

Estimating Heterogeneous Treatment Effects Using Causal Forests:

**Two Applications Concerning Romanian Education System and Incentives
For College Achievement**

ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Bachelor Thesis: International Econometrics and Operations Research

Student name: Cheng Jiang

Student ID number: 448433

Supervisor: Andrea Naghi

Second assessor: Andreas Pick

Date final version: 06/07/2019

**The views stated in this thesis are those of the author and not necessarily those of
Erasmus School of Economics or Erasmus University Rotterdam.**

Abstract

The causal forests algorithm is widely known as a machine learning technique that is powerful in drawing causal effect inferences of a treatment. Meanwhile, linear regression model as a more statistics-based technique can also estimate treatment effects. In this study, we apply both methods to two different datasets concerning the Romanian Baccalaureate exam outcomes and the students' college achievement measures to answer the following two questions. The first one is do the two estimation methods give similar results for estimating average treatment effects, and the second one is can causal forests generate accurate treatment effect estimates if being trained with few observations. The empirical findings show that the two methods can give accurate and similar estimation results if large enough dataset is utilized. Moreover, when being trained with small dataset, the causal forests algorithm cannot correctly give estimates however, it can still capture potential treatment heterogeneity.

Keywords: causal forests, causal effect inference, treatment heterogeneity, linear regression model

Contents

- 1 Introduction** **2**

- 2 Data** **4**
 - 2.1 Data of the Romanian education system application 4
 - 2.2 Data of the incentives for college achievement application 5

- 3 Methodology** **6**
 - 3.1 Difference-in-differences 6
 - 3.2 Causal forests 7
 - 3.2.1 An introduction to causal forests 7
 - 3.2.2 Asymptotic inference of causal forests 11
 - 3.2.3 Cluster analysis 12
 - 3.2.4 Inspecting treatment heterogeneity 12

- 4 Results** **14**
 - 4.1 Results of Romanian education system application 14
 - 4.2 Results of incentives for college achievement application 17

- 5 Conclusion** **20**

- References** **22**

- Appendices** **24**

1 Introduction

As a machine learning technique, causal forests algorithm which extends the random forests algorithm that was first developed by Breiman (2001), is widely considered to be one of the most efficient techniques in respect of drawing causal effect inferences of a treatment. A brief understanding of how causal forest operates is that, the forest generates estimate for the treatment effect by averaging the estimates produced by all the decision trees, which are grown by splitting the data points conditionally on the predictive covariates in such a way to achieve maximum model fit. Generally, causal forests algorithm is data-hungry.

Wager and Athey (2018) suggest that the nonparametric causal forests algorithm can generate pointwise consistent estimate for the true treatment effect. Moreover, several other researchers reveal the fact that the estimates are asymptotically Gaussian distributed which enables us to formally examine the significance of those estimates: (1) potential nearest neighbors construction introduced by Yi and Yongho (2006) and classical analysis tools proposed by Hoeffding (1948) and Hajek (1968) indicate that estimates made by causal forests are Gaussian and unbiased; (2) infinitesimal jackknife for random forests suggested by Efron (2014) and Wager, Hastie, and Efron (2014) implies the way to estimate their asymptotic variance. Furthermore, Chernozhukov et al. (2017) has established the “best linear predictor” method to investigate the accuracy of the mean forest prediction as well as the quality of the treatment heterogeneity. Except causal forests, another difference-in-differences (DD) strategy as a statistical method can also assess the treatment effects. In general, one can simply represent the strategy via a linear regression model where the dependent variable is the outcome being treated, meanwhile a dummy variable which can indicate the existence of treatment forms part of the regressors. On the whole, the most prominent difference between DD and causal forests approaches is that, there exists efficient built-in algorithm in causal forests to detect treatment heterogeneity on covariates, whereas there are no sufficient exploitable properties about the DD strategy that can imply potential treatment heterogeneity, apart from tediously adding covariates into the regression model and testing the significance of their coefficients.

Considering above, we will devote the bulk of our study to investigate and answer the following two major research questions, namely (1) whether the DD and causal forests approaches would result in similar estimates for the treatment effects (for the purpose of making extension on the reference paper, we also devote part of this study to estimate the heterogeneous treatment effects in the first application, we will explicate this later), (2) can causal forests algorithm

give accurate treatment effect estimates and capture heterogeneity using small dataset however, with large number of covariates. To that end, we will apply causal forests to two datasets. The first one is about the Romanian Bacculaureate exam outcomes which is collected and provided by Borcan, Lindahl, and Mitrut (2017), while the second one is about the college achievement measures which is collected and provided by Angrist, Lang, and Oreopoulos (2009). More details on the two applications as well as our findings are given in the next paragraph.

Initiated in 2011, a national anti-corruption campaign was launched in Romanian to fight against the corruption in the Bacculaureate exam by two means: the first one is increasing punishment on students and teachers caught bribing and the second one is installing cameras to monitor students during exams. Thus, we will examine the treatment effects on the students' Bacculaureate exam outcomes by the "camera" treatment. To that end, we employ a dataset containing 100,000 observations that partly consist of 22 covariates. Our empirical findings show that, the causal forests algorithm and DD strategy give rather similar results of the estimated average treatment effects by camera such that both suggest students' exam performance can be undermined under the camera monitor. Moreover, results suggest that covariates such as county id, school id, standardized ability measure and middle school GPA cause remarkable treatment heterogeneity. In the incentive for college achievement application a SFSP programme is thought to influence students' college achievement, therefore, we use causal forests to estimate the treatment effects on those achievement measures by the SFSP. The aim of this application is to study whether causal forests algorithm can accurately estimate treatment effects as well as capturing treatment heterogeneity with small dataset containing large number of covariates, specifically, we employ a dataset containing 1102 observations which partly consist of 26 covariates. In brief, the results indicate that causal forests algorithm cannot correctly generate treatment effect estimates with few observations. Yet, it can still capture potential treatment heterogeneity if being trained with large number of covariates.

This paper mainly focus on extensions, namely answering the two central research questions as mentioned earlier. The replication forms a small part of this paper with the results given in Appendix A B and C. We replicate the reference paper Borcan, Lindahl, and Mitrut (2017) by replicating table 2, 4, and 6 from it, and we obtain the same results as the reference paper.

The remainder of this paper is arranged as follows. Section 2 will describe the data utilized in both of the two applications. Section 3 will explicate the main methodologies which we apply to conduct our analysis. Section 4 will show the main results of our study. Finally, in section 5 we will reveal our conclusions.

2 Data

2.1 Data of the Romanian education system application

We use two datasets to conduct our analysis for the first application. The first one is the administrative data provided by the Ministry of Education documenting the information of students who enrolled at the Bacalaureate exam during the period between year 2009 and 2012. This dataset contains the exam outcomes of students especially the three that are of our main interests, i.e., Romanian written score, Bacalaureate exam pass indicator, and overall Bacalaureate score. Also, the “camera” treatment information is included in the data which consists of dummy variables that equal to one if students were monitored by cameras during the exam and zero otherwise. Moreover, it enables us to obtain students’ background information such as gender, high school GPA, and date of birth, etc. In order to train our causal forests, students’ poverty status is one of our primary considerations for originating heterogeneity in the treatment effects, as the first dataset does not provide the relevant information, we therefore utilize a second one which consists of the information on students’ eligibility for the Money for High School (MHS) public program of financial assistance for high school students from poor households, covering the data of students who took the Bacalaureate exam between year 2009 and 2012. The requirement for a student to be eligible for the financial assistance is that, each household member of a student should have a gross monthly income less than 150 RON (*US\$45*) in the three months before the application to the MHS financial assistance. Correspondingly, we create a dummy variable called *poor*, which equals to one if a student was eligible for the financial assistance for at least one year during high school and zero otherwise. In short, we use 22 covariates in total to train our causal forests, please refer to Appendix D for more detailed explanations of the selected covariates.

Overall, the final merged dataset consists of 731,505 observations containing the information of students coming from 42 different counties, however, due to the computational capacity limit of our computer when running with **R**, we need to effectively reduce the dataset being applied. Therefore, we randomly select observations from the whole dataset to create a subsample as our working dataset. Our final dataset contains 100,000 observations. Table 1 below shows some of the main summary statistics of our working dataset.

Table 1: Summary statistics for the working dataset of the first application.

	2009		2010		2011		2012	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Baccalaureate pass indicator	0.860	0.347	0.767	0.422	0.555	0.497	0.538	0.499
Romanian written score	7.063	1.769	7.341	1.558	6.522	1.986	6.406	2.054
Overall Baccalaureate score	8.251	1.082	7.230	1.591	6.315	1.950	6.221	2.117
Romanian oral score	8.678	1.387	2.549	0.657	2.589	0.648	2.574	0.662
Middle school GPA	8.639	0.918	8.633	0.930	8.671	0.919	8.775	0.872
Standardized ability measure	0.003	0.990	0.015	0.991	0.023	0.984	0.047	0.971
Poor	0.203	0.402	0.220	0.415	0.232	0.422	0.222	0.415
Male	0.455	0.498	0.465	0.499	0.463	0.499	0.446	0.497
Rural	0.043	0.203	0.049	0.216	0.052	0.222	0.053	0.225
Theoretical track	0.509	0.500	0.485	0.500	0.506	0.500	0.548	0.498
Observations	26648		26109		24561		22682	

Notes: For the main exam outcomes, the Romanian written score and overall Baccalaureate score have a 1-10 scale, the Baccalaureate pass indicator is a dummy variable which equals to one if a student passed the overall Baccalaureate exam and zero otherwise.

2.2 Data of the incentives for college achievement application

As mentioned earlier in the introduction section, the purpose of the incentives for college achievement application is to discover whether causal forests algorithm can perform well detecting heterogeneity with small number of observations while with large number of covariates. Hence, we employ a dataset containing 1102 observations with 26 covariates. In comparison with the dataset of the first application which consists of 100,000 observations with 22 covariates, the number of observations is notably smaller and number of covariates is slightly larger in the second dataset.

We work with a dataset containing students' information who were studying in an university in 2005, the university was originally a commuter school. In order to explicate the functions of different variables, we classify the variables pertained to the dataset into three categories. The first one measures students' college achievement, i.e., GPA, good standing and credits earned of both first and second university year. The second one is the SFSP treatment, which is a dummy variable that equals to one if a student received the SFSP service and zero otherwise. The SFSP consists of two components, i.e., a service strategy known as the Student Support Program (SSP), and an incentive strategy known as the Student Fellowship Program (SFP). The SSP provides students with opportunities to receive a peer-advising service and a supplemental instruction service by means of Facilitated Study Groups (FSGs). While the SFP gives students chance to win merit scholarships for obtaining satisfactory grades in the first university year.

Finally, the last category consists of students’ background information, e.g., age, gender, and parents’ education levels, which we use to construct our covariates set. Totally, there are 26 covariates, please refer to Appendix E for more detailed explanations of those covariates. Table 2 below shows some of the main summary statistics of our working dataset.

Table 2: Summary statistics for the working dataset of the second application.

	GPA year 1	GPA year 2	Good standing year 1	Good standing year 2	Credits earned year 1	Credits earned year 2
Mean	1.916	2.096	0.544	0.749	2.576	2.893
SD	0.849	0.877	0.498	0.434	0.817	1.224
	Father’s education	Mother’s education	Female	High school GPA	Age	Work
Mean	4.975	4.516	0.576	78.992	18.249	0.767
SD	2.329	2.271	0.494	4.253	0.551	0.423

Notes: GPA, good standing and credits earned are the three main college achievement measures. GPA has a scale of 1-4, while good standing is a dummy variable which equals to one if a student is in good standing and zero otherwise. For the father’s and mother’s education levels, they are scaled numerically from 1 to 10. The high school GPA is calculated in a hundred mark system. “Female” and “Work” are two dummy variables equaling to one indicating a student is female and a student planned to work while in school separately.

3 Methodology

3.1 Difference-in-differences

We introduce the difference-in-differences (DD) strategy in this subsection. The following formula which is retaken from paper Borcan, Lindahl, and Mitrut (2017) specifies the strategy:

$$y_{ict} = \alpha + \beta T_{ct} + \gamma' X_{ict} + \varphi_t + \theta_c + \theta_c \cdot t + \varepsilon_{ict}, \quad (1)$$

where $i, c,$ and t are the indicators for student, county and year separately. The dependent variable y_{ict} consists of three main outcomes: (i) the standardized written Romanian language exam score, (ii) the overall Bacalaureate exam score, and (iii) the pass indicator which equals to one if a student passed the Bacalaureate exam and zero otherwise. T_{ct} is one of the main concerned variables in our research which acts as a treatment indicator that equals to one if a student is monitored by camera during the exam and zero otherwise. Hence, the estimated $\hat{\beta}$ will capture the pure treatment effect by the “camera” treatment on exam outcomes which may fluctuate over time. Regressor X_{ict} consists of students’ background information, which includes poverty status, study track, gender, and whether coming from a rural municipality. The estimated results of β and γ therefore, can reveal the heterogeneous treatment effects. In

the meantime, φ_t acts as the year dummy variable (in our research we focus on the period from 2009 to 2012), and θ_c is the county dummy variable. We also include $\theta_c \cdot t$ as an interaction term between the county dummies and linear time trend in our DD estimator. We replace the county dummies with a full set of family or school dummies in some regression estimations as suggested by Borcan, Lindahl, and Mitrut (2017). There are 25 out of 42 counties which installed camera monitors in 2011, and all counties had the monitors installed in 2012, it is plausible that the camera treatment implementation is county-wise. Therefore, we cluster the standard errors at the county level (with 42 clusters).

3.2 Causal forests

This section mainly reveals how causal forests algorithm generate estimates, how to test the accuracy of the estimates, how to test the significance of the treatment heterogeneity captured by causal forests and, what are the asymptotic properties of those estimates. Therefore, the remainder of this section is structured as follows. Subsection 3.2.1 will give a brief introduction to causal forests, for instance, how they estimate conditional average treatment effects given the necessary assumptions; subsection 3.2.2 will reveal the asymptotic properties of causal forests which allow us to conduct tests on the significance of the conditional average treatment effects estimates; subsection 3.2.3 will explain how causal forests work with clusters; finally, subsection 3.2.4 will propose the indicators to ascertain whether causal forests successfully capture heterogeneity in the conditional average treatment effects estimations.

3.2.1 An introduction to causal forests

In order to give a smooth explanation of causal forests algorithm, we first introduce the random forests algorithm. The random forests algorithm is known as a statistical learning algorithm which is widely used for non-parametric conditional mean estimation. Specifically, assume that we generate a dataset $(X_i, Y_i) \in \mathcal{X} \times \mathbb{R}$, then we can use random forests to estimate $\mu(x) = \mathbb{E}[Y_i | X_i = x]$ (equation is retaken from paper Athey, Tibshirani, and Wager (2019)). Athey, Tibshirani, and Wager (2019) proposed that an estimate of the expectation at a specific point $X_i = x$ can be obtained by averaging the estimates made by all decision trees in the forest. Generally, we grow each tree by greedy recursive partitioning the data points based on the variables (covariates) on the nodes of the tree to form leaves, one variable at a time. In the end, each leaf consists of a category of subsamples that are classified in a way to maximally enhance model fit. Put differently, we grow each tree by recursively adding those axis-aligned splits to

each tree (Athey, Tibshirani, and Wager (2019)). The following paragraph gives more details on how random forests generate predictions of the expected value $\mu(x)$.

