Erasmus University Rotterdam

Erasmus School of Economics

Bachelor thesis

Econometrics and operational research

Voter polarisation in USA: Do education levels make a difference in the effect of Fox News on voting preferences?

July 4, 2021

Author:

Leonie Kantor

Student number:

485036

Supervisor:

Welz, M.F.O.

Second assessor:

Prof. dr. Franses, PHBF

Abstract

In this paper, we aim to give insight into whether the level of education attained by voters has a role in the current trend towards political polarisation in USA, by focusing on the relationship between Fox News and Republican voters. We research whether the effect of the introduction of Fox News on voting behaviour is heterogeneous along education levels. The effect, originally studied in DellaVigna and Kaplan (2007), is measured as a change in the Republican vote share in the 2000 Presidential elections. We use the recently developed machine learning method of causal forests, which makes it possible to accurately measure heterogeneous treatment effects. Our findings show that the education level has an impact on the effect of Fox news on voter behaviour. Areas with a larger proportion of non college educated people experienced smaller effects on the introduction of Fox news, a finding which may arise if these areas were already largely Republican.

Table of contents

1 Introduction

Over the past few decades, political ideologies in the US have grown more and more divided (Pew Research Center, 2017). One of the main factors that plays a role in this divide seems to be the level of education of the U.S. voter. In the 2016 presidential election, more than two-thirds of white voters who had not completed college voted for Donald Trump, whereas a significant majority of white voters with a four-year college degree or higher level education voted for Hillary Clinton. This trend, for white voters without a college degree to vote Republican, has grown over the past twenty years (Jones, 2019).

The political polarisation is also visible in the relations between voters and media sources. A recent research by Pew Research Center finds that Republicans tend to mistrust rather than trust a large majority of news channels, with the main exception of Fox News, while Democrats trusted most of the available sources(Mitchell, 2020). Furthermore, Democrats received their information from a number of different news channels, whereas, generally, Republicans used just one channel, Fox News.

In contrast to the distance between Republican voters and most of the media, their trust in Fox News is highlighted in these researches. The average viewer of Fox News is more likely to be Republican and conservative than the average U.S. adult (Gramlich, 2021). The channel has often been described as being politically slanted towards the Republican Party and having a conservative bias.[1]

The research by DellaVigna and Kaplan (2007) sheds light on the origins of the relation between Fox News and Republican voters. They find that in the 2000 Presidential elections, only four years after Fox News entered the cable market, the channel caused a significant increase in the Republican vote share. This shows that Fox News already had an important impact on voters when it was a new media platform and its audience was still relatively small and geographically limited. Further research into these results about the early stages of a close relationship between Fox News and Republicans could offer an understanding of the emergence of the voter polarisation currently witnessed in the U.S..

As DellaVigna and Kaplan (2007) do not specifically look at the role of education in their work, and given the increasing trend of political division based on education, insight could be gained by revisiting their research and investigating the relation between education and the uncovered impact of Fox News. Therefore, our paper researches whether the effect of Fox News on voting behaviour in the 2000 Presidential elections is heterogeneous along different levels of

[1] Retrieved 2021, June 15 from https://en.wikipedia.org/wiki/Fox_News_controversies#cite_note-Goldberg2007-5

education. For this, we use the estimation method of causal forests, a machine learning method for estimating heterogeneous treatment effects (Wager and Athey, 2018; Athey et al., 2019).

Accurate research on heterogeneous effects has only recently been made possible, and revisiting the paper by DellaVigna and Kaplan (2007) is interesting, as methods such as causal forests were not yet available when they performed their analysis. Causal forests are able to estimate per-individual treatment effects, which are consistent and asymptotically Gaussian. Therefore, they can be used for statistical inference, without making assumptions about the functional form of the data generating process (DGP). This is an important improvement for causal inference, as estimation methods such as difference-in-difference estimation, which is used in DellaVigna and Kaplan (2007), require strong assumptions. In particular, the assumption that the functional form is linear may be too strong, and findings of causality based on these methods might be biased, nonrobust, or may not be truly causal.

We find that the effect of Fox News on the Republican vote share in the 2000 Presidential election is heterogeneous along different levels of education. We also find that this effect is smaller in towns where the majority of people only have a high-school degree, which is contrary to what we may expect if the trend of Republican voters increasingly not having a college education originated around twenty years ago. These findings show that the magnitude of the impact of Fox News is indeed related to education levels. Further research could refine the understanding of the relationship between media, education levels, and the current political polarisation.

This thesis is structured as follows. Section 2 summarises papers relevant to our research, for both the relation between political media and education, and machine learning in causal inference. Section 3 describes the processing and collection of the data, and contains relevant information about the variables used in the research. In Section 4 we explain the methodology of causal forests, present our modelling decisions, and describe the analysis of heterogeneity. We also briefly denote the estimation methodology used to replicate a relevant table of DellaVigna and Kaplan (2007). The results of our analysis are presented in Section 5, along with a robustness check, and the replication results. Finally, Section 6 contains our conclusions and a discussion of the study, together with ideas for further research.

2 Literature review

This section presents an overview of the literature relevant to this research, which is divided into two sections. First, research into the relations between Fox news, media bias, and education is briefly discussed, and second, a small overview of the development of machine learning in causal inference is given.

2.1 Fox news, media bias, and education

The paper by DellaVigna and Kaplan (2007) researches the effect of (the media bias of) Fox News on voting behaviour. They find that Fox News caused an increase of 0.4 to 0.7 percentage points in the Republican vote share in the Presidential elections between 1996 and 2000, and that it induced 3 to 28 percent of non-Republican viewers to vote Republican.

It is reasonable to surmise that the extent to which parties are representative for people is related to their level of education, as this itself is linked to many other demographic factors that determine political interests. Depending on which values the Republican Party were specifically appealing to in the 2000 elections, this may therefore account for possible differences in the effect of Fox News voter behaviour. However, education may play a role in other ways, and relevant factors to look into are the impact of education on persuasion effects and on the perception of biased media.

Johansen and Joslyn (2008) research the mediating role of education in the face of propaganda, or media slanted by a political agenda. They find that although education can undermine political propaganda, under circumstances of one-sided news it does not form a barrier. In their paper, factual inaccuracies about the Iraq War were found to be equal among the highest and lowest educated levels after exposure to the least war-critical channels at that time, Fox News and CBS.

