ERASMUS UNIVERSITY ROTTERDAM
Erasmus School of Economics

Master's Thesis Econometrics and Management Science (FEM21031)
Business Analytics & Quantitative Marketing

Estimating Heterogeneous Treatment Effects with Instrumental Forests

The Long-run Effect of Historical Immigration on Voting Behavior and the Economic Assimilation of Immigrants

Author: Nadja van 't Hoff (543182)
Supervisor: dr. A.A. Naghi
Second Assessor: dr. A. Baiardi
Date final version: November 23, 2020

Abstract

It is of importance to understand the long-run effect of immigration for efficient policy making, especially since the recent rise in immigration in the United States. The present study investigates the effect of historical immigration during the Age of Mass Migration on the support of the Republican party and the assimilation of recent foreign borns on the labor market. To identify the causal effect, we construct a "leave-out" version of the shift-share instrument and implement a two-stage least squares regression analysis. Then, we allow for a more flexible specification through instrumental forests. The forests show that the average conditional local average treatment effect of historical immigration is insignificant when nonlinearities in the data are accounted for. Potential effect modifiers are identified in a data-driven fashion. The effect of historical immigration is found to be significant for some subgroups in the sample. The results suggest a persistence of the effect of historical immigration through a two-sided assimilation process as well as through the establishment of migrant communities. Although immigrants appear to assimilate well within those communities, our findings might suggest an increased segmentation between natives and foreign borns. This has important implications for policy making.

Contents

1 Introduction **1**

2 Literature **3**

2.1 Background of historical immigration 4

2.2 Political and economic effects of immigration 4

2.3 Assimilation of recent immigrants 6

3 Data **7**

3.1 Data sources 7

3.2 Outcomes 9

3.3 Average immigrant shares 9

3.4 County controls 1870 10

3.5 Demographics 2016 11

3.6 Geographical controls 11

3.7 Summary statistics 11

4 Methodology **13**

4.1 Specification 13

4.2 Instrument 14

4.3 Forest 16

4.3.1 Advantages of forest-based methods 17

4.3.2 Generalized random forests 17

4.3.3 Instrumental forests 19

5 Results **20**

5.1 OLS estimates 20

5.2 TSLS estimates 21

5.3 Instrumental forest 26

5.3.1 Settings and parameters 26

5.3.2 ACLATEs 30

5.3.3 Detected heterogeneity 32

5.3.4 Republican vote share 33

5.3.5 Wage gap 34

5.3.6 Occupational income score gap 35

5.3.7 Unemployment gap 36

5.3.8 Summary 36

5.4 Robustness checks for the instrumental forest 36

5.4.1 Robustness of the model specification 36
5.4.2 Tuning parameters 38
5.4.3 Violation of the overlap assumption 39
5.4.4 Binary treatment variable 39

6 Conclusion **41**

7 Limitations and further research **42**

8 Appendix **48**
8.1 Variable descriptions and corresponding sources 48
8.2 TSLS checks for the Republican vote share outcome 49
8.3 TSLS checks for the wage gap outcome 50
8.4 TSLS checks for the occupational income score gap outcome 51
8.5 TSLS checks for the unemployment gap outcome 52
8.6 Variable importances for the instrumental forest 53
8.7 ACLATEs without significant difference for the Republican vote share . . 56
8.8 ACLATEs without significant difference for the wage gap 58
8.9 ACLATEs without significant difference for the occupational income score gap 60
8.10 ACLATEs without significant difference for the unemployment gap . . . 62
8.11 Results instrumental forest without tuning 64
8.12 Results causal forest with binary treatment 65
8.13 Results causal forest with overlap-weighting 66
8.14 Results causal forest with continuous treatment 67

1 Introduction

During the past decades, immigration to the United States has increased (Abramitzky and Boustan, 2017). The recent rise in the number of immigrants has increased the interest in the effect of immigration on the society, institution, and economy. In the past, historical immigration has led to natives' backlash which resulted in more restrictive policy making (Tabellini, 2020). In the short-term, this restrictive policy making has been shown to reduce natives' wages and economic productivity (Ager and Hansen, 2017). Though past immigration has been found to positively influence economic outcomes, natives are nowadays again concerned that immigration negatively influences their wages and that foreign borns are unable to assimilate either socially or economically (Abramitzky and Boustan, 2017). While the literature mainly focuses on the short-run effects of immigration (Mayda et al., 2018, 2016; Dustmann et al., 2019; Halla et al., 2012), only little research studies the long-run effects of historical immigration on political, social, and economic outcomes (Sequeira et al., 2020; Giuliano and Tabellini, 2020). However, short-run effects might persist, vanish, or change over time, leading to different effects in the long-run. For instance, Sequeira et al. (2020) find that the economic short-run effects of immigration persist over time, while the social cost seems to disappear. This has important implications for policy making. While the current concerns of natives might lead to more restrictive policy making, potential long-run benefits resulting from higher immigration levels might be neglected by the policy makers. Nevertheless, the importance of considering the long-run effects of immigration for efficient policy making cannot be emphasized enough.

Immigration policy was an important topic during the Trump presidential campaign in 2016, since the position of politicians on the topic of immigration policy can make a difference in their election campaign (Abramitzky and Boustan, 2017). The topic of immigration might affect democratic support which influences the implemented immigration policies that potentially change the institution and political voting behavior in the long-run. In addition, the values of immigrants are likely to be passed on to the natives, transforming the ideology of the country in the long-run (Giuliano and Tabellini, 2020). Therefore, it is of interest to understand the long-run effect of immigration on the political voting behavior as well as the underlying mechanisms.

Next to the voting behavior of the natives, the assimilation process of recently-arrived foreign borns to the native population forms an important factor when considering immigration policies. According to Gathmann and Keller (2018), the receiving country experiences several disadvantages when foreign borns encounter more difficulties to assimilate. If foreign borns are unable to adapt, then the social cohesion and American values might be threatened. They further argue that immigrants' ability to assimilate might have economic consequences. A disadvantaged economic position of immigrants potentially reduces the fiscal benefit of the destination country. Hence, it is of importance to understand the mechanisms underlying the economic assimilation process of immigrants nowadays.

In this paper, we investigate the long-run effect of historical immigration during the Age of Mass migration on the voting behavior during the presidential elections of 2016 and the recent economic assimilation of foreign borns on the labor market in the United States. To investigate the mechanisms underlying the long-run effects, we study the heterogeneity in the treatment variable to identify subgroups in the population. To our knowledge, we are the first to investigate the effect of historical immigration on the economic assimilation of recent foreign borns. For this purpose, we combine publicly-available historical data with recent data from different sources and create a unique data set. This data set combines information on the immigrants entering the United States for each decade since 1870, historical county data as well as recent data on voting behavior, current labor market gaps between natives and foreign borns, recent county characteristics, and demographics of the natives and foreign borns.

For our long-run analysis, we first estimate the average immigrant share for our periods of interest. The period of the Age of Mass Migration is defined from 1870 to 1920, followed by a period of restricted migration from 1930 to 1960, and a subsequent period of higher immigration from 1970 to 2010. We then first estimate the effect of historical immigration on the current support of the Republican party and the labor market gaps between natives and foreign borns using an ordinary least squares (OLS) regression.

The estimation of the long-run effect of historical immigration is challenging. Self-selection of historical immigrants into counties particularly forms a concern. Hence, we construct a version of the shift-share instrument introduced by Card (2001) and perform a two-stage least squares (TSLS) analysis. The instrument is based on the finding by Bartel (1989) that immigrants tend to cluster geographically based on their nationality. Next to the exploitation of the variation resulting from the sending countries, we follow Tabellini (2020) and additionally include variation across the decades of each of the three migration periods. To allow for more exogeneity, we exclude the immigrants that eventually settle in a county to obtain a "leave-out" version of the predicted average immigrant share (Tabellini, 2020; Giuliano and Tabellini, 2020).

Finally, to investigate the underlying mechanisms of the causal effect, we study the heterogeneity in the treatment variable, that is, the heterogeneity of the average immigrant share during the Age of Mass Migration. We do this by implementing an instrumental forest as proposed by Athey et al. (2019). This method naturally divides the data into subgroups by maximization of the heterogeneity in the individual trees of the forest. This offers some advantages compared to a two-stage least squares regression, since subgroups in the population do not have to be defined by the researcher. Instead, subgroups are determined in a data-driven way. Instrumental forests allow for considering each control as a potential effect modifier. This is especially beneficial in case of multiple subgroups, since in this case the estimation by two-stage least squares results in a rather complicated specification. Moreover, instrumental forests automatically account for nonlinearities in the data and are especially useful in case of high-dimensional data.