To begin with, assume a random forest consists of B decision trees, and $L_b(x)$ denotes the leaf of the b -th tree which contains the training sample x . Then, as introduced by Athey and Wager (2019), through the following three steps the random forest can effectively make predictions:

- (1) Draw a subsample $\mathcal{S}_b \subseteq \{1, \dots, n\}$ for every tree $b = 1, \dots, B$.
- (2) Recursively split each subsample \mathcal{S}_b , $b = 1, \dots, B$, and add the splits to form leaves to grow a tree.
- (3) Make predictions in the following way:

$$\hat{\mu}(x) = \frac{1}{B} \sum_{b=1}^B \sum_{i=1}^n \frac{Y_i \mathbf{1}(\{X_i \in L_b(x), i \in \mathcal{S}_b\})}{|\{i : X_i \in L_b(x), i \in \mathcal{S}_b\}|}. \quad (2)$$

Equation 2 is retaken from paper Athey, Tibshirani, and Wager (2019). Previous study suggests that causal forests algorithm draw causal effect inferences by using the so called “out-of-bag” predictions which are denoted by an extra superscript $^{-i}$. Consider estimating $\hat{\mu}(x)$ in the above equation, making an “out-of-bag” prediction means that computing $\hat{\mu}(x)$ without using the real observation Y_i . In practice, we estimate $\hat{\mu}^{(-i)}(X_i)$ using only the trees b for which $i \notin \mathcal{S}_b$ for “out-of-bag” predictions.

In addition, Athey, Tibshirani, and Wager (2019) gave a different perspective about random forests by considering it as an adaptive kernel method. As a result, we rewrite equation 2 as follows:

$$\hat{\mu}(x) = \sum_{i=1}^n \alpha_i(x) Y_i, \quad \alpha_i(x) = \frac{1}{B} \sum_{b=1}^B \frac{\mathbf{1}(\{X_i \in L_b(x), i \in \mathcal{S}_b\})}{|\{i : X_i \in L_b(x), i \in \mathcal{S}_b\}|}. \quad (3)$$

We consider $\alpha_i(x)$ as a data-adaptive kernel, which keeps track of the frequency that the i -th training sample falls into the leaf that includes the test point x . The above equations are retaken from paper Athey and Wager (2019).

As mentioned earlier, variables play a crucial role in the splitting scheme, therefore, it is essential to select the suitable variables to train the random forests such that more accurate predictions can be made. Thus, we propose the following method to measure the importance level of a variable ℓ in a decision tree T with $J - 1$ nodes:

$$\mathcal{I}_\ell^2(T) = \sum_{t=1}^{J-1} \hat{i}_t^2 I(v(t) = \ell), \quad (4)$$

where \hat{i}_t^2 denotes the approximated squared error improvement at node t , the function $I(\cdot)$ is used to indicate that a variable $v(t)$ is chosen at node t such that the variable maximally improve the squared error at that node (Hastie, Tibshirani, and Friedman (2009)). Correspondingly, we use

$$\mathcal{I}_\ell^2 = \frac{1}{M} \sum_{m=1}^M \mathcal{I}_\ell^2(T_m) \quad (5)$$

to calculate the importance level of variable ℓ in a random forest consists of M trees. Equations 4 and 5 are retaken from paper O'Neill and Weeks (2018).

We use the explanations on random forests algorithm outlined above as a starting point to elaborate the causal forests algorithm in the rest of this section. To begin with, assume we have a training dataset containing n i.i.d. samples, labelled by $i = 1, \dots, n$, each of which consists of three components, i.e., a feature vector $X_i \in [0, 1]^d$, a treatment indicator $W_i \in \{0, 1\}$ and a response $Y_i \in \mathbb{R}$. Then we hypothesize that, for sample i , $Y_i^{(1)}$ and $Y_i^{(0)}$ are the potential outcomes of the response Y_i with and without the treatment, and define the conditional average treatment effect as

$$\tau(x) = \mathbb{E} \left[Y_i^{(1)} - Y_i^{(0)} | X_i = x \right]. \quad (6)$$

Equation 6 is acquired from paper Wager and Athey (2018). The aim is to get $\tau(x)$, to this end, we need to fit $Y_i^{(1)}$ and $Y_i^{(0)}$ into equation 6, yet the difficulty here is that in practice either $Y_i^{(1)}$ or $Y_i^{(0)}$ can ever be observed instead of both of them. Conventionally, it is impossible to estimate $\tau(x)$ based only on the observations (X_i, Y_i, W_i) . To overcome this problem, we need to further restrict the data-generating distribution (restriction 7 is acquired from paper Wager and Athey (2018)). Sufficiently, we can assume (i) unconfoundedness (Rosenbaum and Rubin (1983)) in our dataset:

$$\left\{ Y_i^{(0)}, Y_i^{(1)} \right\} \perp W_i | X_i, \quad (7)$$

equation 7 explicitly states that the treatment W_i is independent of the possible outcome Y_i conditional on X_i . The assumption of unconfoundedness has the following incentives: under the assumption of continuity, unconfoundedness effectively enables us to consider the nearby observations in the x space as they have come from a randomized experiment, hence, making some of the local methods such as nearest-neighbor matching consistent for $\tau(x)$ (Wager and Athey (2018)). We denote $e(x) = \mathbb{P}[W_i | X_i = x]$ as propensity score and $m(x) = \mathbb{E}[Y_i | X_i = x]$ as expected outcome of the response Y_i marginalizing over treatment. In order to obtain a semiparametrical efficient estimator of τ , in addition we need the following assumptions other than unconfoundedness (Athey and Wager (2019)): (ii) \hat{m} and \hat{e} are $o(n^{-1/4})$ consistent for

m and e respectively in root-mean-square error; (iii) the data is independently and identically distributed; (iv) there is overlap, that is, for some $\varepsilon > 0$ and all $x \in [0, 1]^d$, it holds that $\varepsilon < \mathbb{P}[W_i | X_i = x] < 1 - \varepsilon$ (this equation is acquired from paper Wager and Athey (2018)); (v) $\tau(x) = \tau$ for all $x \in \mathcal{X}$, that is, we assume constant of conditional average treatment effect function. Eventually, we have the following efficient estimator of τ (Chernozhukov et al. (2018); Robinson (1988)):

$$\hat{\tau} = \frac{\frac{1}{n} \sum_{i=1}^n (Y_i - \hat{m}^{(-i)}(X_i)) (W_i - \hat{e}^{(-i)}(X_i))}{\frac{1}{n} \sum_{i=1}^n (W_i - \hat{e}^{(-i)}(X_i))^2}. \quad (8)$$

Conventionally, we will estimate all of the results of conditional average treatment effects based on the “out-of-bag” predictions. Equation 8 is retaken from paper Athey, Tibshirani, and Wager (2019).

In general, equation 8 can only be used to estimate the constant treatment effect, however, as our goal also includes heterogeneous treatment effects analysis, we now introduce an “R-learner” objective function (Nie and Wager (2017)) which can be utilized for heterogeneous treatment effect estimation. We retake the “R-learner” objective function from paper Athey and Wager (2019) and represent it as follows:

$$\hat{\tau}(\cdot) = \operatorname{argmin}_{\tau} \left\{ \sum_{i=1}^n \left((Y_i - \hat{m}^{(-i)}(X_i)) - \tau(X_i) (W_i - \hat{e}^{(-i)}(X_i)) \right)^2 + \Lambda_n(\tau(\cdot)) \right\}, \quad (9)$$

where $\Lambda_n(\tau(\cdot))$ is called a regularizer which is able to control the complexity of the learned $\hat{\tau}(\cdot)$ function. The motivation for introducing the “R-learner” objective function, as discussed by Athey and Wager (2019) is that, in case that the true conditional average treatment effect function $\tau(\cdot)$ is simpler than the propensity function $e(\cdot)$ or the main effect function $m(\cdot)$, the optimized outcome $\hat{\tau}(\cdot)$ from 9 would converge faster than the estimation in which $\hat{m}(\cdot)$ or $\hat{e}(\cdot)$ are used to construct the objective function as in equation 8. Based on 8 and 9, we can utilize the idea of the data-adaptive kernel in 3 to estimate treatment effects in a more advanced way. That is, we first grow a forest to evaluate the weights $\alpha_i(x)$ (data-adaptive kernel), then we get the estimator as shown below (which is retaken from paper Athey and Wager (2019)):

$$\hat{\tau} = \frac{\sum_{i=1}^n \alpha_i(x) (Y_i - \hat{m}^{(-i)}(X_i)) (W_i - \hat{e}^{(-i)}(X_i))}{\sum_{i=1}^n \alpha_i(x) (W_i - \hat{e}^{(-i)}(X_i))^2}. \quad (10)$$

Furthermore, there are some built-in tuning parameters in the causal forests algorithm, we will use cross-validation on the “R-learner” objective to choose those parameters which would effectively minimize the “out-of-bag” estimates of the objective function given in 9.

3.2.2 Asymptotic inference of causal forests

In order to test the significance of the conditional average treatment effects, we can make use of the asymptotic properties of causal forests. In this subsection we will reveal two valuable asymptotic properties of causal forests concerning the consistency and asymptotic distributions of causal forests respectively. As discussed by Wager and Athey (2018), the first property states that the estimated treatment effect with causal forests is consistent with the true treatment effect $\tau(x)$, which requires the following two assumptions:

- (1) Both $\mathbb{E}[Y^{(0)}|X = x]$ and $\mathbb{E}[Y^{(1)}|X = x]$ are Lipschitz continuous,
- (2) There is overlap, i.e., for some $\varepsilon > 0$ and all $x \in [0, 1]^d$, it holds that $\varepsilon < \mathbb{P}[W = 1|X = x] < 1 - \varepsilon$ (as discussed in section 3.2.1).

The second assumption ensures that for large enough sample size n , the treatment and control units near any test point x will be enough to enable the local methods to work.

The second property is about the asymptotic sampling distribution of the estimated treatment effects $\hat{\tau}(x)$. In order to explicate this property, we utilize the potential nearest neighbors construction introduced by Yi and Yongho (2006) and classical analysis tools established by Hoeffding (1948) and Hajek (1968). Assume the subsample size s used to build individual causal trees is scaled as $s \asymp n^\beta$ for some $\beta_{\min} < \beta < 1$, then it holds that asymptotically the predictions made by a causal forest are Gaussian and unbiased as specified as follows (we acquire the following definition from paper Wager and Athey (2018)):

$$(\hat{\tau}(x) - \tau(x))/\sqrt{\text{Var}[\hat{\tau}(x)]} \Rightarrow \mathcal{N}(0, 1). \quad (11)$$

Furthermore, we can use the infinitesimal jackknife for random forests introduced by Efron (2014) and Wager, Hastie, and Efron (2014) to estimate the asymptotic variance of causal forests. The infinitesimal jackknife method requires the following two assumptions (Wager and Athey (2018)):

- (1) The number of trees B is large enough such that the Monte Carlo variability is trivial.
- (2) We only measure the randomness in $\hat{\tau}(x)$.

Once the two assumptions are satisfied, we can consistently estimate the asymptotic variance

(we acquire its representation from paper Wager and Athey (2018)) as below:

$$\widehat{V}_{IJ}(x) = \frac{n-1}{n} \left(\frac{n}{n-s} \right)^2 \sum_{i=1}^n \text{Cov}_* [\hat{\tau}_b^*(x), N_{ib}^*]^2, \quad (12)$$

where $\hat{\tau}_b^*(x)$ denotes the treatment effect estimated by tree b , and $N_{ib}^* \in \{0, 1\}$ shows if the i -th sample is used to train tree b . The covariance in 12 is obtained for all trees $b = 1, \dots, B$ in the forest. Consequently, it holds that $\widehat{V}_{IJ}(x) \sim \text{Var}[\hat{\tau}(x)]$.

3.2.3 Cluster analysis

Section 2.1 shows that the dataset in the Romanian education system application contains information of students coming from 42 different counties, inspired by Athey and Wager (2019), we implement the “cluster-robust random forests” to conduct analysis for this application in order to get more plausible results. We denote $C_i \in \{1, \dots, K\}$ ($K=42$) as county with index i , and Y_i as the outcomes of students in county i , then we have the following estimates which we acquire from paper Athey and Wager (2019):

$$\hat{\mu}_k = \frac{1}{n_k} \sum_{\{i:A_i=k\}} Y_i, \quad \hat{\mu} = \frac{1}{K} \sum_{k=1}^K \hat{\mu}_k, \quad \hat{\sigma}^2 = \frac{1}{K(K-1)} \sum_{k=1}^K (\hat{\mu}_k - \hat{\mu})^2, \quad (13)$$

n_k is the number of students in county k , $\hat{\mu}$ is the global mean and $\hat{\sigma}^2$ denotes the variance. With clusters, we therefore accordingly adjust the running process of random forests algorithm illustrated in section 3.2.1 (Athey and Wager (2019) in the following way. To start with, in the first step we draw a subsample of clusters $\mathcal{K}_b \subseteq \{1, \dots, K\}$, and draw k random samples from every cluster $k \in \mathcal{K}_b$ to generate \mathcal{S}_b . Moreover, we estimate the average treatment effect $\hat{\tau}$ and standard error $\hat{\sigma}^2$ in a different way (we acquire the following equations from paper Athey and Wager (2019)):

$$\begin{aligned} \hat{\tau}_k &= \frac{1}{n_k} \sum_{\{i:A_i=k\}} \hat{\Gamma}_i, \quad \hat{\tau} = \frac{1}{K} \sum_{k=1}^K \hat{\tau}_k, \quad \hat{\sigma}^2 = \frac{1}{K(K-1)} \sum_{k=1}^K (\hat{\tau}_k - \hat{\tau})^2, \\ \hat{\Gamma}_i &= \hat{\tau}^{(-i)}(X_i) + \frac{W_i - \hat{e}^{(-i)}(X_i)}{\hat{e}^{(-i)}(X_i)(1 - \hat{e}^{(-i)}(X_i))} (Y_i - \hat{m}^{(-i)}(X_i) - (W_i - \hat{e}^{(-i)}(X_i)) \hat{\tau}^{(-i)}(X_i)). \end{aligned} \quad (14)$$

3.2.4 Inspecting treatment heterogeneity

After obtaining the estimation results of the average treatment effects, it is worth investigating whether the trained causal forests has precisely detected treatment heterogeneity. To that end,

we propose a test for heterogeneity, namely the “best linear predictor” method, as established by Chernozhukov et al. (2017), which attempts to fit the conditional average treatment effect as a linear function of the “out-of-bag” estimates of treatment effects $\hat{\tau}^{(-i)}(X_i)$. We compose the following two synthetic predictors (which are acquired from paper Athey and Wager (2019)) based on equation 8:

$$C_i = \bar{\tau} \left(W_i - \hat{e}^{(-i)}(X_i) \right), \quad (15)$$

$$D_i = \left(\hat{\tau}^{(-i)}(X_i) - \bar{\tau} \right) \left(W_i - \hat{e}^{(-i)}(X_i) \right), \quad (16)$$

where $\bar{\tau}$ denotes the estimate of the average “out-of-bag” treatment effects. We then regress $Y_i - \hat{m}^{(-i)}(X_i)$ on C_i and D_i , the estimated coefficients of C_i and D_i have next interpretations respectively:

C_i : it is the absorption of the average treatment effect, hence, for C_i a coefficient of one means that the mean forest prediction is correct.

