Bachelor Thesis (Econometrics and Operational Research, Dutch programme)

Searching for heterogeneous treatment effects using generalized random forest

Abstract

One approach to estimate the treatment effect is the generalized random forest method of Athey et al. (2019). I apply this method on the dataset of Borcan et al. (2017) to investigate the possibility of additional heterogeneous effects. Borcan et al. (2017) estimated the effect of the anticorruption campaign on education in Romania. In the aftermath of an anticorruption campaign, mostly poor students were negatively impacted by the policy. In my research I take all available variables into account to access which variable has been mostly affected by the anticorruption campaign. I find that in addition to a significant heterogeneous effect for poor students there is also heterogeneous effects for the (expected) number of students taking the Baccalaureate exam, county's trust score for justice and unemployment rate. Due to the additional heterogeneous findings, both policy makers and researchers should be cautious when using parametric estimation methods for policy inference as it might not capture the heterogeneous effects well enough.

Erasmus University Rotterdam

Erasmus School of Economics

Author

Jon Hao Liu - 409933

Supervisor

Dr. (Andrea) A. A. Naghi

Second accessor

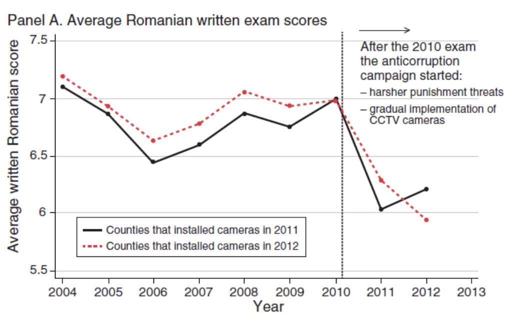
Dr. (Andreas) A. Pick

July 7, 2019

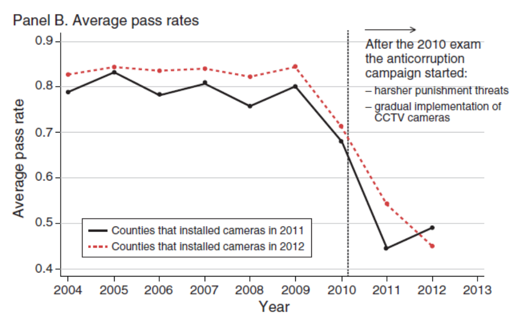
Contents

1 Introduction

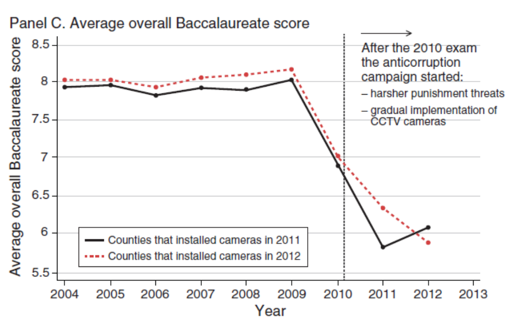
One of the major drives for country's economic growth can be assigned to an optimal functioning education system. Through the educational system, graduates are allocated to the job market by its ability and performance. For this system to be efficient, many hurdles have still to be overcome. Think about tuition fees and segregation, which limits the possibility for student to show their talents on a theoretical and/or practical aspects. But there is another mechanism that increases this inequality: corruption. Because of corruption in education, students can cheat by bribing the examinators and thus gain a higher performance than is possible with their skills. Corruption is often observed in (under)developing countries where inspection lacks or are even part of the corruption scandal but it does not limits to those kind of countries. Even in the US there is corruption in education present. In March 2019, Operation Varsity Blues investigated college admission scams from students in prestigious universities like Harvard University and Stanford University.[1] Corruption can exist in many forms such as bribes and fraud. Since these forms require financial resources, one would say the richer students mostly benefits from this kind of corruption and put the poorer peers at a disadvantage. By implementing policies that fight corruption, the results should favors the poor students but this is not the case in Romania. Borcan et al. (2017) measured the effectiveness of an anticorruption campaign held in Romania in 2011 and 2012. This campaign was proven to be effective since the grades of the written exam and final exam, called the Baccalaureate, dropped after the implementation. The following graphs show the steep drop for both the exam grades and pass rate of the Baccalaureate after the introduction of the campaign.



(a) Romanian written exam score



(b) Baccaulaureate pass rate



(c) Overall Baccalaureate score

Figure 1: Changing average scores and pass rate due to the anticorruption campaign (vertical line around 2010). Note: the change in pass rate and overall Baccalaureate score in 2010 is partly attributed to the change in exam structure which made it more difficult. Source: Borcan et al. (2017) "Fighting Corruption in Education: What Works and Who Benefits?".

But what was more interesting is that on average poorer students were worse off than richer students after the campaign. This raises a paradox because poor students shouldn't have access to bribery since they don't have the financial

[1] https://www.esquire.com/entertainment/a26800556/operation-varsity-blues-explainer/

resources to do so. To get more understanding on how the poor students were affected more negatively than their richer peers, we should investigate other student related variables simultaneously when estimating the heterogeneous treatment effect. The focus of this research is to answer the following question: Does the anticorruption campaign have additional heterogeneous treatment effects beside the poverty status of students? By investigating the heterogeneous impact of the anticorruption policy, more information on the consequences of the policy and whom are affected by these consequences are gained. With this information, the policy construction can be more polished to combat corruption effectively. Corruption hurts the economic and societal progress of a country so having more knowledge is of great importance to tackle corruption.

To estimate the heterogeneous effect, Borcan et al. (2017) decided to create interactions with the variable of interest with the treatment variable and other exogenous variables. Since Borcan et al. (2017) wanted to measure the heterogeneous effect on poor students, their variable of interest is the poverty status of students. After creating the interactions of poverty status with the treatment variable and exogenous variables, they estimated the heterogeneous effect with least square method in a difference-in-difference setting. The estimated treatment effect, the installment of cameras in examination centers, were higher for the interacted variables in comparison to the non-interacted variables which meant that the poorer students were more affected by the anticorruption campaign.

However, this approach of estimating the heterogeneous treatment effects comes with some problems. First problem is the amount of interaction variables needed when investigating significant heterogeneous effects with the difference-in-difference strategy. Since Borcan et al. (2017) only wanted to estimate the heterogeneous effect on poor, the amount of interaction variables required is not too much. But in this research I like to take all available variables to look for additional heterogeneous effects so the amount of variables I have to put in the equation is too much to handle. Another problem that comes with this is the multiple hypothesis testing problem because we simultaneously test for all interacted variables (Shaffer, 1995). Altogether, to avoid these problems I have to use a different method that both investigate heterogeneous effects and at the same time keep it computational efficient and have valid inference for the parameters. One method that fulfills these requirements is the generalized random forest (grf) from Athey, Tibshirani and Wager (2018). Different from the parameter estimation method of Borcan et al. (2017), this method uses a non-parametric way of estimation by creating numerous trees in a forest. For each tree, the grf uses a subsample of the dataset to calculate the treatment effects. To do so, the subsample which is taken is divided into two evenly parts with observations containing the features (X_i), treatment outcome (Y_i) and treatment assignment (W_i). Only one part is taken to calculate the treatment effects while the other part is used to measure the weight of the tree estimate by looking at the prediction accuracy. Our final estimate thus consist of the combination of the estimates of the individual tree and the weight the grf assigns to these trees. How the grf calculates the tree estimates and weight values will be more elaborated in the methodology section.

Since the focus of this paper is to find the heterogeneous treatment effects, the final estimate of the forest (which is the average treatment effect) is not of main interest. The focus lies more on the trees themselves because they contain information how the estimated treatment effect changes depending how the splits are done. Our splits are axis-aligned meaning the branches are made according to the different features (X_i). Moreover, the splits are made according to a splitting rule which goal is to maximize heterogeneity in the parameters. The branches are continued to branch with another X-aligned split till we arrive at the leafs. These leafs represent a subgroup of the dataset and the grf estimates the treatment effect on those subgroups. Additional to the gained treatment effects I'd to ensure our trees capture the heterogeneous effects, so I only take the most important variables for the forest creation. This means I first put all available feature variables (X_i) in the causal forest and calculate for each variable its importance. For each variable, its importance depends how many times the variable is used at the X-aligned splits and I only use the variables which importance is higher than the mean importance of all variables. The final causal forest is then made with this set of important variables. Lastly, I test the final causal forests for heterogeneity with help of the best linear predictor from Chernozhukov, Demirer, Duo, and Fernandez-Val (2018) to access whether the forest contains heterogeneity.

