Case Study IV Final Report

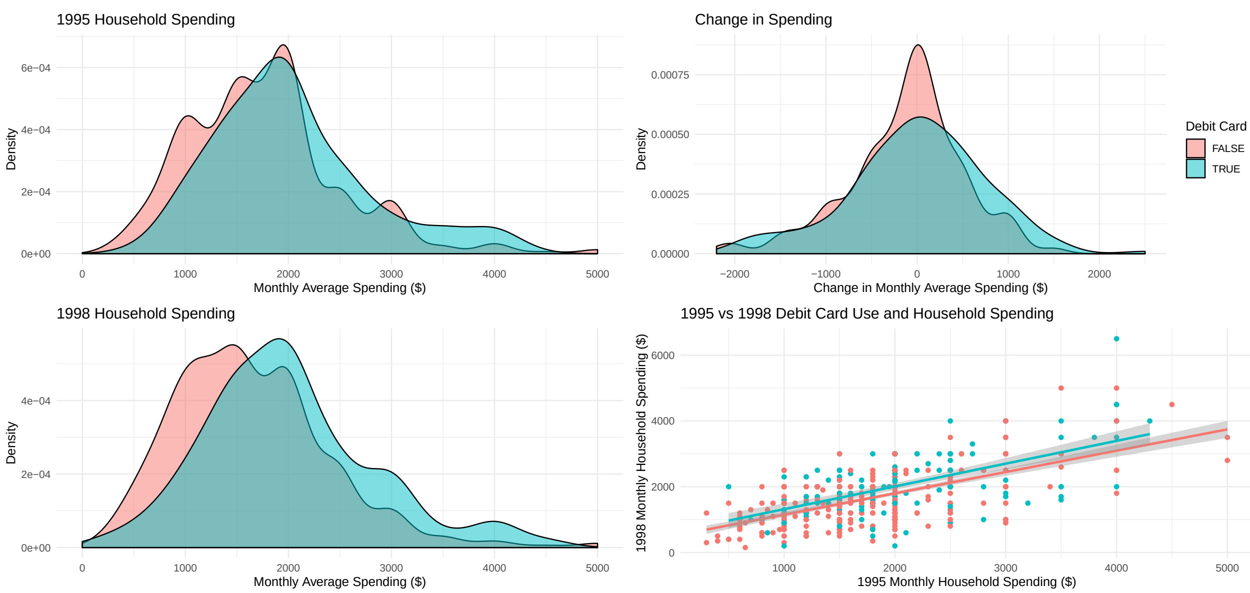
Jake Epstein, Daniel Spottiswood, Man-Lin Hsiao, Michael Tan, Sahil Patel

11/05/2019

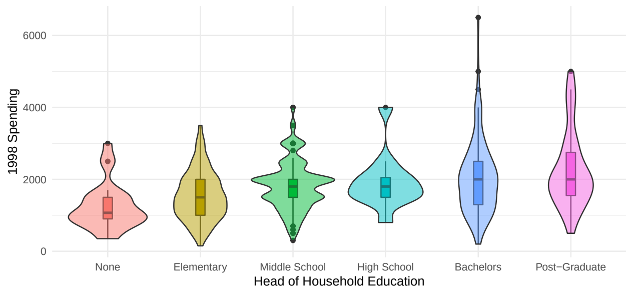
Introduction

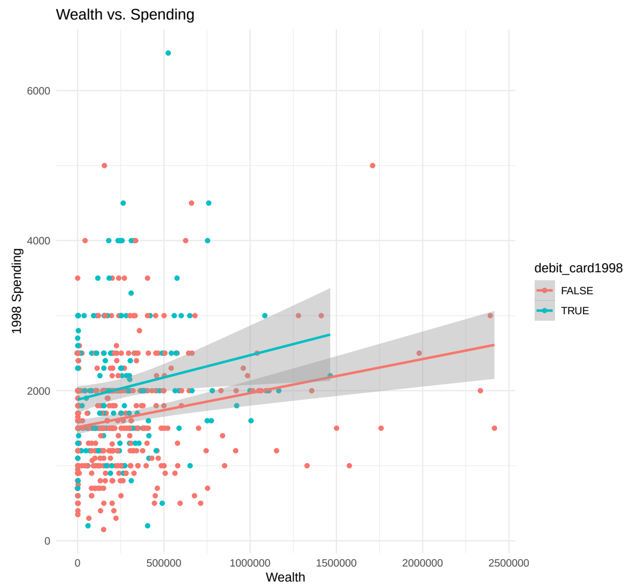
The goal of this case study is to evaluate the causal impact of debit card ownership on household spending. The data come from the Italy Survey on Household Income and Wealth (SHIW), a 1995-1998 survey of 584 Italian households. The dataset includes 1995 and 1998 monthly household spending, whether the 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.