The TSLS results suggest a significant, negative effect of historical immigration during the Age of Mass Migration on the support of the Republican party as well as a significant decrease in the economic assimilation of foreign borns in terms of wages. However, the instrumental forest does not find these effects to be significant. Our research provides evidence that when allowing for more nonlinearities in the model specification, there is no significant long-run average conditional local average treatment effect (ACLATE) of historical immigration. Instead, the forests allows for identification of subgroups for which the ACLATE is significant. The results provide evidence for two mechanisms that potentially cause the persistence of the effect of historical immigration on voting behavior and the assimilation of immigrants on the labor market. Firstly, the results support the statement of Giuliano and Tabellini (2020) that the assimilation of immigrants is a two-sided process. Thus, the liberal values of the European immigrants were transmitted to the native population, leading to a more liberal population nowadays. Secondly, we find evidence that counties with higher historical immigration have more established migrant communities that offer support on the labor market, supporting the finding of Portes and Zhou (1993). This simultaneously might indicate an increased segmentation between the natives and the foreign borns in counties with high historical immigration.

Finally, this study has some policy indications. Immigrants appear to be able to assimilate on the labor market within the established migrants communities. They might, however, need more support when they have to compete on the labor market with the native labor force. In addition, the segmentation between the natives and the foreign borns has social implications. Policies should lead to a decrease in the segmentation to encourage the "melting-pot" society and allow for the two-sided transmission of values (Giuliano and Tabellini, 2020).

2 Literature

This section first provides a background of historical immigration (Section 2.1). We then describe the political and economic short-run and long-run effects of immigration (Section 2.2). We conclude the literature review with a discussion on the assimilation process of immigrants and potential mechanisms underlying this process (Section 2.3).

2.1 Background of historical immigration

The history and development of the United States has been shaped by its immigrants. The arrival of European immigrants peaked during the Age of Mass Migration between 1850 and 1920 (Abramitzky and Boustan, 2017). This was due to revolutions in shipping technology and an increased network of migrant finance, resulting in lower migration costs. The high number of immigrants led to natives' backlash. Therefore, a literacy test was introduced in 1917, followed by strict immigration quotas in 1921. These were the two main factors that led to the end of the Age of Mass migration (Tabellini, 2020; Abramitzky and Boustan, 2017). A period of restricted immigration followed the Age of Mass Migration until the immigration quotas were adjusted with the Immigration and Nationality Act in 1965. The adjustment brought about a renewed increase in the entry of immigrants during the past decades, enforced by the Immigration Act of 1990 (Abramitzky and Boustan, 2017). However, the composition of sending countries differed with the majority of the immigrants now coming from Latin America and Asia.

2.2 Political and economic effects of immigration

The wave of immigrants between 1850 and 1920 led to different short-run effects during the Age of Mass Migration, and influenced political decisions amongst others. Tabellini (2020) finds that immigration during the Age of Mass migration led to natives' backlash and hostile political reactions, leading to more conservative representatives and legislation hostile to immigrants. This backlash found its roots in social reasons, since the immigrants did not negatively impact the county's economy. On the contrary, immigrants delivered workforce and provided important skills for industry and agriculture (Tabellini, 2020; Sequeira et al., 2020). The high immigration rates increased the industrial and agricultural production and hence increased the employment of the native population at the beginning of the twentieth century. Sequeira et al. (2020) finds that the latter even holds when taking into account the heterogeneity of different skill groups.

Interestingly, the recent short-run effect of immigration in the United States is found to differ substantially from the short-run effect of immigraiton in the past as well as from the recent effect of immigration estimated for other Western countries. While in Europe, more immigration generally implies more right-wing voting (Halla et al., 2012; Dustmann et al., 2019), in the United States the recent immigration is often found to decrease the Republican vote share (Mayda et al., 2018, 2016). Mayda et al. (2018) argues that this is mainly due to the indirect effect of a preference shift of the native voters instead of the direct effect through the naturalization of immigrants, since naturalized immigrants are less likely to vote.

While literature exists on the past and current short-run effects of immigration in the United States, limited research describes the long-run effect of immigration on political and economic outcomes. Though natives might react positively or negatively to high immigration levels in the short-run, these effects might vanish or flip sign over time. For instance, Tabellini (2020) finds that the social cost of historical immigration did not persist over time. Sequeira et al. (2020) confirms that past immigration does not influence social cohesion nowadays. So though the social cost of immigration might not persist and influence voting behavior nowadays, the transmission of liberal views might have shaped political decision making.

Next to cultural values that are transferred from immigrants to natives, positive economic effects appear to persist over time. In Brasil, high-skilled immigrants shifted the economy to the skill-intensive sector at the beginning of the 20th century, leading to higher levels of current income and schooling in regions with high historical immigration (Rocha et al., 2017). In Argentina, the increase in human capital brought by the European immigrants in 1914 led to counties with a higher GDP in 1994 (Droller, 2018). Sequeira et al. (2020) study the economic long-run effects of historical migration in the United States and find evidence that the short-run benefits of high historical immigration have brought about an increase in current income as well as education, a decrease in poverty and unemployment, and the development of urban areas. Immigration during the Age of Mass migration might even be the most important factor that has led to lasting positive effects amongst the county characteristics of that period (Rodríguez-Pose and Von Berlepsch, 2014).

The positive long-run effects of historical immigration during the Age of Mass Migration might have persisted through different channels. Two main channels are the political and the institutional channels (Acemoglu et al., 2005). However, culture and human capital are equally potential channels that might lead to the persistence of effects. Tabellini (2010) finds that past cultural institutions influence the current economic development in Europe. These institutions might also be likely to influence political decisions. Moreover, Giuliano and Tabellini (2020) argue that the cultural values of past immigrants are probably passed on to the natives, transforming the American ideology to a more liberal ideology in the long-run.

Summarizing, the short-run effects of immigration are different in the past and present, and might vanish or change over time. The different response to the immigrants in the United States compared to European countries is rather puzzling. Therefore, it would be of interest to understand the underlying mechanisms of the current voting behavior. For instance, some short-run effects of the past might have persisted through one of the potential channels, shaping the current economy and politics of the United States.

2.3 Assimilation of recent immigrants

A wide-spread concern of natives remains that immigrants are unable to assimilate. However, this concern has not proven to be true historically. The straight-line assimilation model states that ethnic differences between migrants and natives vanished over time (Alba and Nee, 1997). This has likely been a two-sided process where immigrants assimilated to the natives, and the natives equally adapted values of the immigrants (Tabellini, 2020). Though historical immigrants appear to have assimilated, recent foreign borns struggle to assimilate due to changes in labor market conditions as well as changes in the sending countries (Abramitzky and Boustan, 2017; Portes and Zhou, 1993). There might even be downward assimilation for some of the immigrant groups (Portes and Zhou, 1993).

The economic assimilation of foreign borns is very important, especially since income has been found to affect multiple outcomes for the subsequent generations, suggesting that the ability to assimilate might have severe consequences in the long-run (Maurin, 2002; Akee et al., 2018; Duncan et al., 2014). While immigrants in the past held similar occupations and were able to assimilate fast in wages, the mean wages of immigrants compared to the mean wages of the native population have declined during the past thirty years (Abramitzky and Boustan, 2017; Butcher and DiNardo, 2002). Borjas (2015) does not find any economic assimilation for the immigrants who arrived in the 1990s. They argue that this is mainly due to the decline in skills of the recent immigrants. Although this finding is supported by some literature, the decline in human capital brought to the United States by immigrants does not completely explain the difference in labor market outcomes between natives and foreign borns. For instance, Villarreal and Tamborini (2018) find that differences in earnings for Hispanic immigrants cannot completely be explained by differences in human capital. Grand and Szulkin (2002) find that in Sweden, immigrants from Asia and from Latin America equally struggle to assimilate on the labor market. They suspect that this is due to labor market discrimination next to lower labour quality. This is supported by Smith and Fernandez (2015), who argue that the wage gap for natives and immigrants who possess the same set of skills is due to discrimination.

Though ethnic differences might increase discrimination, the differences in earnings can also be a result from the experience of immigrants upon arrival (Villarreal and Tamborini, 2018). The experience of immigrants arriving in gateways with more historical immigration might differ from the experience of immigrants arriving in gateways with little historical immigration (Zúñiga and Hernández-León, 2005). In gateways with less historical immigration, foreign borns might have more freedom in defining their position whereas their position is more fixed in well-defined gateways. In gateways with high historical immigration, there might be more institutional arrangements established for new immigrants. On the one hand, this might increase social support if there is an immigrant community, encouraging economic assimilation (Portes and Zhou, 1993). On the other hand, ethnic communities that support subsequent generations might hinder assimilation on the labor market, as they might lead to an increase in segmentation between the foreign borns and natives in the population (Alba and Nee, 1997). Hence, in addition to discrimination, historical immigration might affect the ability of current immigrants to assimilate economically. Since especially labor migrants settle in established communities, the assimilation process is slower for those migrants as opposed to human capital migrants (Alba and Nee, 1997). The labor immigrants are also less likely to invest in human capital skills. Borjas (2015) finds that there is a general decline in the take-up speed of the English language skills. When immigrants settle in established communities and find a job within those communities, the investment in skills such as English language proficiency is less profitable.

Altogether, foreign borns nowadays encounter difficulties to assimilate due to lower skill levels and discrimination. Furthermore, historical immigration might have led to established communities which enable foreign borns to assimilate well within those communities while the assimilation outside of those communities remains complicated. Within the communities, the potential decrease in human capital and English proficiency of recent foreign borns does not affect their assimilation on the job market, whereas human capital migrants who often settle outside of those communities might find themselves in a disadvantaged position for instance due to discrimination when competing on the same job markets as the natives.