D_i : it is used to indicate the quality of the estimates of treatment heterogeneity (Athey and Wager (2019)), a coefficient of one suggests that the causal forest has captured heterogeneity, in other words, the treatment heterogeneity estimates are properly calibrated. One can use the p-value of the coefficient to test the null hypothesis that heterogeneity has been successfully captured.

Once the causal forests have captured heterogeneity in conditional average treatment effects, we can assess the heterogeneous treatment effects. To that end, we first split our sample into certain number of twofold subsamples based on the selected important covariates as suggested in section 3.2.1, e.g., consider the middle school GPA covariate in the first application, we split the entire sample into two: the first one contains observations whose middle school GPA are above the median and the second one contains observations whose middle school GPA are below the median. Secondly, we calculate the conditional average treatment effects of the subsamples in each twofold subsample group, and regard them as the heterogeneous treatment effects. Furthermore, we employ Student’s t-test to investigate whether the the two estimated average treatment effects of each twofold subsample group are statistically the same, if the null hypothesis of same treatment effects is rejected, we conclude that the two treatment effects are heterogeneous.

4 Results

4.1 Results of Romanian education system application

We demonstrate the central results of the conditional average treatment effects (CATEs) estimation on the three exam outcomes in table 3 below. As marked with asterisks, all three CATEs are significant using a significance level of 0.001, meaning that our trained causal forests are statistically accurate evaluating the CATEs. We observe that the “camera” treatment effectively undermines students’ Bacculaureate exam performance. That is, when students are monitored by the cameras in the exam hall, the probability for them to pass the exam is lowered by 0.086, furthermore, their written exam scores tends to reduce by 0.216 and the overall Bacculaureate scores would decrease by 0.487 approximately, as compared to a scenario when they are not being monitored.

Table 3: Conditional average treatment effects on three outcomes.

	Baccalaureate Pass Indicator	Romanian Written Score	Overall Bacculaureate Score
Conditional average treatment effect (CATE)	-0.086*** (0.015)	-0.216*** (0.045)	-0.487*** (0.061)
Conditional average treatment effect (high)	-0.062*** (0.013)	-0.050* (0.020)	-0.254*** (0.031)
Conditional average treatment effect (low)	-0.111*** (0.022)	-0.391*** (0.078)	-0.709*** (0.092)
95% CI for the CATE	(-0.115, -0.057)	(-0.303, -0.129)	(-0.606, -0.368)
95% CI for the difference in CATE	(-0.002, 0.098)	(0.184, 0.498)	(0.263, 0.645)
Observations	100000	100000	100000
Number of covariates	22	22	22

Notes: In this table we show the estimated results of the average treatment effects by “camera” treatment on the three outcomes as indicated in the first row, as well as the means of the treatment effects which are higher and lower than the median value of the whole set of estimated treatment effects (standard errors are in the brackets). We use ***, ** and * to indicate that a CATE is significant at a significance level of 0.001, 0.01 and 0.05 respectively (same for table 8). For the sake of completeness, we also include in the table the 95% confidence intervals for both the CATEs and the difference between the the higher and lower CATE.

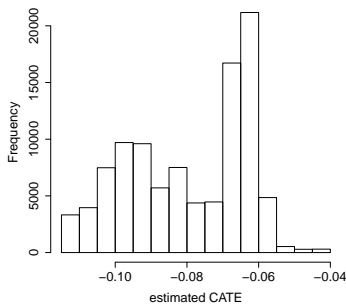


Figure 1: Bacculaureate pass indicator.

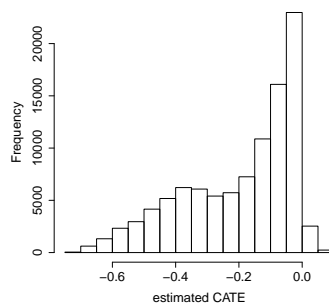


Figure 2: Romanian written score.

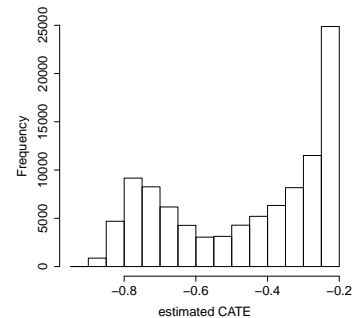


Figure 3: Overall bacculaureate score.

Except estimating the CATEs, causal forests are also able to capture the potential treatment heterogeneity, hence, we will focus on the heterogeneity analysis from now on. Above figures vaguely give a first impression of existence of heterogeneity by showing strong variations exhibited by CATE estimates of all three causal forests. In order to formally detect whether the causal forests have captured potential heterogeneity in the treatment effects, we employ calibration tests. As elaborated in table 4, the p-values indicate that the treatment effects on the written scores and overall scores are significantly heterogeneous, whereas no evidence shows there exists heterogeneity in the treatment effect on the pass indicator given a large p-value of 0.853. Additionally, the p-values of the omnibus tests show that the means of the “out-of-bag” predictions of the three causal forests are correct.

Table 4: Calibration tests for heterogeneity.

	Baccalaureate Pass Indicator		Romanian Written Score		Overall Baccalaureate Score	
	D_i	C_i	D_i	C_i	D_i	C_i
Value	0.222	0.973	0.954	1.015	1.142	1.002
Standard error	1.198	0.268	0.345	0.385	0.289	0.192
t-value	0.186	3.625	2.765	2.633	3.958	5.222
p-value	0.853	< 0.001***	0.006**	0.008**	7.581e-5***	1.772e-7***

Notes: The *** denotes that the p value is significantly small such that the null hypothesis of the omnibus test is rejected, with a significance level of 0.001, whereas ** corresponds to a significance level of 0.01 (same for table 9). Please refer to section 3.2.4 for detailed explanations of coefficients D_i and C_i .

After successfully discovering the existence of heterogeneity, it is worthy to ascertain which covariates cause the heterogeneity. Table 5 below displays the selected covariates to train the causal forests, along with their corresponding weights. Covariates are selected by the built-in `variable_importance()` function in the `grf` package. As can be seen, county id and school id are selected by all three causal forests especially in the pass indicator, in which the two covariates are adopted as the most important ones, totally the causal forest spends around 30% of splits on them. The reasons for that are rather straight forward: there could exist some unobserved school-level features that may affect the treatment effects, e.g., new study programs established by some schools to help students to perform better in the exams when it is not possible to cheat; as for the county level, we use it to form clusters when we train our forests since many of the covariates used are county-level properties such as the county population and county poor rate. Moreover, middle school GPA and standardized ability measure are also selected by all three causal forests, particularly by the forests for the written score and overall score in which the two covariates together possess roughly 40% split weights. Other selected covariates can be found in the table.

Table 5: Selected covariates for training the causal forests and their weights.

Baccalaureate Pass Indicator		Romanian Written Score		Overall Baccalaureate Score	
selected covariates	splits weighted by depth (%)	selected covariates	splits weighted by depth (%)	selected covariates	splits weighted by depth (%)
county_id	17.199	midschool_gpa	23.285	midschool_gpa	20.389
school_id	13.446	ability_std	14.906	ability_std	16.966
ability_std	11.691	orrom	12.204	county_id	14.028
county_shromanian	10.204	county_id	11.287	county_shromanian	8.580
midschool_gpa	8.618	county_shromanian	8.369	orrom	6.631
county_pop	8.465	school_id	6.023	school_id	6.089
total_population	6.528	urban_population	5.196	total_population	6.053
urban_population	6.139	total_population	4.903	urban_population	5.785
rep_bac	5.863			county_pop	5.473

Notes: The meaning of the abbreviations of the covariates in this table can be found in Appendix D. This table shows the “selected” covariates by each causal forest, as well as the split weights on each covariate. The more weights being assigned to a covariate, the more important the covariate is in contributing to heterogeneity. We sort the weights in a descending order (same for table 10).

Table 6: Main heterogeneous treatment effects with t-test results.

	Baccalaureate Pass Indicator		Romanian Written Score		Overall Baccalaureate Score	
	HTE	test statistics	HTE	test statistics	HTE	test statistics
High midschool_gpa	-0.037** (0.012)	t-value = 358.010	-0.003 (0.035)	t-value = 405.570	-0.186*** (0.040)	t-value = 567.200
Low midschool_gpa	-0.115*** (0.029)	p-value <2.2e-16	-0.382*** (0.115)	p-value <2.2e-16	-0.745*** (0.134)	p-value <2.2e-16
High rep_bac	-0.018 (0.016)	t-value = 216.470				
Low rep_bac	-0.086*** (0.024)	p-value <2.2e-16				
High ability_std	-0.039*** (0.012)	t-value = 356.320	-0.004 (0.035)	t-value = 407.570	-0.188*** (0.040)	t-value = 571.190
Low ability_std	-0.114*** (0.029)	p-value <2.2e-16	-0.383*** (0.116)	p-value <2.2e-16	-0.747*** (0.135)	p-value <2.2e-16
Large county_pop	-0.089** (0.034)	t-value = -18.304			-0.462*** (0.147)	t-value = 17.929
Small county_pop	-0.074*** (0.024)	p-value <2.2e-16			-0.481*** (0.107)	p-value <2.2e-16
Large total_population	-0.091** (0.035)	t-value = -17.248	-0.275* (0.117)	t-value = -16.695	-0.485*** (0.147)	t-value = 20.132
Small total_population	-0.072** (0.024)	p-value <2.2e-16	-0.170 (0.087)	p-value <2.2e-16	-0.468*** (0.107)	p-value <2.2e-16
Large urban_population	-0.101** (0.034)	t-value = -9.437	-0.290* (0.115)	t-value = -19.097	-0.540*** (0.138)	t-value = 20.414
Small urban_population	-0.068** (0.025)	p-value <2.2e-16	-0.167 (0.087)	p-value <2.2e-16	-0.442** (0.109)	p-value <2.2e-16
High county_shromanian	-0.057* (0.028)	t-value = 25.469	-0.111 (0.089)	t-value = 7.252	-0.375*** (0.115)	t-value = 4.119
Low county_shromanian	-0.094*** (0.029)	p-value <2.2e-16	-0.281** (0.108)	p-value = 4.148e-13	-0.556*** (0.132)	p-value = 3.814e-5
High orrom			-0.176*** (0.040)	t-value = 14.577	-0.471*** (0.034)	t-value = 4.313
Low orrom			-0.196** (0.071)	p-value <2.2e-16	-0.495*** (0.064)	p-value = 1.613e-5

Notes: The meaning of the abbreviations of the covariates in this table can be found in Appendix D. Furthermore, we indicate different subsamples by giving them different names, e.g., “High midschool_gpa” and “Low midschool_gpa” represent the subsamples of students with middle school GPA that is higher than the median value and lower than the median value respectively. “HTE” is short for heterogeneous treatment effect. We use ***, ** and * to indicate that the heterogeneous treatment effect is significant when using a significance level of 0.001, 0.01 and 0.05 respectively. (same for table 11 and Appendix F)

Table 6 above shows the heterogeneous treatment effects corresponding to each selected covariate as well as test statistics of the tests for heterogeneity. Please refer to section 3.2.4 for complete explanations on how heterogeneous treatment effects are calculated and how the tests work. From the table we can see that all the p-values of all the tests are extremely small such that the null hypothesis of no heterogeneity is rejected. One possible reason for that is the degrees of freedom for all tests are quite large due to large sample size, namely 100,000.

Finally, we show the results of the comparison between treatment effects estimated by OLS estimation and causal forest estimation (we use OLS estimation to replicate the results from Borcan, Lindahl, and Mitrut (2017), and give the replicated outcomes in Appendix A). From table 7 we can observe that the estimated results from two different estimation techniques are quite similar, this further suggests that the exam outcomes are affected negatively by the camera monitors.

Table 7: Comparison of the “camera” treatment effects between OLS estimation and causal forests estimation.

	Baccalaureate Pass Indicator	Romanian Written Score	Overall Baccalaureate Score
OLS estimates	-0.076 (0.030)	-0.246 (0.108)	-0.430 (0.144)
Causal forests estimates	-0.086 (0.015)	-0.216 (0.045)	-0.487 (0.061)

4.2 Results of incentives for college achievement application

The results from the above section has verified that causal forests algorithm is able to accurately estimate CATEs and in the meantime capture heterogeneity when it works with large enough number of observations. For now, we examine the results of the second application to ascertain if we will draw the same conclusion in a scenario where causal forests are trained with much fewer observations.

The following table shows the main results of the estimated CATEs on the three outcomes: GPA, good standing and credits earned for the first and second university years of sampled students. For the purpose of comparison, we also include the estimated HTEs for male and female students in the table to keep pace with table 6 from Angrist, Lang, and Oreopoulos (2009).

Surprisingly, all the CATEs are statistically insignificant at a 0.05 significance level. On the other hand, as for the gender HTEs, only the one concerning female students’ first year good standing is significant at a significance level of 0.01, which numerically indicates that the SFSP treatment affects female students’ first year good standing negatively by a level of 0.180. Thus,

we conclude that with very few observations, causal forests algorithm cannot properly estimate CATEs.

Table 8: Conditional average treatment effects on main outcomes.