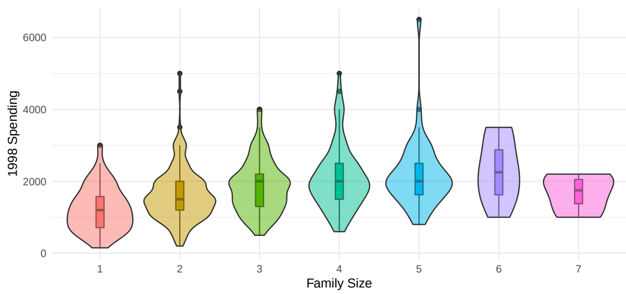
Exploratory Data Analysis



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.

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.

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. Since propensity score methods are sensitive to large outliers we trim the extreme data points outside the common support and also in the nonoverlapping regions.

We should note that the observational nature of this study means that the assumption of unconfoundedness is not necessarily met. However, it is difficult to rigorously test this assumption, so we move forward under the premise that this is met.

In order to improve the covariate balance so that we can proceed with causal inference, we performed matching and weighting methods on the data. Namely, we performed 1:1 matching with replacement, 1:3 matching with replacement, weighting with a basic logistic model (treatment ~ all the other covariates besides response), weighting with a random forest, and weighting with a classification and regression tree (CART). We created loveplots to illustrate the difference in ASD across covariates before and after adjustment. In each case, applying a matching or weighting method decreases ASD significantly. Using .1, a commonly used threshold for ASD, as our benchmark for balance, we see that most of our covariates exhibit at least a satisfactory level of balance.

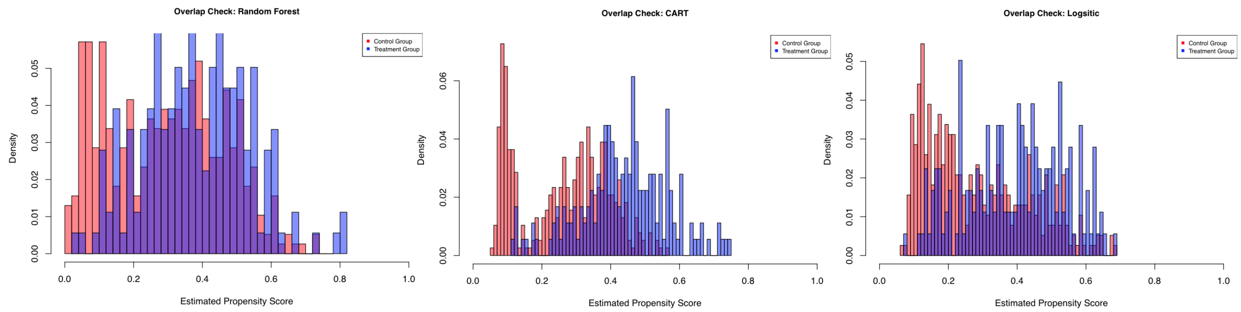

Overlap Check: Random Forest
Overlap Check: CART
Overlap Check: Logsitic
Control Group
Treatment Group
Density
Estimated Propensity Score

Figure 1:


