

Case Study IV, Interim Report II

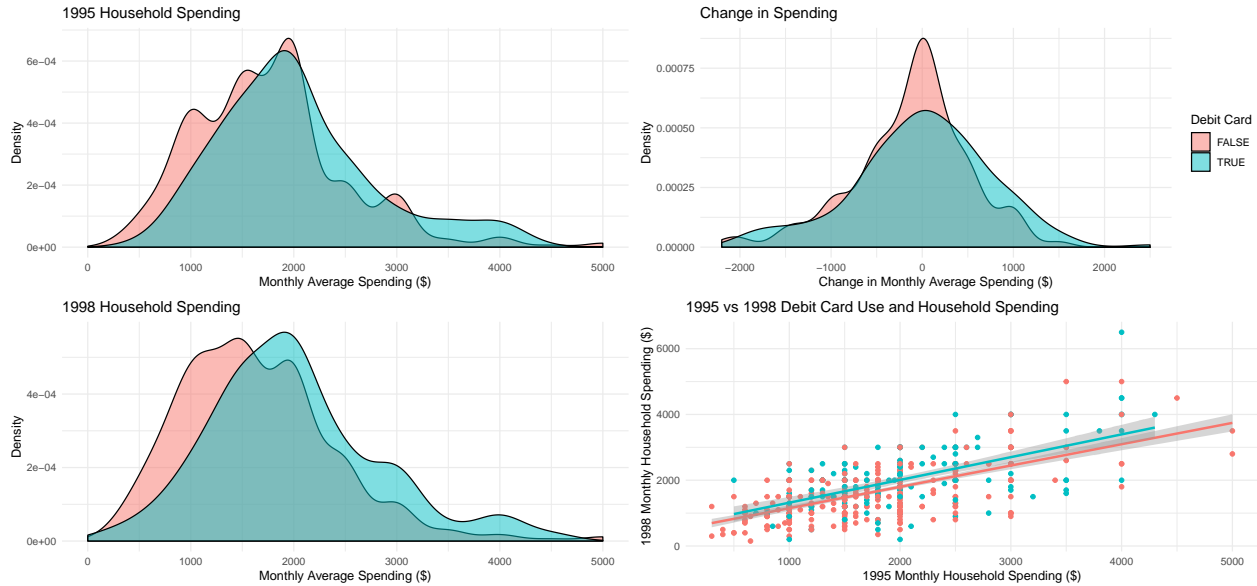
Jake Epstein, Daniel Spottiswood, Man-Lin Hsiao, Michael Tan, Sahil Patel

