University of St. Gallen

Data Analytics II

Self Study Documentation

Monte-Carlo Simulation

Arbian Halilaj
16-609-828

Lecturer: Prof. Michael Lechner

May 1, 2022

1 Setting

This documentation examines the performance of two different estimators, **OLS** and **DoublyRobust**, and applying a Monte Carlo simulation for the **Selection-on-Obervables study design**. To operate this strategy we need the following identifying assumptions:

Assumption 1 (Conditional Independence) $(Y_i(0), Y_i(1))\, W_i \mid X_i = x, \forall x \in \text{supp}\,(X_i)$.
Assumption 2 (Common Support) $0 < P\,[W_i = 1 \mid X_i = x] < 1, \forall x \in \text{supp}\,(X_i)$.
Assumption 3 (SUTVA) $Y_i = W_i \cdot Y_i(1) + (1 - W_i) \cdot Y_i(0)$.
Assumption 4 (Exogeneity) $X_i(0) = X_i(1)$.

The performance measurement happens through the Mean Squared Error (*MSE*), the *Bias* and the *Variance* of the estimated **Average Treatment Effect** (*ATE*), which is the parameter of interest. The two estimators differ in their identification strategies. On one hand, the mean outcome regression from the OLS estimation gives the *ATE*, on the other hand, the *ATE* is estimated (1) through weightings from the propensity score estimated with logit and (2) also includes the mean outcome regression, which is the main difference between this and an IPW estimator. Therefore, this *ATE* estimation technique yields a doubly robust property. This means that in case we misspecify one of those two parameters, we still receive unbiased *ATE*. Furthermore, I want to mention that we initially wanted to compare the DoublyRobust estimator with the Double ML approach of Chernozhukov et al. (2018). The idea was to show that "fancy" ML is not always better than traditional methods. Unfortunately, we could not find any DGP's in which the performance of the Double ML approach exceeded the DoublyRobust approach. Reason for this are that either our algorithm was misspecified or that our GDP's were rather low-dimensional in order to not get issues with long during calculation times.

The performance of the estimated *ATE* will be examined under three different settings of Data Generating Processes (DGP). and the Monte Carlo simulation will run 1'000 replications.

2 Monte Carlo Simulation Design

In a Monte Carlo Study the data used for the analysis is being simulated. The model baseline model of interest can be described as follows:

$$y = dy^1 + (1 - d)y^0$$

DGP 1

In the first DGP the model of interest can be described as follows:

$$\begin{aligned}
&y^0 \sim \mathcal{N}(0, 1) \\
&y^1 = -1 + g(\mathbf{X}^{N\times K}) + u, \quad \text{where} \quad \mathbf{X}^{N\times K} \sim \mathcal{N}(0, 1)\ \&\ u \sim \mathcal{N}(0, 1) \\
&g(\mathbf{X}^{N\times K}) = \mathbf{X}\text{'}^{N\times K}\beta \quad \text{where} \quad \beta = (1/K, ..., K/K) \\
&d \sim \mathcal{B}(p = 0.5)
\end{aligned}$$

For this simple DGP all identifying assumptions are fulfilled such that we can estimate an unbiased ATE as $E[y_i^1 - y_i^0] = -1$. We expect that both estimators yield approx. the same result. The reason for this is that OLS is embedded in the DoublyRobust estimator. As far as I understand OLS can never outperform DoublyRobust but vice versa is possible. This case is presented in the DGP2.

DGP 2

The second DGP is similar to the first DGP. The only difference is that we allow for a heterogeneous treatment effect. This can be done by changing the treatment probability p to one different from $p = 0.5$. Mathematically:

$$d \sim \mathcal{B}(p = 0.8)$$

Here, we expect the OLS to perform worse than DoublyRobust. The reason is that we violate OLS homogeneity assumption which results in an increase of the standard errors. The DoublyRobust approach, since it incorporates also propensity scores, is expected to stay unbiased.

DGP 3

The goal of the final DGP should be to allow one of the identifying assumptions to fail. More concretely, we let the CIA assumption fail. This assumption states that the potential outcomes y^0, y^2 are independent of the treatment assignment d once we control for the confounders $\mathbf{X}^{N \times K}$. Thus, a violation involves unobserved confounders which affect both potential outcomes and the treatment assignment. Implementing this is straightforward. By including a new unobserved variable v which we let interact with the treatment probability and which we include in the outcome equation y. This increases the ATE by the mean of $v = 0.5$ to 1.5.

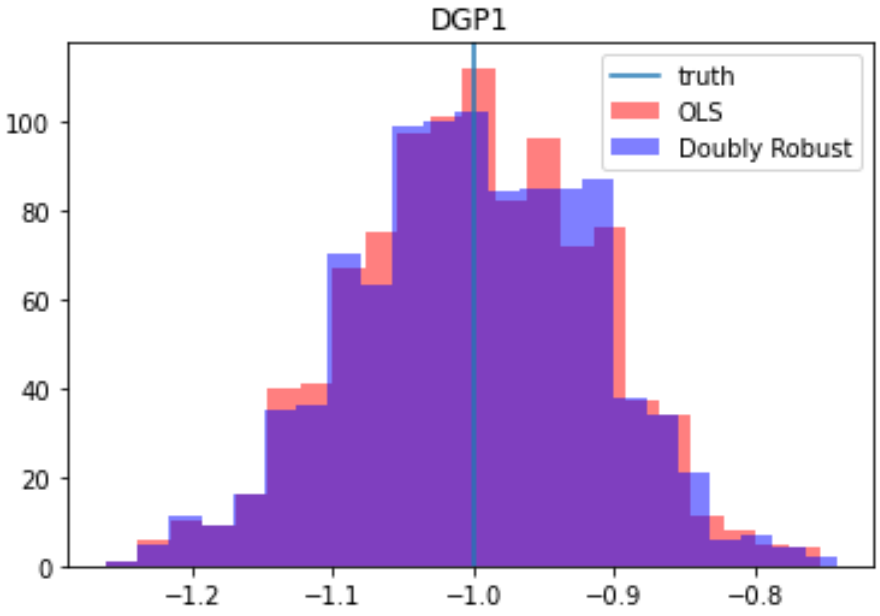
$$y = dy^1 + (1 - d)y^0 + v, \quad \text{where} \quad v \sim \mathcal{U}(0, 1)$$