1:1 Matching
1:3 Matching
spending1995
average_age_65+*
householder_age_65+*
cash_inventory
geograph_area_North*
family_size_1*
family_size_4*
average_age_<30*
average_age_31−40*
family_size_2*
num_of_earners_1*
householder_age_31−40*
geograph_area_South and Islands*
wealth
householder_age_41−50*
householder_education_Bachelors*
num_of_earners_2*
income
householder_education_Elementary*
family_size_3*
average_age_41−50*
interest_rate
geograph_area_Central Italy*
householder_education_None*
average_age_51−65*
householder_education_Middle School*
family_size_5*
householder_age_<30*
num_of_inhabitants_40−500k*
num_of_earners_3*
num_of_inhabitants_<20k*
num_of_banks
family_size_6*
num_of_earners_4*
num_of_inhabitants_500k+*
householder_education_Post−Graduate*
num_of_earners_5*
family_size_7*
householder_education_High School*
householder_age_51−65*
num_of_inhabitants_20−40k*
Absolute Standardized Mean Differences
Sample
Unadjusted
Adjusted
Basic Logistic Weighting
Random Forest Weighting
spending1995
average_age_65+*
householder_age_65+*
cash_inventory
geograph_area_North*
wealth
family_size_4*
family_size_1*
average_age_<30*
average_age_31−40*
family_size_2*
num_of_earners_1*
householder_age_31−40*
geograph_area_South and Islands*
householder_age_41−50*
householder_education_Bachelors*
householder_education_Elementary*
interest_rate
num_of_earners_2*
income
average_age_41−50*
family_size_3*
average_age_51−65*
geograph_area_Central Italy*
family_size_5*
householder_education_Middle School*
householder_age_<30*
num_of_inhabitants_40−500k*
householder_education_None*
num_of_earners_3*
num_of_inhabitants_<20k*
family_size_6*
num_of_earners_4*
num_of_banks
householder_education_Post−Graduate*
householder_age_51−65*
num_of_inhabitants_500k+*
num_of_earners_5*
family_size_7*
Sample
Unadjusted
Adjusted

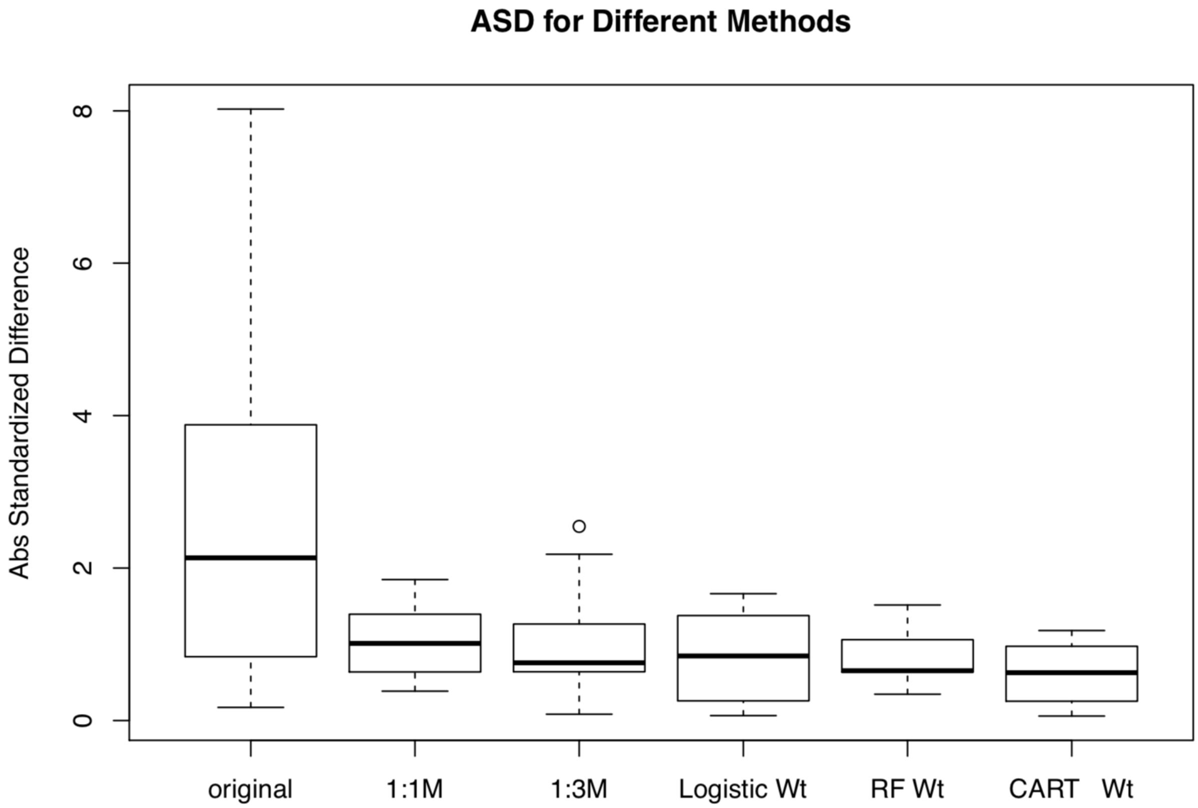


Figure 2:

With weighting, we used propensity scores that resulted from the basic logistic model, random forest, and CART. We created overlapped histograms or density plots of the estimated propensity scores of treatment vs. control groups to provide good summaries of the balance of the covariates between groups. Then, we made sure to only conduct analysis on the overlapped regions, as causal inference is impossible in the non-overlapped regions.