The paper by Joslyn and Haider-Markel (2014) offers two alternative effects of higher levels of education on the effect of biased media. They highlight that on the one hand, more educated people have been exposed to varying models of thought, and have developed cognitive skills which allow them to assess and weigh different sources of information. On the other hand, education can reinforce targeted information processing, or confirmation bias, where information sources supporting an opinion are favoured, and evidence in opposition of the sought conclusion are disregarded.

2.2 Machine learning in causal inference

With the use of machine learning, methods have been (and continue to be) developed which greatly improve causal estimation. Whereas with traditional econometric methods the focus is on estimating effects with models mainly based on statistical theory, machine learning methods use the data to determine the best fitting models (Athey, 2019).

Examples of estimation frameworks that use machine learning include Double Machine Learning (DML) (Chernozhukov et al., 2018a), which uses doubly robust estimation methods for the average treatment effect (ATE) in high-dimensional cases, and Bayesian Additive Regression Trees (BART), which estimates the ATE based on a Bayesian form of boosted regression trees (Chipman et al., 2010). Causal forests, the method used in our paper, is an application of the framework of generalized random forests (Athey et al., 2019). Generalized random forests expand on the random forest algorithms by Breiman (2001), but they have some important differences. Athey et al. (2019) make use of local moment conditions to build non-parametric estimators. These estimators are flexible in the functional form, which means they do not have to assume for example linearity, and the developed asymptotic theory for these forests make them widely applicable to statistical research. Causal forests are proposed by Wager and Athey (2018), and greatly improved by the generalized random forests framework. This method estimates per-individual effects and is focused on estimating heterogeneity.

3 Data

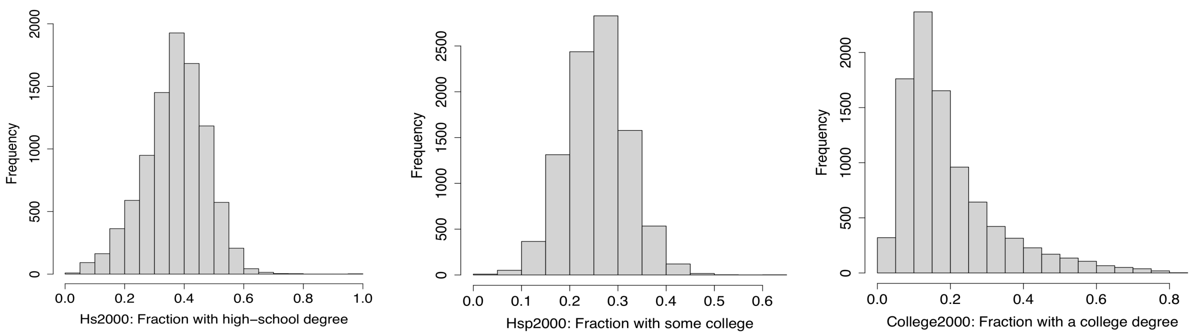
This paper is based on dataset from DellaVigna and Kaplan (2007). We also use the same preprocessing methods as found in their paper. The data is obtained from a number of sources containing information about cable news, U.S Presidential elections, and demographics on the town-level. The availability of channels before the 2000 Presidential election and other information about local cable companies is found in the Television and Cable Factbook, 2001 edition (Warren, 2001). Election data is collected from the Election Division of the Secretary of State of each state and from the Federal Election Project (Lublin and Voss, 2001). The voting data is aggregated to the town level, and then contains voting information for the Presidential elections of 1996 and 2000 for 28 states, with 26,710 towns in total. Finally, demographic information is obtained from the United States 1990 and 2000 Census for 27,064 towns. All the collected data is then matched by town, county, and state, which gives a dataset containing 10,126 towns. 289 of these towns were removed if they had multiple cable companies of which at least one did not offer Fox News, as it is then not guaranteed that those who watched cable news had access to Fox News. 324 additional towns were removed which did not offer CNN, as this indicated that the offered programming was not comparable to other towns. Finally, 257 towns were left out as they were likely to have inaccurate voting data. This leaves 9,256 towns in the final dataset, which are a part of 813 counties and 235 congressional district. The data contains information about 65.9 percent of the U.S. population and 68.6 percent of the total votes cast for 28 states in the 2000 Presidential elections.

Table 1: Relevant variables

Variable name	*Description*
Foxnews2000	Binary Availability of Fox News Channel in 2000
Repvoteshare1996	Two-party Republican vote share in 1996 Presidential elections
Repvoteshare2000	Two-party Republican vote share in 2000 Presidential elections
hs2000	Fraction with high-school degree in 2000 Census
hsp2000	Fraction with some college in 2000 Census
college2000	Fraction with college degree in 2000 Census
Variable groups	
$\mathbf{X}_{k,2000}$	Demographic controls from the 2000 Census
$\mathbf{X}_{k,00-90}$	Change in demographic controls between the 2000 and 1990 Census
$\mathbf{C}_{k,2000}$	Deciles in the number of provided channels and in the potential subscribers

Notes: This table contains descriptions of the most relevant variables and variable groups in this research. For a full overview of the variables and their statistics, see DellaVigna and Kaplan (2007)

The research contains an outcome variable, a treatment variable, and 42 control variables. Table 1 describes the most relevant variables and groups of variables in this paper. All of these variables are at town-level. Further, the two-party Republican vote share means the Republican share of Republican and Democratic votes. Also seen in the table are the three education variables, which are contained in the variable group $X_{k,2000}$, but are mentioned separately, as they are the main interest in this paper. For these variables, we include histograms of their distributions in Figure 1. For an overview of all the variables and relevant statistics, we refer to DellaVigna and Kaplan (2007).



(a) Hs2000 histogram (b) Hsp2000 histogram (c) College2000 histogram

Figure 1: Histograms for the three education variables.

4 Methodology

This research analyses heterogeneous treatment effects using the machine learning method of causal forests Athey et al. (2019); Wager and Athey (2018). As mentioned in Section 2.1, this method has a number of advantages. Firstly, as it is a machine learning method, it does not make assumptions about the functional form of the DGP. Secondly, it is designed to capture heterogeneity, and is able to estimate per-individual treatment effects. And finally, the reason this method is particularly promising is because, under a few conditions, the causal forests predictions are asymptotically Gaussian and unbiased, and they allow for valid confidence intervals. This means that we can perform statistical tests on the predictions, which was previously not possible. Therefore, this method is suitable for this research, and is preferred over other machine learning methods such as DML. The causal forests are programmed in the statistical software program `R`, using the generalized random forest package `grf` (Tibshirani et al., 2020; R Core Team, 2020).