Using the grf method, I found that not only the poor indicator of students is important when accessing the treatment effect but also historical educational performance of students, the number of students 4 years before the Baccalaureate exams, the county's trust in justice and unemployment rate. This suggests that the method of Borcan et al. (2017) to investigate heterogeneous treatment effects didn't captured the underlying effects well. Although grf finds heterogeneous treatment effects on the poverty status of students too, this effect has less impact than the above mentioned variables. With the list of most important variables I construct a tree with average treatment effect (ATE) estimates and standard errors at the end of the branches to show the changing treatment effect for each axis-aligned splits. These trees give a general view that students with low ability were more impacted by the campaign. Moreover, the number of students taking the Baccalaureate test matters as well since a lower number of students requires less resources for the monitoring to be effective. Additionally the county's people trust in justice was found to have a significant impact on estimated treatment effect. A possible explanation would be that a high trust in justice is reflected as a non-corrupt county meaning there were not much corruption present. The effect of an anticorruption campaign would be smaller in counties with high justice trust opposed to counties with low trust in justice. Lastly is the unemployment variable which is county based. The monitoring effect was found to be higher if the unemployment rate were high hinting that during economic depression, employers would choose their new employees based by their final grades thus making cheating irresistible.

The remaining of this paper is as follows. Section 2 provides related literature on generalized random forest and anticorruption campaign. Section 3 presents the data and any transformations applied. Section 4 presents information of the methodologies used in this research. Section 5 presents both the replication results of Borcan et al. (2017) and the estimated heterogeneous treatment effects. Lastly, section 6 concludes.

2 Related literature

The paper of Borcan et al. (2017) is one of the many papers which contributes to the literature about fighting corruption efficiently and how corruption has its impact on society. By combining the threat of punishment and increased probability of detection, corruption can be reduced because it is made more costly (Becker and Stigler, 1974). Borcan et al. (2017) used this notion to investigate to what extend the combination between monitoring and threats helped to fight corruption in Romania compared to punishment threats alone. In Borcan et al. (2017), the data set consist of all counties that faced punishment threats from 2011 onwards but not all counties had circuit-closed TV installed in their examination rooms. This changed in 2012 when the installment of cameras was compulsory so the difference between 2011 and 2012 would be gained effect of combining threats and monitoring. Other methods to decrease corruption through monitoring are official audits (Ferraz and Finan 2008, 2011; Di Tella and Schargrodsky 2003). Ferraz and Finan (2008) examined the effect of audit monitoring in Brazil before the elections of 2004. When the audits exposed the corrupt doings of politicans, this affected their electoral performance. Another method of monitoring is community-based monitoring interventions (Duflo, Hanna, and Ryan 2012; Reinikka and Svensson 2004, 2005; Olken 2007). This type of monitoring asks the community to monitor in stead of the governments. The effectiveness of this kind of monitoring is varying. Reinikka and Svensson (2004) concluded in their research that active communities can decrease the amount of corruption involving public money grants for schools in Uganda whereas Olken (2007) found that grassroot participation had only little average impact on reducing missing expenditures of road projects in Indonesia. An interesting question is how corruption impacts the society and the changes in society due to the anticorruption policy. The inequality in one country is often positively correlated the size of corruption as argued by You and Khagarm (2005) but little is known about the consequences of reducing corruption. Borcan et al. (2017) tried to estimate the effects of an anticorruption for students from poor backgrounds with controls for students ability. The results were not favoring the poor students as the anticorruption campaign decreased their chances to enter tertiary education. In the period after the campaign, evidence in form of applications to a top Romanian university supported this statement. But it is still not clear why poorer students are the victims of this anticorruption campaign as one would say you need money to bribe the teachers. Since Borcan et al. (2017) estimated the heterogeneous effect for poverty but did not take all variables into account, the full heterogeneous effects of the campaign are not shown. My research is to investigate if there are more heterogeneous effects. In addition to Borcan et al. (2017) working sample, I also take variables that hasn't been used at estimations for treatment effect. To search for heterogeneous treatment effect, I use the generalized random forest method.

The generalized random forest method based on the random forest of Breiman (2001). Random forest is a non-parametric estimation method for heterogeneous treatment effect. Older non-parametric estimation methods for calculating heterogeneous treatment effect are nearest-neighborhood matching, kernel and series estimation (Crump et al., 2008; Lee, 2009; and Willke et al., 2012). The random forest of Breiman (2001) differs from these methods by how well it performs under large covariate dimension. Similar to nearest-neighborhood matching and kernel methods, random forest also makes predictions using a weighted average of nearby observations but here the weights are calculated in a data-driven way such that it avoids the curse of dimensionality. Although the performance of random forest at prediction and classification were successful, it was not directly useful for causal inference since the asymptotics of random forest were still unknown. In their work, Wager and Athey (2018) established the asymptotic theory for random forest to do statistical inference and created a new forest method: causal forest. Using this method would give the users predictions which are both asymptotically unbiased and normal distributed. The theory is useful for statistical inference since we don't know the ground truth so the theory plays a central role in evaluation of the noise in estimates of causal effects. Another work which contributes to the literature of random forest is generalized random forest (Athey et al., 2018). The main contribution of generalized random forest is to create a general framework for computationally efficient and problem specific splitting rule which priority is to capture the heterogeneity in the parameter of interest. Since the main focus on my research is to capture all possible heterogeneous effects while having a large covariate space, this method is appropriate to use in my quest to find heterogeneous effects.

3 Data

The data at my disposal for the anticorruption campaign from Borcan et al. (2017) consist of 44 variables for about 700,000 students. This is freely downloadable at the American Economic Association website.[2] A descriptive list of the variables and information can be found in the appendix A. For the replication, only the dependent variables Romanian final written exam score, pass or fail dummy variable and overall Baccalaureate score is important. Our treatment variable is the camera installment which is a dummy variable with value 1 if camera has been installed in the county c during year t. Other exogenous variables are students' properties (gender, school track, living in rural area), year dummies, county dummies and linear time trend interacted with counties. There are also data available with university admissions which I use to investigate how large the effect of the anticorruption campaign on poor students chance to enter university. In this data set the poor indicator is set as endogenous variable, camera installment as treatment variable and a set of student properties in addition to the year, county and linear time trend dummies as exogenous variables.

Table 1 summarizes the key statistics of the main variables used for replication. From the table we see that three endogenous variables are steeply decreasing after 2010 and stabilizing in 2012. This means that the introduction of the anticorruption measures after the Baccalaureate in 2010 left its mark on the student's educational results, both in short-term and long-term perspective.

Table 1: Statistics for full sample size

	2009		2010		2011		2012	
	Mean	Std. Dev	Mean	Std. Dev	Mean	Std. Dev	Mean	Std. Dev
Written Romanian score	6.813	1.819	7.017	1.664	6.147	2.102	6.143	2.138
Baccalaureate pass	0.813	0.390	0.692	0.462	0.482	0.500	0.482	0.500
Overall Baccalaureate score	8.057	1.150	6.969	1.647	6.033	1.998	6.049	2.142
Poor	0.166	0.372	0.175	0.380	0.185	0.388	0.201	0.401
Male	0.483	0.500	0.490	0.500	0.480	0.500	0.463	0.499
Theoretical track	0.447	0.497	0.434	0.496	0.447	0.497	0.469	0.499
Rural	0.057	0.232	0.065	0.246	0.067	0.250	0.059	0.236
Low ability	0.509	0.500	0.514	0.500	0.500	0.500	0.468	0.499
Observations	196,687		195,755		182,939		156,124	

Low ability variable is constructed as indicator variable. If students middle school GPA is higher than the median, his/her low ability dummy value would be zero. This variable is only available for a subset of around 70% of the total sample size.Source: Borcan et al. (2017) "Fighting Corruption in Education: What Works and Who Benefits?".