3 Data

We now turn to the data sources and the aggregation process used to construct a unique data set for the estimation of the long-run effect of historical immigration (Section 3.1). In Sections 3.2 to 3.6, we describe the different variables contained in the data set. Finally, we conclude this section with the summary statistics of our data set (Section 3.7).

3.1 Data sources

To estimate the long-run effect of the historical immigration during the Age of Mass Migration on electoral and economic assimilation outcomes, we create a unique data set by combining several data sets from public sources. We obtain county-level data for the Republican vote share of the 2016 presidential elections from MIT Election Data and Science Lab (2018). For each of the four Virginia counties Fairfax, Franklin, Richmond, and Roanoke, two observations with different information codes are reported. We sum the values of the observations on these four counties to obtain data at the county level.

For the study of the economic assimilation of the foreign borns, we include three different labor market outcomes consisting of the gap between natives and foreign borns in wage, the occupational income score, and the unemployment. Data for the wage gap outcome, the occupational income score gap outcome, and the unemployment gap outcome are obtained from the American Community Survey 2016 sample of the Integrated Public Use Microdata Series (IPUMS) which consists of a 1-in-100 national random sample of the U.S. population [1]. From the same sample, we retrieve the characteristics of the natives and foreign borns residing in a county in the year 2016 as well as the total number of years that a foreign-born has been living in the United States. To investigate the economic assimilation of recent immigrants, the sample of the foreign borns is restricted to the persons who arrived to the United States between 1970 and 2016 and are still residing there in the year 2016. Since we study the assimilation on the labor market, we restrict the sample to observations on the natives and foreign borns between the ages of 25 and 60. The data for the mentioned outcomes and controls are only available at an individual level. Therefore, we aggregate the data at the county-level by averaging the observations over all available persons for each county.

County-level controls for the year 1870 are mainly obtained from the 1870 census data of the National Historical Geographic Information System (NHGIS) of IPUMS (Manson et al., 2019). The fraction of married persons in 1870 is obtained from the vital statistics data of NHGIS, and information on the railway connection is obtained from the 1860 census data. The county-level controls for the year 2016 are obtained from the 2016 American Community Survey based on 5-year data (2012-2016). The fraction of married persons in 2016 is obtained from the 2016 American Community Survey that is based on 1-year data. The present urban rate is obtained from the 2010 census data, as this data is not available in the 2016 community surveys. NHGIS further provides census data on the number of immigrants entering the United States at the beginning of each decade and their corresponding sending countries. Finally, geographical data on the longitude and latitude values of the center of the countries in the year 2000 is retrieved from the United States Census Bureau[2].

The data obtained from NHGIS is matched with the IPUMS data set that contains the assimilation outcomes and demographic controls for the year 2016. The GISJOIN identifier is used to link the county-level observations of these two data sets. We then match the resulting data set with the geographical data based on the same identifier. The GISJOIN identifier is easily constructed for the geographical data by combining the two-character state Federal Information Processing Standard (FIPS) code with the three-character county FIPS code. Finally, we merge this combined data set with the data on the presidential elections by matching the state and county names. Altogether, the matches of the data sets reduce the sample for the Republican vote share outcome to 1854

[1] This data is available at https://usa.ipums.org.

[2] This data is available at https://data.census.gov.

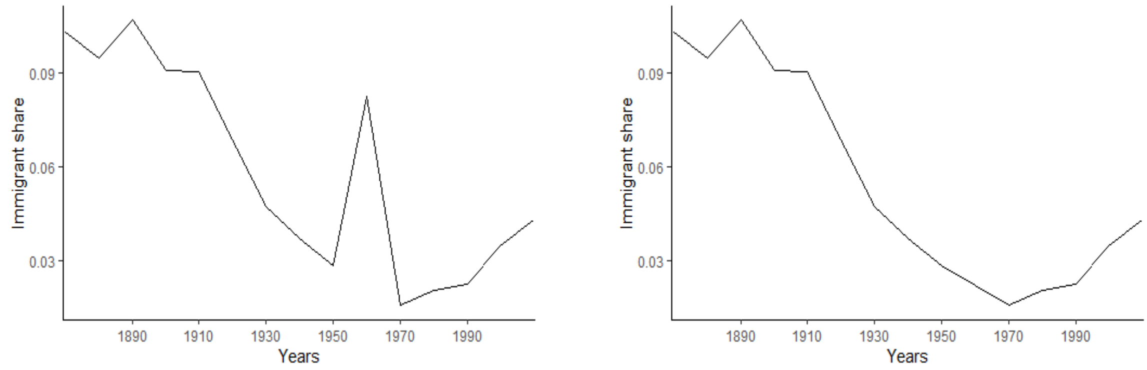
complete observations, for the wage gap outcome to 1856 complete observations, for the occupational income score gap to 1856 complete observations, and for the unemployment gap to 1854 complete observations. An overview of the controls and their corresponding sources can be found in Table 12 of Appendix 8.1.

3.2 Outcomes

We estimate the long-run effect of the historical immigration on four different outcomes. The outcomes of interest are the Republican vote share, the wage gap between natives and foreign borns, the occupational income score gap between natives and foreign borns, and the unemployment gap between natives and foreign borns for the year 2016. The Republican vote share variable consists of the candidate votes for the candidate of the Republican party divided by the total votes. For each county, the outcome for the wage gap is constructed by subtracting the logarithm of the average yearly wage per capita of a county in 1000 U.S. dollars of the foreign borns from the logarithm of the average wage in 1000 U.S. dollars from the natives: log(average wage of the natives) - log(average wage of the foreign borns). In a similar fashion, we construct the outcome for the occupational income score gap. This outcome is constructed by subtracting the logarithm of the average occupational income score of the foreign borns from the logarithm of the average score of the natives. Individuals who do not have a wage income or do not have an occupation are excluded from the data set. For these two latter outcomes, it is not possible to take the logarithm of the difference, as the subtraction of the average outcome of the foreign borns from the average outcome of the natives might lead to negative values for counties where the averages are higher for the immigrants. Finally, we construct the unemployment gap outcome by subtracting the average unemployment rate of the foreign borns from the average unemployment rate of the natives.

3.3 Average immigrant shares

The U.S. history has been shaped by three different migration waves since the mid-19th century. Our period of interest is the historical immigration during the Age of Mass Migration. Following Ager and Brückner (2013), we focus on the period from 1870 to 1920 for this first migration wave. They state three reasons for the definition of this period. Firstly, the American Civil War from 1861 to 1865, which forms an unusually large shock, is excluded. Secondly, there is more data of the U.S. Census available from 1870 onward. Data on the sending countries of the immigrants was collected consistently by the U.S. Census on the county level starting in 1870. The availability of this data is of importance for the construction of our instrument. Finally, the Emergency Quota Act was released in 1921 and hence can be seen as the starting point of the decrease in the



(a) Average immigrant shares including the 1960 data

(b) Average immigrant shares excluding the 1960 data

Figure 1: Average immigrant shares at the beginning of each decade

entry of immigrants.

The period of high immigration rates was followed by a period with low immigration from 1930 to 1960. A large amount of the values on the immigrants are missing for the year 1960 and in addition, the values that are reported seem to be very large, indicating a measurement error. Figure 1a confirms that the data for the year 1960 is likely to contain measurement errors. Figure 1b excludes this decade and shows high immigrant shares for the Age of Mass Migration, a decrease in the share of immigrants in the subsequent decades until the immigrant shares are increasing again from 1970 onwards. The third period of increased recent migration is defined from 1970 to 2010.

For the first and the third migration period, we construct the average immigrant share by dividing the total amount of immigrants in a county at the beginning of a decade by the total population in that county at that time. Then, the average immigrant share for a period is obtained by averaging the immigrant shares over the decades of the specified periods. Data are not available for all decades within a period for some of the counties. In these cases, the average immigrant share is constructed based on the available decades.

3.4 County controls 1870

To control for the possible correlation between the settlement of immigrants with covariates that led to short-run effects which potentially influenced the outcome variables in the long-run, we include county controls for the beginning of the period of Historical Mass Migration that are similar to the ones included by DellaVigna and Kaplan (2007). Proxies are included for some of the controls, since these data are not available for the year 1870. We include the share of males in the total population, the share of Hispanics and Latinos, the share of African Americans, the illiteracy rate as a proxy for educational attainment, the urban rate, the average of the county's manufacturing wages, the average of the county's agricultural wages, and the average personal estate as a proxy for the personal income as well as the number of persons employed in manufacturing as a proxy for the employment rate. In addition to these variables, we control for whether the county had a railway connection in 1860, since counties experienced a higher inflow of immigrants when they provided a railway network according to Sequeira et al. (2020). We include the 1860 control for the railway connection since these data are not available for the year 1870.

3.5 Demographics 2016

We include data on the characteristics of the natives and foreign borns for each county in the year 2016. This includes the share of male natives, the share of male foreign borns, the average age and its squared term of the natives, the age and its squared term of the foreign borns, the share of married natives, the share of married foreign borns, the average years of schooling of natives as well as the average years of schooling of the foreign borns, the share of natives with a college degree or higher, the share of foreign borns with a college degree or higher, and the share of natives and foreign borns who do not speak English respectively.