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

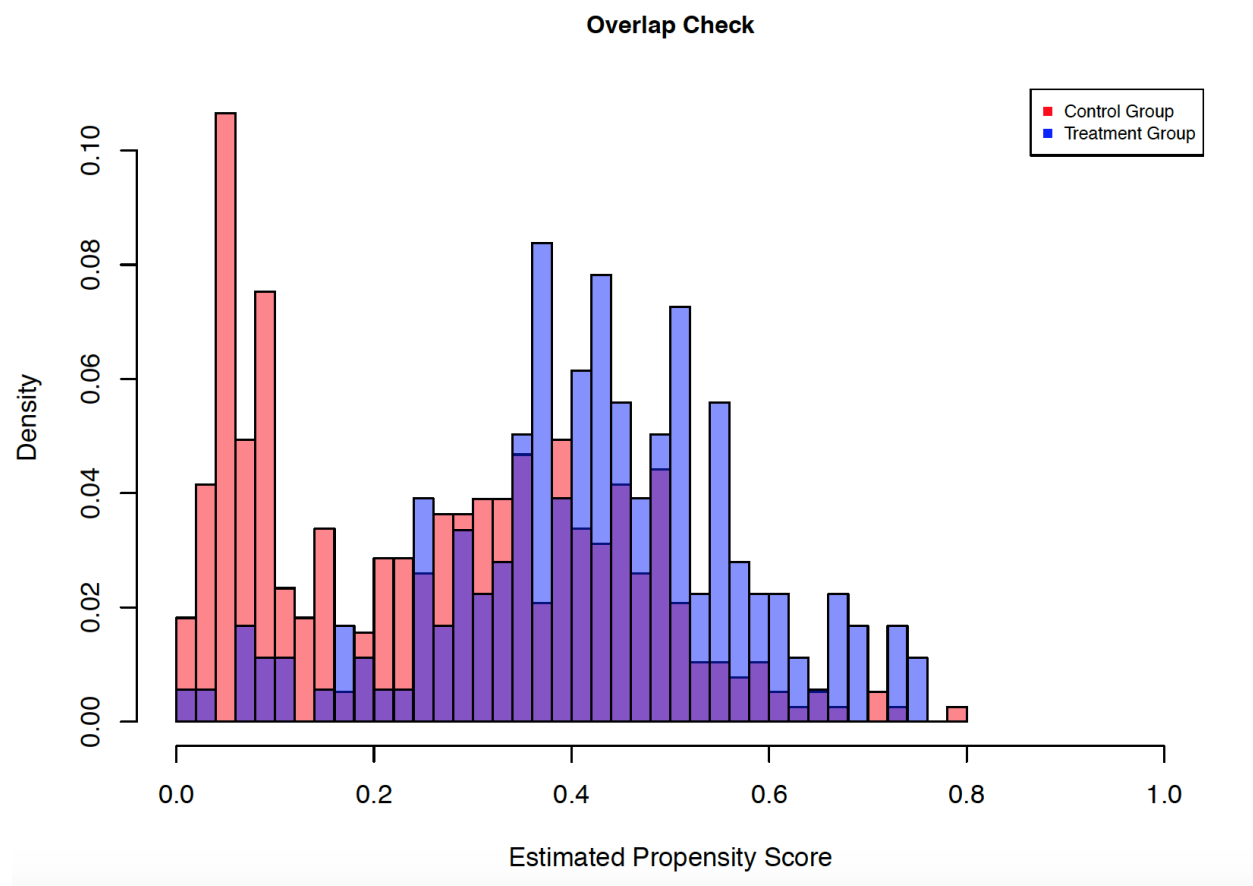


We begin our exploratory data analysis by looking at spending. In 1995 and 1998, households with debit cards tended to spend more than households without. The distribution of difference in household spending is centered at around 0, indicating most households spent about the same amount in 1998 as they 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.