After the replication I like to capture the heterogeneous effect of the full sample size but regrettably there has to be some cleaning of data done beforehand. To avoid complications with the generalized random forest method,

[2]https://www.aeaweb.org/articles?id=10.1257/pol.20150074

observations which contain a NaN value in at least one variable will be removed. This leaves me with a workable sample size of about 100,000 students. In this sub-sample there are no observations from 2009 but this shouldn't be a problem since the anticorruption campaign effectively started at the beginning of 2011 Baccalaureate (for some counties) and 2012 nationwide. So we can still measure the impact of the anticorruption campaign by setting the 2010 observations as baseline.

Table 2 summarizes the statistics of the variables after cleaning. Similar to the full sample size, the written Romanian score, Baccalaureate pass and overall Baccalaureate score decreased after 2010. Proportionally, the share of poor students increased in the adjusted sample size but this might allows the grf method to capture the effects on poor students better. The other variables in the adjusted sample size are somewhat equal to the full sample.

Table 2: Statistics for adjusted sample size

	2010		2011		2012	
	Mean	Std. Dev	Mean	Std. Dev	Mean	Std. Dev
Written Romanian score	7.358	1.538	6.510	1.999	6.348	2.069
Baccalaureate pass	0.770	0.421	0.552	0.497	0.525	0.499
Overall Baccalaureate score	7.255	1.576	6.292	1.957	6.150	2.134
Poor	0.257	0.437	0.260	0.439	0.251	0.433
Male	0.460	0.498	0.458	0.498	0.435	0.496
Theoretical track	0.475	0.499	0.485	0.500	0.518	0.500
Rural	0.038	0.192	0.038	0.191	0.038	0.192
Low ability	0.501	0.500	0.495	0.500	0.458	0.498
Observations	36,593		34,177		29,834	

Different from table 1, low ability variable is available for all observations in this sub-sample.

4 Methodology

4.1 Difference-in-difference

First we look at the effect which the installment of the camera in examination centers has. Following the method of Borcan et al. (2017), I replicate the estimated effects of the anticorruption campaign by means of the DID strategy. The model specification is as follows:

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

To replicate the main results of Borcan et al. (2017), y_{ict} can be any of the three dependent variables mentioned in the data section. The subscript i, c and t stands for student number, county and time respectively. β measures the effect of the installed cameras ($T_{ct} = 1$, if camera was installed). γ measures the student characteristics X_{ict} on his/her dependent variable. For the remaining of this paper, the student characteristics is denoted as 'controls'. The variable φ_t and θ_c are dummy variables for time and county respectively and the last dummy variable ($\theta_c \cdot t$) represent the linear time trend interacted with the counties. ε_{ict} denotes the residuals. The parameters α, β and γ are estimated with OLS estimation.

Secondly, the heterogeneous effect of this campaign by poverty and ability is investigated (table 4) but requires additional dummy variables to our model specification (1) to let the poor and ability indicator interact with other variables. This leads to the following equation:

$$y_{ict} = \alpha + \beta T_{ct} + \gamma' X_{ict} + \varphi_t + \theta_c + \theta_c \cdot t + \lambda_1' Z_{ict} x_{1,i} + \lambda_2' Z_{ict} x_{2,i} + \varepsilon_{ict} \tag{2}$$

where the added $x_{1,i}$ is an indicator variable for poor for student i and $x_{2,i}$ an indicator function for low ability. Z_{ict} contains column vectors with treatment, time, student and control dummy variables.

Lastly, the consequences of the anti-corruption campaign on university admissions for students from a poor background. Since the anticorruption campaign affected the students' score, I like to access whether this has effect on the university admissions. Using admission data from an elite university and combining it with the original data, this results in the following equation:

$$p_{ict} = \alpha + \beta T_{ct} + \gamma' X_{ict} + \varphi_t + \theta_c + \theta_c \cdot t + \varepsilon_{ict} \tag{3}$$

where p_{ict} is the poor indicator for the corresponding student i, X_{ict} the matrix with controls containing column vectors for gender, track and dummy indicator if student has taken the Baccalaureate exam at least 1 year before his/her admission to university.

4.2 Generalized random forest

Second part of my research is to investigate whether there are more heterogeneous effects present. For this examination I use the random forest method proposed by Athey et al. (2018). The core mechanisms of this methodology are similar to Breimans (2001).

The procedure of generalized random forest is as follows: First the dataset is divided in a set of samples $i = 1, ..., n$ which are independent and identically distributed. These samples contain information of the treatment outcome Y_i, the treatment assignment W_i and some auxiliary covariates X_i. The ultimate goal of grf is to estimate $\theta(x)$ which is characterized by a local estimation equation $E[\psi_{\theta(x),\nu(x)}(O_i)|X_i = x] = 0$ where O_i is the set of observables $\{Y_i, W_i\}$, $\theta(x)$ our parameter of interest and $\nu(x)$ a nuisance parameter. To estimate the parameter of interest in a non-parametric way, Athey et al. (2018) relied on local solutions which is represented as:

$$\sum_{i=1}^{n} \alpha(x; X_i)\psi_{\hat{\theta}(x),\hat{\nu}(x)}(O_i) = 0 \tag{4}$$

where the $\alpha(x, X_i)$ are the weights for the local estimation. Historically, the weights are obtained by kernel method but this method is prone to curse of dimensionality. To counter this problem, Athey et al. (2018) chose a method which let the weight calculation be data-adaptive. This leads to an idea averaging tree-based neighborhoods based on Meinhausen (2006). The tree weights are calculated as follows:

Let there be a number of B trees indexed as $b = 1, ..., B$ and for each tree define $L_b(x)$ as the set of samples falling in the same leaf as x. Then $\alpha_i(x)$ can be calculated as:

$$\alpha_{bi}(x) = \frac{\mathbf{1}(\{X_i \in L_b(x)\})}{|L_b(x)|}, \qquad \alpha_i(x) = \frac{1}{B}\sum_{b=1}^{B} \alpha_{bi}(x). \tag{5}$$

In short, generalized random forest calculates $\theta(x)$ by solving (4) with weights from (5) but the real challenges of grf are constructing trees such that the $\theta(x)$ estimates contain maximum heterogeneity and making the algorithm computationally efficient.

4.2.1 Maximum heterogeneity by splitting rule

When we split the parent node into two children nodes, we like to make an axis-aligned cut such that it improves the accuracy of the $\theta(x)$ estimate. Following Breiman et al. (1984) CART method, grf make use of greedy splits too. To maximize heterogeneity in the parameters we maximize the following equation during each split:

$$\Delta(C_1, C_2) = n_{C_1} n_{C_2} \left(\hat{\theta}_{C_1} - \hat{\theta}_{C_2}\right)^2 \tag{6}$$

where n_{C_1} and n_{C_2} are the size of the split samples in children node 1 and 2 and $\hat{\theta}_{C_1}$ and $\hat{\theta}_{C_2}$ are the solutions to estimation equation. For each node. the estimation equation is:

$$\left(\hat{\theta}_P, \hat{\nu}_P\right)(J) \in argmin_{\theta,\nu} \left\{ \left\| \sum_{\{i \in J: X_i \in P\}} \psi_{\theta,\nu}(O_i) \right\|_2 \right\}. \tag{7}$$

where $\left(\hat{\theta}_P, \hat{\nu}_P\right)(J)$ stands for the estimated parameter of interest and nuisance parameter with (sub)-sample J.