3.6 Geographical controls

We follow Rocha et al. (2017) and add data on the longitude and latitude values of the centers of the counties as measured in the year 2000. These controls are necessary since immigrants were likely to settle in the North-East and Mid-West Abramitzky and Boustan (2017). The data set additionally includes current county controls consisting of the share of urban population in the year 2010, as data on the urban population for the year 2016 are not available.

3.7 Summary statistics

The summary statistics in Table 1 show that natives earn more than foreign borns and hold higher-earning occupations on average. However, this varies to a great extent amongst the counties. In some counties the foreign borns appear to outperform the natives on the job market. Interestingly, the wage and occupational income score are more dispersed for the foreign borns than for the natives. This might be due to the fact that we only include registered immigrants in our sample, excluding illegal immigrants. The registered immigrants might already have found a job in the United States before moving. If we were to include undocumented immigrants as well, we would likely find larger differences in the gaps between foreign borns and natives. However, these data are not available.

Table 1: Summary statistics

Variable	Mean	Std. Dev.	Min.	Max.	N
Outcome variables					
repvoteshare2016	0.630	0.159	0.008	0.960	3156
wagenatives2016	57116	12021	1800	398000	3314
wagefb2016	53026	25211	4	583000	3120
wagegap2016	4.056	25.381	-524.65	126.533	3119
occscorenatives2016	29.125	1.653	11	52	3315
occscorefb2016	27.862	3.997	9	80	3139
occscoregap2016	1.297	4.138	-47.269	22.5	3139
unempnatives2016	0.047	0.034	0	1	3315
unempfb2016	0.042	0.063	0	1	3150
unempgap2016	0.005	0.066	-0.98	0.333	3150
Immigrant shares					
avgimmshare1870_1920	0.083	0.104	0	0.535	2155
avgimmshare1930_1950	0.034	0.046	0	0.288	2155
avgimmshare1970_2010	0.025	0.034	0	0.359	2155
County controls 1870					
male1870	0.521	0.055	0.443	1	2132
hispaniclatinos1870	0.006	0.025	0	0.333	2152
africanamericans1870	0.158	0.218	0	0.928	2131
illiteracy1870	0.198	0.161	0	0.849	2131
urban1870	0.0005	0.003	0	0.095	2132
manufacturingwages1870	23.711	56.346	0	853	2129
agriculturalwages1870	145421	197396	0	1979768	2112
manufacturingemployed1870	956.825	5236.549	0	137496	2133
personalestate1870	199.628	196.624	0	3510	2152
railway1860	0.385	0.487	0	1	1888
Demographics 2016					
malenatives2016	0.511	0.059	0	1	3314
malefb2016	0.54	0.172	0	1	3120
agenatives2016	42.77	1.854	26	60	3314
agefb2016	43.081	3.619	26	60	3120
marriednatives2016	0.576	0.084	0	1	3314
marriedfb2016	0.658	0.178	0	1	3120
educnatives2016	14.880	0.450	10.333	20	3315
educfb2016	13.921	1.636	0	22	3155
highdegreenatives2016	0.453	0.083	0	1	3315
highdegreefb2016	0.430	0.171	0	1	3155
noenglishnatives2016	0	0.006	0	0.333	3314
noenglishfb2016	0.049	0.079	0	1	3120
Geographical controls					
longitude2000	1837.765	728.045	3	3225	2155
latitude2000	1589.19	833.782	97	3150	2155
urban2010	0.428	0.296	0	1	2155

The share of immigrants from 1930 to 1950 is higher than the share from 1970 to 2010. This is not surprising, since the definition of an immigrant is a foreign-born person and the high levels between 1930 and 1950 include immigrants who moved to the United States during the Age of Mass Migration. The relatively low average immigrant share from 1970 to 2010 might be explained by the fact that our dataset only contains registered foreign borns and does not contain illegal immigrants.

4 Methodology

In this section we will describe our empirical strategy. First, we introduce our baseline estimating equation (Section 4.1). Then, we will describe the construction of a "leave-out" version of the shift-share instrument for the immigration periods of our interest (Section 4.2). In Section 4.3, information will be provided on generalized random forests, their advantages, and on the necessary adjustments of the generalized method in order to obtain an instrumental forest.

4.1 Specification

To investigate the long-term effect of historical immigration on the Republican vote share and the assimilation gap outcomes, we estimate a specification as follows:

$$y_{is} = \alpha_s + \theta \text{ avgimmshare1870\_1920}_{is} + \beta X_{is} + \varepsilon_{is} \tag{1}$$

where y_{is} is the outcome in county i in state s in the year 2016. We include state fixed effects α_s, since persistent economic, cultural and institutional county characteristics might affect the outcomes and the covariates over time. We further control for X_{is} which includes the 1870 county characteristics, the 2016 demographics of the natives as well as those of the recent immigrants, and the geographical controls. It further includes the average immigrant share between 1970 and 2010. The standard errors are clustered at the state level because of two reasons that both hold in our application (Abadie et al., 2017). Firstly, we assume heterogeneity in the treatment variables. Secondly, we include data that was gathered at an individual level and hence does not represent the total population. We follow DellaVigna and Kaplan (2007) and weight the data by total votes for the Republican vote share outcome, because the precision of the dependent variable is increasing by the number of votes cast. Similarly, we weight the observations for the assimilation on the labor market by the persons in the sample, following Card (2001) and Dustmann et al. (2019). The treatment effect and all controls are standardized by subtracting their respective mean and dividing by their standard deviation. This standardization facilitates the interpretation of the coefficients.

4.2 Instrument

A common problem with observational studies is confounding with observed and unobserved factors. The assumption of no confounding often is unlikely to hold due to selection into treatment. In our setting, immigrants may sort into counties where they want to work and live based on demand-pull factors leading to positive selection. Conversely, immigrants might sort into counties based on costs and discrimination leading to negative selection (Giuliano and Tabellini, 2020). Reverse causality leads to endogenous regressors. When the variable of interest and the error term are mutually correlated and the exogeneity assumption does not hold, ordinary least squares regressions are no longer consistent. In this case, the effect of the variable of interest cannot be distinguished.

A two-stage least squares regression with instrumental variables (IV) solves this reverse causality and the selection problem by exploiting exogenous variation in the treatment variable. This substitutes the random assignment of the treatment, allowing for identification of the causal effect. More specifically, this allows for estimation of the local average treatment effect (LATE), that is, the average treatment effect in the group of compliers.

To obtain an unbiased estimate, the choice of the instrument is important since it defines the causal effect that is estimated (Heckman and Vytlacil, 2005). A widely used instrument in immigration economics is the shift-share instrument (Ager and Brückner (2013); Giuliano and Tabellini (2020); Mayda et al. (2018); Tabellini (2020)). This instrument is introduced by Card (2001) and originates in the finding by Bartel (1989) that immigrants are geographically concentrated based on their nationality. The instrument exploits the cross-sectional variation coming from the geographical distribution of the immigrants entering the United States based on their network. Therefore, it is exogenous to the county-specific demand shocks of each county at that time. Next to exploiting cross-sectional variation, we follow Tabellini (2020) and construct the instrument such that it additionally allows for the time-series variation across decades that is induced by, for instance, World War I and the Immigration Acts of the 1920s, lowering the serial correlation. This boils down to predicting the average number of immigrants entering the United States for each decade separately instead of predicting the average number of immigrants over the entire period. Finally, to add more exogeneity to the instrument, we construct a "leave-out" version of the shift-share instrument (Tabellini, 2020; Giuliano and Tabellini, 2020). This means that for each county, we exclude the immigrants that eventually settle in that county.

To construct the instrument, we multiply the fraction of immigrants from a particular sending country at the beginning of a decade by the total number of immigrants from that country who enter the United States during the decade. Then, we sum over all sending countries to obtain the predicted number of immigrants for a specific decade on the county level:

$$\hat{I}_{cd} = \sum_{g} \lambda_{gcd} \cdot M_{gd} \tag{2}$$

with M_g the immigrants from country g who enter the United States between decade $(d-1)$ and decade d, and λ_{gc} the fraction of immigrants from country g in county c at the beginning of decade d, excluding the immigrants that actually move to county c. We then average over the corresponding decades for each of the two time periods of interest to obtain the predicted average number of immigrants who enter the United States during that time period. Finally, we scale the instrument by the total population at the beginning of the migration periods. Thus, the instrument is scaled by the baseline population in the years 1870, 1930, and 1970 respectively, to obtain the predicted average immigrant share.

The countries that are used to construct the instruments are selected based on the data availability for each period. This results in 14 different countries for the first migration period and in 20 different countries for the third period. The countries used for each period and their initial fractions of immigrants for the years 1870, 1930, 1970, and 2010 are given in Table 2. The Table shows that in 1870, the majority of the immigrants came from Europe, while the majority of immigrants in 2010 are Mexican. The table clearly shows that the distribution of the sending countries changes between 1870 and 1930. The change in the distribution of the sending countries also holds for the period of restricted migration and the period of recent migration. This is the cross-sectional variation we exploit for our instruments.

