CS396: Final Report

Jose Cordova, Amir Dhami, Dennis Wu

June 8 2023

1 Academic Integrity

The work described in your project report must contain substantial original work. You are of course welcome to use published libraries and code, but you must cite those resources. Your report should make it clear what parts of your project was copied or adapted from existing resources, and which parts were original work. Do not take credit for work you copied from other resources. Misrepresenting your work or failing to cite your sources will result in point deductions and may be grounds for a report of academic misconduct. If you have questions or concerns, please ask for help! You may use ChatGPT or another LLM to help write your report, but if you do, you must include at least two sentences describing how you used it. You are responsible for all content submitted for grading; if we can't understand your report, saying that someone or something else wrote it is not an excuse. If it fails to cite sources or otherwise violates academic integrity policies, you will be held responsible. Thus, please triple-check that any text generated on your behalf accurately describes your work.

2 Code and Documentation

data.pynb: This is our only file, it includes the bootstrap and backdoor functions taken from HW2, calculations using theses functions, and the double ML estimators used (from econml package) along with some code for ploting our results.

3 Estimation Implementation

Cell 4 includes calculations done for the project update and general exploratory analysis.

Cell 5 includes the backdoor bootstrap estimation

Cell 7 Includes fitting the double ML estimators and the bootstrap estimates for their average treatment effect.

Cell 8 includes code to plot shapley values for the ML models.

We decided to implement Double Machine Learning (DML). This type of estimates use 2 types of nested models, hence double. One fits the treatment to the features and the other fits the outcome to the features. The reason behind this is twofold, first it reduces the inherent bias of ML algorithms and second, it allows for the benefits of using ML algorithms over traditional statistical models (e.g., OLS). This benefits include ease to handle high dimensional data and no constraints when fitting the models (i.e., non parametric models can be used). In our case it is mainly the second one that attracted us to this approach. In our update we used a backdoor estimator relying in a OLS regression but we believe that the relationships between the variables can't be captures by this model e.g., the interaction between age/educ and salary is probably not linear and the interactions between races and age/educ might not be reproducible by ols formulas. We implemented 2 DML models, one with a linear final stage (i.e., the last stage model that fits outcomes to features is linear) and a non parametric model based on random forests.

The coefficients in the linear model of the constant marginal treatment effect of the the linear DML estimator are presented below

Table 1: Coefficients

	Black	Hispanic	Marriage	Education	Age
Linear Model	1854.9291	872.355	1449.483	410.8943	47.118

4 Changes Since the Update

Our main analysis of the dataset has remained the same since the project update. We shifted our goal from measuring the possible unobserved confounding caused by criminal records to using double machine learning to account for possible measurement errors. We were unable to find a proper estimator for the new DAG to account for unobserved confounding. To account for the possibility of using an incorrect model to regress the data upon, we used double machine learning (the final cell in data.ipynb).

5 Interpreting Your Results

In the project update assignment, we asked you to interpret your results: What does your estimate of the causal effect mean for your problem? Tie this back to the reasons why you were interested in this dataset in the first place. If you were able to make policy decisions based on your analysis of this data, what would these results tell you? How does the uncertainty in your estimate influence your interpretation? Answer these questions again in the subsequent subsections: both before and after the additional challenges you considered since your project update.

5.1 Before

People who underwent job training were expected to earn nearly 1.4 times more than those who didn't. The training program generally leads to higher expected income, particularly benefiting individuals with at least a high school degree. Although the program has a smaller but still positive effect on individuals without a high school degree, having a high school degree is generally advantageous for higher income. The impact of the job training program on earnings is greater than that of a high school degree alone, but having a high school degree influences the effectiveness of the program. One potential improvement is to provide opportunities for individuals without a high school degree to obtain an equivalent degree while participating in the job training program. Although the findings cannot be directly applied to the current job market due to the prevalence of high school degrees, they underscore the importance of education on future earnings and the effects of job training programs. Therefore, an appropriate interpretation is to compare high school degrees in the 70s to college degrees today.

5.2 After

The previous findings validate the observations made in the update, particularly when considering the means. The training program exhibits a positive effect overall, as evidenced by an Average Treatment Effect (ATE) of 1560.381 and 1616.845 in the entire population, as estimated by the linear and non-parametric Double Machine Learning estimators. This implies that program participants earned an average of 1560.381 more than non-participants. Furthermore, the influence of the program is more pronounced among individuals with a high school degree. The ATE for this subgroup is nearly triple the effect observed for individuals without a high school degree.