To explain causal forests, we first explain the potential outcomes framework, and the related assumptions on which causal forests are based. We then elaborate on the theory of generalized random forest, and how this is implemented in causal forests. This is followed by an outline of our modelling decisions. Next, we describe the methodology of our heterogeneity analysis. Finally, we add a brief explanation about the method for the replication of the Presidential election results in DellaVigna and Kaplan (2007).

4.1 Potential outcomes framework and assumptions

In the potential outcomes framework (Imbens and Rubin, 2015), $\mathrm{Y}_i^{(0)}$ and $\mathrm{Y}_i^{(1)}$ denote the potential outcomes if a unit i does not receive treatment and if it does receive treatment, respectively. The treatment effect is then the distance between these two potential outcomes. However, at a given time, a unit i can either receive or not receive treatment, and only one of the potential outcomes can be measured. Therefore, causal inference is per definition a problem of missing data. As it is impossible to directly measure the treatment effect $\tau(x)$, the aim of causal inference is to estimate the treatment effect. At the basis of methods used for estimation is the assumption of unconfoundedness (Rosenbaum and Rubin, 1983). In observational studies the assignment probability of a treatment is unknown before it is observed. However, the unconfoundedness assumption states that the assignment probability is independent of the potential outcomes. Furthermore, it is assumed that the treatment assignment is a probability, bounded between 0 and 1, and that this probability is independent for each unit. With these assumptions, treatment assignment for units within subpopulations which have the same covariate values can be viewed as random.

4.2 Causal forests

The unconfoundedness assumption is also the basis for causal forests as developed by Wager and Athey (2018). Given n independent and identically distributed training samples $i = 1, ..., n$, with for each sample a feature vector $X \in [0,1]^d$, outcome $Y_i \in \mathbb{R}$, and treatment indicator $W_i \in \{0,1\}$, they define the conditional average treatment effect as

$$\tau(x) = \mathbb{E}[Y_i^{(1)} - Y_i^{(0)} \mid X_i = x], \tag{1}$$

and the unconfoundedness assumption as

$$Y_i^{(0)}, Y_i^{(1)} \perp\!\!\!\perp W_i \mid X_i. \tag{2}$$

The assumption means that treatment is assigned independently of the outcome, given features X. As described in the potential outcomes framework, this allows treatment assignment for observations with similar features to be viewed as random, which is necessary for consistency of estimation methods of $\tau(x)$. In this context of trees and forests, the assumption implies that observations within the same leaf on the decision tree can be treated as randomly assigned.

4.2.1 Generalized Random Forest

In Breiman (2001)'s random forests, the outcomes predicted by the forest are the average of the predictions in all the trees. Athey et al. (2019) suggest that it is better for statistical analysis to use weights which adaptively find the best local estimation, by averaging neighbourhoods which are implied by the trees. Therefore, estimation of the treatment effect starts by defining weights $\alpha_i(x)$, which effectively measure the frequency with which each observation in the training sample falls in the same leaf as x. With a set of B trees $b = 1, ..., B$, the weights $\alpha_i(x)$, which sum to 1, are defined as

$$\alpha_{bi}(x) = \frac{\mathbf{1}(X_i \in L_b(x))}{|L_b(x)|}, \quad \alpha_i(x) = \frac{1}{B}\sum_{b=1}^{B} \alpha_{bi}(x), \tag{3}$$

where L_b is the set of training observations that fall in the same leaf as x. With these weights, Athey et al. (2019) use orthogonalized moment conditions to arrive at the treatment effect estimator. Following notation from the 2018 Atlantic Causal Inference Conference analysis (Athey and Wager, 2019), we denote the treatment estimator $\hat{\tau}$ as

$$\hat{\tau} = \frac{\sum_{i=1}^{n} \alpha_i(x)(Y_i - \hat{m}^{(-i)}(X_i))(W_i - \hat{e}^{(-i)}(X_i)}{\sum_{i=1}^{n} \alpha_i(x)(W_i - \hat{e}^{(-i)}(X_i))^2}, \tag{4}$$

where the conditional probability of receiving treatment, or the propensity score, is written as $e(x) = \mathbb{P}[W_i = 1 \mid X_i = x]$, and $m(x) = \mathbb{E}[Y_i \mid X_i = x]$ is the conditional expected outcome. The superscript $^{(-i)}$ is used to indicate that the variables are predicted without the use of outcome observations with which the forest was trained, also referred to as out-of-bag predictions. This relates to the concept of honest trees, which is later explained.

Generalized random forests are grown from trees that use greedy splitting, which means that each split is done to directly improve the fit of the tree as much as possible. While the random forests developed by Breiman (2001) focus on minimising the mean squared error, the forests in this research are focused on heterogeneity in the treatment effect.

Further, following Wager and Athey (2018), all the trees in the causal forest use subsamples of the data to train the forest, and the trees are all honest. This means that for each observation in the training subsample, outcome Y_i is used either to estimate the treatment effect, or to determine where the splits are placed, and not both. This is realised by dividing the training data into two subsets, where one is used to grow the tree, and the other is used to estimate the predictions. The division into two subsets of the training data is randomized for each tree, which makes this an efficient way to use the training data. With the proper use of subsamples and honesty, the estimates of the treatment effect are consistent and asymptotically normal. This allows us to perform statistical tests on the causal forest predictions.

In our implementation of the causal forest, we follow the methods of Athey and Wager (2019). Similarly, we first train a forest on all the feature variables, which gives the variables which are split on the most, and then grow the final forest on only these most important variables. This is motivated by Basu et al. (2018), and could improve the causal forest's precision.

4.3 Model selection

The variables used to construct the causal forest closely follow the variable choices as made in DellaVigna and Kaplan (2007). The outcome of interest is the change in the Republican vote share between the 1996 and the 2000 Presidential elections ($Repvoteshare2000 - Repvoteshare1996$), and the binary treatment variable is the availability of Fox News in 2000 ($Foxnews2000$). The control variables consist of the 2000 Census variables ($X_{k,2000}$), the change in Census variables between the 2000 and 1996 Census ($X_{k,00-90}$), as well as the deciles in the number of channels provided and in the potential subscribers ($C_{k,2000}$). DellaVigna and Kaplan (2007) analyse which factors have an influence on receiving treatment, and find that after controlling for the mentioned feature variables, and after weighting towns by size, the treatment assignment is uncorrelated with political variables, and essentially random within an area, such as county or district.