4.2.2 Using gradient-based approximations for better computational performance

Although the splitting rule results in maximum heterogeneity for the parameter θ, it might be computational expensive to get solutions for $\hat{\theta}_{C_1}$ and $\hat{\theta}_{C_2}$. This is because the CART method checks for every possible split and for each split the algorithm has to calculate $\hat{\theta}_{C_1}$ and $\hat{\theta}_{C_2}$ again. To solve this issue, grf makes use of gradient-based approximations which is similar to gradient boosting method from Friedman (2001). The idea is to get an approximate solution for $\hat{\theta}_C$ with help of the gradient of the ψ-function. The formula for the approximated $\tilde{\theta}_C$ is as follows:

$$\tilde{\theta}_C = \hat{\theta}_P - \frac{1}{|\{i : X_i \in C\}|} \sum_{\{i:X_i \in C\}} \xi^T A_P^{-1} \psi_{\hat{\theta}_P, \hat{\nu}_P}(O_i) \tag{8}$$

where ξ represent a vector that picks the θ-coordinate from (θ, ν) and given that the ψ-function is continuously differentiable the derivative matrix A_P is defined as:

$$A_P = \frac{1}{|\{i : X_i \in P\}|} \sum_{\{i:X_i \in P\}} \nabla \psi_{\hat{\theta}_P, \hat{\nu}_P}(O_i) \tag{9}$$

There are cases where the ψ-function is non-differentiable like in quantile regression but since we use the causal forest as estimation method, this differentiable problem shouldn't apply.

4.2.3 grf algorithm

We combine the above equations to get an algorithm for grf. Starting from the beginning we split our dataset in n i.i.d training samples. After we choose how many trees the causal forest should make, the function draws a subsample of size s from the available training samples without replacement. We continue by splitting this subsample in two evenly-sized halves without overlaps of the observations. The first half of this subsample is taken for the tree construction while the latter half is put on hold to determine the weight of this tree at the end.

For the tree construction, it follows the splitting rule of maximizing equation (6). Since this rule has no computational constraint, it might be prohibitive expensive to follow this rule hence the use of gradient-based approximations. At the start of the tree (the parent node), we calculate A_P^{-1} and the parameters $\hat{\theta}_P$ and $\hat{\nu}_P$ from equations (7) and (9) respectively to get our pseudo-outcomes $\rho_i = -\xi^T A_P^{-1} \psi_{\hat{\theta}_P, \hat{\nu}_P}(O_i) \in \mathbb{R}$. With this outcome we can make our CART regression split by maximizing the approximated criterion:

$$\tilde{\Delta}(C_1, C_2) = \sum_{j=1}^{2} \frac{1}{|\{i : X_i \in C_j\}|} \left(\sum_{\{i:X_i \in C_j\}} \rho_i \right)^2 \tag{10}$$

which then result in two new children nodes. In these children nodes we follow the same steps again as we were in the parent node. After some iterations we end up with a tree and we use this to compute our weights with (5). When this is finished the algorithm continue with constructing the next tree till the number of created trees reaches the desired number.

4.2.4 Properties of forest

We like to know whether the forest captured the heterogeneity well enough so I run a test calibration on the created forests. Before doing this, I only take the important variables when creating the forest. For a variable to be marked as important, the number of when the variable is chosen to be split during the CART regression must be above the average number of splits of all variables. After I created the forest with the important variable, I run the best linear predictor on it. This method from Chernozhukov et al. (2017) computes the best linear fit of the parameter θ using forest predictions and mean forest predictors as regressors. If the estimate for mean.forest.prediction is close to 1 and significant different from zero, then the mean forest predictions are correct but my focus goes to the estimates for differential.forest.prediction. If this estimate is significant different from zero, then we reject the null hypothesis of no heterogeneity.

Lastly, I like to know the heterogeneous effects for changing covariates. Because there are too many trees to choose from and there are no certainty for choosing the best tree, I randomly choose one tree of the forest to show the changing treatment effects for the leafs. This is done by solving equation (7) at the end nodes.

5 Results

5.1 Findings of Borcan et al. (2017)

In the first part of the result section, I summarize the findings of Borcan et al. (2017) in their research on the impact of the anticorruption campaign in education. Since these results are replicated I won't give too much detailed information and only highlight the main findings. Table 3 shows the results of the least squares estimation of the difference-in-difference strategy. As seen in the table the treatment of monitoring, denoted as "Camera", decreased the students Romanian written score (column 2,3 and 4), passing rate (column 6,7 and 8) and the overall Baccalaureate score (column 10,11 and 12). Because the focus lies on the heterogeneous effect, for simplicity I continue the summary with the results containing all dummy variables and controls which are column 4, 8 and 12. The effect of monitoring was an average decrease in scores for students by 0.353 and 0.512 for the Romanian written exam and Baccalaureate exam respectively. It also lowered their Baccaulaureate pass rate by 0.095%. With the t-test we reject the hypothesis of the monitoring effect being zero on 5% significance level so we can infer that anticorruption campaign did have an impact on the corrupt education in Romania.

Table 3: Main results of the effect of anticorruption campaign in Romania

	Romanian written score				Baccalaureate pass				Overall Baccalaureate score			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Camera		-0.246	-0.251	-0.353		-0.076	-0.076	-0.095		-0.430	-0.439	-0.512
		(0.108)	(0.106)	(0.106)		(0.030)	(0.029)	(0.025)		(0.144)	(0.142)	(0.137)
Year '12	-0.874	-0.628	-0.716	-0.463	-0.211	-0.135	-0.148	-0.082	-0.923	-0.492	-0.579	-0.323
	(0.065)	(0.087)	(0.078)	(0.087)	(0.024)	(0.025)	(0.023)	(0.017)	(0.092)	(0.115)	(0.106)	(0.094)
Year '11	-0.875	-0.713	-0.743	-0.597	-0.211	-0.161	-0.166	-0.129	-0.943	-0.660	-0.690	-0.547
	(0.058)	(0.070)	(0.071)	(0.081)	(0.022)	(0.019)	(0.018)	(0.016)	(0.088)	(0.091)	(0.090)	(0.088)
Year '09	-0.205	-0.205	-0.237	-0.311	0.121	0.121	0.115	0.093	1.087	1.086	1.055	0.967
	(0.054)	(0.054)	(0.057)	(0.033)	(0.011)	(0.011)	(0.011)	(0.012)	(0.042)	(0.042)	(0.040)	(0.040
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 x Yearly trends	No	No	No	Yes	No	No	No	Yes	No	No	No	Yes
Observations	712,298	712,298	712,298	712,298	731,505	731,505	731,505	731,505	706,895	706,895	706,895	706,895
R^2	0.060	0.060	0.275	0.289	0.102	0.103	0.239	0.253	0.204	0.206	0.417	0.432

Least squares estimation results of the difference-in-difference strategy where y_{ict} is either Romanian written score (columns 1 to 4), Baccalaureate pass (columns 5 to 8) or overall Baccalaureate score (columns 9 to 12). The treatment variable is camera and the control variables include student's poverty status, gender and if they follow the theoretical track. All standard errors are county level clustered.
All estimates are at least significant on 5% level.
Source: Borcan et al. (2017) "Fighting Corruption in Education: What Works and Who Benefits?".

Next is the summary on heterogeneous effect on this campaign on poor students. If we take the ability interactions into account then from table 4 column 1,4 and 7 it occurs that the effect of camera installation in exam centers did not favor the poor students compared to the richer students. An average score decrease of 0.515 for the written Romanian exam, 0.779 score decrease for the Baccalaureate score and a lowered pass rate of 0.143. When we also control for ability (column 3,6 and 9), then the decreases are lower but still significant. However, this result only took a few exogenous variables into account so it might be possible that there are underlying effects present.