	Year 1			Year 2		
	All (1)	Men (2)	Women (3)	All (4)	Men (5)	Women (6)
Panel A: GPA						
Conditional average treatment effect (CATE)	0.047 (0.098)	0.046 (0.172)	0.046 (0.119)	-0.013 (0.100)	-0.129 (0.188)	0.060 (0.114)
95% CI for the CATE	(-0.145, 0.239)	(-0.291, 0.383)	(-0.186, 0.278)	(-0.210, 0.184)	(-0.497, 0.239)	(-0.163, 0.283)
Observations	1102	467	635	1102	467	635
Number of covariates	26	-	-	26	-	-
Panel B: good standing						
Conditional average treatment effect (CATE)	-0.082 (0.050)	0.078 (0.078)	-0.180** (0.065)	0.010 (0.030)	0.021 (0.060)	0.002 (0.031)
95% CI for the CATE	(-0.180, 0.016)	(-0.075, 0.231)	(-0.306, -0.054)	(-0.048, 0.068)	(-0.096, 0.138)	(-0.059, 0.063)
Observations	1102	467	635	1102	467	635
Number of covariates	26	-	-	26	-	-
Panel C: credits earned						
Conditional average treatment effect (CATE)	-0.148 (0.096)	-0.196 (0.176)	-0.118 (0.112)	-0.193 (0.147)	-0.147 (0.253)	-0.221 (0.180)
95% CI for the CATE	(-0.337, 0.041)	(-0.540, 0.148)	(-0.337, 0.101)	(-0.482, 0.096)	(-0.643, 0.349)	(-0.574, 0.132)
Observations	1102	467	635	1102	467	635
Number of covariates	26	-	-	26	-	-

The conclusion is also in line with the results given by table 9 below, in which nearly all the p-values of the omnibus tests for C_i s suggest that the means of “out-of-bag” predictions of the causal forests are incorrect except the one for the first year credits earned. On the other hand, the p-values of the tests for heterogeneity imply that four out of six causal forests successfully find heterogeneity in the CATEs, namely first year GPA and credits earned, as well as second year good standing and credits earned. Moreover, figures 4, 5 and 6 below illustrate that there exist variations in the estimated CATEs on the first year study outcomes, the illustrations for the second year outcomes can be found in Appendix G.

Table 9: Calibration tests for heterogeneity.

Year 1	GPA		Good standing		Credits earned	
	D_i	C_i	D_i	C_i	D_i	C_i
Value	-61.233	-1.874	-0.035	1.317	-57.546	1.461
Standard error	5.638	1.884	0.991	1.082	4.513	0.454
t-value	-10.862	-0.995	-0.035	1.216	-12.750	3.216
p-value	<2e-16***	0.320	0.972	0.224	<2.2e-16***	0.001**
Year 2	GPA		Good standing		Credits earned	
	D_i	C_i	D_i	C_i	D_i	C_i
Value	0.416	0.409	-73.217	0.827	-83.264	1.028
Standard error	1.202	5.772	7.518	1.856	8.059	0.546
t-value	0.346	0.071	-9.739	0.446	-10.332	1.882
p-value	0.729	0.944	<2e-16***	0.656	<2e-16***	0.060

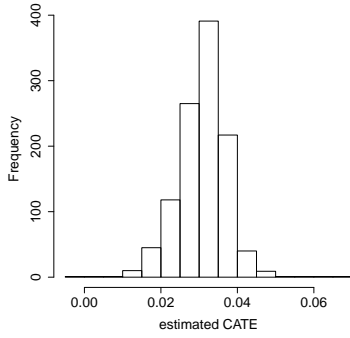


Figure 4: GPA.

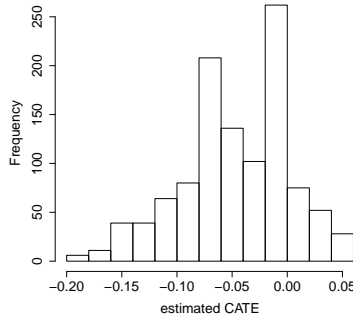


Figure 5: good standing.

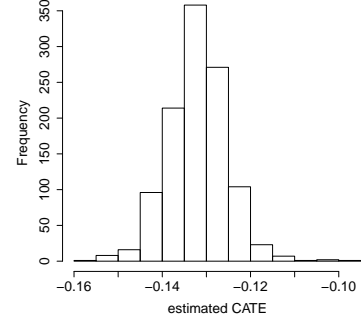


Figure 6: credits earned.

Table 10 shows the important covariates that are selected by each causal forest as well as their weights. It is notable that all six causal forests select the next four covariates: high school GPA, a dummy variable indicating whether a student has emailed an advisor, father’s education level and mother’s education level. Accordingly, we can have the following interpretations: high school GPA reflects a student’s academic ability, while parents’ education level genetically affect a student’s academic ability, therefore, they both play essential roles in affecting a student’s college achievement; also, the study advice from an advisor effectively helps students to achieve more in college.

Table 10: Selected covariates for training the causal forests and their weights.

Year 1	GPA		Good standing		Credits earned	
	selected covariates	splits weighted by depth (%)	selected covariates	splits weighted by depth (%)	selected covariates	splits weighted by depth (%)
	gpa0	20.386	gpa0	28.009	gpa0	18.723
	used_adv	14.913	used_adv	18.617	used_adv	13.209
	dad_edn	10.796	mom_edn	10.575	mom_edn	12.590
	lm_rarely	9.649	female	7.286	dad_edn	10.820
	mom_edn	8.701	dad_edn	5.470	lm_rarely	7.363
	lastmin	5.374	numcourses_nov1	4.305	numcourses_nov1	4.465
Year 2	GPA		Good standing		Credits earned	
	selected covariates	splits weighted by depth (%)	selected covariates	splits weighted by depth (%)	selected covariates	splits weighted by depth (%)
	gpa0	20.774	gpa0	23.438	gpa0	20.021
	used_adv	14.249	used_adv	14.337	used_adv	16.105
	dad_edn	10.412	dad_edn	9.847	mom_edn	9.942
	lm_rarely	9.349	prob_year1	8.596	dad_edn	8.947
	lastmin	7.054	mom_edn	7.313	numcourses_nov1	7.960
	mom_edn	5.857	lm_rarely	5.069	lm_rarely	5.263
	dad2	5.302	numcourses_nov1	4.522	lastmin	4.676
	female	3.883				

Notes: The meaning of the abbreviations of the covariates in this table can be found in Appendix E.

Based on the results from table 10, we can assess the HTEs for the selected covariates, table 11 below presents the estimates of all HTEs on the first year outcomes and their test statistics, please refer to Appendix F for the second year results. Again, we observe that roughly all the estimated HTEs are insignificant except the two for credits earned of the students whose fathers’ education level is below the median and of the students who enroll fewer courses than the median, which are significant at a significance level of 0.05, along with the one for the good

standing of the female students which is significant at a significance level of 0.01. Meanwhile, the p-values evidence that the HTEs of most twofold subsamples are statistically different.

Table 11: Main heterogeneous treatment effects with t-test results.

Year 1	GPA		Good standing		Credits earned	
	HTE	test statistics	HTE	test statistics	HTE	test statistics
High dad_edn	0.370 (0.234)	t-value = -1.567	-0.096 (0.109)	t-value = -8.447	0.159 (0.180)	t-value = 6.110
Low dad_edn	-0.062 (0.106)	p-value = 0.1179	-0.075 (0.056)	p-value = 3.44e-16	-0.250* (0.115)	p-value = 1.9e-9
High gpa0	0.103 (0.152)	t-value = -7.622	-0.120 (0.075)	t-value = -14.118	-0.154 (0.124)	t-value = 5.161
Low gpa0	-0.009 (0.129)	p-value = 5.414e-14	-0.050 (0.068)	p-value <2.2e-16	-0.136 (0.145)	p-value = 2.911e-7
High lastmin	-0.109 (0.181)	t-value = -11.773				
Low lastmin	0.124 (0.115)	p-value <2.2e-16				
Lm_rarely	-0.252 (0.189)	t-value = -12.035			-0.348 (0.210)	t-value = -6.400
Non-lm_rarely	0.177 (0.113)	p-value <2.2e-16			-0.053 (0.100)	p-value = 5.666e-10
High mom_edn	0.158 (0.158)	t-value = -1.401	-0.151 (0.081)	t-value = -15.204	-0.031 (0.134)	t-value = -1.100
Low mom_edn	-0.042 (0.121)	p-value = 0.161	-0.038 (0.062)	p-value <2.2e-16	-0.252 (0.132)	p-value = 0.271
Used_adv	0.105 (0.143)	t-value = -0.896	-0.137 (0.080)	t-value = -14.753	-0.156 (0.138)	t-value = 0.536
Non-used_adv	-0.024 (0.134)	p-value = 0.373	-0.021 (0.058)	p-value <2.2e-16	-0.140 (0.136)	p-value = 0.594
Male			0.078 (0.078)	t-value = -22.651		
Female			-0.178** (0.064)	p-value <2.2e-16		
High numcourses_nov1			-0.016 (0.085)	t-value = -0.368	0.034 (0.201)	t-value = -3.555
Low numcourses_nov1			-0.109 (0.060)	p-value = 0.713	-0.225* (0.106)	p-value <0.001

Notes: The meaning of the abbreviations of the covariates in this table can be found in Appendix E. Furthermore, we indicate different subsamples by giving them different names, e.g., “High dad_edn” and “Low dad_edn” represent the subsamples of students whose fathers’ education level is higher than the median and lower than the median respectively.

5 Conclusion

In this paper, we conduct empirical analysis to explore two research questions. The first one is whether the machine-learning-based causal forests algorithm and the statistics-based difference-in-differences strategy have similar ability to estimate conditional average treatment effects, specifically, do they generate similar estimates of the average treatment effects. The second one is can the causal forests algorithm correctly estimate average treatment effects meanwhile capture the existing treatment heterogeneity when running with small dataset which contains large number of covariates. In order to answer the two questions, we implement the causal forests algorithm with two datasets which were collected and provided by Borcan, Lindahl, and Mitrut (2017) and Angrist, Lang, and Oreopoulos (2009). We elaborate the two applications as follows.

In the first application, we try to estimate the treatment effects on students' Bacalaureate exam outcomes, namely Bacalaureate pass rate, Romanian written score and overall Bacalaureate score, by the "camera" treatment which was promoted by the national anti-corruption campaign in Romanian. We do so by means of both the causal forests algorithm and difference-in-differences strategy. The results show that the two different approaches give rather similar estimates of the average treatment effects which numerically suggest that students' exam performance would be significantly undermined by the camera monitor. Moreover, significant treatment heterogeneity has been captured by the causal forests on the Romanian written score and overall Bacalaureate score. Overall, county id, school id, standardized ability measure and middle school GPA are the most important covariates in originating heterogeneity in the treatment effects on all the three exam outcomes.

In the second application, we try to estimate the treatment effects on students' college achievement by a SFSP treatment which was used to increase students incentives. In order to answer the second research question, we use a dataset which contains only 1102 observations. Surprisingly, we find that the average treatment effect estimates are rather incorrect in a sense that they are all statistically insignificant. Nevertheless, treatment heterogeneity has been captured. Specifically, there exists significant heterogeneity in the treatment effects on first year GPA and credits earned, along with the second year good standing and credits earned.

To sum up, as a machine learning technique, causal forests algorithm can draw accurate causal effect inferences of a treatment only when it is trained with a large dataset containing enough observations, otherwise it may perform quite poorly. Also, the algorithm is powerful in detecting potential treatment heterogeneity even if being trained with small dataset. Furthermore, the difference-in-differences strategy can generate accurate treatment effect estimates that are similar to the ones generated by the causal forest algorithm. However, due to the fact that the causal forests algorithm can efficiently test the accuracy of the estimated treatment effects as well as detecting heterogeneity, we suggest other researchers who may also want to study causal effect inferences of a treatment to utilize the causal forests algorithm other than the DD strategy.

References

- Angrist, Joshua, Daniel Lang, and Philip Oreopoulos (2009). “Incentives and Services for College Achievement: Evidence from a Randomized Trial”. en. In: *American Economic Journal: Applied Economics* 1.1, pp. 136–163. ISSN: 1945-7782, 1945-7790. DOI: [10.1257/app.1.1.136](https://doi.org/10.1257/app.1.1.136). URL: <http://pubs.aeaweb.org/doi/10.1257/app.1.1.136> (visited on 06/28/2019).
- Athey, Susan, Julie Tibshirani, and Stefan Wager (2019). “Generalized random forests”. en. In: *The Annals of Statistics* 47.2, pp. 1148–1178. ISSN: 0090-5364. DOI: [10.1214/18-AOS1709](https://doi.org/10.1214/18-AOS1709). URL: <https://projecteuclid.org/euclid.aos/1547197251> (visited on 06/13/2019).
- Athey, Susan and Stefan Wager (2019). “Estimating Treatment Effects with Causal Forests: An Application”. en. In: *arXiv:1902.07409 [stat]*. arXiv: 1902.07409. URL: <http://arxiv.org/abs/1902.07409> (visited on 06/08/2019).
- Borcan, Oana, Mikael Lindahl, and Andreea Mitrut (2017). “Fighting Corruption in Education: What Works and Who Benefits?” en. In: *American Economic Journal: Economic Policy* 9.1, pp. 180–209. ISSN: 1945-7731, 1945-774X. DOI: [10.1257/pol.20150074](https://doi.org/10.1257/pol.20150074). URL: <http://pubs.aeaweb.org/doi/10.1257/pol.20150074> (visited on 05/31/2019).
- Breiman, Leo (2001). “Random Forests”. In: *Machine Learning* 45.1, pp. 5–32. ISSN: 08856125. DOI: [10.1023/A:1010933404324](https://doi.org/10.1023/A:1010933404324). URL: <http://link.springer.com/10.1023/A:1010933404324> (visited on 06/08/2019).
- Chernozhukov, Victor et al. (2017). “Generic Machine Learning Inference on Heterogenous Treatment Effects in Randomized Experiments”. In: *arXiv:1712.04802 [econ, math, stat]*. arXiv: 1712.04802. URL: <http://arxiv.org/abs/1712.04802> (visited on 06/09/2019).
- Chernozhukov, Victor et al. (2018). “Double/debiased machine learning for treatment and structural parameters”. en. In: *The Econometrics Journal* 21.1, pp. C1–C68. ISSN: 1368-4221, 1368-423X. DOI: [10.1111/ectj.12097](https://doi.org/10.1111/ectj.12097). URL: <https://academic.oup.com/ectj/article/21/1/C1/5056401> (visited on 06/09/2019).
- Efron, Bradley (2014). “Estimation and Accuracy After Model Selection”. en. In: *Journal of the American Statistical Association* 109.507, pp. 991–1007. ISSN: 0162-1459, 1537-274X. DOI: [10.1080/01621459.2013.823775](https://doi.org/10.1080/01621459.2013.823775). URL: <https://www.tandfonline.com/doi/full/10.1080/01621459.2013.823775> (visited on 06/13/2019).
- Hajek, Jaroslav (1968). “Asymptotic Normality of Simple Linear Rank Statistics Under Alternatives”. en. In: *The Annals of Mathematical Statistics* 39.2, pp. 325–346. ISSN: 0003-4851. DOI: [10.1214/aoms/1177698394](https://doi.org/10.1214/aoms/1177698394). URL: <http://projecteuclid.org/euclid.aoms/1177698394> (visited on 06/13/2019).