Table 2: Backdoor Bootstrap Intervals

	a = 0	a=1
$E[Y^a]$	$[3944.451, 5272.808]$	$[5249.235, 7468.709]$
$E[Y^a \vert N = 1]$	$[3243.197, 5854.330]$	$[6528.325, 10429.6092]$
$E[Y^a \vert N = 0]$	$[3766.817, 4480.426]$	$[5221.072, 7032.931]$

Table 3: ATE LinearDML

	Mean	Interval
Overall	1560.3811	$[-1498.8913, 4619.6536]$
Given degree	3075.2911	$[-4893.3486, 11043.9308]$
Given no degree	1176.3742	$[-2575.8534, 4928.6020]$

Table 4: ATE NonParamDML

	Mean	Interval
Overall	1616.8459	$[-2664.2672, 5897.9591]$
Given degree	3081.8558	$[-1320.6522, 7484.3638]$
Given no degree	1161.0876	$[-3348.2886, 5670.46397]$

The following figures represent the shapley values for the DML estimators, the distribution of the shapley values of the different features highlight the key difference between these. For example, binary features (black, hipanic, marriage) have only two values associated with them in the linear DML but there are multiple values associated with them in the non parametric estimator. We can only see how the shapley values of age in the nonparametric model don't have a clear trend like in the linear DML, this might show some support for our hypothesis that the relationship with the outcome are not linear. On the other hand, the education years still show a clear positive correlation with the outcome.

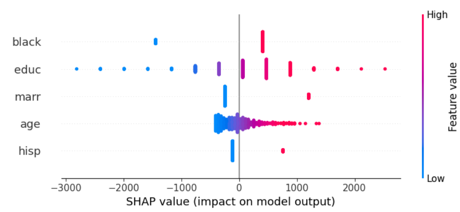

Figure 1: Shapley Values Linear DML

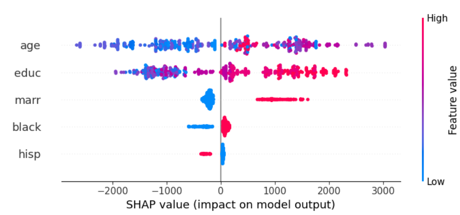

Figure 2: Shapley Values Non Parametric Estimator

6 Reflections

This section is your chance to reflect on this project. You should write at least a few sentences for each of these questions. However, try to be concise; a longer answer is not necessarily a better answer. If you want to write several paragraphs about something you're excited about, that's great! On the other hand, don't just write several paragraphs listing small implementation details; if we fall asleep reading your report, you might lose points.

6.1 What was interesting?

Something that we learned from this project that wasn't covered in class was how to adapt and adjust the focus of our project when certain data wasn't available. The lectures on estimators, d-separation, and machine learning were most relevant to our project. The reading on conditional exchangeability was also relevant.

6.2 What was difficult?

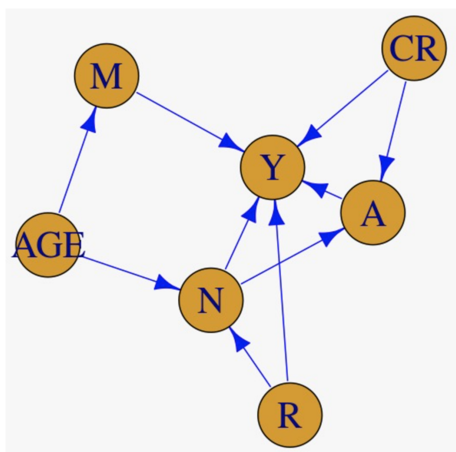

Figure 3: New DAG

The most difficult part of this project was not being able to account for possible unobserved confounding. We were unable to specify a new estimator for the following DAG that would account for participants' criminal records (CR). If someone were to start a similar project from scratch, I would recommend them to spend more time thinking about possible unobserved confounders before diving into the dataset.

6.3 Unaddressed challenges

As mentioned previously, participants' criminal records, or lack thereof, possibly played a role in their average real earnings. Our project is unable to account for this unobserved confounding. Additionally, in order for our backdoor estimator to work, we assume conditional exchangeability and consistency. Since the initial study the data is retrieved from was not an RCT, this is not an empirically-backed assumption. Based on existing knowledge, criminal records have a highly correlative effect on securing a job. Because we cannot account for this relationship, the possible unobserved confounding likely has a bigger effect on our results

6.4 What's left to do?

If we had to spend another month working full-time on this project, we would try to account for unobserved confounding. We would test different estimators on our redefined DAG to account for participants' criminal records. Additionally, since we don't have access to which participants had criminal records, we would consider identifying instrumental variables that can help address the unobserved confounding and explore their inclusion in our analysis.

If Northwestern gave our group 1,000,000 USD to run an experiment, we would consider running a quasi-randomized control trial, while accounting for participants' criminal records. This would remove that instance of unobserved confounding, and allow us to better answer whether our treatment has a causal effect on our outcome.