Table 4: Heterogeneous effect of campaign on poverty with ability and poverty interaction variables

	Written Romanian score			Baccalaureate pass			Overall Baccalaureate score		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Camera	-0.302	-0.255	-0.131	-0.081	-0.088	-0.077	-0.433	-0.434	-0.308
	(0.110)	(0.082)	(0.056)	(0.026)	(0.023)	(0.017)	(0.141)	(0.113)	(0.076)
Poor x camera	-0.213	-0.257	-0.214	-0.062	-0.056	-0.051	-0.346	-0.350	-0.306
	(0.063)	(0.053)	(0.052)	(0.015)	(0.012)	(0.013)	(0.078)	(0.061)	(0.063)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County FE x 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	712,298	547,447	547,447	731,505	553,903	553,903	706,895	545,121	545,121
R^2	0.291	0.356	0.459	0.256	0.310	0.394	0.435	0.504	0.613

Least squares estimation results of heterogeneous effects on poor. Columns (1), (4) and (7) uses the full sample size of 2009-2012 and the remaining columns uses a sub sample where the average middle school score of students are available. The control variables include students poor status, gender, if they follow the theoretical track and if they live in rural area. The interactions are the variables poor and ability multiplied with control variables, year, county and linear time trend dummy variables.
All estimates are at least significant on 5% level.
Source: Borcan et al. (2017) "Fighting Corruption in Education: What Works and Who Benefits?".

Table 5: Effect of anti-corruption campaign on admission into an elite university using students ranked at the top 20 percent

	Poor admitted to an elite university		Poor in an elite university tuition-exempt (top students)		Poor in an elite university tuition-paying (good students)		Poor in top 20 percent at Baccalaureate	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Camera	-0.025	-0.025	-0.025	-0.026	-0.025	-0.026	-0.022	-0.019
	(0.013)	(0.013)	(0.015)	(0.015)	(0.031)	(0.030)	(0.006)	(0.006)
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 x 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

Dependent variable is students poor status. Data set only includes observation where students admissions to university is available. Control variables for the first six columns are students gender, if they follow the theoretical track and if they took the Baccalaureate one year prior. Control variables for the last 2 columns are students gender, track and if they live in a rural area.
Source: Borcan et al. (2017) "Fighting Corruption in Education: What Works and Who Benefits?".

The last finding of Borcan et al. (2017) is the changing university admission composition due to the anticorruption campaign (table 5). In this table we see a significant drop of poor students admitting to an elite university in Romania. When we include the controls, we see from column 2 that the share of poor students decreased by about 2.5% in 1 year after the nationwide anticorruption campaign. This university uses a rule that exempt tuition fees for admitted students who are ranked in the top 55 to 65 of the overall score. The composition of (non) tuition-exempt students follows the same direction as poor students admitted to the elite university: a drop of 2.6% poor students for both cases albeit in the tuition paying case, the estimates are insignificant. If we look at the last column, we see that the percentage of poor students that are in the top 20% percentile of the Baccalaureate score dropped also. All in all we can conclude that the impact of the anticorruption campaign is larger for poor students than rich students. This applies to both test scores and passing rate of the Baccalaureate and future application to university. Next part of this paper is to see whether there are other attributes of students which the campaign has a high effect on beside the poor indicator variable.

5.2 Variable importancy of the average treatment effect

The last part of the research is devoted to find additional heterogeneous treatment effects. Here I take all 44 variables which are described in Appendix A and put them in the causal forest function of the grf. After creating the forest I calculate for each variable its importancy. This gives us the following result (Table 6) for the three endogenous variables:

Table 6: Top 10 important variables of causal forest

Rank	Romanian written exam	Baccalaureate pass	Overall Baccalaureate score
1	pcrank_wrrom - 23.3%	pcrank_wrrom - 10.5%	pcrank_wrrom - 9.6%
2	midschool_gpa - 11.4%	midschool_gpa - 9.1%	cty_hs_enrolled - 8.3%
3	ability_std - 10.8%	ability_std - 8.6%	cty_midschool_grad - 8.2%
4	cty_hs_enrolled - 5.7%	trust_justice - 7.1%	county_shromania - 7.2
5	cty_midschool_grad - 5.4%	cty_midschool_grad - 7.0%	unempl_apr - 7.1%
6	trust_justice - 4.0%	cty_hs_enrolled - 6.3%	midschool_gpa - 7.1%
7	unempl10 - 4.0%	unempl_apr - 6.2%	trust_justice - 7.0%
8	orrom - 3.9%	county_shromania - 5.8%	ability_std - 6.7%
9	pcrank_oral - 3.9%	unempl10 - 5.4%	unempl10 - 5.9%
10	urban_population - 3.7%	county_pop - 4.9%	county_pop - 5.6%

Refer to Appendix A for the explanation of the abbreviations

On first sight we see that the exams score and pass rate is highly dependent on school performance and capabilities of students like percentile ranking of the written Romanian exam (*pcrank_wrrom*), average grade during middle school period (*midschool_gpa*) and standardized ability score (*ability_std*). If we continue with the rankings of important variable we notice that the number of students enrolled in high school (*cty_hs_enrolled*) and number of middle school graduates (*cty_midschool_grad*) is also of importance for its grade or pass during the campaign years and especially for the overall Baccalaureate score (8.3% and 8.2% for high school and middle school respectively). It might be because some county's high schools has lower amount of students thus making monitoring more effective as it only have to focus on a small number of students. Another variable of interest is the indicator for trust in justice in its own county (*trust_justice*). Maybe if the trust is low, students are tempted to cheat during the exams because they know they won't be caught or consequences won't be that harsh. The last variable of interest is the unemployment rate. When there is scarcity of jobs in the country, employees choose the best candidate for the job. For beginners without job experience the employees might choose based on best grades which pushes the students into cheating because they need to get the highest grades.

Next we use the causal forest again but only take the variables which are important. For a variable to classify as important, its contribution percentage has to be above the average contribution of all variables. A ranking of variable importance of this causal forest is given in table 7. Before we access the results we like to know whether the causal forest captured the heterogeneity correctly. Therefore, I ran the test calibration function on the three forest (each endogenous variable) and test whether the estimate thereof is close to 1 and significant. The results of the best linear predictor are given in Appendix B, figure 5. For the first forest (Romanian written exams) the estimate is not close to 1 and insignificant too so the null of no heterogeneity is not rejected. On the other hand we have the last two forests (Baccalaureate pass and overall Baccalaureate score) which rejects the null of no heterogeneity on 5% significant level for the Baccalaureate pass and 10% significant level for the overall Baccalaureate score. In this case we can say the heterogeneity is present in the causal forest of the last two endogenous variables.

Table 7: Top 10 important variables of causal forest after cleaning

Rank	Romanian written exam	Baccalaureate pass	Overall Baccalaureate score
1	pcrank_wrrom - 42.0%	pcrank_wrrom - 20.4%	pcrank_wrrom - 25.7%
2	midschool_gpa - 16.6%	ability_std - 14.3%	midschool_gpa - 12.8%
3	ability_std - 10.2%	midschool_gpa - 12.8%	ability_std - 10.0%
4	rep_bac - 3.8%	trust_justice - 7.0%	trust_justice - 6.5%
5	cty_midschool_grad - 3.3%	county_shromanian - 5.8%	county_shromanian - 6.4%
6	cty_hs_enrolled - 3.2%	unempl10 - 4.8%	cty_midschool_grad - 5.9%
7	trust_justice - 3.0%	cty_midschool_grad - 4.8%	unempl_apr - 5.8%
8	unempl10 - 2.6%	unempl_apr - 4.5%	cty_hs_enrolled - 4.7%
9	orrom - 2.6%	top20pcent_rom - 4.2%	unempl10 - 4.3%
10	county_shromanian - 2.5%	urban_population - 2.9%	rep_bac - 2.7%

Refer to Appendix A for the explaination of the abbreviations

The results of table 7 are quite similar to the previous table. What is interesting compared to the heterogeneous effect findings of Borcan et al. (2017) is that the poor indicator variable didn't pass the threshold to be important for the Romanian written exam. And although it did pass for the Baccalaureate pass and overall score, it only ranked 12 out of 22 and 17 out of 21 for the pass and overall score respectively. So the difference in treatment effect for the poor and rich students could be affected by variables more important than the poor indicator only. We continue with the research by constructing a tree such that it shows the treatment effect after each axis-aligned branching.