- Hastie, Trevor, Robert Tibshirani, and Jerome Friedman (2009). *The Elements of Statistical Learning*. Springer Series in Statistics. New York, NY: Springer New York. ISBN: 978-0-387-84857-0 978-0-387-84858-7. DOI: [10.1007/978-0-387-84858-7](https://doi.org/10.1007/978-0-387-84858-7). URL: <http://link.springer.com/10.1007/978-0-387-84858-7> (visited on 06/30/2019).
- Hoeffding, Wassily (1948). “A Class of Statistics with Asymptotically Normal Distribution”. en. In: *The Annals of Mathematical Statistics* 19.3, pp. 293–325. ISSN: 0003-4851. DOI: [10.1214/aoms/1177730196](https://doi.org/10.1214/aoms/1177730196). URL: <http://projecteuclid.org/euclid.aoms/1177730196> (visited on 06/13/2019).
- Nie, Xinkun and Stefan Wager (2017). “Quasi-Oracle Estimation of Heterogeneous Treatment Effects”. In: *arXiv:1712.04912 [econ, math, stat]*. arXiv: 1712.04912. URL: <http://arxiv.org/abs/1712.04912> (visited on 06/13/2019).
- O’Neill, Eoghan and Melvyn Weeks (2018). “Causal Tree Estimation of Heterogeneous Household Response to Time-Of-Use Electricity Pricing Schemes”. en. In: *arXiv:1810.09179 [econ, q-fin]*. arXiv: 1810.09179. URL: <http://arxiv.org/abs/1810.09179> (visited on 06/30/2019).
- Robinson, P. M. (1988). “Root-N-Consistent Semiparametric Regression”. In: *Econometrica* 56.4, p. 931. ISSN: 00129682. DOI: [10.2307/1912705](https://doi.org/10.2307/1912705). URL: <https://www.jstor.org/stable/1912705?origin=crossref> (visited on 06/09/2019).
- Rosenbaum, Paul R. and Donald B. Rubin (1983). “The central role of the propensity score in observational studies for causal effects”. en. In: *Biometrika* 70.1, pp. 41–55. ISSN: 0006-3444, 1464-3510. DOI: [10.1093/biomet/70.1.41](https://doi.org/10.1093/biomet/70.1.41). URL: <https://academic.oup.com/biomet/article-lookup/doi/10.1093/biomet/70.1.41> (visited on 06/09/2019).
- Wager, Stefan and Susan Athey (2018). “Estimation and Inference of Heterogeneous Treatment Effects using Random Forests”. en. In: *Journal of the American Statistical Association* 113.523, pp. 1228–1242. ISSN: 0162-1459, 1537-274X. DOI: [10.1080/01621459.2017.1319839](https://doi.org/10.1080/01621459.2017.1319839). URL: <https://www.tandfonline.com/doi/full/10.1080/01621459.2017.1319839> (visited on 06/08/2019).
- Wager, Stefan, Trevor Hastie, and Bradley Efron (2014). “Confidence Intervals for Random Forests: The Jackknife and the Infinitesimal Jackknife”. eng. In: *Journal of machine learning research: JMLR* 15.1, pp. 1625–1651. ISSN: 1532-4435.
- Yi, Lin and Jeon Yongho (2006). “Random Forests and Adaptive Nearest Neighbors”. In: *American Statistical Association, Taylor & Francis, Ltd.* 101, pp. 578–590. ISSN: 01621459. URL: <http://www.jstor.org/stable/27590719>.

Appendices

A The effects of anti-corruption campaign on students' exam outcomes

	Baccalaureate Pass Indicator				Romanian Written Score				Overall Baccalaureate Score			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Camera		-0.076 (0.030)	-0.076 (0.029)	-0.095 (0.025)		-0.246 (0.108)	-0.251 (0.106)	-0.353 (0.106)		-0.430 (0.144)	-0.439 (0.142)	-0.512 (0.137)
Year 12	-0.211 (0.024)	-0.135 (0.025)	-0.148 (0.023)	-0.082 (0.017)	-0.874 (0.065)	-0.628 (0.087)	-0.716 (0.078)	-0.463 (0.087)	-0.923 (0.092)	-0.492 (0.115)	-0.579 (0.106)	-0.323 (0.094)
Year 11	-0.211 (0.022)	-0.161 (0.019)	-0.166 (0.018)	-0.129 (0.016)	-0.875 (0.058)	-0.713 (0.070)	-0.743 (0.071)	-0.597 (0.081)	-0.943 (0.088)	-0.660 (0.091)	-0.690 (0.090)	-0.547 (0.088)
Year 09	0.121 (0.011)	0.121 (0.011)	0.115 (0.011)	0.093 (0.012)	-0.205 (0.054)	-0.205 (0.054)	-0.237 (0.057)	-0.311 (0.033)	1.087 (0.042)	1.086 (0.042)	1.055 (0.040)	0.967 (0.040)
Male			-0.109 (0.003)	-0.109 (0.003)			-0.852 (0.016)	-0.852 (0.016)			-0.590 (0.014)	-0.590 (0.013)
Poor			-0.045 (0.004)	-0.044 (0.004)			-0.224 (0.022)	-0.222 (0.022)			-0.263 (0.020)	-0.260 (0.020)
Theoretic track			0.318 (0.013)	0.318 (0.012)			1.458 (0.049)	1.457 (0.049)			1.559 (0.052)	1.559 (0.051)
Rural municipality			-0.137 (0.020)	-0.137 (0.020)			-0.662 (0.067)	-0.666 (0.067)			-0.652 (0.085)	-0.656 (0.085)
County FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	No	Yes	Yes	No	No	Yes	Yes	No	No	Yes	Yes
County FE × yearly trends	No	No	No	Yes	No	No	No	Yes	No	No	No	Yes
Observations	731,505	731,505	731,505	731,505	712,298	712,298	712,298	712,298	706,895	706,895	706,895	706,895
R ²	0.102	0.103	0.239	0.253	0.060	0.060	0.275	0.289	0.204	0.206	0.417	0.432

Notes: This table replicates table 2 from Borcan, Lindahl, and Mitrut (2017). The table demonstrates the results of the OLS estimates based on the difference-in-differences strategy. The dependent variables include Baccalaureate exam pass rates, Romanian Written exam scores and overall Baccalaureate scores, which are regressed on the year dummies in columns (1), (5), (9), also on the “camera” treatment in columns (2), (6), (10), and on all the regressors including dummies for gender, poverty, study track and municipality status in the rest columns, displaying the estimated coefficient of each regressor, with the standard error in the parentheses underneath the coefficient. The “camera” treatment is equal to one for student from county that implemented the CCTV monitoring in 2011 and in 2012, and zero otherwise. The standard errors are clustered at the county level (42 clusters).

B The heterogeneous effects of the anti-corruption campaign on poor students

	Baccalaureate Pass Indicator			Romanian Written Score			Overall Baccalaureate Score		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Camera	-0.081 (0.026)	-0.088 (0.023)	-0.077 (0.017)	-0.302 (0.110)	-0.255 (0.082)	-0.131 (0.056)	-0.433 (0.141)	-0.434 (0.113)	-0.308 (0.076)
Poor × camera	-0.062 (0.015)	-0.056 (0.012)	-0.051 (0.013)	-0.213 (0.063)	-0.257 (0.053)	-0.214 (0.052)	-0.346 (0.078)	-0.350 (0.061)	-0.306 (0.063)
Year fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County fixed effect × yearly trends	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Poor interactions	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Ability interactions	No	No	Yes	No	No	Yes	No	No	Yes
Observations	731,505	553,903	553,903	712,298	547,447	547,447	706,895	545,121	545,121
R ²	0.256	0.310	0.394	0.291	0.356	0.459	0.435	0.504	0.613

Notes: This table replicates table 4 from Borcan, Lindahl, and Mitrut (2017). The table demonstrates the results of the OLS estimates based on the difference-in-differences strategy which regress the Baccalaureate exam pass rates, Romanian Written exam scores and overall Baccalaureate scores on the “camera” treatment and and the interaction terms between poverty status and all variables for the 2009–2012 school years. Columns (1), (4), and (7) include only poverty status interactions with all variables, including county fixed effects and county trends. Columns (2), (5), and (8) include only poverty status interactions. While columns (3), (6), and (9) include all interactions between an ability indicator and all variables, including county fixed effects and county trends. We control for students of low ability in columns (3), (6), (9). The controls include: poor, gender, theoretical track, and rural indicators. The standard errors are clustered at the county level (42 clusters).

C The effects of the anti-corruption campaign on the admission rates to elite university of poor students

	Poor students admitted to elite university		Poor students in an elite university with tuition-exempt (top students)		Poor students in an elite university with tuition-paying (good students)		Poor students in top 20 percent at Baccalaureate exam	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Camera	-0.025 (0.013)	-0.025 (0.013)	-0.025 (0.015)	-0.026 (0.015)	-0.025 (0.031)	-0.026 (0.030)	-0.022 (0.006)	-0.019 (0.006)
Year 09	-0.006 (0.004)	-0.005 (0.004)	-0.007 (0.006)	-0.006 (0.006)	-0.012 (0.009)	-0.013 (0.009)	-0.032 (0.004)	-0.031 (0.004)
Year 11	0.038 (0.011)	0.038 (0.011)	0.043 (0.012)	0.043 (0.012)	0.036 (0.028)	0.035 (0.027)	0.036 (0.004)	0.035 (0.004)
Year 12	0.055 (0.011)	0.056 (0.011)	0.055 (0.012)	0.055 (0.012)	0.063 (0.027)	0.061 (0.027)	0.060 (0.005)	0.058 (0.005)
Year fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County fixed effect×yearly trends	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Observations	15,821	15,821	10,023	10,023	5,798	5,798	142,214	142,214
R ²	0.043	0.048	0.043	0.050	0.064	0.069	0.048	0.066

Notes: This table replicates table 6 from Borcan, Lindahl, and Mitrut (2017). The table demonstrates the results of the OLS estimates based on the difference-in-differences strategy for a formation of students admitted at an elite university in the period between 2009 and 2012. The dependent variable from column (1) to column (6) are the poverty status of the students who were admitted into an elite university, with dummy=1 indicating poor and dummy=0 otherwise. For column (7) and (8), the dependent variables are the poverty status for students who scored in the top 20 percent in the working sample of the Baccalaureate exam, with dummy=1 indicating poor and dummy=0 otherwise. The control variables for the university admission sample are as follows: student's gender, study track and dummy indicator for students who took the Baccalaureate before the year of university admission; control variables for the Baccalaureate top 20 percent sample are as follows: student's gender, study track, and a municipality status indicator. The standard errors are clustered at the county level (42 clusters).

D Definitions of covariates in the Romanian education system application

Y		Three exam outcomes, i.e., the overall Bacalaureate exam score, the Romanian written score and the dummy that equals to 1 if a student pass the exam
W		The “camera” treatment, which equals to 1 if a student is monitored by CCTV during the exam
X_1	school_id	School id
X_2	county_id	County id
X_3	male	Dummy which equals to 1 if a student is male, and 0 otherwise
X_4	orrom	Romanian oral score
X_5	rur	Municipality status of the place which the student comes from, e.g., rural and non-rural
X_6	yr09	Dummy for year 2009
X_7	yr10	Dummy for year 2010
X_8	yr11	Dummy for year 2011
X_9	yr12	Dummy for year 2012
X_{10}	poor	Dummy which equals to 1 if a student is poor, and 0 otherwise
X_{11}	midschool_gpa	Student’s middle school GPA
X_{12}	theoretic	Dummy which equals to 1 for theoretic study track, and 0 otherwise
X_{13}	soros_2	Dummy which equals to 1 if a student comes from a more corrupt county and 0 otherwise
X_{14}	rep_bac	How many times a student retakes Bacalaureate exam
X_{15}	top20pcent	Dummy which equals to 1 if a student scores top 20% of the overall exam and 0 otherwise
X_{16}	top20pcent_rom	Dummy which equals to 1 if a student scores top 20% of the Romanian written exam and 0 otherwise
X_{17}	ability_std	Standardized ability measure
X_{18}	fam_2child	Dummy which equals to 1 if a student comes from a family with 2 children and 0 otherwise
X_{19}	county_pop	County population
X_{20}	county_shromanian	Poor rate of a county
X_{21}	total_population	Total population
X_{22}	urban_population	Urban population

E Definitions of covariates in the incentives for college achievement application

Y		The main achievement measures in college of samples, i.e., GPA, good standing and credits earned of year 1 and year 2
W		The SFSP treatment, which equals to 1 if a student receives the SFSP service and 0 otherwise
X_1	age	The age of a student
X_2	chooseUTM	Dummy which equals to 1 if a student at his first choice school and 0 otherwise
X_3	composure	Dummy which equals to 1 if a student completed the survey and 0 otherwise
X_4	control	Dummy which equals to 1 if a student is one of the controlled sample and 0 otherwise
X_5	dad1	Dummy which equals to 1 if a student’s father graduated from high school and 0 otherwise
X_6	dad2	Dummy which equals to 1 if a student’s father graduated from college and 0 otherwise
X_7	dad_edn	Students’ fathers’ education level
X_8	english	Dummy which equals to 1 if a student’s mother tongue is English and 0 otherwise
X_9	female	Dummy for gender which equals to 1 for female and 0 for male
X_{10}	finish4	Dummy which equals to 1 if a student intends to finish college study within 4 years and 0 otherwise
X_{11}	gpa0	Student’s high school GPA
X_{12}	graddeg	Dummy which equals to 1 if a student wants more than a Bachelor degree and 0 otherwise
X_{13}	hcom	Dummy which equals to 1 if a student lives at home and 0 otherwise
X_{14}	hsgroup	Student’s high school group
X_{15}	lastmin	How often does a student leave studying until the last minute for tests and exams
X_{16}	lm_never	Dummy which equals to 1 if a student never puts of studying for tests and 0 otherwise
X_{17}	lm_rarely	Dummy which equals to 1 if a student rarely puts of studying for tests and 0 otherwise
X_{18}	mom1	Dummy which equals to 1 if a student’s mother graduated from high school and 0 otherwise
X_{19}	mom2	Dummy which equals to 1 if a student’s mother graduated from college and 0 otherwise
X_{20}	mom_edn	Students’ fathers’ education level
X_{21}	numcourses_nov1	Number of courses a student enrolled as of fall 2005
X_{22}	used_adv	Dummy which equals to 1 if a student ever emailed an advisor and 0 otherwise
X_{23}	used_fsg	Dummy which equals to 1 if a student ever attended FSGs and 0 otherwise
X_{24}	work1	Dummy which equals to 1 if a student plans to work while in school and 0 otherwise
X_{25}	prob_year1	Dummy which equals to 1 if a student is on probation in year 1 and 0 otherwise
X_{26}	prob_year2	Dummy which equals to 1 if a student is on probation in year 2 and 0 otherwise

F Second year heterogeneous treatment effects results of the incentives for college achievement application

