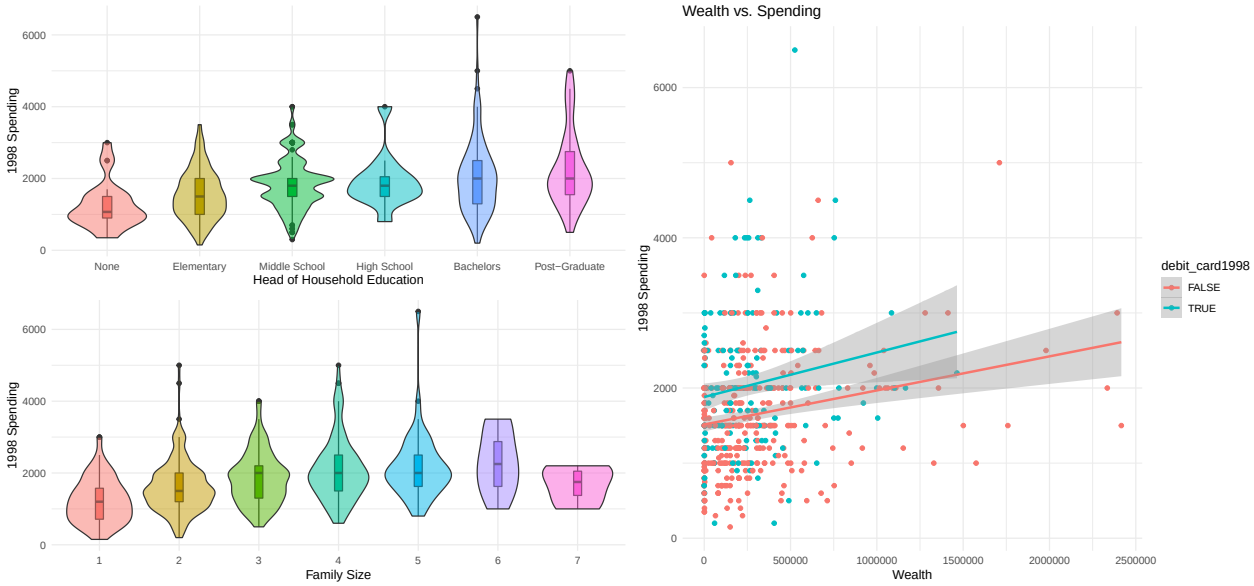
Introduction

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 slighly 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.

Data Balancing

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 treatment groups.

Using propensity scores we went ahead with matching and weighting to help acheive balance in the covariates across control and treatment groups. We performed matching with and without replacement, although our prior is that with replacement will be superior as we can create larger more balanced samples. Comparing the absolute standardized differences for the original data and the balanced data given our three different methods we see that weighting acheives the most balanced data. While it seems that both forms of matching should record improvements in balance over the original data, it is not unlikely that due to variation they may perform worse.