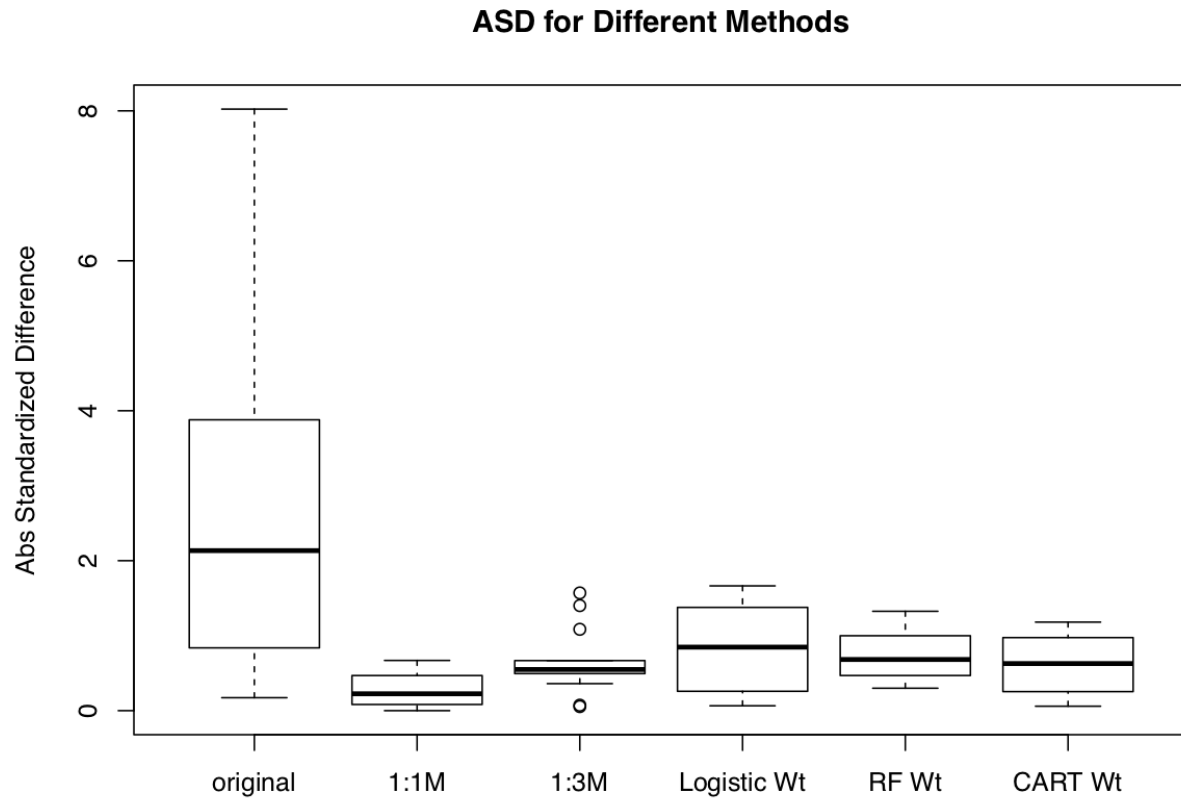


We then examined relationships between demographic characteristics and 1998 spending. Some selected plots are shown above. Our initial analysis indicates that family size is positively associated with spending, which is intuitive given the cost of raising children. Additionally, families with household 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.

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 debit_card1998 groups; however, we expect this can be improved upon.



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 debit_card1998 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 more interpretable, we will report our conclusions using the results from this model. However, further exploration should be done to determine whether a more flexible model, such as Xgboost might in general yield more accurate results.

```
## 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 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 family_size has new levels 7
## ERROR: factor num_of_earners has new levels 5
## ERROR: factor num_of_earners has new levels 5
## [1] "== 10 =="
## [1] "== 11 =="
## ERROR: factor num_of_earners has new levels 5
```

```

## ERROR: factor family_size has new levels 7
## [1] "== 14 =="
## ERROR: factor family_size has new levels 7
## 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] "== 21 =="
## ERROR: factor num_of_earners has new levels 5
## [1] "== 23 =="
## ERROR: factor num_of_earners has new levels 5
## ERROR: factor family_size has new levels 7
## [1] "== 26 =="
## ERROR: factor num_of_earners has new levels 5
## ERROR: factor householder_education has new levels None
## [1] "== 29 =="
## ERROR: factor family_size has new levels 7
## [1] "== 31 =="
## [1] "== 32 =="
## ERROR: factor num_of_earners has new levels 5
## [1] "== 34 =="
## [1] "== 35 =="
## [1] "== 36 =="
## ERROR: factor householder_education has new levels None
## ERROR: factor num_of_earners has new levels 5
## ERROR: factor family_size has new levels 7
## [1] "== 40 =="
## ERROR: factor family_size has new levels 7
## 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
## [1] "== 45 =="
## ERROR: factor householder_education has new levels None
## [1] "== 47 =="
## 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 5
## ERROR: factor num_of_earners has new levels 5
## ERROR: factor num_of_earners has new levels 5
## ERROR: factor householder_education has new levels None
## ERROR: factor householder_education has new levels None
## ERROR: factor num_of_earners has new levels 5
## ERROR: factor num_of_earners has new levels 5
## [1] "== 57 =="
## 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 5
## [1] "== 61 =="
## ERROR: factor num_of_earners has new levels 5
## ERROR: factor householder_education has new levels None
## ERROR: factor num_of_earners has new levels 5
## ERROR: factor num_of_earners has new levels 5
## [1] "== 66 =="