Year 2	GPA		Good standing		Credits earned	
	HTE	test statistics	HTE	test statistics	HTE	test statistics
High dad_edn	-0.045 (0.184)	t-value = -17.121	-0.009 (0.048)	t-value = 1.316	-0.067 (0.284)	t-value = -0.333
Low dad_edn	-0.009 (0.117)	p-value <2.2e-16	0.014 (0.036)	p-value = 0.189	-0.231 (0.173)	p-value = 0.739
High gpa0	-0.024 (0.157)	t-value = -4.207	0.035 (0.033)	t-value = 0.325	-0.098 (0.205)	t-value = -1.732
Low gpa0	0.006 (0.125)	p-value = 2.81e-5	-0.011 (0.048)	p-value = 0.746	-0.280 (0.211)	p-value = 0.084
High lastmin	-0.155 (0.176)	t-value = -13.473			-0.320 (0.238)	t-value = -2.851
Low lastmin	0.065 (0.120)	p-value <2.2e-16			-0.113 (0.185)	p-value = 0.005
Lm_rarely	-0.227 (0.188)	t-value = -13.040	-0.007 (0.064)	t-value = 1.274	-0.360 (0.263)	t-value = -2.905
Non-lm_rarely	0.085 (0.117)	p-value <2.2e-16	0.017 (0.031)	p-value = 0.204	-0.110 (0.176)	p-value = 0.004
High mom_edn	-0.083 (0.143)	t-value = -15.187	0.044 (0.047)	t-value = -0.728	-0.055 (0.221)	t-value = -1.965
Low mom_edn	0.043 (0.135)	p-value <2.2e-16	-0.019 (0.037)	p-value = 0.467	-0.316 (0.201)	p-value = 0.050
Used_adv	-0.042 (0.163)	t-value = -0.77263	0.023 (0.043)	t-value = -0.107	-0.248 (0.246)	t-value = 0.300
Non-used_adv	0.020 (0.114)	p-value = 0.442	-0.005 (0.042)	p-value = 0.915	-0.139 (0.162)	p-value = 0.769
Male	-0.121 (0.184)	t-value = 6.127				
Female	0.057 (0.114)	p-value = 1.309e-9				
High numcourses_nov1			-0.038 (0.059)	t-value = -0.443	0.166 (0.265)	t-value = 0.699
Low numcourses_nov1			0.025 (0.036)	p-value = 0.658	-0.357* (0.180)	p-value = 0.485
Dad2	-0.264 (0.144)	t-value = -51.336				
Non-Dad2	0.185 (0.134)	p-value <2.2e-16				
Prob_year1			0.048 (0.164)	t-value = 0.346		
Non-prob_year1			-0.002 (0.018)	p-value = 0.730		

Notes: The meaning of the abbreviations of the covariates in this table can be found in Appendix E. Furthermore, we indicate different subsamples by giving them different names, e.g., “High dad_edn” and “Low dad_edn” represent the subsamples of students whose fathers’ education level is higher than the median and lower than the median respectively.

G Plots illustrating heterogeneity in CATEs in the incentives for college achievement application (second year)

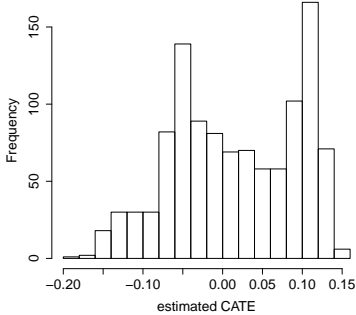


Figure 7: GPA.

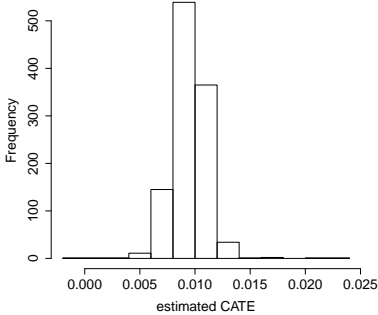


Figure 8: good standing.

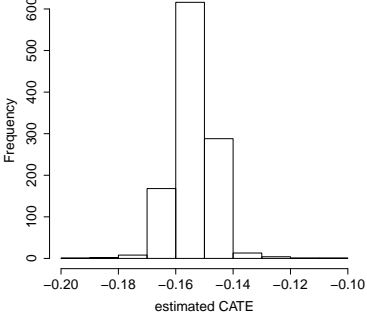


Figure 9: credits earned.

H R code for the Romanian education system application

```
1 library(grf)
2 if (packageVersion("grf") < '0.10.2') {
3   warning("This script requires grf 0.10.2 or higher")
4 }
5 library(sandwich)
6 library(lmtest)
7 library(Hmisc)
8 library(ggplot2)
9
10 # get county id covariate
11 county_id = as.numeric(covariates$county_id)
12
13 # generate Y.forest and W.forest, and estimate Y.hat and W.hat
14 Y.forest = regression_forest(X, Y, clusters = county_id)
15 Y.hat = predict(Y.forest)$predictions
16 W.forest = regression_forest(X, W, clusters = county_id)
17 W.hat = predict(W.forest)$predictions
18
19 # grow a temporary causal forest, and get the important covariates causing heterogeneity
20 cf.raw = causal_forest(X,
21                       Y,
22                       W,
23                       Y.hat = Y.hat,
24                       W.hat = W.hat,
25                       clusters = county_id)
26 varimp = variable_importance(cf.raw)
27 selected.idx = which(varimp > mean(varimp))
28
29 # grow the final version of a causal forest based on the selected covariates, and estimate tau hat for each observation
30 cf = causal_forest(
31   X[, selected.idx],
32   Y,
33   W,
34   Y.hat = Y.hat,
35   W.hat = W.hat,
36   clusters = county_id,
37   tune.parameters = TRUE
38 )
39 tau.hat = predict(cf)$predictions
40
41 #
42 # Estimate average treatment effect (ATE)
43 #
44
45 ATE = average_treatment_effect(cf)
46 paste("95% CI for the ATE:", round(ATE[1], 3),
47       "+/-", round(qnorm(0.975) * ATE[2], 3))
48
49 #
50 # Omnibus tests for heterogeneity
51 #
52
53 test_calibration(cf)
54
55 # Compare regions with high and low estimated CATEs
56 high_effect = tau.hat > median(tau.hat)
57 ate.high = average_treatment_effect(cf, subset = high_effect)
58 ate.low = average_treatment_effect(cf, subset = !high_effect)
59 paste(
60   "95% CI for difference in ATE:",
61   round(ate.high[1] - ate.low[1], 3),
62   "+/-",
```

```

63   round(qnorm(0.975) * sqrt(ate.high[2] ^ 2 + ate.low[2] ^ 2), 3)
64 )
65
66 #
67 # try test significance of the heterogeneity for students with high middle school GPA and low middle school GPA
68 #
69
70 middle_school_gpa <- covariates$midschool_gpa
71 high_gpa = middle_school_gpa > median(middle_school_gpa)
72
73 # get heterogeneous treatment effect on students with high middle school gpa by "camera" treatment
74 hete_treatment_high_gpa = average_treatment_effect(cf, target.sample = "overlap", subset = high_gpa)
75
76 # get heterogeneous treatment effect on students with low middle school gpa by "camera" treatment
77 hete_treatment_low_gpa = average_treatment_effect(cf, target.sample = "overlap", subset = !high_gpa)
78
79 # t-test heterogeneity in middle school gpa
80 t.test(tau.hat[high_gpa], tau.hat[!high_gpa])
81
82 #
83 # try test the significance of the heterogeneity for students with high Oral exam score and low Oral exam score
84 #
85
86 Oral_exam_score <- covariates$orrom
87 high_Oral_exam_score = Oral_exam_score > median(Oral_exam_score)
88
89 # get heterogeneous treatment effect on students with high Oral score by "camera" treatment
90 hete_treatment_high_Oral = average_treatment_effect(cf, target.sample =
91   "overlap", subset = high_Oral_exam_score)
92 # get heterogeneous treatment effect on students with low Oral score by "camera" treatment
93 hete_treatment_low_Oral = average_treatment_effect(cf,
94   target.sample = "overlap",
95   subset = !high_Oral_exam_score)
96
97 # t-test heterogeneity in Oral score
98 t.test(tau.hat[high_Oral_exam_score], tau.hat[!high_Oral_exam_score])
99
100 #
101 # try test the significance of the heterogeneity for male students and female students
102 #
103 gender <- covariates$male
104 male = gender > 0
105
106 # get heterogeneous treatment effect for male students by "camera" treatment
107 hete_treatment_male = average_treatment_effect(cf, target.sample = "overlap", subset = male)
108
109 # get heterogeneous treatment effect for female students by "camera" treatment
110 hete_treatment_female = average_treatment_effect(cf, target.sample = "overlap", subset = !male)
111
112 # t-test heterogeneity for male and female students
113 t.test(tau.hat[male], tau.hat[!male])
114
115 #
116 # try test the significance of the heterogeneity for rural and nonrural municipality
117 #
118
119 municipality_status <- covariates$rur
120 rural = municipality_status > 0
121
122 # get heterogeneous treatment effect for students from rural municipality by "camera" treatment
123 hete_treatment_rural = average_treatment_effect(cf, target.sample = "overlap", subset = rural)
124
125 # get heterogeneous treatment effect for students from nonrural municipality by "camera" treatment
126 hete_treatment_nonrural = average_treatment_effect(cf, target.sample = "overlap", subset = !rural)
127
128 # t-test heterogeneity for rural and nonrural students

```



```

129 t.test(tau.hat[rural], tau.hat[!rural])
130
131 #
132 # try test the significance of the heterogeneity for poor students and nonpoor students
133 #
134
135 poverty_status <- covariates$poor
136 poor = poverty_status > 0
137
138 # get heterogeneous treatment effect for poor students by "camera" treatment
139 hete_treatment_poor = average_treatment_effect(cf, target.sample = "overlap", subset = poor)
140
141 # get heterogeneous treatment effect for nonpoor students by "camera" treatment
142 hete_treatment_nonpoor = average_treatment_effect(cf, target.sample = "overlap", subset = !poor)
143
144 # t-test heterogeneity for poor and nonpoor students
145 t.test(tau.hat[poor], tau.hat[!poor])
146
147 #
148 # try test the significance of the heterogeneity for students taking theoretic track and other track
149 #
150
151 study_track <- covariates$theoretic
152 theoretic = study_track > 0
153
154 # get heterogeneous treatment effect for theoretic track students by "camera" treatment
155 hete_treatment_theoretic = average_treatment_effect(cf, target.sample =
156                                     "overlap", subset = theoretic)
157 # get heterogeneous treatment effect for other track students by "camera" treatment
158 hete_treatment_nontheoretic = average_treatment_effect(cf, target.sample =
159                                     "overlap", subset = !theoretic)
160
161 # t-test heterogeneity for theoretic track students and other track students
162 t.test(tau.hat[theoretic], tau.hat[!theoretic])
163
164 #
165 # try test the significance of the heterogeneity for students coming from more corrupt and less corrupt counties
166 #
167
168 corruption_status <- covariates$soros_2
169 more_corrupt = corruption_status > 0
170
171 # get heterogeneous treatment effect for students coming from more corrupt counties by "camera" treatment
172 hete_treatment_more_corrupt = average_treatment_effect(cf, target.sample =
173                                     "overlap", subset = more_corrupt)
174 # get heterogeneous treatment effect for students coming from less corrupt counties by "camera" treatment
175 hete_treatment_less_corrupt = average_treatment_effect(cf, target.sample =
176                                     "overlap", subset = !more_corrupt)
177
178 # t-test heterogeneity for students coming from more corrupt and less corrupt counties
179 t.test(tau.hat[more_corrupt], tau.hat[!more_corrupt])
180
181 #
182 # try test the significance of the heterogeneity for students with high and low ability index
183 #
184
185 standardized_ability <- covariates$ability_std
186 high_ability = standardized_ability > median(standardized_ability)
187
188 # get heterogeneous treatment effect for students with high ability by "camera" treatment
189 hete_treatment_high_ability = average_treatment_effect(cf, target.sample =
190                                     "overlap", subset = high_ability)
191 # get heterogeneous treatment effect for students with low ability by "camera" treatment
192 hete_treatment_low_ability = average_treatment_effect(cf, target.sample =
193                                     "overlap", subset = !high_ability)
194

```

```

195 # t-test heterogeneity for students with high and low ability index
196 t.test(tau.hat[high_ability], tau.hat[!high_ability])
197
198 #
199 # try test the significance of the heterogeneity for students coming from counties with high and low populations
200 #
201
202 county_population <- covariates$county_pop
203 high_population_county = county_population > median(county_population)
204
205 # get heterogeneous treatment effect for students coming from counties with high population by "camera" treatment
206 hete_treatment_high_population_county = average_treatment_effect(cf, target.sample =
207                                     "overlap", subset = high_population_county)
208 # get heterogeneous treatment effect for students coming from counties with low population by "camera" treatment
209 hete_treatment_low_population_county = average_treatment_effect(cf,
210                                     target.sample = "overlap",
211                                     subset = !high_population_county)
212
213 # t-test heterogeneity for students coming from counties with different populations
214 t.test(tau.hat[high_population_county], tau.hat[!high_population_county])
215
216 #
217 # try test the significance of the heterogeneity for students coming from areas with high and low total populations
218 #
219
220 total_population <- covariates$total_population
221 high_population_total = total_population > median(total_population)
222
223 # get heterogeneous treatment effect for students coming from areas with high total populations by "camera" treatment
224 hete_treatment_high_population_total = average_treatment_effect(cf, target.sample =
225                                     "overlap", subset = high_population_total)
226 # get heterogeneous treatment effect for students coming from areas with low total populations by "camera" treatment
227 hete_treatment_low_population_total = average_treatment_effect(cf,
228                                     target.sample = "overlap",
229                                     subset = !high_population_total)
230
231 # t-test heterogeneity for students coming from areas with different total populations
232 t.test(tau.hat[high_population_total], tau.hat[!high_population_total])
233
234 #
235 # try test the significance of the heterogeneity for students coming from areas with high and low urban populations
236 #
237
238 urban_population <- covariates$urban_population
239 high_population_urban = urban_population > median(urban_population)
240
241 # get heterogeneous treatment effect for students coming from areas with high urban populations by "camera" treatment
242 hete_treatment_high_population_urban = average_treatment_effect(cf, target.sample =
243                                     "overlap", subset = high_population_urban)
244 # get heterogeneous treatment effect for students coming from areas with low urban populations by "camera" treatment
245 hete_treatment_low_population_urban = average_treatment_effect(cf,
246                                     target.sample = "overlap",
247                                     subset = !high_population_urban)
248
249 # t-test heterogeneity for students coming from areas with different urban populations
250 t.test(tau.hat[high_population_urban], tau.hat[!high_population_urban])
251
252 #
253 # try test the significance of the heterogeneity for students coming from counties with high and low poor rates
254 #
255
256
257 shromanian_population <- covariates$county_shromanian
258 high_population_shromanian = shromanian_population > median(shromanian_population)
259