Since this random variable v is not observable, CIA is violated. We expect no particular breakdown in the estimation procedure since our estimators are not violated. Thus, we expect similar to DGP2 since OLS has an increased bias due to heterogeneity, but the ATE of both estimators will be biased.

3 Results

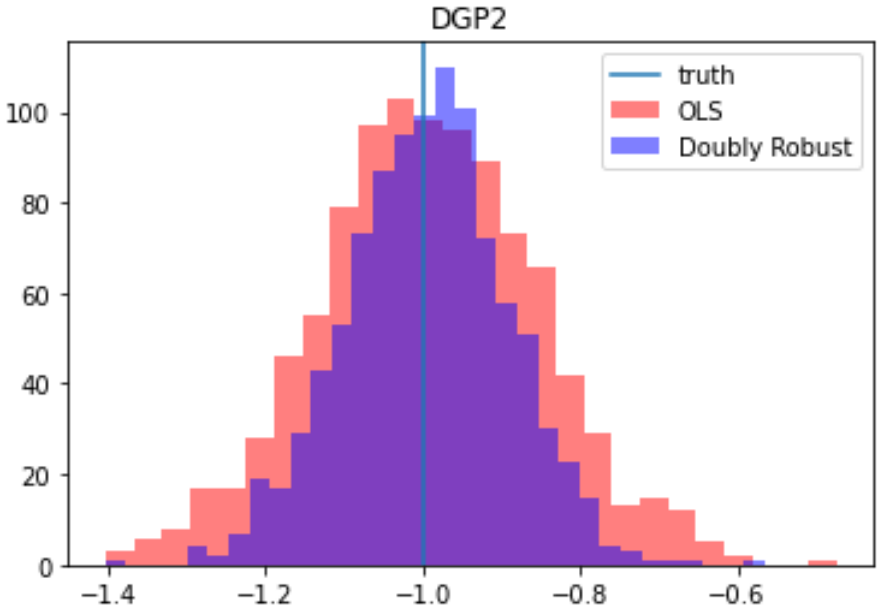
DGP 1

The following histograms show the empirical distribution of our estimated ATE for both OLS and DoublyRobust. Please note that we set K, the number of observable confounders to 10 and the number of observations to $N = 1'000$. As expected, OLS and DoublyRobust accurately capture the ATE of DGP1.



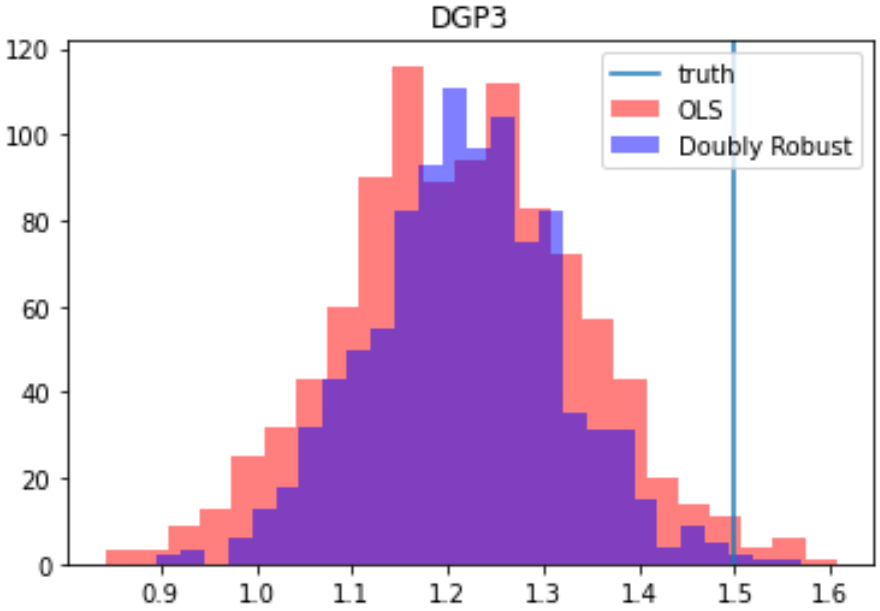
DGP 2

We have also met our expectations with DGP2. Due to heterogeneity the asymptotic ATE is slightly biased due to the variance. The DoublyRobust approach accurately captured the effect due to its inherent propensity score estimate.



DGP 3

Finally, also this histogram is in line with our expectations. DoublyRobust performs better than OLS due to heterogeneity and both estimators cannot capture the true *ATE* due to unobserved confounders.This shows the importance of doing one's best to comply with the CIA.



The following table summarizes the findings in numbers by means of the performance measures stated in section one.

	DGP1		DGP2		DGP3	
	OLS	DR	OLS	DR	OLS	DR
Bias	0	0	0.007	0.005	-0.287	-0.287
Variance	0.008	0.007	0.02	0.011	0.015	0.01
MSE	0.008	0.007	0.02	0.011	0.098	0.09