This result also leads to the choice of clustering the standard errors in the causal forest. As the availability of Fox News depends on the area, the assignment mechanism seems to be clustered, and this is a valid reason to cluster standard errors based on area (Abadie et al., 2017). In DellaVigna and Kaplan (2007), counties and congressional districts are used to group areas for fixed effects, and we follow this by analysing a causal forest with clustering standard errors by county, and a forest which has district clustered standard errors. We note that the analysis by (DellaVigna and Kaplan, 2007) clusters standard errors by cable company owners. However, as there are 2992 cable owners, an average cluster contains around 3 towns. These are very small clusters, and they may lead to a lot of noise in the individual treatment effects. Therefore, we do not include this clustering in our analysis.

Another modelling decision is whether the clusters should all be weighted equally. This would be the case if the aim of the analysis is to make the results generalisable to other districts, counties, or towns with certain cable owners. However, as the aim of this research is to accurately investigate the effect that Fox News when the channel was still relatively new, we do not adjust the weights of the clusters.

Further, following DellaVigna and Kaplan (2007), we weight the observations by the votes cast in 1996 to better represent the average voter, and because towns with more voters can show more precise changes in vote share.

The variables of interest in this research are those related to education. These are $hs2000$, the fraction of people with a high-school degree, $hsp2000$, the fraction of people who went to college, and $college2000$, the fraction who obtained a four-year college degree. Those who have some college experience or who have obtained a college degree also have a high-school degree. Therefore, the $hs2000$ variable can be interpreted as the fraction that did not go on to do higher education after high-school.

4.4 Analysing heterogeneity

With the causal forests, we obtain estimates of the treatment effects for each town, allowing us to investigate the presence of heterogeneity. For this, we use methods that assess the accuracy of the causal forests and their general indications about heterogeneity, after which we zoom in on the analysis along the education variables.

First, following the method from Athey and Wager (2019), we perform a calibration test of the causal forests, which gives an indication of the predictive accuracy of the causal forest and whether it captures treatment heterogeneity. It is based on the "best linear predictor" method from Chernozhukov et al. (2018b), and tries to fit the conditional average treatment effect as a linear function of the causal forest individual treatment effect estimates. This allows us to test whether heterogeneity in the estimates is associated with heterogeneity in the treatment effect. The function gives output for the mean forest prediction, where a coefficient of 1 indicates that the forest is good at predicting the out-of-bag sample, and for the differential forest prediction, where a value of 1 indicates heterogeneity over the treatment effects in general. If the associated p-value is significant, this is evidence to reject the null hypothesis of no heterogeneity. However, this is only an indication, and results are not definitive.

Then, we look at heterogeneity along each of the education variables, $hs2000$, $hsp2000$, and $college2000$. Plots of the individual treatment effect estimates against the education variables can give a first indication of heterogeneity along education.

To further analyse this, we construct hypothesis tests. For each variable, the towns are divided into two groups, based on whether they have higher or lower values than the median of the variable. Then, we compare the estimated individual treatment effects of the towns in the two groups. We refer to the subsets of estimates of the towns with higher and lower than the median of the education variables as $\hat{\tau}_{high}$ and $\hat{\tau}_{low}$, respectively. A t-test in `R` determines whether the treatment effects differ for towns with higher values of the variable and for towns with lower values. The null hypothesis states that the means of these estimates are equal, and the alternative hypothesis is that they are not. However, this test may not be sufficient to indicate whether heterogeneity is present, as the test assumes that the estimated individual effects are fixed, thereby neglecting estimation uncertainty.

Therefore, we further analyse the groups of higher and lower variables by calculating the ATE's for both subsets of towns with the `average_treatment_effect` function in `R`. We denote the ATE's for the group of towns with higher than median values of the specific education variable as ATE_{high}, and with lower values as ATE_{high}. Using a manually calculated t-test, we compare the ATE's of the two groups. The null hypothesis for this test is that the ATE's are equal. If the p-value of the t-statistic is significant, there is evidence to reject this null hypothesis and accept the alternative hypothesis, which states that the average effects are different, indicating heterogeneous treatment effects based on the education variable. Yet, the results from this method must also be interpreted with caution, as it assumes that the ATE's for the two groups are independent, which is not necessarily true.

As the causal forests implement clusters based on county and district, it is also interesting to look at within-area heterogeneity. We create cluster-level education variables, and calculate per-cluster treatment effect estimates. For this, we use the estimators for average treatment effect ($\hat{\tau}$) and standard error ($\hat{\sigma}$) underlying the `average_treatment_effect` function to manually find robust estimates of the per-cluster treatment effects. Following notation from Athey and

Wager (2019), these estimators are written as

$$\hat{\tau}_j = \frac{1}{n_j}\sum_{i:A_i=j}\hat{\Gamma}_i, \quad \hat{\tau} = \frac{1}{J}\sum_{j=1}^{J}\hat{\tau}_j, \quad \hat{\sigma}^2 = \frac{1}{J(J-1)}\sum_{j=1}^{J}(\hat{\tau}_j-\hat{\tau})^2,$$
$$\hat{\Gamma}_i = \hat{\tau}^{-i}(X_i) + \frac{W_i - \hat{e}^{(-i)}(X_i)}{\hat{e}^{(i)}(X_i)(1-\hat{e}^{(}-i)(X_i))}(Y_i - \hat{m}^{(-i)}(X_i) - (W_i - \hat{e}^{(-i)}(X_i))\hat{\tau}^{(-i)}(X_i)). \tag{5}$$

Using this to manually calculate the doubly robust treatment effect estimates per cluster $\hat{\tau}_j$, we can compare treatment effects from clusters with education variables above the median to those from clusters with values below the median. Again, a t-test indicates whether the null hypothesis of equal treatment effects is to be accepted, or whether there is evidence of heterogeneity.

4.5 Replication methodology

DellaVigna and Kaplan (2007) have published replication files for their paper, containing the necessary data and files to run the regressions in Stata. Of relevance to this research are the results on the Presidential elections of the year 2000, which are found in their paper in Table *iv*. DellaVigna and Kaplan (2007) research whether towns which broadcast Fox News experienced an increase in Republican vote share, using the estimation method difference-in-difference. The baseline equation for the estimation is written as

$$\begin{aligned} Repvoteshare2000 - Repvoteshare1996 &= \alpha + \beta_F Foxnews2000 \\ &+ \Gamma_{2000}X_{k,2000} + \Gamma_{00-90}X_{k,00-90} + \Gamma_C C_{k,2000} + \mathcal{E}_k, \end{aligned} \tag{6}$$

where $Foxnews2000$ is again the dummy variable of receiving treatment. The observations are weighted by the votes cast in 1996, and the standard errors are clustered at at the level of the local cable company.