There are several assumptions that the instrument needs to meet to be valid. First, the independence assumption needs to be met. This assumption states that the treatment offer is not selective. In our setting, the independence assumption is very likely to be met since the country of origin is comparable to a lottery and therefore randomly assigned. Secondly, the instruments should only affect the dependent variable through the treatment variable. This assumption is referred to as the exclusion assumption. The exclusion assumption is likely to be met since the network patterns only influence the voting behavior and economic assimilation of recent foreign borns through the immigrant share. The third assumption is the first stage assumption which states that the instrument should have a meaningful effect on the treatment variable. The first stage assumption is likely to hold since networks influence each other. When these assumptions are not met, IV estimation may lead to more biased results compared to the results of an ordinary least squares regression (Heckman and Vytlacil, 2005).

Table 2: Fraction of immigrants per sending county living in the United States in 1870, 1930, 1970, and 2010

Country	Fraction in 1870	Fraction in 1930	Fraction in 1970	Fraction in 2010
All countries	100.00%	100.00%	100.00%	100%
Atlantic Islands	0.06%			
Austria	0.41%	3.55%	2.89%	0.27%
Canada		12.22%	10.95%	4.26%
China			2.32%	10.56%
Cuba			5.92%	5.17%
Czechoslovakia		4.69%	2.17%	0.37%
Denmark	0.57%	1.71%		
Finland		1.33%		
France	2.85%	1.28%	1.42%	0.79%
Germany	41.39%	15.39%	11.23%	3.24%
Great Britain	0.10%	11.67%		
Greece		1.56%	2.39%	0.74%
Hungary		2.56%	2.47%	0.40%
Ireland	45.42%	7.12%	3.39%	0.66%
Italy	0.18%		13.60%	1.98%
Japan			1.62%	1.80%
Lithuania		1.74%		
Mexico	0.90%	0.14%	10.24%	59.65%
Netherlands	1.06%	1.24%	1.49%	0.46%
Northern Ireland		1.71%		
Poland		12.12%	7.39%	2.39%
Portugal	0.02%		1.23%	0.97%
Romania		1.28%		
Russia		11.03%	6.25%	1.99%
Spain	0.03%			
Sweden		5.70%	1.71%	0.24%
Sweden and Norway	5.18%			
Switzerland	1.84%			
United Kingdom			9.25%	3.53%
Yugoslavia		1.99%	2.07%	0.52%

4.3 Forest

In this section, the benefits of forests for causal effect estimation are stated (Section 4.3.1), then we introduce the generalized random forest (Section 4.3.2), and finally we specify how to adjust the generalized random forest for treatment effect estimation with instrumental variables (Section 4.3.3).

4.3.1 Advantages of forest-based methods

We now turn to the instrumental forest which is based on the generalized random forest. This method is proposed by Athey et al. (2019) and is based on the random forest as introduced by Breiman (2001). Instrumental forests offer two major advantages compared to a two-stage least squares regression. First, the estimates of econometric methods such as two-stage least squares can easily be biased if the model is not correctly specified. Instrumental forests are a data-driven method and give nearby observations more weight. This is especially useful in case of high-dimensional data when the data contains irrelevant covariates or when there are complex interactions between the covariates (Wager and Athey, 2018). For instance, Mayda et al. (2016) suggest that the effect of immigrants may be non-linear. Instrumental forests automatically account for such nonlinearities without having to consider this for each variable, therefore decreasing the risk of introducing bias into the model. Secondly, the effect of immigration is often found to be influenced by effect modifiers (Dustmann et al. (2019); Halla et al. (2017); Mayda et al. (2018)). Determining the correct subgroups that influence the treatment effect ex ante becomes infeasible, especially in the case of high-dimensional data. Selecting subgroups of covariates ex post neither offers a solution, since this may lead to overfitting. Moreover, the specification in case of two-stage least squares becomes complicated when the treatment effect is influenced by multiple effect modifiers. Instrumental forests naturally divide the data into subgroups with different estimates conditional on the covariates and hence allow for the estimation of the conditional local average treatment effects (CLATEs).

4.3.2 Generalized random forests

The generalized random forest is a method proposed by Athey et al. (2019) that allows for the estimation of the parameter of interest $\theta(x)$ using local moment conditions. Define $Y_i \in \mathbb{R}$ as the dependent variable and $W_i \in \{0,1\}$ as the treatment variable. Given the data $(X_i, O_i) \in \mathcal{X} \times \mathcal{O}$ with $O_i = \{Y_i, W_i\}$, $\hat{\theta}(x)$ can be obtained by solving the moment equation

$$\mathbb{E}[\phi_{\theta_x,\nu_x}(O_i)|X_i = x] = 0 \quad \text{for all} \quad x \in \mathcal{X} \tag{3}$$

where $\phi(\cdot)$ is a chosen scoring function and $\nu(x)$ a nuisance parameter.

The fundamental problem of causal inference is that the outcome for an observation can only be observed in case of treatment or no treatment, but never for both cases simultaneously. Generalized random forests circumvent this fundamental problem by using an adaptive weighting function that is as sensitive as possible to heterogeneity. Moreover, treating forests as a type of adaptive locally weighted estimators alleviates the bias caused by noisy solutions to equation 3. Similarity weights $\alpha_i(x)$ that capture the importance of the i-th training example when estimating $\theta(\cdot)$ for a specific x are used to estimate θ_x. These weights are obtained by first growing a forest that returns a weighted set of neighbors for each observation x. For high-dimensional data, weights that account for the frequency that the i^{th} training example shares the same leaf as x are most appropriate. When B trees are grown with indices $b = 1, ..., B$, define $L_b(x)$ as the set of training examples that share the same leaf as x. Then, the weights are defined as

$$\alpha_{bi} = \frac{\mathbf{1}(\{X_i \in L_b(x)\})}{|L_b(x)|}, \qquad \alpha_i(x) = \frac{1}{B}\sum_{b=1}^{B} \alpha_{bi}(x). \tag{4}$$

These weights can subsequently be used in the empirical estimating equation

$$\left(\hat{\theta}(x), \hat{\nu}(x)\right) \in \text{argmin}_{\theta,\nu} \left\{ \left\| \sum_{i=1}^{n} \alpha_i(x)\phi_{\theta,\nu}(O_i) \right\|_2 \right\}. \tag{5}$$

To allow for faster computation of the causal effect, generalized random forests use a gradient-based approximation to solving equation 5. Treating the generalized random forest as an adaptive nearest neighbor estimator further allows for statistical extensions (Athey et al., 2019; Wager and Athey, 2018). There are two reasons why this is especially important in the setting of causal inference. First, confidence intervals are necessary for decision making of policy makers. Secondly, the model's performance in estimating the treatment effect is difficult to evaluate through other methods such as for example cross-validation, since only one of the potential outcomes is observed in practice. This impedes the use of a test set for evaluation. Causal forests offer a major advantage as opposed to other machine learning methods, since asymptotic consistency, Gaussianity of the estimates of $\theta(x)$, and asymptotic confidence intervals are established (Athey et al., 2019).

The statistical behavior is ensured by training each tree of the forest on a subsample of the data and by introducing honesty in the tree. Honesty means that the subsample used for each tree is further divided into two samples. Then, one sample is used for the tree construction and the other half of the subsample is used to fill the leaves of the tree. Next to ensuring good statistical behavior, honesty reduces bias. Honesty is arguably inefficient in the case only one tree is constructed and used for estimation of the causal effect. However, when establishing a forest consisting of a large number of trees, eventually all observations will be used. This should linger the concerns about potential inefficiency induced by honesty.

Generalized random forests are similar to Breiman's forest. Firstly, they include subsampling where each tree is grown on a different, randomly selected subset of the full data set. Secondly, it uses greedy recursive partitioning, i.e. each split is chosen such that the model fit is instantaneously maximized. Recursive partitioning stops when a leaf node contains the specified minimum number of data points. Thirdly, random split selection restricts the number of variables considered at each split. To construct a tree, the parent node $P \subseteq \mathcal{X}$ is split into two non-overlapping child nodes $C_1, C_2 \subseteq \mathcal{X}$ such that heterogeneity of the parameter of interest is maximized. Subsampling and random split selection contribute to randomization of the trees.

The random forest needs to be adapted slightly in case of clustered data (Athey and Wager, 2019). Instead of drawing a subsample of observations, a subsample of clusters $\mathcal{J}_b \subseteq 1, ..., J$ is drawn. Then, k observations are randomly selected from each cluster $j \in \mathcal{J}_b$ of the subsample for the construction of the individual trees. Clustering further influences the out-of-bag predictions in that an observation i is only considered out-of-bag when its cluster was not drawn for the tree at hand, that is, if $A_i \not\subset \mathcal{J}_b$. This approach of clustering compares to estimating a functional random effects model of the form

$$Y_i = m_{A_i}(X_i) + W_{i\theta_{A_i}} + \varepsilon_i, \qquad \theta(x) = \mathbb{E}[\theta_j(x)].$$

Therefore, each cluster has its own treatment effect function next to the main treatment effect function.