In the end, the results were outputted in the following graphic. Given that 1:1 matching yields a strong ASD, and has strong covariate balance we will proceed with this subset of data with further regression adjustment. We chose this method as it is more simplistic and produces strong results.

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

Regression Adjustment

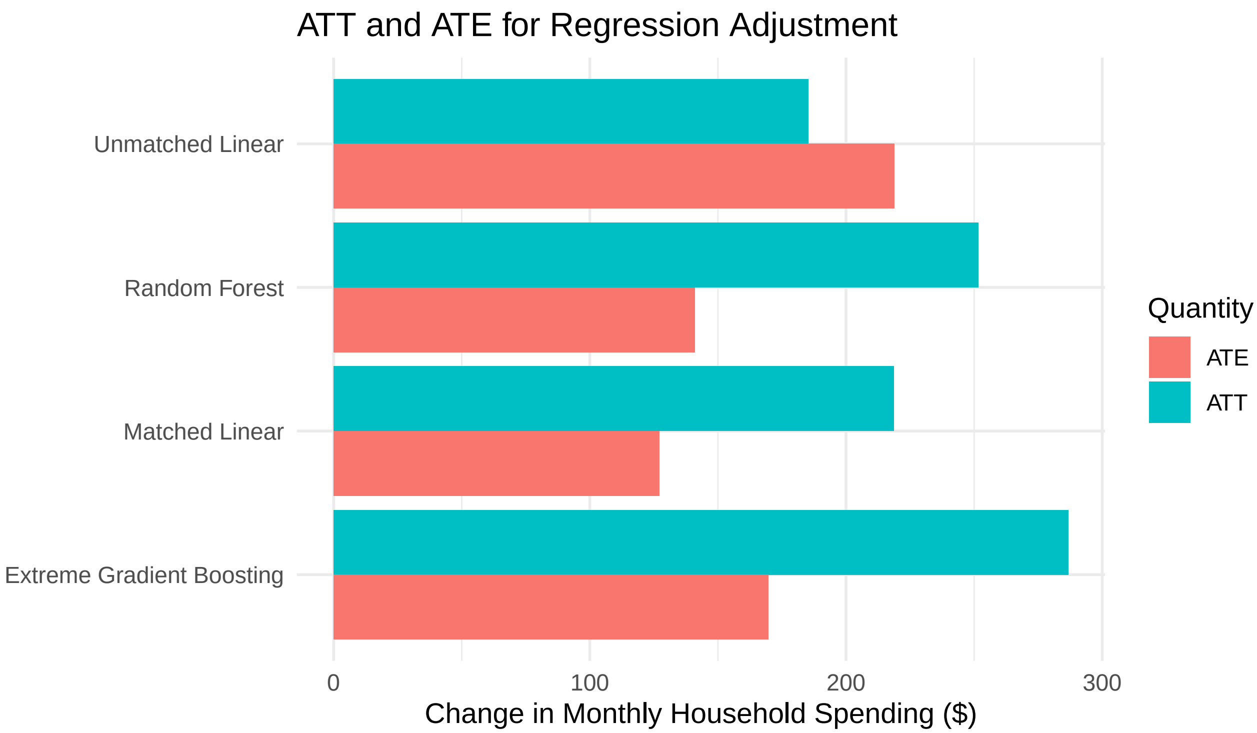
We attempt to fit models on top of the 1:1 matching with replacement as well as on our original dataset. We first use a basic linear model before fitting more flexible models, random forest and extreme gradient boosting.

The motivation for regression adjustment is to impute missing potential outcomes by fitting models to the unobserved counterparts of subsets of the observed data. We have elected to include a regression adjustment in addition to our propensity score matching as it allows us to mitigate the sensitivity of our models to the model specification used to impute these unobserved values.