5 Results

In this section, we first report the results for the heterogeneity analysis for the two causal forests which are clustered either by county or by congressional district. We discuss the calibration tests, indicating prediction accuracy and testing the hypothesis of no heterogeneity, and follow this by discussing the plots and the various t-tests, which are all performed on the three education variables: $hs2000$, $hsp2000$, and $college2000$. The following subsection describes a check for robustness, based on unweighted observations. Finally, we report the results from the partial replication of the paper by DellaVigna and Kaplan (2007).

5.1 Main results

The results in Table 2 give an indication of whether the causal forests are successful in accurately estimating the treatment heterogeneity. As the estimates of the differential forest predictions are significant on a 1% significance level for both forests, there is strong evidence to reject the null hypothesis of no heterogeneity. Furthermore, with a value of 1.01, the mean forest prediction estimate indicates that the district clustered forest has accurate out-of-bag predictions. For the forest based on county clustering, this estimate is slightly smaller, but still close to 1 (0.930), and a t-test indicates that it does not differ significantly from 1, suggesting a correct average prediction.

Table 2: Calibration tests for the causal forests clustered by county and by congressional district.

	County		District	
	Mean prediction	Differential prediction	Mean prediction	Differential prediction
Estimate	0.930	0.860	1.01	0.825
p-value	0.0661*	0.00907***	0.0890*	0.00356***

Notes: Asterisks denote a significance level of ****0.001, ***0.01, **0.05, and *0.1.

Figure 2 shows the estimates from the causal forest clustered by county plotted against the three education variables, $hs2000$, $hsp2000$, and $college2000$. All three plots show varying sizes of treatment effect estimates, which indicates heterogeneity along education. For $hs2000$, effect magnitudes seem to vary especially across the towns with smaller fractions of people with only a high-school degree. This is mirrored in the plot for $hsp2000$, and similarly but less strongly in the plot for $college2000$. As the plots look similar for the district clustered causal forest estimates, we include these in Appendix A.

Table 3 presents the results of t-tests on the estimates of individual treatment effects grouped by values of the education variables. For each variable and both ways of clustering in the causal forest, the outcome of the t-test comparing $\hat{\tau}_{high}$ and $\hat{\tau}_{low}$ gives a significant (on a 0.1% significance level) p-value. Therefore, for each variable, we can reject the null hypothesis of the estimates in $\hat{\tau}_{high}$ and those in $\hat{\tau}_{low}$ being equal. This would suggest heterogeneity along every variable. We note that the t-value for the test along $hs2000$ is negative. This indicates that the towns with larger fractions of people with a high-school degree experience smaller treatment effects, which suggests that Fox News being available here does not cause the Republican vote share to increase as much, when compared to towns with smaller fractions of people with a high-school degree. In general, if a town has a larger fraction of people with a high-school degree, it has a smaller fraction of people with college education. Similarly, the positive t-values for

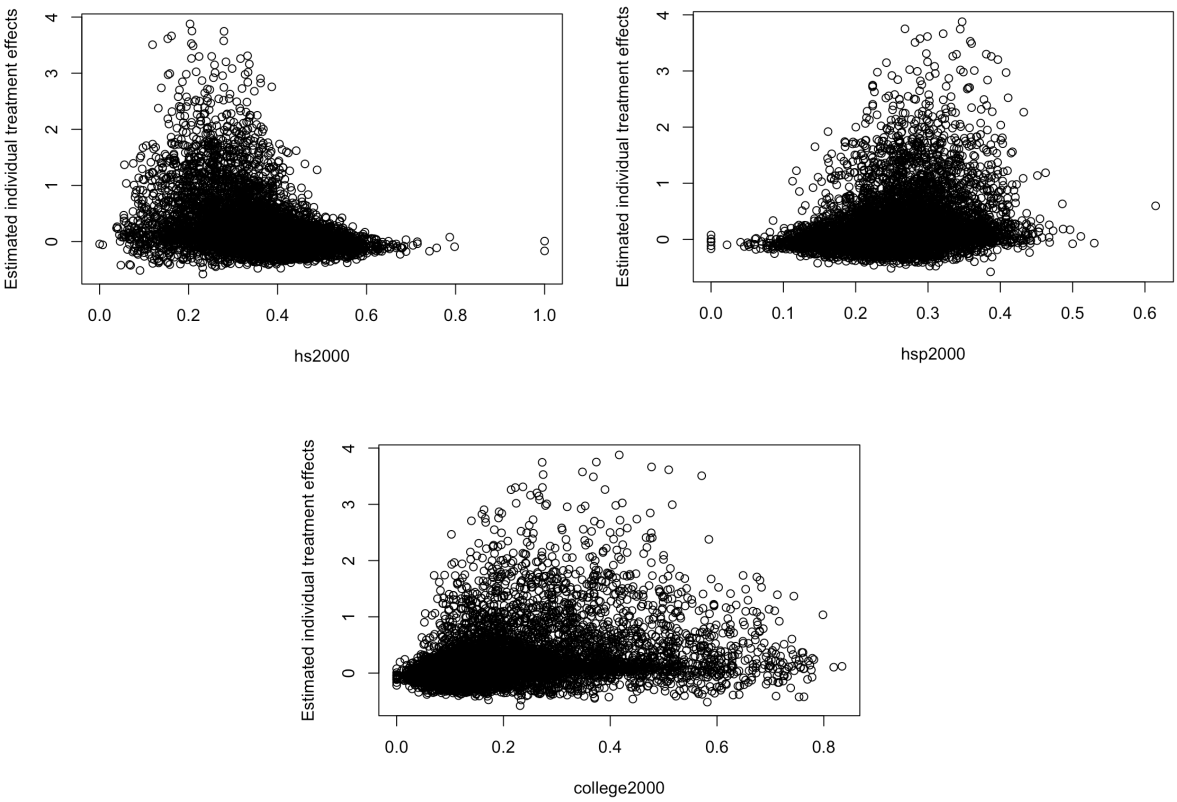


Figure 2: The figures plot the estimated individual treatment effects against the three education variables. The estimates are obtained from the causal forest which uses clustering by county.

$hsp2000$ and $college2000$ suggest that Fox News had a bigger impact on the Republican vote share in towns with larger fractions of higher educated people. However, as explained in Section 4.4, these results do not take into account estimation uncertainty, and we compare these results with those from the further analysis to assess the evidence for heterogeneity.