4.3.3 Instrumental forests

We can adjust the generalized random forest such that it allows for an instrumental variable regression. The resulting forest is called instrumental forest. In our setting, we suppose that the outcomes Y_i and the treatment W_i are related through the equation

$$Y_i = \mu(X_i) + \theta(X_i)W_i + \varepsilon_i, \tag{6}$$

with $\theta(X_i)$ the treatment effect of W_i on Y_i, $\mu(x)$ is the intercept and nuisance parameter, and ε_i an error term that may be correlated with the treatment indicator. We introduce an instrument $Z_i \in \{0, 1\}$. Analogous to two-stage least squares regression, we can then solve the following moment equations

$$\begin{aligned} \mathbb{E}[Z_i(Y_i - W_i\theta(x) - \mu(x))|X_i = x] = 0 \\ \mathbb{E}[Y_i - W_i\theta(x) - \mu(x)|X_i = x] = 0 \end{aligned} \tag{7}$$

to obtain the average treatment effect

$$\theta(x) = \frac{Cov[Y_i, Z_i|X_i = x]}{Cov[W_i, Z_i|X_i = x]}. \tag{8}$$

If we assume heterogeneity and want to obtain the conditional local average treatment effect, then the splitting rule for maximizing heterogeneity when splitting a parent node $P \subseteq \mathcal{X}$ into the two child nodes $C_1, C_2 \subseteq \mathcal{X}$ also needs to be adjusted. A detailed description of this adjustment can be found in Athey et al. (2019). The estimated individual causal effects $\hat{\theta}(x)$ are subsequently obtained by the gradient-based approximation to solving equations 7 with weights 4. As proposed by Athey and Wager (2019), the average conditional local average treatment effect can then be estimated by a doubly robust estimator based on Chernozhukov et al. (2016) which is similar to the variant of augmented inverse-propensity weighting by Robins et al. (1994):

$$\hat{\theta}_j = \frac{1}{n_j} \sum_{\{i:A_i=j\}} \hat{\Gamma}_i, \qquad \hat{\theta} = \frac{1}{J} \sum_{j=1}^{J} \hat{\theta}_j,$$
$$\hat{\Gamma}_i = \hat{\theta}^{(-i)}(X_i) + \frac{c * (Z_i - \hat{o}^{(-i)}(X_i))}{\hat{o}^{(-i)}(X_i)(1 - \hat{o}^{(-i)}(X_i))} \left(Y_i - \hat{m}^{(-i)}(X_i) - \left(W_i - \hat{e}^{(-i)}(X_i)\right) \hat{\theta}^{(-i)}(X_i)\right) \tag{9}$$

where $\hat{e}^{(-i)}(X_i)$, $\hat{m}^{(-i)}(X_i)$, and $\hat{o}^{(-i)}(X_i)$ are the out-of-bag estimates of the expected outcome $e(x) = \mathbb{P}[W_i|X_i = x]$, the treatment propensities $m(x) = \mathbb{E}[Y_i|X_i = x]$, and the instrument propensities $o(x) = \mathbb{P}[Z_i|X_i = x]$ respectively. "Out-of-bag" means that for instance Y_i is not used in the computation of $\hat{m}^{(-i)}$. The compliance scores c are predicted by a causal forest with the treatment W as the dependent variable and the instrument Z as the independent variable. When the unconfoundedness assumption is met, the doubly robust estimator gives the conditional treatment effect in the group of compliers.

5 Results

We now turn to our results. In Section 5.1, we discuss the results from the ordinary least squares analysis, followed by a discussion of the two-stage least squares estimates in Section 5.2. Next, we turn to our instrumental forests, describing first the settings and then the resulting estimates and detected subgroups (Section 5.3). Section 5.4 concludes with several robustness checks for the forests.

5.1 OLS estimates

We now turn to the OLS estimation of the baseline specification with the full set of variables. The OLS estimate for the election outcome of specification 1, including 1870 controls, 2016 demographics, geographical controls, state fixed effects, clustered standard errors, and weighting by the total votes, is given in column (1) of Table 3. We find a significant negative relationship between the average immigrant share from 1870 to 1920 and the Republican vote share in 2016. The 50th percentile of the distribution of the average immigrant share between 1870 and 1920 is 0.032. Therefore, these results suggest that moving a county with no historical immigration to the 50th percentile of the distribution, which is a change of 3.21 percent, is associated with a decrease in the vote share of the Republican party by 0.076 percentage points. This is a decrease by 0.12 percent relative to the sample mean.

Columns (2) to (4) of Table 4 show that the estimated effects of historical immigration on the three economic assimilation outcomes are rather small. For example, the results suggest that moving a county with no historical immigration to the 50th percentile increases the ratio of the natives' wages to the wages of the foreign borns by 0.005 percent. The estimates for the economic assimilation outcomes are not significant. Hence, there appears to be no significant long-run effect of historical immigration on the assimilation of foreign borns on the labor market based on these results.

Specification 1 includes the demographics of the natives and foreign borns in the year 2016, the urban rate of the year 2010, and the average immigrant share during the recent period of higher immigration from 1970 to 2010. However, these controls are potentially influenced by the treatment variable, since these were not fixed at the time when the treatment variable was shaped, and might bias the estimates (Angrist and Pischke, 2008). Therefore, we also estimate our baseline specification excluding these controls. The OLS results for the election outcome of specification 1, excluding the controls that potentially bias the results, are given in column (1) of Table 4. Similar as before, we find a significant negative long-run effect of the average immigrant share from 1870 to 1920 on the support of the Republican party in 2016. These results indicate that moving a county with no historical immigration to the 50th percentile of the distribution is associated with a decrease in the fraction of the Republican vote share by 0.28 percentage points, which is a decrease by 0.46 percent relative to the sample mean.

Columns (2) to (4) of Table 4 contain the results for the economic assimilation outcomes when excluding the 2016 controls. In contrast to the specification that includes the current controls, we now obtain opposite signs for the estimates of the wage gap and occupational income score gap. The effect of historical immigration is highly significant for the wage gap outcome. The result suggests that moving a county with no historical immigration to the 50th percentile decreases the ratio of the natives' wages to the wages of the foreign borns by 0.07 percent. This effect again is rather small. The effects for the occupational income score gap and the unemployment gap are not significant.

Altogether, the OLS estimates indicate a significant negative effect of historical immigration on the support of the Republican party nowadays and provide some evidence for a decrease in the wage gap between natives and foreign borns.

5.2 TSLS estimates

Due to the possibility of self-selection, we now turn to the two-stage least squares estimation. Since historical immigrants were not randomly sorted into counties and might have settled in counties based on its prospects, endogeneity might be induced in the treatment variable of historical immigration. On one hand, negative selection might lead the OLS estimates to be biased downwards. Negative selection occurs when immigrants settle in less prosperous counties due to for example discrimination or high

Table 3: OLS estimates of the effect of historical immigration on voting behavior and economic assimilation including recent controls

	(1) Republican vote share, 2016	(2) Wage gap, 2016	(3) Occupational income score gap, 2016	(4) Unemployment gap, 2016
avgimmshare1870_1920	-0.024***	0.002	0.003	0.001
	(0.0072)	(0.0056)	(0.0033)	(0.0017)
county controls 1870	yes	yes	yes	yes
demographics 2016	yes	yes	yes	yes
geographical controls	yes	yes	yes	yes
avgimmshare1970_2010	yes	yes	yes	yes
state fixed effects	yes	yes	yes	yes
clustered standard errors	yes	yes	yes	yes
weighted by total votes 2016	yes	no	no	no
weighted by persons 2016	no	yes	yes	yes
N	1854	1856	1856	1854
R^2	0.781	0.209	0.249	0.074
AIC	-4025.0	-393.2	-5280.8	-7937.7
BIC	-3848.2	-232.9	-5120.6	-7777.5

Standard errors in parentheses. An observation is a county. The variable of interest is the average immigrant share between 1870 and 1920. The method of estimation is least squares. The county controls from the year 1870 consist of the share of males, the share of African Americans, the share of Hispanics and Latinos, the share of illiterates, the share of urban population, the logarithm of the average manufacturing wage, the logarithm of the average agricultural wage, the number of people employed in manufacturing, the logarithm of the average amount of personal estate, and the connection to a railroad. The demographics from the year 2016 consist of the average age of the natives and its squared term, average age of the foreign borns and its squared term, the share of males in the native population, the share of males in the foreign-born population, the share of natives who do not speak English, the share of foreign borns who do not speak English, the share of married natives, the share of married foreign borns, the average years of schooling of the natives, the average years of schooling of the foreign borns, the share of natives with a college degree or higher, and the share of foreign borns with a college degree or higher. For the Republican vote share outcome, the following controls are additionally included: logarithm of the natives' wage, logarithm of the foreign borns' wage, the unemployment rate of the natives, and the unemployment rate of the foreign borns. The geographical controls consist of the longitude and latitude values of a county, and the urban rate in 2010. We further control for the average immigrant share between 1970 and 2010. The treatment variable and controls are standardized.
Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4: OLS estimates of the effect of historical immigration on voting behavior and economic assimilation excluding recent controls