5.3 Heterogeneous treatment effect

By visualizing one tree of the forest, we can see how the treatment effect differs for each branching. Since there are many trees constructed during causal forest, we ought to choose the tree which contains the most information of the forest. Unfortunately this algorithm has not been implemented yet in the first release of the grf package so I arbitrary choose one tree from the different forests.

Beginning with figure 2, we see that some subgroups of the data exhibits positive treatment effects. Most of them are insignificant on 5% level except for the large positive estimate at the second branching. This treatment effect of 0.564 (p = 0.0070) is for the subgroup whom trust justice is lower than 1.66 and whom percentile rank is lower than 0.08 at the written Romanian exam. Explanation for it is still left in the dark since the anticorruption campaign should have affected the less intelligent students negatively as they are the ones profiting from educational corruption.

Aside from this outlier the rest of the tree estimations shows quite nicely who the campaign affects the most. If the educational abilities are low, then on average the monitoring (and thus the campaign) effect decreased the students' Romanian written exam score. A few examples where this is visualized are the different ATE estimates at the 6th ($ability\_std <= -0.09$) and 7th level ($pcrank\_wrrom <= 0.24$). Although these estimates are not significant it still gives us the general overview how the monitoring affected the student's grade on average.

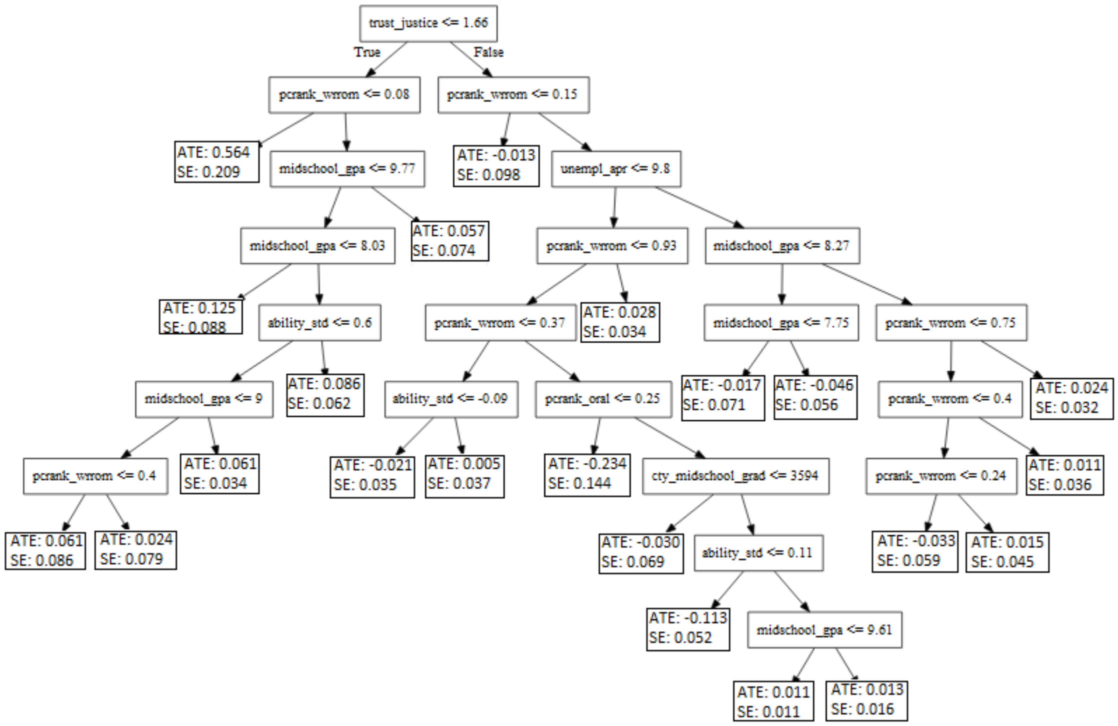


Figure 2: Visualization tree of causal forest (Romanian written exam).
Left arrow denotes true while right arrow denotes statement is false

Next I like to know how the campaign affected the students in poverty. Since the first causal forest failed to capture the poor variable in the forest (because of low importancy), we take a look at the last two causal forests. Figure 3 shows one tree of the causal forest for Baccalaureate pass. Similar to the first causal forest for Written Romanian exam scores, if the students' ability is lower then the (negative) effect is higher and this is shown by different ATE estimates at the 6th (*pcrank_wrrom* <= 0.6) and 8th level (*ability_std* <= 0.79 and *pcrank_wrrom* <= 0.8) of the tree. This tree also captures the effect of the students' poverty status given a few attributes. In the 5th level of the tree (right side) we split the sample size with on one hand the poor students and the other hand the non-poor students. For the poor students, on average the pass rate decreased by 47.1% (p = 0.0701) but what's more interesting is that this decrease is much lower than richer students with lower educational ability (*pcrank_wrrom* <= 0.6, 6th level). The decrease for richer students with same attributes as the poor is 61.5% (p = 0.0003) if their percentile ranking of the written Romanian exam is lower than 0.60 while richer students with higher percentile ranking only has an average decrease of 10.6% (p = 0.180). This means that although the students with a poor background have a lower pass rate for the Baccalaureate after the campaign, the effect is more dependent on the educational capabilities of the students than the poverty status.

Another interesting finding is the different in ATE estimates for the amount of middle school graduates (*cty_mid school_grad* <= 3698, 5th level) and trust justice (*trust_justice* <= 1.81, 8th level). Starting with the middle school graduates, we see that if the county's middle school graduates are higher than 3698, then the ATE effect is 0.043 (p = 0.0318) against 0.031 (p = 0.140) when the amount of middle school graduates is less than 3698. Notice that in this branch, our sample consist of students with low scores of the Romanian written exam scores (*pcrank_wrrom* <= 0.13, 4th level) which are most subject to cheating during the exam. However, if the amount the graduates is small then monitoring them shouldn't be a problem since it might take less resources to get a complete overview. Hence the reason why the effect of camera treatment will lower the pass rate of the county with less graduates than county with higher amount of graduates. The question of why the estimate is positive is interesting on its own and might require additional future research.

The last interesting finding is the county's trust justice score. From the ATE results of the 8th level, counties with higher trust in justice sees their students' average Baccalaureate pass rate decreasing more than lower trust justice counties (decrease of 40.7% against increase of 7.8%). Explanation for this difference is probably the effectiveness of the anticorruption policy. Counties with higher trust justice will enforce the policy more than counties with lower trust justice as their inspectors are more willing combat corruption. The effect of lower trust justice in this tree is insignificant so I cannot infer that if a county has low trust justice then the campaign has an opposite effect but it is clear that a higher trust justice results in a larger decrease of the Baccalaureate pass rate.

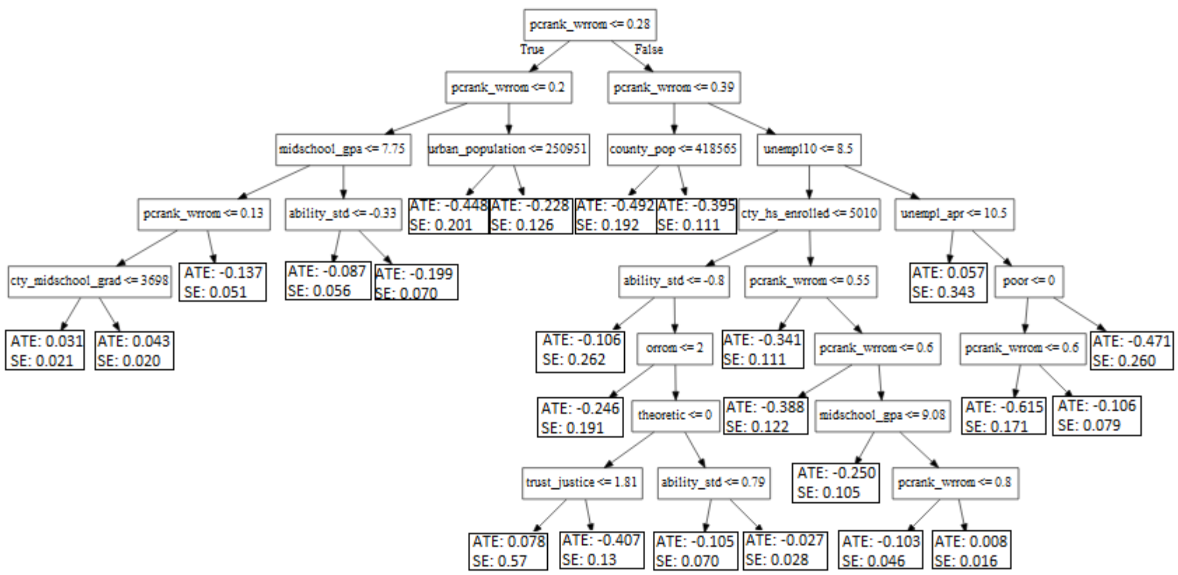


Figure 3: Visualization tree of causal forest (Baccalaureate pass).
Left arrow denotes true while right arrow denotes statement is false