Table 3: Results of t-tests for heterogeneous treatment effects of Fox News on the Republican vote share in the 2000 Presidential election along education.

	hs2000		hsp2000		college2000	
	County	District	County	District	County	District
t-value	-33.5	-31.4	18.3	18.5	28.6	28.7
p-value	0.000****	0.0000****	0.000****	0.000****	0.000****	0.000****

Notes: Asterisks denote a significance level of ****0.001, ***0.01, **0.05, and *0.1.

Table 4 shows the ATE$_{high}$ and ATE$_{low}$, the ATE of the towns with higher than the median values for the specific variable and that of the towns with lower than the median values. It also presents the results from manually performing t-tests to determine whether these ATE's differ significantly. As the p-values of the t-tests for the variable $hs2000$ are significant on a 5% and 10% significance level for the county and district clustering model, respectively, this suggests that the null hypothesis of equal ATE's should be rejected. This means that the treatment effects are heterogeneous for towns with larger and smaller fractions of people with high-school degrees. Specifically, as the ATE_{high}'s are -0.0456 and -0.0325 for the county and district clustered forests, respectively, and the ATE_{low}'s are higher with values of 0.395 and 0.411, towns with larger fractions of people who have no further education after high-school experienced on average significantly smaller or no increase in the Republican vote share.

These results support the heterogeneous results for $hs2000$ found in Table 3, but do not provide evidence for heterogeneity along the other two variables. However, as these tests assume independence of the ATE's, which may not hold true, the results must again be interpreted with caution.

Table 4: Results of t-tests for heterogeneity in ATE_{high} and ATE_{low}

.

	hs2000		hsp2000		college2000	
	County	District	County	District	County	District
ATE_{high}	-0.0456	-0.0325	0.284	0.309	0.254	0.281
	(0.107)	(0.112)	(0.135)	(0.194)	(0.177)	(0.196)
ATE_{low}	0.395	0.411	0.0498	0.0505	0.0798	0.0809
	(0.170)	(0.208)	(0.137)	(0.126)	(0.0919)	(0.113)
t-value	-2.24	-1.81	1.2	1.06	0.899	0.776
p-value	0.0251**	0.0703*	0.230	0.289	0.369	0.438

Notes: Standard errors are reported in parentheses ATE_{high} (ATE_{low}) refers to the ATE of towns with higher (lower) values for the education variable than the median. The ATE's are based on weighted observations. Asterisks denote a significance level of ****0.001, ***0.01, **0.05, and *0.1.

Finally, Table 5 shows the results of looking at within-county and within-district heterogeneity for the three education variables. The t-tests finds evidence to reject the null hypothesis of no heterogeneity across the district-level high-school degree fraction, at a significance level of 1%. This suggests that the magnitude of the per-district treatment effect differs significantly across larger and smaller fractions of people with only a high-school degree. Further, the negative sign of the t-value points to the effect Fox news has on voter behaviour being smaller in districts with a larger fraction of people with high-school degrees.

Table 5: Results of t-tests for heterogeneity in clustered treatment effects for high and low values of education variables.

	hs2000		hsp2000		college2000	
	County	District	County	District	County	District
t-value	-0.934	-2.68	1.29	1.32	-0.934	1.47
p-value	0.350	0.00789***	0.196	0.187	0.351	0.144

Notes: Asterisks denote a significance level of ****0.001, ***0.01, **0.05, and *0.1.

5.2 Robustness

Following DellaVigna and Kaplan (2007), the observations in this research are weighted by the votes cast in the 1996 Presidential elections, with the aim of increasing accuracy and better representing the average voter. However, this may mask effects from towns with a smaller population, and important associated differences. Specifically, educational attainment depends on proximity to schools and colleges, and residents of rural counties are less likely to obtain a college degree (Campbell, 2019). As rural towns typically have smaller populations, weighting by votes cast may obscure important results of this analysis. Therefore, we investigate the robustness of the results by performing the same heterogeneity analysis on unweighted observations.

The resulting tables are included in Appendix B. These results are similar to the main analysis for the calibration tests and the t-tests on per-individual treatment effect estimates, but they differ for the t-tests on the ATE's and on the per-cluster estimates. For the unweighted observations, while the t-tests on per-individual treatment effects again all strongly suggest heterogeneity, there is no indication of treatment effect heterogeneity when we look at the ATE's of towns with higher and lower values of the education variables. Further, at 10% significance, the results suggest heterogeneity in the estimates of the per-district treatment effects along $hsp2000$. As the t-value is positive, this means that districts with larger fractions of people who attend college after high-school (excluding those who obtain a degree) experienced a larger effect of Fox News on the Republican vote share in the 2000 Presidential elections.

The observed differences in the results for weighted and unweighted observations show that the decision to weight by the total votes cast does have an influence on our findings.

The reasons for this are not clear from this research. We do however note that the plot in Figure 3 shows that most of the towns with a larger than median fraction with high-school degrees have a smaller number of votes cast. DellaVigna and Kaplan (2007) find that the impact of Fox News is smaller in more rural towns. The plots of the estimated treatment effects in Figure 2 in Section 5.1 indicate that the heterogeneity we observe is mainly in the towns with smaller fractions of people with high-school degrees. Therefore, when we do not weight the observations, the heterogeneity may be less observable.

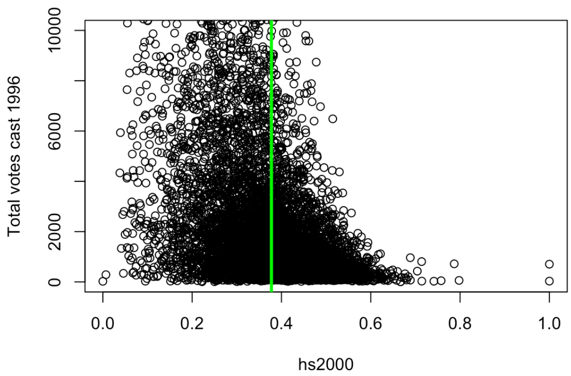


Figure 3: Plot of total votes cast in 1996 against $hs2000$

Notes: This figure plots the total votes cast per town in the 1996 Presidential elections against the education variable $hs2000$. The green line indicates the median value of $hs2000$.

5.3 Replication results

Table 6 shows the results from replicating the effect of Fox News on the change in the two-party Republican vote share between the 1996 and 2000 Presidential elections, using the methods from DellaVigna and Kaplan (2007). First, estimation is performed with the difference-in-difference equation, without controls. Then, in column (2) and (3) the demographic controls and the cable size control variables are added consecutively. These are followed by the benchmark estimations, which are district fixed effects (column (4)) and county fixed effects (column (5)). In the final two columns, the fixed effect estimations are replicated, but with the variable of the change in the Republican vote share between 1988 and 1992 added to control for political trends. This reduces the sample to 3,772 observations.