	(1) Republican vote share, 2016	(2) Wage gap, 2016	(3) Occupational income score gap, 2016	(4) Unemployment gap, 2016
avgimmshare1870_1920	-0.088***	-0.021***	-0.002	0.00001
	(0.0112)	(0.0072)	(0.0041)	(0.0013)
county controls 1870	yes	yes	yes	yes
demographics 2016	no	no	no	no
geographical controls	yes	yes	yes	yes
avgimmshare1970_2010	no	no	no	no
state fixed effects	yes	yes	yes	yes
clustered standard errors	yes	yes	yes	yes
weighted by total votes 2016	yes	no	no	no
weighted by persons 2016	no	yes	yes	yes
N	1860	1857	1858	1856
R^2	0.608	0.077	0.158	0.051
AIC	-2993.6	-140.1	-5103.5	-7930.9
BIC	-2921.7	-68.28	-5031.6	-7859.1

Standard errors in parentheses. An observation is a county. The variable of interest is the average immigrant share between 1870 and 1920. The method of estimation is least squares. The county controls from the year 1870 consist of the share of males, the share of African Americans, the share of Hispanics and Latinos, the share of illiterates, the share of urban population, the average manufacturing wage, the average agricultural wage, the number of people employed in manufacturing, the average amount of personal estate, and the connection to a railroad. The geographical controls consist of the longitude and latitude values of a county. The treatment variable and controls are standardized.
Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

migration costs. On the other hand, positive selection into more prosperous counties might lead the OLS estimates to be biased upwards. Immigrants might for instance select the county where they settle based on the prevailing ideology of the natives in that county. Finally, there is also the possibility of measurement errors which might bias the OLS estimates towards zero. To overcome the problem of potential endogeneity in the treatment variable, we estimate the causal effect with a two-stage least squares regression where we instrument with the average predicted immigrant share as specified in Section 4.2.

The first stage results of specification 1 are reported in Table 5. A very strong first stage is estimated for all outcomes. The Sanderson-Windmeijer (SW) F-statistics are above the conventional threshold, except for the unemployment gap outcome in column (4). Nevertheless, the latter statistic is very close to the threshold of ten. Hence, a bias resulting from a weak instrument is not a concern in this analysis. The difference in the first stage results for the economic assimilation outcomes in columns (2) to (4) are due to the inclusion of different observations in the regressions, since data on different counties are available for each outcome.

The signs of the two-stage least squares estimates for the outcomes in columns (1) to (3) of Table 5 align with the signs of the OLS results with the same specification (see Table 3). The sign of the unemployment gap outcome in column (4) does not align with the OLS estimate. The OLS estimates are smaller in magnitude compared to the TSLS estimates, indicating that the OLS estimates might be biased due to negative selection of the historical immigrants and that the TSLS estimates are likely to be causal. The OLS bias can be induced by unobserved factors such as economic shocks that simultaneously influence the treatment and outcome variable, resulting in omitted variable bias. Another explanation for the difference in estimates is that TSLS returns the LATE, that is, the treatment effect in the group of compliers. Since our instrument is based on the ethnic network of the immigrants, the immigrants who eventually did settle in counties with similar ethnic groups are different from those who did not settle in a county based on social ties.

Column (1) of Table 5 contains the result for the Republican vote share. The effect of the average immigrant share from 1870 to 1920 on the support of the Republican party remains negative and significant as in the TSLS analysis, with standard errors that are now more than twice as large. These results suggest that moving a county with no historical immigration to the 50th percentile of the distribution decreases the Republican vote share by 0.242 percentage points. This is a decrease by 0.39 percent relative to the sample mean. This effect is smaller than the effect found by Giuliano and Tabellini (2020) who study the long-run effect of the historical average immigrant share between 1910 and 1930 on reporting a liberal ideology. They find that increasing the average historical immigration to roughly 40 percent of the interquartile range results in a 1.2

percent higher probability of being liberal, relative to the sample mean. They further find that this increase results in a 6.2 percent higher likelihood of identifying with the Democratic Party, relative to the sample mean.

Column (2) shows that the effect of the average immigrant share between 1870 and 1920 on the wage gap outcome is highly significant. Moving a county with no historical immigration to the 50th percentile of the distribution decreases the ratio of the natives' wages to the wages of the foreign-borns by 0.26 percent. Although the effect is larger in magnitude than in the OLS estimation, it remains rather small. The effects for the occupational income score gap and the unemployment gap in columns (3) and (4) remain insignificant.

We again estimate specification 1 without the controls that might be influenced by historical immigration in the long run. Table 6 anew shows strong first stages with Kleibergen-Paap F-statistics that lie above the conventional threshold of ten. Compared to the OLS regression with comparable specification (see Table 4), the estimate for the Republican vote share outcome remains significant and increases in magnitude. The results suggest that moving a county with no historical immigration to the 50th percentile of the distribution is associated with a decrease in the fraction of the Republican vote share by 0.242 percentage points. This is a decrease by 0.65 percent relative to the sample mean. All estimates of the economic assimilation outcomes switch sign compared to the OLS estimates. Compared to the TSLS that includes the current controls (see Table 5), all signs of the effects remain unchanged. The estimate for the wage gap outcome now is insignificant, while the estimate for the unemployment gap is significant. For the unemployment gap we find that moving a county with no historical immigration to the 50th percentile of the distribution decreases the difference between the unemployment rate of the natives and the unemployment rate of the foreign borns by 0.030. However, the R-squared for the regression of the unemployment gap is negative, indicating a poor model specification.

To linger concerns about omitted variable bias and misspecification of the model, we perform several robustness checks. First, we estimate the effect when no controls are included in the model, that is, when the model only contains the treatment. Then, for the baseline model specification as well as for the specification excluding the current controls, we assess the effect of weighting the observations by the total votes or the number of persons in a county as well as the effect of clustering the standard errors at the state level. Moreover, we check both specifications when defining the period of historical immigration from 1900 to 1920, as immigration peaked during these decades (Abramitzky and Boustan, 2017). Finally, we construct a binary treatment based on the median value of the average immigrant share from 1870 to 1920. Reassuringly, Tables 13 to 16 in Appendices 8.2 to 8.5 overall show that the sign of the effect for each outcome remains the same, indicating that the results are robust to different specifications. In case of the occupational income score gap and the Republican vote share, not including any controls leads to a different sign of the effect of the average immigrant share. For the unemployment gap outcome, the R-squared is always negative when the current controls are excluded from the model specification, indicating a poor model specification in this case. The R-squared is also negative when the treatment is binary.

Overall, historical immigration has a significant negative effect on the Republican vote share in the long run. For the assimilation outcomes, we find an increase in the wage gap when including the full set of controls. When excluding the controls that are potentially influenced by historical immigration, this effect becomes insignificant. In the next section we will elaborate on the possible mechanisms leading to these results.

5.3 Instrumental forest

By means of an instrumental forest, we now study the results when allowing for more nonlinearities in the data. First, we will specify the settings and parameters of the instrumental forest (Section 5.3.1). Once the settings are optimally specified, we estimate the ACLATE for each of the four outcomes, and perform a test to investigate whether the instrumental forest successfully detects heterogeneity (Section 5.3.2). In Section 5.4, potential effect modifiers are identified. This allows for explaining the mechanisms underlying the long-run effect of historical immigration on the support of the Republican party as well as the economic assimilation of foreign borns. Finally, we perform several robustness checks.

5.3.1 Settings and parameters

Currently, there exists no literature on the estimation of the average conditional local average treatment effect with an instrumental forest in case of a continuous instrument. Therefore, we construct a binary instrument by splitting the continuous version of the predicted average immigrant share between 1870 and 1920 based on its median. A value of one is assigned to the upper 50 percent quantile of the distribution, and the variable takes on the value of zero otherwise.

Since forest-based methods automatically account for nonlinearities, the instrumental forest includes the controls without logarithmic transformation and excludes the squared term of age. Moreover, we include the full, unstandardized set of controls. State fixed effects are not included as one-hot encoded dummies, as assuming the states to have an additive effect is rather restrictive (Athey and Wager, 2019). Furthermore, including the 33 states in the data set as one-hot encoded dummies in a forest-based method is rather wasteful, as this adds a lot of low-signal variables (Johannemann et al., 2019). Instead, clustering the observations on the state level allows each state to have its own treatment effect, which might be a better approach to capturing differences on the state-level.