In the tree of the last causal forest (figure 4), I highlight the change in ATE estimate for the unemployment rate (*unempl_apr* <= 8.2, 4th level). If the unemployment rate is lower than 8.2%, then the average treatment effect would be -1.247 (p = 0.0001) against -1.686 (p = 0.0001) if there were a higher unemployment rate than 8.2%. This clearly shows that the unemployment rate in one county also mattered when calculating the impact of the anticorruption campaign. The monitoring effect in counties with higher unemployment rate is larger since there were many corruption ongoing. The exact reason is still to be found but it might be because if there are low amount of job vacancies then companies choose their trainees by their academical performance. Getting a higher grade on the Baccalaureate exam by cheating or fraud was common on those counties because a high grade was necessary for the job market. So after the introduction of the anticorruption campaign, this resulted in a sharp drop of the overall Baccalaureate score of the students.

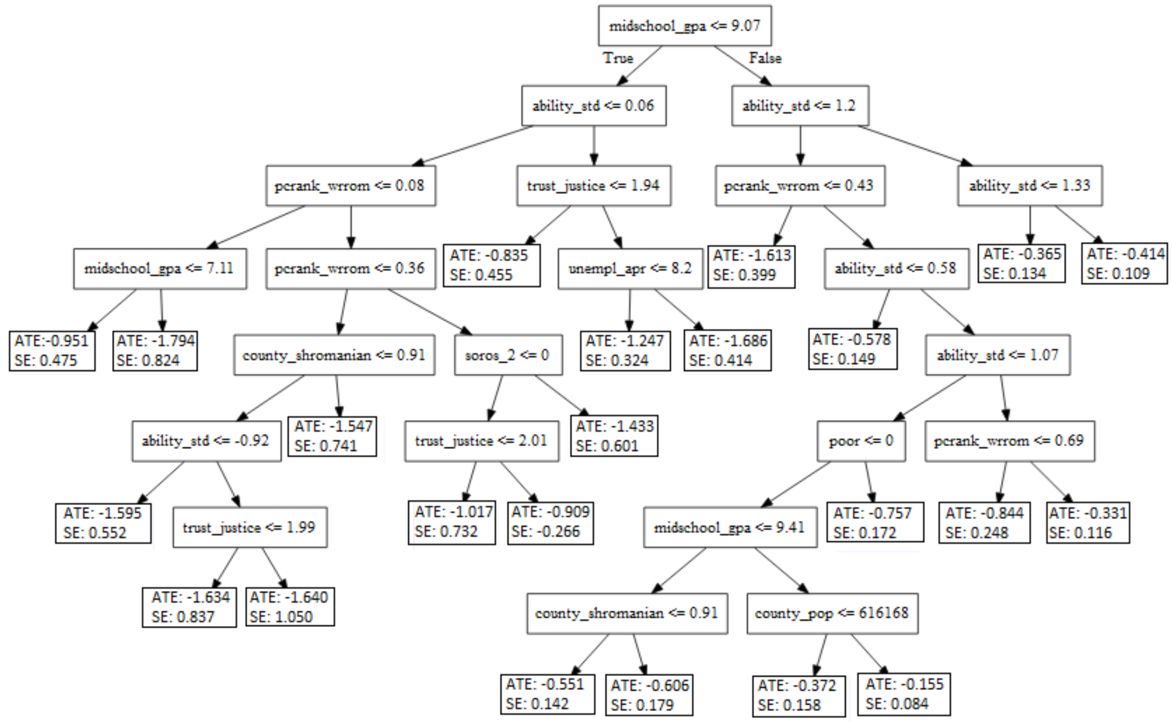


Figure 4: Visualization tree of causal forest (overall Baccalaureate score).
Left arrow denotes true while right arrow denotes statement is false

Although some treatment estimates are not significant, it still shows the general pattern which variables are important for estimated heterogeneous treatment effect. By the look of the three causal forests for different endogenous variable, we see that the ability of the student, the county's amount of middle school graduates and high school enrollments 4 years before the Baccalaureate test, average county's trust in justice and unemployment rate matters.

6 Conclusion

When applying generalized random forest on the dataset of Borcan et al. (2017), we found evidence for additional heterogeneous effects besides the poor indicator for students. Although the causal forest for Romanian written exam did not capture the heterogeneous well enough according to the best linear predictor estimate, I could reject the null of no heterogeneity for the other causal forests. The causal forests for Baccalaureate pass and overall Baccalaureate score did find significant presence of heterogeneous treatment effects and they were more important than the poor indicator variable thus hints that the heterogeneous treatment effect findings of Borcan et al. (2017) could be overshadowed by other variables.

Borcan et al. (2017) found significant heterogeneous treatment effect on poor students which lowered their score and pass rates. This result suggests that the anticorruption campaign targeted the wrong subgroup as the poor students to do have many monetary resources to bribe teachers or examinators to begin with. However, during the variable importancy measurement, the poor indicator variable was not ranked high which allude that the result of Borcan et al. (2017) did not take the underlying effect into account. By using the causal forest, I took all variables and this allowed me to find more heterogeneous effects than poor alone. Across the three endogenous variables, the most interesting variables are the students' ability and performance, the county's amount of high school attendees and middle school graduates, the county's trust in justice and the unemployment rate.

To establish higher credibility for the results, few problems might need to be solved beforehand. First is the extraction of data without any NA values for causal forest. Because the observations of year 2009 weren't taken into account and a huge decrease in the number of observations, comparison of the heterogeneous treatment effect of generalized random forest and average treatment effect of difference-in-difference method might be less valid. I could have omitted a few variables such that the sub sample size increased and contained observations from 2009 but generalized random forest performs better when there are many covariates so I did not consider this option. Another small problem is that there are no asymptotic results available yet that justify the inference of the best linear predictor method (Athey and Wager, 2019) and lastly I chose one tree of the causal forests at random which might not contain all essential information of the forest.

Suggestions for future work are, but not limited to, creating dummy variables for high schools to see if there are some schools with high moral values and teach their students not to cheat. As the amount of schools present in the data were large, time-constraint withheld me from doing so. Another suggestion is to group the counties (and possibly schools) by its geographical location since the south of Romania is arguably more corrupt than the north which might result in different monitoring effect (Duţulescu and Nişulescu-Ashrafzadeh, 2016).

Reference

Athey, S., and S. Wager (2019). "Estimating Treatment Effect with Causal Forest: An Application", *Observational Studies*, Upcoming.

Athey, S., Tibshirani, J., and S. Wager (2018). "Generalized Random Forest." *The Annals of Statistics*, 47 (2), 1148-1178.

Becker, G. S., and G. J. Stigler (1974). "Law Enforcement, Malfeasance, and Compensation of Enforces." *Journal of Legal Studies*, 3 (1), 1-18.

Borcan, O., Lindahl, M. and A. Mitrut (2017). "Fighting Corruption in Education: What Works and Who Benefits?" *American Economic Journal: Economic Policy*, 9 (1): 180-209.

Breiman, L. (2001). "Random Forest." *Machine Learning*, 45, 5-32.