```

```

260 # get heterogeneous treatment effect for students coming from areas with high shromanian populations by "camera"
      treatment
261 hete_treatment_high_population_shromanian = average_treatment_effect(cf, target.sample =
262                                     "overlap", subset = high_population_shromanian)
263 # get heterogeneous treatment effect for students coming from areas with low shromanian populations by "camera"
      treatment
264 hete_treatment_low_population_shromanian = average_treatment_effect(cf,
265                                     target.sample = "overlap",
266                                     subset = !high_population_shromanian)
267
268 # Bonferroni test heterogeneity for students coming from areas with different shromanian populations
269 t.test(tau.hat[high_population_shromanian], tau.hat[!high_population_shromanian])
270
271 #
272 #
273 # try test the significance of the heterogeneity for retaking different times of exams
274 #
275
276 retake_times <- covariates$rep_bac
277 high_retake_times = retake_times > median(retake_times)
278
279 # get heterogeneous treatment effect for students coming from areas with high shromanian populations by "camera"
      treatment
280 hete_treatment_high_retake_times = average_treatment_effect(cf, target.sample =
281                                     "overlap", subset = high_retake_times)
282 # get heterogeneous treatment effect for students coming from areas with low shromanian populations by "camera"
      treatment
283 hete_treatment_low_retake_times = average_treatment_effect(cf,
284                                     target.sample = "overlap",
285                                     subset = !high_retake_times)
286
287 # t-test heterogeneity for students coming from areas with different shromanian populations
288 t.test(tau.hat[high_retake_times], tau.hat[!high_retake_times])
289
290 #
291 # Make the CATE frequency plot
292 #
293
294 pdf("tauhat_hist.pdf")
295 pardef = par(
296   mar = c(5, 4, 4, 2) + 0.5,
297   cex.lab = 1.5,
298   cex.axis = 1.5,
299   cex.main = 1.5,
300   cex.sub = 1.5
301 )
302 hist(tau.hat, xlab = "estimated CATE", main = "")
303 dev.off()

```

I R code for the incentives for college achievement application

```
1 library(grf)
2 if (packageVersion("grf") < '0.10.2') {
3   warning("This script requires grf 0.10.2 or higher")
4 }
5 library(sandwich)
6 library(lmtest)
7 library(Hmisc)
8 library(ggplot2)
9
10
11 # train Y.forest and W.forest, and estimate Y.hat and W.hat
12 Y.forest = regression_forest(X, Y)
13 Y.hat = predict(Y.forest)$predictions
14 W.forest = regression_forest(X, W)
15 W.hat = predict(W.forest)$predictions
16
17 # grow a temporary causal forest, and get the important covariates causing heterogeneity
18 cf.raw = causal_forest(X, Y, W, Y.hat = Y.hat, W.hat = W.hat)
19 varimp = variable_importance(cf.raw)
20 selected.idx = which(varimp > mean(varimp))
21
22 # grow the final version of a causal forest based on the selected covariates, and estimate tau hat for each observation
23 cf = causal_forest(X[, selected.idx],
24                   Y,
25                   W,
26                   Y.hat = Y.hat,
27                   W.hat = W.hat,
28                   tune.parameters = TRUE)
29 tau.hat = predict(cf)$predictions
30
31 #
32 # Estimate average treatment effect (ATE)
33 #
34
35 ATE = average_treatment_effect(cf, target.sample = "treated")
36 paste("95% CI for the ATE:", round(ATE[1], 3),
37       "+/-", round(qnorm(0.975) * ATE[2], 3))
38
39 #
40 # Omnibus tests for heterogeneity
41 #
42 test_calibration(cf)
43
44
45 # get gender covariates
46 gender <- covariates$female
47 female <- gender > 0
48
49 # get the heterogeneous treatment effects along gender
50 hetero_female <-
51   average_treatment_effect(cf, target.sample = "treated", subset = female)
52 hetero_male <-
53   average_treatment_effect(cf, target.sample = "treated", subset = !female)
54
55 # get the 95% confidence interval of the treatment effects for female
56 paste(
57   "95% CI for the hetero_female:",
58   round(hetero_female[1], 3),
59   "+/-",
60   round(qnorm(0.975) * hetero_female[2], 3)
61 )
62
```

```

63 # get the 95% confidence interval of the treatment effects for male
64 paste(
65   "95% CI for the hetero_male:",
66   round(hetero_male[1], 3),
67   "+/-",
68   round(qnorm(0.975) * hetero_male[2], 3)
69 )
70
71 # t-test for the significance of the heterogeneity
72 t.test(tau.hat[female], tau.hat[!female])
73
74 #
75 # try to test significance of the heterogeneity along father's education level
76 #
77
78 dad_edu <- covariates$dad_edn
79 high_dad_edu <- dad_edu > median(dad_edu)
80 hete_high_dad_edu <-
81   average_treatment_effect(cf, target.sample = "treated", subset = high_dad_edu)
82 hete_low_dad_edu <-
83   average_treatment_effect(cf, target.sample = "treated", subset = !high_dad_edu)
84
85 # t-test the signigicance of the heterogeneity
86 t.test(tau.hat[high_dad_edu], tau.hat[!high_dad_edu])
87
88 #
89 # try to test significance of the heterogeneity along students' high school GPA
90 #
91
92 highschool_GPA <- covariates$gpa0
93 high_GPA <- highschool_GPA > median(highschool_GPA)
94 hete_high_GPA <-
95   average_treatment_effect(cf, target.sample = "treated", subset = high_GPA)
96 hete_low_GPA <-
97   average_treatment_effect(cf, target.sample = "treated", subset = !high_GPA)
98
99 # t-test the signigicance of the heterogeneity
100 t.test(tau.hat[high_GPA], tau.hat[!high_GPA])
101
102 #
103 # try to test significance of the heterogeneity along the lastmin covariate
104 #
105
106 lastmin <- covariates$lastmin
107 high_lastmin <- lastmin > median(lastmin)
108 hete_high_lastmin <-
109   average_treatment_effect(cf, target.sample = "treated", subset = high_lastmin)
110 hete_low_lastmin <-
111   average_treatment_effect(cf, target.sample = "treated", subset = !high_lastmin)
112
113 # t-test the signigicance of the heterogeneity
114 t.test(tau.hat[high_lastmin], tau.hat[!high_lastmin])
115
116 #
117 # try to test significance of the heterogeneity along the lm_rarely covariate
118 #
119
120 lm_rarely <- covariates$lm_rarely
121 rarely <- lm_rarely > 0
122 hete_rarely <-
123   average_treatment_effect(cf, target.sample = "treated", subset = rarely)
124 hete_non_rarely <-
125   average_treatment_effect(cf, target.sample = "treated", subset = !rarely)
126
127 # t-test the signigicance of the heterogeneity
128 t.test(tau.hat[rarely], tau.hat[!rarely])

```

```

129
130 #
131 # try to test significance of the heterogeneity along mother's education level
132 #
133
134 mom_edu <- covariates$mom_edn
135 high_mom_edu <- mom_edu > median(mom_edu)
136 hete_high_mom_edu <-
137   average_treatment_effect(cf, target.sample = "treated", subset = high_mom_edu)
138 hete_low_mom_edu <-
139   average_treatment_effect(cf, target.sample = "treated", subset = !high_mom_edu)
140
141 # t-test the signigicance of the heterogeneity
142 t.test(tau.hat[high_mom_edu], tau.hat[!high_mom_edu])
143
144 #
145 # try to test significance of the heterogeneity along the used_adv covariate
146 #
147
148 used_adv <- covariates$used_adv
149 used <- used_adv > 0
150 hete_used_adv <-
151   average_treatment_effect(cf, target.sample = "treated", subset = used)
152 hete_non_used_adv <-
153   average_treatment_effect(cf, target.sample = "treated", subset = !used)
154
155 # t-test the signigicance of the heterogeneity
156 t.test(tau.hat[used], tau.hat[!used])
157
158 #
159 # try to test significance of the heterogeneity along the the numcourses_nov1 covariate
160 #
161
162 numcourses_nov1 <- covariates$numcourses_nov1
163 high_numcourses_nov1 <- numcourses_nov1 > median(numcourses_nov1)
164 hete_high_numcourses_nov1 <-
165   average_treatment_effect(cf, target.sample = "treated", subset = high_numcourses_nov1)
166 hete_low_numcourses_nov1 <-
167   average_treatment_effect(cf,
168     target.sample = "treated",
169     subset = !high_numcourses_nov1)
170
171 # t-test the signigicance of the heterogeneity
172 t.test(tau.hat[high_numcourses_nov1], tau.hat[!high_numcourses_nov1])
173
174 #
175 # try to test significance of the heterogeneity along the dad2 covariate
176 #
177
178 dad2 <- covariates$dad2
179 dad_college_gra <- dad2 > 0
180 hete_dad_college_gra <-
181   average_treatment_effect(cf, target.sample = "treated", subset = dad_college_gra)
182 hete_dad_non_college_gra <-
183   average_treatment_effect(cf,
184     target.sample = "treated",
185     subset = !dad_college_gra)
186
187 # t-test the signigicance of the heterogeneity
188 t.test(tau.hat[dad_college_gra], tau.hat[!dad_college_gra])
189
190 #
191 # try to test significance of the heterogeneity along the prob_year1 covariate
192 #
193
194 prob_fir_year <- covariates$prob_year1

```

```
195 prob <- prob_fir_year > 0
196 hete_prob <-
197   average_treatment_effect(cf, target.sample = "treated", subset = prob)
198 hete_non_prob <-
199   average_treatment_effect(cf, target.sample = "treated", subset = !prob)
200
201 # t-test the signigicance of the heterogeneity
202 t.test(tau.hat[prob], tau.hat[!prob])
203
204 #
205 # Make the CATE frequency plot
206 #
207
208 pdf("tauhat_hist.pdf")
209 pardef = par(
210   mar = c(5, 4, 4, 2) + 0.5,
211   cex.lab = 1.5,
212   cex.axis = 1.5,
213   cex.main = 1.5,
214   cex.sub = 1.5
215 )
216 hist(tau.hat, xlab = "estimated CATE", main = "")
217 dev.off()
```

J Stata code for the replication which is provided by Borcan, Lindahl, and Mitrut (2017), the authors of this code are: Oana Borcan, Andreea Mitrut, Mikael Lindahl

```

1  *Authors: Oana Borcan, Andreea Mitrut, Mikael Lindahl
2  *Last modified: 21/06/2016
3
4  /*****
5
6  Replication File for Tables 2, 4, 6 of Borcan, Lindahl and Mitrut (2016)
7      "Fighting Corruption in Education: What Works and Who Benefits?"
8
9  *****/
10
11 /*
12
13
14 Table 2:
15
16     Datasets: BAC_camera_paper_dataset_1015.dta
17
18     Additional Variables:
19
20     new_cam: treatment dummy=1 if student treated with camera surveillance
21     t: linear time trend
22     dcounty1-dcounty42: dummies for each county
23     yr09-yr12: dummy=1 for 2009, 2010, 2011 and 2012
24     cam: dummy=1 for counties treated with cameras in 2011.
25           These counties are: Arges, Alba, Bucuresti, Iasi, Bihor, Bacau, Vrancea, Giurgiu, Neamt, Prahova,
26                               Buzau, Calarasi, Caras Severin, Braila, Dolj, Galati, Brasov, Maramures,
27                               Timis, Mures, Tulcea, Dambovita, Arad, Hunedoara, Mehedinti
28     new_cam: Treatment variable. dummy=1 for counties that installed early in year 2011, and for all counties in 2012.
29
30 Table 4:
31
32     Datasets: BAC_camera_paper_dataset_1015.dta
33
34     Additional Variables:
35
36     poor_new_cam: poor interacted with camera treatment
37     poor09-poor12: poor interacted with year dummies
38     poor_male: poor interacted with male
39     poor_theoretic: poor interacted with theoretical track
40     poor_rur: poor interacted with rural
41     poort1-poort42: poor inteacted with county specific linear time trends
42     poordcounty1-poordcounty42: poor interacted with county indicators
43     lo_new_cam: low ability interacted with camera treatment
44     lo09-lo12: low ability interacted with year dummies
45     lo_male: low ability interacted with male
46     lo_theoretic: low ability interacted with theoretical track
47     lo_rur: low ability interacted with rural
48     lo_poor: low ability interacted with poor
49     lot1-lot42: low ability inteacted with county specific linear time trends
50     lodcounty1-lodcounty42: low ability interacted with county indicators
51
52 Table 6:
53
54     Datasets: BAC_elite_uni.dta, BAC_camera_paper_dataset_1015.dta
55
56     Additional Variables:

```



```

57
58     fee_exempt: dummy=1 for students exempt from tuition fees at an elite university (top students)
59     top20percent: a dummy=1 for students in the top 20% of the overall Baccalaureate score distribution in each year
60 */
61
62 /*****
63 Variable definition
64 *****/
65 clear all
66 set more off
67
68
69 cd "C:\2015-0074_replication" /*change path*/
70
71 use "2015-0074_data\BAC_main_2009_2012.dta"
72
73 *create county dummies
74 tab county_id, gen(dcounty)
75
76 * create county specific linear time trends
77 forvalues j = 1/42 {
78     gen t'j'=t*dcounty'j'
79 }
80
81 /* generate indicator for low ability students*/
82 su midschool_gpa, det
83 *8.81 median
84 gen lo_ability=1 if midschool_gpa<8.81&midschool_gpa!=.
85 replace lo_ability=0 if midschool_gpa>=8.81&midschool_gpa!=.
86
87
88 /*****
89 Create Table 2: The impact of the anti-corruption campaign: main results
90 *****/
91
92 reg finwrrom yr12 yr11 yr09 dcounty*, cluster(county_id)
93 outreg2 yr12 yr11 yr09 using "Tables\Table2_main.rtf",dec(3) nocons ctitle("") addtext("County FE, Yes, Controls,
94     No, County FE x Yearly Trends, No") word replace
95 reg finwrrom new_cam yr12 yr11 yr09 dcounty*, cluster(county_id)
96 outreg2 new_cam yr12 yr11 yr09 using "Tables\Table2_main.rtf",dec(3) nocons ctitle("") addtext("County FE, Yes,
97     Controls, No, County FE x Yearly Trends, No") word append
98 reg finwrrom new_cam yr12 yr11 yr09 dcounty* male poor theoretic rur, cluster(county_id)
99 outreg2 new_cam yr12 yr11 yr09 using "Tables\Table2_main.rtf",dec(3) nocons ctitle("") addtext("County FE, Yes,
100     Controls, Yes, County FE x Yearly Trends, Yes") word append
101
102 reg pass yr12 yr11 yr09 dcounty*, cluster(county_id)
103 outreg2 yr12 yr11 yr09 using "Tables\Table2_main.rtf",dec(3) nocons ctitle("") addtext("County FE, Yes, Controls,
104     No, County FE x Yearly Trends, No") word append
105 reg pass new_cam yr12 yr11 yr09 dcounty*, cluster(county_id)
106 outreg2 new_cam yr12 yr11 yr09 using "Tables\Table2_main.rtf",dec(3) nocons ctitle("") addtext("County FE, Yes,
107     Controls, No, County FE x Yearly Trends, No") word append
108 reg pass new_cam yr12 yr11 yr09 dcounty* male poor theoretic rur, cluster(county_id)
109 outreg2 new_cam yr12 yr11 yr09 using "Tables\Table2_main.rtf",dec(3) nocons ctitle("") addtext("County FE, Yes,
110     Controls, Yes, County FE x Yearly Trends, Yes") word append
111
112 reg final_score yr12 yr11 yr09 dcounty*, cluster(county_id)
113 outreg2 yr12 yr11 yr09 using "Tables\Table2_main.rtf",dec(3) nocons ctitle("") addtext("County FE, Yes, Controls,
114     No, County FE x Yearly Trends, No") word append
115 reg final_score new_cam yr12 yr11 yr09 dcounty*, cluster(county_id)