Our replication provides almost identical results to the original research. However, very slight differences can be observed in the standard errors in the fixed effect regressions, seen in columns (4) through (7). We note that these differences do not affect the significance of the variables, or the R^2 of the models. A possible explanation for the minimal differences in results is that the program used to run the estimations, Stata, may have slightly changed the fixed effects calculations since the original paper in 2007. In that year, version 10 was the most recent release, and we perform the replication in version 16.

Table 6: The Effect of Fox News on the 2000-1996 Vote Share Change

	Republican two-party vote share change between 2000 and 1996 pres. elections						
Dep. var.	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Availability of Fox News via cable in 2000	-0.0025 (0.0037)	0.0027 (0.0024)	0.0080 (0.0026)***	0.0042 (0.0015)***	0.0069 (0.0015)***	0.0037 (0.0022)*	0.0048 (0.002)**
Pres. Rep. vote share change 1988-1992						0.0229 (0.0218)	0.0514 (0.0234)**
Constant	0.0347 (0.0017)***	-0.0280 (0.0245)	-0.0255 (0.0236)	0.0116 (0.0156)	0.0253 (0.0198)	-0.0377 (0.0261)	0.0081 (0.0335)
Control variables							
Census controls: 1990 and 2000	-	X	X	X	X	X	X
Cable system controls	-	-	X	X	X	X	X
U.S. House district fixed effects	-	-	-	X	-	X	-
County fixed effects	-	-	-	-	X	-	X
R^2	0.0007	0.5207	0.5573	0.7533	0.8119	0.7528	0.8244
N	9,256	9,256	9,256	9,256	9,256	3,722	3,722

Notes: This table reports the results from replicating *table iv* from DellaVigna and Kaplan (2007). The standard errors are in parentheses. Asterisks denote a significance level of ****0.001, ***0.01, **0.05, and *0.1. The mark X denotes whether the control variables where included in the model.

6 Conclusion

In this paper, we research whether the effect of Fox News on voting behaviour in the 2000 Presidential elections is heterogeneous along education levels. Based on the study by DellaVigna and Kaplan (2007), we analyse the change in the Republican vote share between 1996 and 2000. For this, we use the recently developed causal inference method of causal forests, which makes it possible to accurately estimate heterogeneous treatment effects (Athey et al., 2019; Wager and Athey, 2018). We aim to give insight into the development of the current political polarisation, by investigating whether education level was a factor in the early development of the relationship between Fox News and Republican voters.

Our findings indicate significant heterogeneity along education levels in the effect of the availability of Fox News on an increase in the Republican vote share in the 2000 Presidential elections. In addition, we specifically find that the effect is smaller in towns where more people only have a high-school degree. We find similar heterogeneity along education at the district level.

These findings seem to be contrary to the current trend of people without higher education increasingly supporting the Republican Party. However, as the impact of Fox News is measured here as an increase in the Republican vote share, a smaller effect may be observed in towns and areas that were already largely Republican before Fox News was introduced. In such places, it is conceivable that there are fewer non-Republican voters in the areas susceptible to influence from Fox News. We consider it possible therefore that our findings may in fact reflect populations with largely high-school degrees having a pre-existing preference for the Republican Party.

In their study, DellaVigna and Kaplan (2007) find smaller effects of Fox News in small and mostly rural towns. As in our dataset the towns with less highly educated people mostly have smaller voting populations, and residents of smaller and more rural areas are less likely to attend college (Campbell, 2019), our findings may not indicate that education is causal for the heterogeneity that we observe. Further research would therefore be needed to examine the causality of education in our results.

As we observe that the heterogeneity seems to be especially present across the towns with more highly educated people, and that this is associated with larger voting populations and possibly more urban areas, the deeper analysis into the heterogeneity along education could benefit from being focused on these populations. This could also give results that are more robust to our decision to weight the observations, as we find that a lack of robustness could be due to the above mentioned factors.

Finally, we note that the methods to analyse heterogeneous effects are each flawed to some extent. The t-tests on individual treatment effect estimates do not take estimation uncertainty into account, and the t-tests on average treatment effects assume that the averages are independent, which may not in this case be true. It would greatly improve the assessments of heterogeneity in treatment effect estimates if more valid tests for this were developed in future research.

References

A. Abadie, S. Athey, G. W. Imbens, and J. Wooldridge. When should you adjust standard errors for clustering? Technical report, National Bureau of Economic Research, 2017.

S. Athey. 21. the impact of machine learning on economics. In *The economics of artificial intelligence*, pages 507–552. University of Chicago Press, 2019.

S. Athey and S. Wager. Estimating treatment effects with causal forests: An application. *arXiv preprint arXiv:1902.07409*, 2019.

S. Athey, J. Tibshirani, S. Wager, et al. Generalized random forests. *Annals of Statistics*, 47 (2):1148–1178, 2019.

S. Basu, K. Kumbier, J. B. Brown, and B. Yu. Iterative random forests to discover predictive and stable high-order interactions. *Proceedings of the National Academy of Sciences*, 115(8): 1943–1948, 2018.

L. Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

C. Campbell. Those left behind gaps in college attainment by race and geography. 2019.

V. Chernozhukov, D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins. Double/debiased machine learning for treatment and structural parameters, 2018a.

V. Chernozhukov, M. Demirer, E. Duflo, and I. Fernandez-Val. Generic machine learning inference on heterogenous treatment effects in randomized experiments. Technical report, National Bureau of Economic Research, 2018b.

H. A. Chipman, E. I. George, R. E. McCulloch, et al. Bart: Bayesian additive regression trees. *The Annals of Applied Statistics*, 4(1):266–298, 2010.

S. DellaVigna and E. Kaplan. The fox news effect: Media bias and voting. *The Quarterly Journal of Economics*, 122(3):1187–1234, 2007.

J. Gramlich. 5 facts about fox news. https://www.pewresearch.org, 2021.

G. W. Imbens and D. B. Rubin. *Causal inference in statistics, social, and biomedical sciences.* Cambridge University Press, 2015.

M. S. Johansen and M. R. Joslyn. Political persuasion during times of crisis: the effects of education and news media on citizens' factual information about iraq. *Journalism & Mass Communication Quarterly*, 85(3):591–608, 2008.

J. Jones. Non-college whites had affinity for gop before trump. https://news.gallup.com, 2019.

M. R. Joslyn and D. P. Haider-Markel. Who knows best? education, partisanship, and contested facts. *Politics & Policy*, 42(6):919–947, 2014.