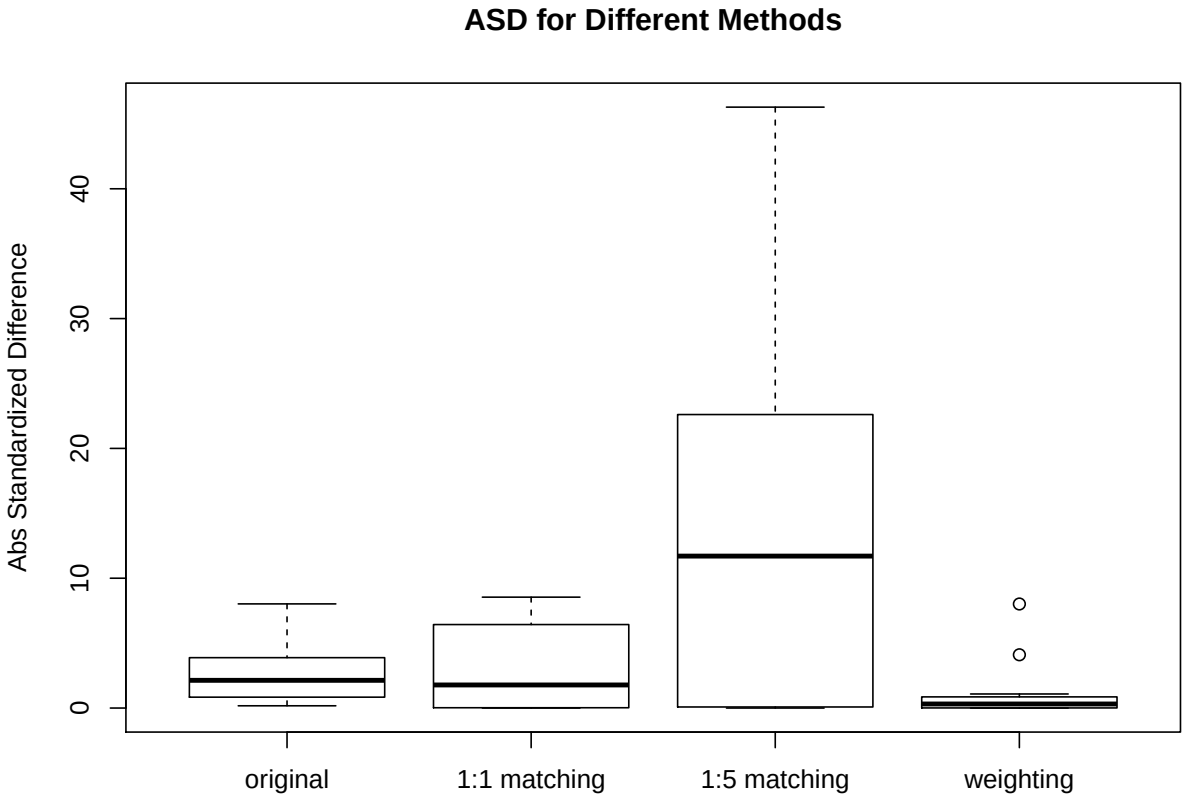
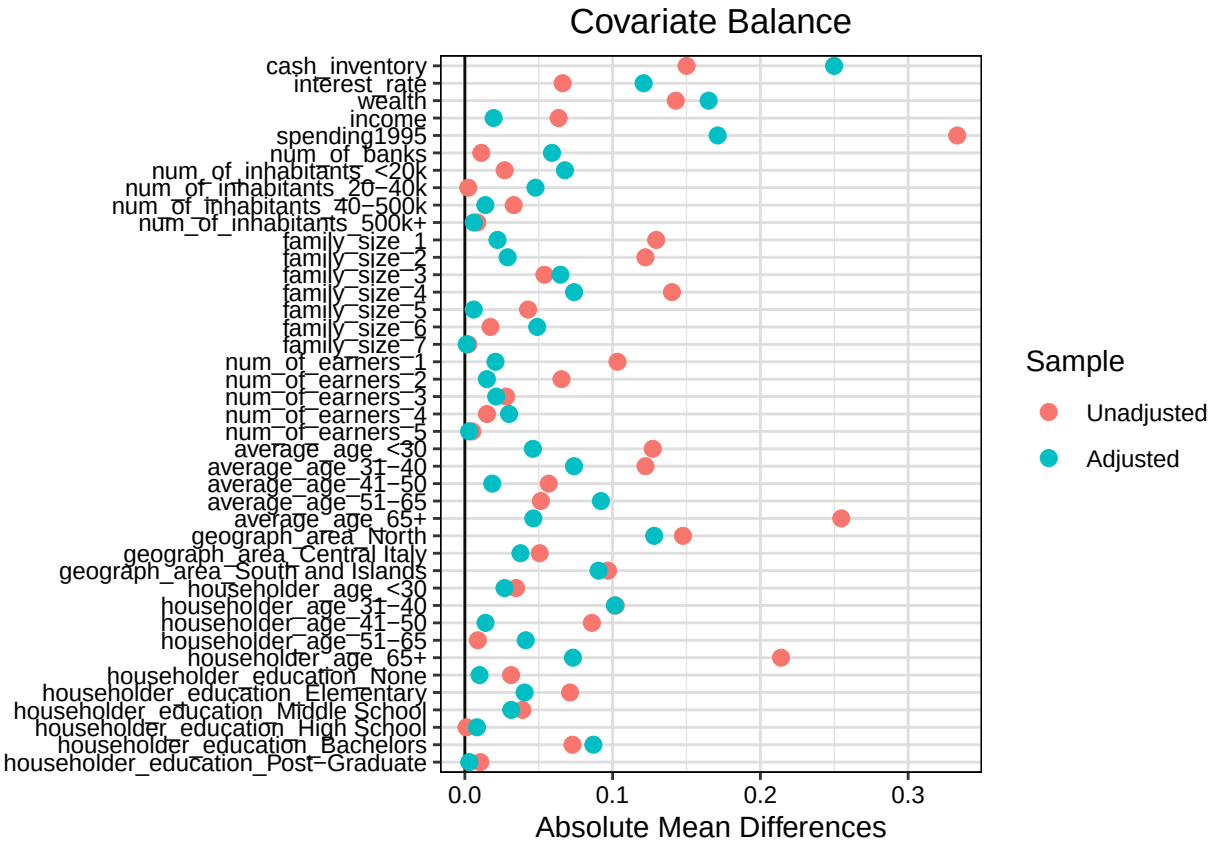

ASD for Different Methods
Abs Standardized Difference
0
10
20
30
40
original
1:1 matching
1:5 matching
weighting

Figure 1:



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

Model Selection

We considered two models to find the treatment effect of debit card on spending: a simple linear model as well as a more flexible XGBoost Model. Both models yield very similar results, and their outputs are shown below. Since the linear model is simpler, we will report our conclusions using the results from this model.

```
##   ATT_mix_lin ATE_mix_lin ATT_mix_boost ATE_mix_boost
## 1    460.8285    159.0414      445.0504      169.5686
```

Conclusions