```

```

113  outreg2 new_cam yr12 yr11 yr09 using "Tables\Table2_main.rtf",dec(3)  nocons ctitle("") addtext("County FE, Yes,
      Controls, No, County FE x Yearly Trends, No") word append
114  reg final_score new_cam yr12 yr11 yr09 dcounty* male poor  theoretic rur , cluster(county_id)
115  outreg2 new_cam yr12 yr11 yr09 using "Tables\Table2_main.rtf",dec(3)  nocons ctitle("") addtext("County FE, Yes,
      Controls, Yes, County FE x Yearly Trends, No") word append
116  reg final_score new_cam yr12 yr11 yr09 male poor  theoretic rur  dcounty* t1-t41, cluster(county_id)
117  outreg2 new_cam  yr12 yr11 yr09 using "Tables\Table2_main.rtf",dec(3)  nocons ctitle("") addtext("County FE, Yes,
      Controls, Yes, County FE x Yearly Trends, Yes") word append

118
119  /*****
120  Create Table 4: Heterogeneous effects of the anti-corruption
121          campaign by poverty: a fully interacted model
122  *****/
123
124  /* generate interactions between poor and treatment/time/student
125  and school characteristics*/
126
127  gen poor_new_cam=poor*new_cam
128  gen poor09=poor*yr09
129  gen poor11=poor*yr11
130  gen poor12=poor*yr12
131  gen poor_male=poor*male
132  gen poor_theoretic=poor*theoretic
133  gen poor_rur=poor*rur
134
135  foreach var of varlist t1-t42 {
136    generate poor`var' = `var' * poor
137  }
138  foreach var of varlist dcounty1-dcounty42 {
139    generate poor`var' = `var' * poor
140  }
141
142  /* generate interactions between low ability and treatment/time/student
143  and school characteristics*/
144
145  gen lo_new_cam=lo_ability*new_cam
146  gen lo09=lo_ability*yr09
147  gen lo11=lo_ability*yr11
148  gen lo12=lo_ability*yr12
149  gen lo_male=lo_ability*male
150  gen lo_theoretic=lo_ability*theoretic
151  gen lo_rur=lo_ability*rur
152  gen lo_poor=lo_ability*poor
153
154  foreach var of varlist t1-t42 {
155    generate lo`var' = `var' * lo_ability
156  }
157  foreach var of varlist dcounty1-dcounty42 {
158    generate lo`var' = `var' * lo_ability
159  }
160
161  *declare variables lists
162
163  global poorinteractions "poor_new_cam poor12  poor11 poor09  poor_male poor_theoretic poor_rur
      poordcounty1-poordcounty37 poordcounty39-poordcounty42 poort1-poort37 poort39-poort42"
164  global lo_ability_interactions "lo_new_cam  lo12 lo11 lo09  lo_male lo_theoretic lo_rur lo_poor lodcounty1-lodcounty37
      lodcounty39-lodcounty42 lot1-lot37 lot39-lot42"
165
166  *create table
167
168  reg finwrrom new_cam yr12 yr11 yr09  $poorinteractions  poor male theoretic rur  dcounty1-dcounty37
      dcounty39-dcounty42 t1-t37 t39-t42 , cluster(county_id)
169  outreg2 new_cam poor_new_cam  using "Tables\Table4_interactions.rtf",dec(3)  nocons ctitle("") addtext("Year FE,
      Yes, County FE, Yes, County FE x Yearly Trends, Yes, Controls, Yes, Poor Interactions, Yes, Ability
      Interactions, No") word replace

```

```

170 reg finwrrom new_cam yr12 yr11 yr09 $poorinteractions poor male theoretic rur dcounty1-dcounty37
      dcounty39-dcounty42 t1-t37 t39-t42 if lo_ability!=., cluster(county_id)
171 outreg2 new_cam poor_new_cam using "Tables\Table4_interactions.rtf",dec(3) nocons ctitle("") addtext("Year FE,
      Yes, County FE, Yes, County FE x Yearly Trends, Yes, Controls, Yes, Poor Interactions, Yes, Ability
      Interactions, No") word append
172 reg finwrrom new_cam yr12 yr11 yr09 $poorinteractions $lo_ability_interactions poor lo_ability male theoretic rur
      dcounty1-dcounty37 dcounty39-dcounty42 t1-t37 t39-t42 if lo_ability!=., cluster(county_id)
173 outreg2 new_cam poor_new_cam using "Tables\Table4_interactions.rtf",dec(3) nocons ctitle("") addtext("Year FE, Yes,
      County FE, Yes, County FE x Yearly Trends, Yes, Controls, Yes, Poor Interactions, Yes, Ability Interactions,
      Yes") word append
174
175 reg pass new_cam yr12 yr11 yr09 $poorinteractions poor male theoretic rur dcounty1-dcounty37 dcounty39-dcounty42
      t1-t37 t39-t42 , cluster(county_id)
176 outreg2 new_cam poor_new_cam using "Tables\Table4_interactions.rtf",dec(3) nocons ctitle("") addtext("Year FE, Yes,
      County FE, Yes, County FE x Yearly Trends, Yes, Controls, Yes, Poor Interactions, Yes, Ability Interactions,
      No") word append
177 reg pass new_cam yr12 yr11 yr09 $poorinteractions poor male theoretic rur dcounty1-dcounty37 dcounty39-dcounty42
      t1-t37 t39-t42 if lo_ability!=., cluster(county_id)
178 outreg2 new_cam poor_new_cam using "Tables\Table4_interactions.rtf",dec(3) nocons ctitle("") addtext("Year FE, Yes,
      County FE, Yes, County FE x Yearly Trends, Yes, Controls, Yes, Poor Interactions, Yes, Ability Interactions,
      No") word append
179 reg pass new_cam yr12 yr11 yr09 $poorinteractions $lo_ability_interactions poor lo_ability male theoretic rur
      dcounty1-dcounty37 dcounty39-dcounty42 t1-t37 t39-t42 if lo_ability!=., cluster(county_id)
180 outreg2 new_cam poor_new_cam using "Tables\Table4_interactions.rtf",dec(3) nocons ctitle("") addtext("Year FE, Yes,
      County FE, Yes, County FE x Yearly Trends, Yes, Controls, Yes, Poor Interactions, Yes, Ability Interactions,
      Yes") word append
181
182 reg final_score new_cam yr12 yr11 yr09 $poorinteractions poor male theoretic rur dcounty1-dcounty37
      dcounty39-dcounty42 t1-t37 t39-t42 , cluster(county_id)
183 outreg2 new_cam poor_new_cam using "Tables\Table4_interactions.rtf",dec(3) nocons ctitle("") addtext("Year FE,
      Yes, County FE, Yes, County FE x Yearly Trends, Yes, Controls, Yes, Poor Interactions, Yes, Ability
      Interactions, No") word append
184 reg final_score new_cam yr12 yr11 yr09 $poorinteractions poor male theoretic rur dcounty1-dcounty37
      dcounty39-dcounty42 t1-t37 t39-t42 if lo_ability!=., cluster(county_id)
185 outreg2 new_cam poor_new_cam using "Tables\Table4_interactions.rtf",dec(3) nocons ctitle("") addtext("Year FE, Yes,
      County FE, Yes, County FE x Yearly Trends, Yes, Controls, Yes, Poor Interactions, Yes, Ability Interactions,
      No") word append
186 reg final_score new_cam yr12 yr11 yr09 $poorinteractions $lo_ability_interactions poor lo_ability male theoretic rur
      dcounty1-dcounty37 dcounty39-dcounty42 t1-t37 t39-t42 if lo_ability!=., cluster(county_id)
187 outreg2 new_cam poor_new_cam using "Tables\Table4_interactions.rtf",dec(3) nocons ctitle("") addtext("Year FE, Yes,
      County FE, Yes, County FE x Yearly Trends, Yes, Controls, Yes, Poor Interactions, Yes, Ability Interactions,
      Yes") word append
188
189 /*****
190 Create Table 6: The composition in terms of poverty at the admission
191 into an elite university and in the top 20% at the
192 Baccalaureate
193 *****/
194 clear all
195 set more off
196 cd "C:\2015-0074_replication" /*change path accordingly*/
197
198 use "2015-0074_data\Bac_elite_uni.dta"
199
200 ***define the camera indicator***
201 gen cam=1 if county_id==35 | county_id==22 | county_id==1 | county_id==26 | county_id==3 | county_id==37 | /*
202 */county_id==11 |county_id==9 |county_id==40 |county_id==24 |county_id==5 | county_id==14 |county_id==30
      |county_id==12 |county_id==18 /*
203 */county_id==21 |county_id==29 |county_id==19 |county_id==13 |county_id==27 |county_id==10 |county_id==4 |
      county_id==25 |county_id==7 | county_id==23
204 replace cam=0 if cam==.
205
206 gen new_cam=1 if (cam==1&year==2011)|year==2012
207 replace new_cam=0 if year<2011|(year==2011&cam==0)
208
209 ***define years and year trend var***

```

```

210 gen yr09=year==2009
211 gen yr10=year==2010
212 gen yr11=year==2011
213 gen yr12=year==2012
214
215 gen male=sex=="M"
216
217 *generate county dummies
218 tab county_id, gen(dcounty)
219
220 *generate linear time trend
221 gen t=1 if year==2009
222 replace t=2 if year==2010
223 replace t=3 if year==2011
224 replace t=4 if year==2012
225
226 *generate county specific linear time trends
227 forvalues j = 1/42 {
228     gen t'j'=t*dcounty'j'
229 }
230
231
232 *the mean admitted at elite university is in the top 20% in the bac score distribution of written Romanian score at
    BAC*
233
234 *generate dummy for tuition exempt students (the top admitted students)
235 gen fee_exempt=forma=="B"
236 *generate dummy for students who followed the theoretical track in high school
237 gen theoretic=(filiera=="TEO")
238
239 *cannot control for rur, almost no rural students
240 reg poor new_cam yr09 yr11 yr12 t1-t41 dcounty* , cluster(county_id)
241 outreg2 poor new_cam using "Tables\Table6_elite.rtf", dec(3) nocons ctitle("") addtext(Year FE, Yes, County FE, Yes,
    County FE x Yearly Trends, Yes, Controls, No) word replace
242 reg poor new_cam yr09 yr11 yr12 year_diferit male theoretic t1-t41 dcounty*, cluster(county_id)
243 outreg2 poor new_cam using "Tables\Table6_elite.rtf", dec(3) nocons ctitle("") addtext(Year FE, Yes, County FE, Yes,
    County FE x Yearly Trends, Yes, Controls, Yes) word append
244
245 reg poor new_cam yr09 yr11 yr12 dcounty* t1-t41 if fee_exempt==1, cluster(county_id)
246 outreg2 poor new_cam using "Tables\Table6_elite.rtf", dec(3) nocons ctitle("") addtext(Year FE, Yes, County FE, Yes,
    County FE x Yearly Trends, Yes, Controls, No) word append
247 reg poor new_cam yr09 yr11 yr12 year_diferit male theoretic dcounty* t1-t41 if fee_exempt==1, cluster(county_id)
248 outreg2 poor new_cam using "Tables\Table6_elite.rtf", dec(3) nocons ctitle("") addtext(Year FE, Yes, County FE, Yes,
    County FE x Yearly Trends, Yes, Controls, Yes) word append
249
250 reg poor new_cam yr09 yr11 yr12 dcounty* t1-t41 if fee_exempt==0, cluster(county_id)
251 outreg2 poor new_cam using "Tables\Table6_elite.rtf", dec(3) nocons ctitle("") addtext(Year FE, Yes, County FE, Yes,
    County FE x Yearly Trends, Yes, Controls, No) word append
252 reg poor new_cam yr09 yr11 yr12 year_diferit male theoretic dcounty* t1-t41 if fee_exempt==0, cluster(county_id)
253 outreg2 poor new_cam using "Tables\Table6_elite.rtf", dec(3) nocons ctitle("") addtext(Year FE, Yes, County FE, Yes,
    County FE x Yearly Trends, Yes, Controls, Yes) word append
254
255
256 /*Add columns 7 and 8, on poor composition in top 20% at Bacalaureate*/
257
258 clear all
259 set more off
260 use "2015-0074_data\BAC_main_2009_2012.dta"
261
262
263 *generate county dummies
264 tab county_id, gen(dcounty)
265
266 * create county specific linear time trends
267 forvalues j = 1/42 {
268     gen t'j'=t*dcounty'j'

```

```

269 }
270
271 /*find the 80th percentile in the overall Baccalaureate score for
272 each year */
273
274 sum final_score, det
275 pctl9 = final_score if year==2009, nq(10)
276 list pctl9 in 1/10
277
278 pctl10 = final_score if year==2010, nq(10)
279 list pctl10 in 1/10
280
281 pctl11 = final_score if year==2011, nq(10)
282 list pctl11 in 1/10
283
284 pctl12 = final_score if year==2012, nq(10)
285 list pctl12 in 1/10
286
287 /*generate an indicator for students above the 80th percentile in the
288 overall Baccalaureate score distribution*/
289
290 gen top20percent=1 if final_score!=. & ((final_score>9.15 & year==2009)|(final_score>8.51 & year==2010)|
    final_score>7.95 & year==2011)|(final_score>8.13 & year==2012))
291 replace top20percent=0 if top20percent==.
292
293 *create columns 7 and 8 for table 6
294
295 reg poor new_cam yr12 yr11 yr09 dcounty* t1-t41 if top20percent==1, cluster(county_id)
296 outreg2 new_cam using "Tables\Table6_elite.rtf", dec(3) nocons ctitle("") addtext(Year FE, Yes, County FE, Yes,
    County FE x Yearly Trends, Yes, Controls, No) word append
297
298 reg poor new_cam yr12 yr11 yr09 male theoretic rur dcounty* t1-t41 if top20percent==1, cluster(county_id)
299 outreg2 new_cam using "Tables\Table6_elite.rtf", dec(3) nocons ctitle("") addtext(Year FE, Yes, County FE, Yes,
    County FE x Yearly Trends, Yes, Controls, Yes) word append

```