As described above, our 1:1 matched data has demonstrated the most compelling distribution of absolute standardised differences and therefore yields the strongest results for the overlap of data through a propensity score methodology. Consequently, we partitioned the data into treatment and control groups through the 1:1 matching method, and then imputed unobserved control values for the observed treatment values, and vice versa for the observed control values.

We began testing our models by running a simple linear (OLS) model on the original unmatched dataset. We expect this method to have high bias due to the imbalance of the covariates in the initial dataset. This model suggested that the effect of individuals who were observed to own debit cards in 1998 spent on average 185 more than those who did not. Furthermore, the unmatched model suggests that after imputing control values for those who did own a debit card in 1998, that the implied treatment effect increased to 218 in spending.

We then performed regression adjustment on top of the matched data set in order to reduce sensitivity caused by an imbalance of the covariates across groups. By coupling this with flexible models, we have utilized both design and anlaysis based approaches to reducing model sensitivity. The output from some of the models we tested is shown below.



In terms of the flexibility of our three models, the machine learning models (the gradient boost and random forest) are intrinsically more flexible than our OLS model as they learn from previous errors in order to fine tune their ability to predict and are just a collection of weak classifiers. That being said, it somewhat cumbersome to ascertain which of our machine learning models is in fact more flexible mathematically. The forest is useful as it is easier to fit as it only depends on the number of trees and the number of factors selected (the boosted model depends on more features); in addition, the forest does not overfit the data as much as the boosted model. Nevertheless, the boosted model is also compelling as it is capable of optimising a broader selection of objective function than the forest, and typically leads to more accuracy with fewer trees (the forest tends to prefer smaller trees leading to lesser accuracy in some cases). The complexity of the data at hand has led us to prefer our extreme gradient boosting model; however, either would be reasonable choices in this scenario and lead to similar conclusions. Moving forward, we will report more detailed results for each of these models, save random forest, as well as results from the doubly robust method.

	R Adj. w/out Match	Weighting	Double Robustness	R Adj. Linear	R Adj. XGBoost
ATE	218.937	112.097	126.725	127.194	169.612
ATE SE	32.669	53.429	68.215	83.780	53.908
ATT	185.354	169.085	258.143	218.625	286.751
ATT SE	67.318	64.036	66.849	65.464	78.035
ATT/AOT	0.091	0.083	0.127	0.108	0.141
P-value	0.003	0.004	0.000	0.000	0.000

Figure 3:

Bootstrapping

In order to estimate the error and significance of our models for ATT and ATE, we generated 500 bootstrap samples of the data. Using these bootstrap samples, we recalculated the propensity scores and estimated the resulting causal effects for each dataset accordingly. In our bootstrapping process, we evaluated both the ATT and ATE for weighting, weighting with double robustness, regression adjustment with a linear model, 1:1 matching with a linear regression adjustment, and 1:1 matching with a gradient boosted regression adjustment.

Although bootstrapping is computationally intensive, we chose it over the more efficient delta method in order to avoid underestimating the uncertainty.

The results of bootstrapping support our aforementioned belief in the gradient boosted model. As portrayed in the table above, the boosted model showed significant (as estimated by the model's p-value) impact of the treatment and had the lowest estimated standard deviation (via bootstrapping) for ATE across all models explored. Interestingly, the boosted model did have the highest bootstrap estimate for ATE standard deviation, however given the desirability of the model's other metrics and the flexibility of the model holistically, we continue to endorse the gradient boosted model. This is not meant discredit any of the other methods as it is very promising to see the consistency and significance across several different methods.

Conclusions

Overall, our investigation into the effects of debit card ownership on spending has demonstrated that there is in fact a positive treatment effect associated with owning a debit card in 1998. While it may prove difficult to extrapolate the results of this study to other geographies and time periods (for example), we believe that through implementing 1:1 matching and regression adjustments with flexible machine learning models, like gradient boosting, we have achieved the ability to draw causal inference. We see that that the methods suggest the increase in monthly consumption for a household with a debit card ranges from 8.3% to 14.1% and 169 to 286, with each of these models showing definitively that there is a statistically significant increase. The AOT, average outcome for the treated, is approximately 2028 in this dataset.

These results support psychological research in this area and our general intuition that debit cards reduce the emotional connection to money causing us to be more willing to spend. We hope that further research continues to explore this connection as technology continues to make it easier and more efficient to transfer, spend, and earn money. It is important for us to understand the psychological impact that digital assets have on our spending/saving habits.