Breiman, L., Friedman, J., Stone, C. J., and R. A. Olshen (1984). *Classification and Regression Trees*, CRC press.

Chernozhukov, V., Demirer, M., Duo, E., and I. Fernandez-Val (2018). "Generic machine learning inference on heterogenous treatment effect in randomized experiments.", Technical report, *National Bureau of Economic Research*.

Crump, R. K., Hotz, V. J., Imbens, G. W. and O. A. Mitnik (2008). "Non-parametric Tests for Treatment Effect Heterogeneity.", *The Review of Economics and Statistics*, 90, 389-405.

Di Tella, R., and E. Schargrodsky (2003). "The Role of Wage and Auditing during a Crackdown on Corruption on the City of Buenos Aires." *Journal of Law and Economics*, 46 (1), 269-292.

Duflo, E., Hanna, R., and S. P. Ryan (2012). "Incentives Work: Getting Teachers to Come to School." *American Economic Review*, 102 (4), 1241-1278.

Duţulescu, S. and I. Nişulescu-Ashrafzadeh (2016). "The main causes of corruption in Romania.", *Audit Financiar*, 14, 918-926.

Ferraz, C., and F. Finan (2008). "Exposing Corrupt Politicians: The Effects of Brazil's Publicly Released Audits on Electoral Outcomes." *Quaterly Journal of Economics*, 123 (2), 703-745.

Ferraz, C., and F. Finan (2011). "Electoral Accountability and Corruption: Evidence from the Audits of Local Governments." *American Economic Review*, 101 (4), 1274-1311.

Friedman, J. H. (2001). "Greedy function approximation: A gradient boosting machine.", *Ann. Statist.*, 1189-1232.

Lee, M. -J. (2009). "Nonparametric Test for Distributional Treatment Effect for Randomly Censored Responses.", *Journal of the Royal Statistical Society, Series B*, 71, 243-264.

Meinhausen, N. (2006). "Quatile Regression Forest." *JMLR*, 7 ,983-999.

Olken, B. A. (2007). "Monitoring Corruption: Evidence from a Field Experiment in Indonesia." *Journal of Political Economy*, 115:2, 200-249.

Reinikka, R., and J. Svensson (2004). ”Local Capture: Evidence from a Central Government Transfer Program in Uganda.” *Quaterly Journal of Economics*, 119 (2), 679-705.

Reinikka, R., and J. Svensson (2005). ”Fighting Corruption to Improve Schooling: Evidence from a Newspaper Campaign in Uganda.” *Journal of the European Economic Association* 3 (2), 259-267.

Shaffer, J. P. (1995). ”Multiple Hypothesis Testing”, *Annual Review of Psychology*, 46, 561-584.

Wager, S. and S. Athey (2018), ”Estimation and Inference of Heterogeneous Treatment Effects using Random Forests.”, *Journal of the American Statistical Association*, 133:523, 1228-1242.

Willke, R. J., Zheng, Z., Subedi, P., Althin, R. and C. D. Mullins (2012). ”From Concepts, Theory, and Evidence of Heterogeneity of Treatment Effects to Methodological Approaches: A Primer.”, *BMC Medical Reseach Methodology*, 12, 185

You, J.-S., and S. Khagram (2005). ”A Comparative Study of Inequality and Corruption.” *American Sociological Review*, 70 (1), 136-157.

Appendix A, descriptive list variables of Borcan et al. (2017)

Variable	Additional information
student_id	ID of student
school_id	ID of high school which the student attends
county_id	ID of county where the students attend school
fam_id	ID family of student
male	Dummy variable = 1, if student is male
orrom	Romanian oral exam score
finwrrom	Romanian written exam score
elim	Dummy variable = 1, if student is eliminated from exam due to cheating
abs	Dummy variable = 1, if student is absent from exam
pass	Dummy variable = 1, if student passed exam
year	Year of observation
rur	Dummy variable = 1, if student lives in rural municipality
yr11	Dummy variable = 1, if year = 2011
yr10	Dummy variable = 1, if year = 2010
yr09	Dummy variable = 1, if year = 2009
poor	Dummy variable = 1, if student is egliable for MHS beneficiary
midschool_gpa	Proxy for ability of student (its average score in 5-8th grade)
adm_year	Students year of admission to highschool
cam	Dummy variable = 1, if county is part of the counties which installed camera early
new_cam	Dummy variable = 1, if county has camera installed (in 2011 or 2012)
yr12	Dummy variable = 1, if year is 2012
t	linear time trend
theoretic	Dummy variable = 1, if student follows theoretical track in highschool
soros_2	County average level of perceived corruption from the Public Opinion Barometer (Soros Fundation, 2007)
rep_bac	Number of Baccalaureate retakes of student
first_bac	Year of first Baccalaureate exam taken
pcrank_oral	Percentile rank Romanian oral exam
pcrank_wrrom	Percentile rank Romanian written exam
placebo_cam	Dummy variable = 1 for counties in 2010 and 2011 that installed camera in 2011
final_score	Final average Baccalaureate score
top20pcent	Dummy variable = 1, if student belongs to top 20% by Baccalaureate score
top20pcent_rom	Dummy variable = 1, if student belongs to top 20% by written Romanian score
ability_std	Standardized ability measure
fam_2child	Dummy variable = 1, if student had one sibling in Baccalaureate dataset
adm_year1	Year of admission to highschool if student took Bac exam in 4 years after enrollment
cty_hs_enrolled	Number of students enrolled in highschool 4 years prior to each Baccalauareate year
cty_midschool_grad	Number of students graduated from middle school 4 years prior to each Baccalauareate year
umempl_apr	Unemployement rate as measured in April 2009 and 2010 (Statistics Romania), county level
unempl10	Unemployment in 2010, county-level
trust_justice	County average score on the ”Trust in justice” question in the Public Opinion Barometer (Soros Fundation, 2007)
county_pop	Population of county
county_shromanian	Share Romanians in county population
total_population	
urban_population	

Appendix B, Test results for best linear prediction

```
#test_calibration(cf1)
#>Best linear fit using forest predictions (on held-out data)
#>as well as the mean forest prediction as regressors, along
#>with one-sided heteroskedasticity-robust (HC3) SEs:

#>                               Estimate Std. Error t value  Pr(>t)
#>mean.forest.prediction         1.97291    1.28375  1.5368 0.06217 .
#>differential.forest.prediction 0.18739    1.20419  0.1556 0.43817
#>---
#>  Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#----------------------------------------------------------------------
#test_calibration(cf2)
#>Best linear fit using forest predictions (on held-out data)
#>as well as the mean forest prediction as regressors, along
#>with one-sided heteroskedasticity-robust (HC3) SEs:

#>                               Estimate Std. Error t value    Pr(>t)
#>mean.forest.prediction         1.04850    0.30402  3.4488 0.0002816 ***
#>differential.forest.prediction 0.97931    0.44691  2.1913 0.0142154 *
#>---
#>  Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#----------------------------------------------------------------------
#test_calibration(cf3)
#>Best linear fit using forest predictions (on held-out data)
#>as well as the mean forest prediction as regressors, along
#>with one-sided heteroskedasticity-robust (HC3) SEs:

#>                               Estimate Std. Error t value    Pr(>t)
#>mean.forest.prediction         1.05707    0.24840  4.2554 1.044e-05 ***
#>differential.forest.prediction 0.64240    0.43778  1.4674 0.07114 .
#>---
#>  Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#----------------------------------------------------------------------
```

Figure 5: Estimates of the best linear predictor on the cleaned causal forests

Appendix C, List of code files in zip folder

File name	Description function of code
Tables_1-5_Replication.do	Replicate the first five tables of this paper
Main.m	Cleans the dataset from Borcan et al. (2017) of any NaN values
grf - hte.R	Creates the causal forests for the three endogenous variables together with its treatment effect and variable importance
leaf estimation.R	Calculates the heterogeneous treatment effects for one tree of the causal forest (since I have three endogenous variables, there are estimations for three trees in total)