D. Lublin and D. S. Voss. Federal elections project. *American University, Washington, DC and the University of Kentucky, Lexington, KY*, 2001.

W. Mitchell, Shearer. U.s. media polarization an the 2020 election: a nation divided. https://www.pewresearch.org/, 2020.

Pew Research Center. Partisan divides over political values widen. https://www.pewresearch.org, 2017.

R Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2020. URL https://www.R-project.org/.

P. R. Rosenbaum and D. B. Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55, 1983.

J. Tibshirani, S. Athey, and S. Wager. *grf: Generalized Random Forests*, 2020. URL https://CRAN.R-project.org/package=grf. R package version 1.2.0.

S. Wager and S. Athey. Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523):1228–1242, 2018.

Warren. *Television & Cable Factbook*, volume 2. Television Digest, Incorporated, 2001.

Appendices

A Plots for causal forest clustered by district

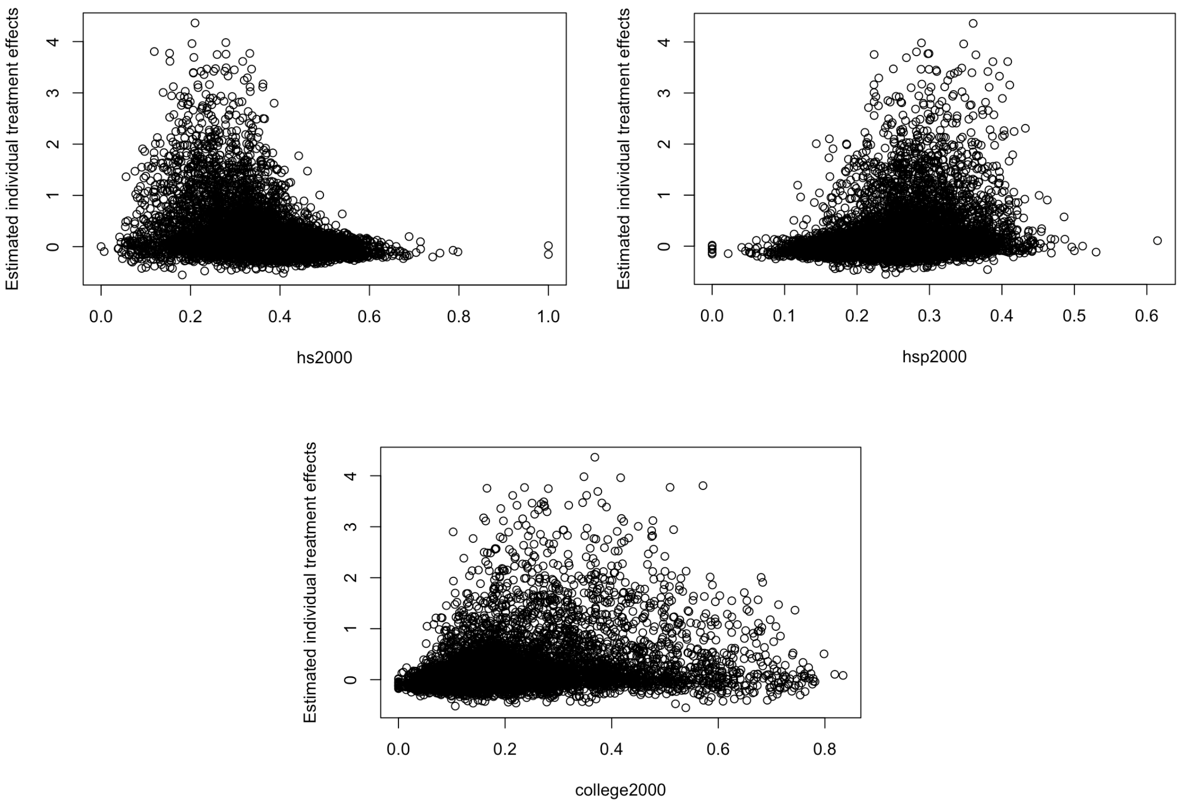


Figure 4: Plots of the estimated individual treatment effects against the three education variables. The estimates are obtained from the causal forest which uses district clustering.

B Robustness results

Table 7: Results of calibration tests for the causal forests clustered by county and by congressional district, based on unweighted observations.

	County		District	
	Mean prediction	Differential prediction	Mean prediction	Differential prediction
Estimate	0.934	0.875	1.02	0.875
p-value	0.131	0.00517**	0.174	0.00180**

Notes: Asterisks denote a significance level of ****0.001, ***0.01, **0.05, and *0.1.

Table 8: Results of t-tests for individual treatment effect estimate subsets $\hat{\tau}_{high}$ and $\hat{\tau}_{low}$ of unweighted observations .

	hs2000		hsp2000		college2000	
	County	District	County	District	County	District
t-value	-32.4	-24.7	21.9	19.1	21.6	16.2
p-value	0.000****	0.000****	0.000****	0.000****	0.000****	0.000****

Notes: Asterisks denote a significance level of ****0.001, ***0.01, **0.05, and *0.1.

Table 9: Results of t-tests on ATE_{high} and ATE_{low} for unweighted observations.

	hs2000		hsp2000		college2000	
	County	District	County	District	County	District
ATE_{high}	-0.000590	-0.000405	0.00467	0.00474	0.00247	0.00259
	(0.00292)	(0.0033)	(0.00307)	(0.00432)	(0.00309)	(0.00350)
ATE_{low}	0.00618	0.00637	0.000626	0.000808	0.00269	0.00291
	(0.00333)	(0.00443)	(0.00326)	(0.00360)	(0.00322)	(0.00427)
t-value	-1.53	-1.22	0.903	0.704	-0.0484	-0.0653
p-value	0.126	0.223	0.367	0.481	0.961	0.948

Notes: ATE_{high} (ATE_{low}) refers to the ATE of towns with higher (lower) values for the education variable than the median. The standard errors are reported in parentheses. Asterisks denote a significance level of ****0.001, ***0.01, **0.05, and *0.1.

Table 10: Results of t-tests for clustered treatment effects for high and low values of education variables, based on unweighted observations.

	hs2000		hsp2000		college2000	
	County	District	County	District	County	District
t-value	-0.316	-0.619	0.697	1.88	-1.33	0.228
p-value	0.752	0.537	0.486	0.0610*	0.183	0.820

Notes: Asterisks denote a significance level of ****0.001, ***0.01, **0.05, and *0.1.