```

```

## 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 5
## [1] "== 70 =="
## 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 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
## 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 5
## ERROR: factor num_of_earners has new levels 5
## ERROR: factor family_size has new levels 7
## [1] "== 83 =="
## 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] "== 88 =="
## 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 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 7
## [1] "== 96 =="
## ERROR: factor family_size has new levels 7
## [1] "== 98 =="
## ERROR: factor num_of_earners has new levels 4, 5
## ERROR: factor num_of_earners has new levels 5
## [1] "== 101 =="
## ERROR: factor family_size has new levels 7
## [1] "== 103 =="
## ERROR: factor family_size has new levels 7
## ERROR: factor num_of_earners has new levels 5
## [1] "== 106 =="
## [1] "== 107 =="
## ERROR: factor family_size has new levels 7
## ERROR: factor family_size has new levels 7
## ERROR: factor num_of_earners has new levels 5
## [1] "== 111 =="
## [1] "== 112 =="
## [1] "== 113 =="
## [1] "== 114 =="
## 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 5
## [1] "== 119 =="
## ERROR: factor family_size has new levels 7

```

```

## ERROR: factor num_of_earners has new levels 5
## [1] "==" 122 ==
## ERROR: factor householder_education has new levels None
## [1] "==" 124 ==
## ERROR: factor family_size has new levels 7
## ERROR: factor num_of_earners has new levels 5
## ERROR: factor householder_education has new levels None
## ERROR: factor householder_education has new levels None
## ERROR: factor family_size has new levels 7
## [1] "==" 130 ==
## ERROR: factor num_of_earners has new levels 5
## ERROR: factor num_of_earners has new levels 5
## ERROR: factor householder_education has new levels None
## [1] "==" 134 ==
## ERROR: factor family_size has new levels 7
## ERROR: factor family_size has new levels 7
## [1] "==" 137 ==
## ERROR: factor num_of_earners has new levels 5
## ERROR: factor householder_education has new levels None
## ERROR: factor num_of_earners has new levels 5
## ERROR: factor householder_education has new levels Post-Graduate
## 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 5
## 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
## [1] "==" 148 ==
## 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 5
## ERROR: factor family_size has new levels 7
## [1] "==" 153 ==
## ERROR: factor num_of_earners has new levels 5
## [1] "==" 155 ==
## [1] "==" 156 ==
## ERROR: factor num_of_earners has new levels 5
## ERROR: factor householder_education has new levels None
## [1] "==" 159 ==
## ERROR: factor num_of_earners has new levels 5
## ERROR: factor householder_education has new levels None
## ERROR: factor family_size has new levels 7
## [1] "==" 163 ==
## ERROR: factor family_size has new levels 7
## [1] "==" 165 ==
## [1] "==" 166 ==
## [1] "==" 167 ==
## 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 5
## [1] "==" 171 ==
## 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

```

```

## ERROR: factor num_of_earners has new levels 5
## ERROR: factor family_size has new levels 7
## [1] "== 177 =="
## 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
## [1] "== 182 =="
## ERROR: factor num_of_earners has new levels 5
## [1] "== 184 =="
## [1] "== 185 =="
## ERROR: factor num_of_earners has new levels 5
## 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 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 num_of_earners has new levels 5
## ERROR: factor num_of_earners has new levels 5
## ERROR: factor num_of_earners has new levels 5
## ERROR: factor householder_education has new levels None
## 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 5
## ERROR: factor family_size has new levels 7

```

Conclusions

Model	ATT	ATE
Linear Regression	460.8	159.0
Extreme Gradient Boosted Trees	445.1	169.6

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.