Table 5: TSLS estimates of the effect of historical immigration on voting behavior and economic assimilation including recent controls

	(1) Republican vote share, 2016	(2) Wage gap, 2016	(3) Occupational income score gap, 2016	(4) Unemployment gap, 2016
avgimmshare1870_1920	-0.075***	0.081***	0.002	-0.009
	(0.018)	(0.031)	(0.012)	(0.007)
First stages with average immigrant share 1870 - 1920 as dependent variable				
pred_immshare1870_1920	1.564***	1.027***	1.025***	1.115***
	(0.326)	(0.300)	(0.303)	(0.362)
county controls 1870	yes	yes	yes	yes
demographics 2016	yes	yes	yes	yes
geographical controls	yes	yes	yes	yes
avgimmshare1970_2010	yes	yes	yes	yes
state fixed effects	yes	yes	yes	yes
clustered standard errors	yes	yes	yes	yes
weighted by total votes 2016	yes	no	no	no
weighted by persons 2016	no	yes	yes	yes
N	1854	1856	1856	1854
R^2	0.663	0.132	0.116	0.011
AIC	-3871.2	-345.4	-5272.5	-7885.4
BIC	-3694.4	-185.1	-5112.3	-7725.2
Kleibergen Paap F-statistic	11.204	5.665	5.541	4.773
SW F-statistic avgimmshare1870_1920	23.87	13.06	12.61	9.82
SW F-statistic avgimmshare1970_2010	21.34	10.59	10.42	10.00

Standard errors in parentheses. An observation is a county. The variable of interest is the average immigrant share between 1870 and 1920. The method of estimation is two-stage least squares. The county controls from the year 1870 consist of the share of males, the share of African Americans, the share of Hispanics and Latinos, the share of illiterates, the share of urban population, the logarithm of the average manufacturing wage, the logarithm of the average agricultural wage, the number of people employed in manufacturing, the logarithm of the average amount of personal estate, and the connection to a railroad. The demographics from the year 2016 consist of the average age of the natives and its squared term, average age of the foreign-borns and its squared term, the share of males in the native population, the share of males in the foreign-born population, the share of natives who do not speak English, the share of foreign-borns who do not speak English, the share of married natives, the share of married foreign-borns, the average years of schooling of the natives, the average years of schooling of the foreign-borns, the share of natives with a college degree or higher, and the share of foreign-borns with a college degree or higher. For the Republican vote share outcome, the following controls are additionally included: logarithm of the natives' wage, logarithm of the foreign-borns' wage, the unemployment rate of the natives, and the unemployment rate of the foreign-borns. The geographical controls consist of the longitude and latitude values of a county, and the urban rate in 2010. We further control for the average immigrant share between 1970 and 2010. The treatment variable and controls are standardized.
Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 6: TSLS estimates of the effect of historical immigration on voting behavior and economic assimilation excluding recent controls

	(1) Republican vote share, 2016	(2) Wage gap, 2016	(3) Occupational income score gap, 2016	(4) Unemployment gap, 2016
avgimmshare1870_1920	-0.124***	0.030	0.002	-0.009***
	(0.028)	(0.029)	(0.011)	(0.003)
First stages with average immigrant share 1870 - 1920 as dependent variable				
pred_immshare1870_1920	1.709***	1.566***	1.563***	1.704***
	(0.376)	(0.416)	(0.421)	(0.515)
county controls 1870	yes	yes	yes	yes
demographics 2016	no	no	no	no
geographical controls	yes	yes	yes	yes
avgimmshare1970_2010	no	no	no	no
state fixed effects	yes	yes	yes	yes
clustered standard errors	yes	yes	yes	yes
weighted by total votes 2016	yes	no	no	no
weighted by persons 2016	no	yes	yes	yes
N	1860	1857	1858	1856
R^2	0.429	0.001	0.012	-0.015
AIC	-2940.1	-115.3	-5101.8	-7878.3
BIC	-2868.3	-43.48	-5029.9	-7806.4
Kleibergen Paap F-statistic	20.625	14.161	13.823	10.937

Standard errors in parentheses. An observation is a county. The variable of interest is the average immigrant share between 1870 and 1920. The method of estimation is two-stage least squares. The county controls from the year 1870 consist of the share of males, the share of African Americans, the share of Hispanics and Latinos, the share of illiterates, the share of urban population, the logarithm of the average manufacturing wage, the logarithm of the average agricultural wage, the number of people employed in manufacturing, the logarithm of the average amount of personal estate, and the connection to a railroad. The geographical controls consist of the longitude and latitude values of a county. The treatment variable and controls are standardized.
Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

We first estimate the expected outcome $\mathbf{E}[Y_i|X_i = x]$, the treatment propensity scores $\mathbf{P}[W_i|X_i = x]$, and the instrument propensities $\mathbf{P}[Z_i|X_i = x]$ using regression forests with 2000 trees each. Instead of tuning these forests, the default values of the parameters are used, as these values generally perform relatively well (Athey and Wager, 2019). The resulting estimates are subsequently included in the instrumental forest.

With the aim of reducing noise in the estimation of the average treatment effects with causal forests, Athey and Wager (2019) implement a variable selection step in their application. They first implement a causal forest with the full set of controls. Then, they include only those variables that have a variable importance above the mean value of all variable importances in a second causal forest, hence excluding low-signal features. We refer to this approach as the double forest and to the approach containing the full set of controls as the single forest. In addition to reducing noise, Wager and Athey (2018) finds that the confidence intervals improve with an increase of the number of variables up to approximately ten variables, using a simulation method on a data set consisting of 10,000 observations. When more than ten variables are included in the causal forest, no nominal coverage is achieved. Thus, the performance of the instrumental forest might improve when selecting only a subset of the variables based on their variable importance, and subsequently running a second instrumental forest including only this subset of variables. Therefore, in our application, we follow this approach and implement a double forest for the estimation of the ACLATEs.

The controls to be included are chosen based on the variable importance. The importances of the first forest are quite uniformly distributed amongst the variables (see Table 17 in Appendix 8.6). This implies that selecting a subset of variables based on the mean or median importance results in a rather large subset of variables. Instead, we focus on the graphs of the variables sorted by their importance to decide on the cut-off point (see Figure 5 in Appendix 8.6). Since these figures are similar to scree plots, we choose the number of variables according to the "elbow" of the graph. We choose the cut-off point such that the increase in accumulative importance of including an additional variable is relatively small, indicating that the additional variable might rather be adding noise instead of contributing to the model.

To further improve the prediction performance of the instrumental forest, the number of trees in the forest is increased from the default value of 2000 to 5000. The parameters of the instrumental forest are tuned with cross-validation. Seven parameters are tuned in total. We tune the sample fraction used to construct each tree, the number of variables considered at each split, the minimum number of observations in a leaf node, the fraction of data used for the honesty of the tree, whether empty leaves should be pruned, the maximum imbalance of a split, and how harshly imbalanced splits are penalized. Since instrumental forests are known to be sensitive to the choice of tuning parameters, we increase the default values for the tuning specifications of the cross-validation. The number of trees for each mini forest used to fit the tuning model is set to 1000. Then, 250 mini forests are created to fit the tuning model, considering 5000 parameter values that are randomly drawn. Finally, the optimal parameter settings are chosen such that the debiased error is minimized.

The parameters for the first forest are given in Table 7, and for the second forest in Table 8. The tuned parameter for the sample fraction used for the construction of an individual tree is very low for both forests of the unemployment gap outcome. For the first forest of the occupational income score gap, the parameters are set to the default values, because the debiased error of the best possible set of parameters selected with cross-validation is higher than the error obtained by using the default values. We manually set the parameter for the number of variables considered at each split to $\lfloor K/3 \rfloor$. This is commonly done in literature to decorrelate the trees, which results in a decrease in the variance of the estimates. In the second forest of the occupational income score gap, parameter settings that result in a lower error than the default values are found by cross-validation, and hence these tuned parameters are used to build the final instrumental forest.

For all of our subsequent analyses we use the *R* package *grf*, version 1.2.0 (R Core Team, 2017; Tibshirani et al., 2020).

5.3.2 ACLATEs

The doubly robust ACLATE for the Republican vote share outcome as estimated by 9 is reported in column (1) of Table 9. The ACLATE suggests that the Republican vote share in 2016 is on average 8.1 percentage points lower for counties that experienced above median levels of historical immigration. The negative long-run effect of the immigration during the Age of Mass migration is in line with the result of the TSLS regression. However, the instrumental forest does not find this effect to be significant.

The doubly robust ACLATE for the ratio of the average yearly wage of the natives and

Table 7: Parameter tuning for the first forest

Parameter	(1) Republican vote share, 2016	(2) Wage gap, 2016	(3) Occupational income score gap, 2016	(4) Unemployment gap, 2016
sample.fraction	0.39	0.46	0.50	0.10
mtry	21.00	17.00	8.00	3.00
min.node.size	1.00	1.00	5.00	57.00
honesty.fraction	0.59	0.55	0.50	0.57
honesty.prune.leaves	0.00	0.00	1.00	1.00
alpha	0.11	0.01	0.05	0.14
imbalance.penalty	1.19	0.35	0.00	0.64

Table 8: Parameter tuning for the second forest

Parameter	(1) Republican vote share, 2016	(2) Wage gap, 2016	(3) Occupational income score gap, 2016	(4) Unemployment gap, 2016
sample.fraction	0.39	0.49	0.50	0.05
mtry	6.00	3.00	5.00	7.00
min.node.size	1.00	1.00	1.00	35.00
honesty.fraction	0.59	0.63	0.50	0.52
honesty.prune.leaves	0.00	0.00	0.00	1.00
alpha	0.11	0.05	0.10	0.21
imbalance.penalty	1.19	0.06	0.69	0.28

the average yearly wage of the foreign borns is reported in column (2) of Table 9. The sign of the treatment effect does not align with the effect found by the TSLS estimation and the results suggest that high levels of historical immigration in a county lead to a decrease in the ratio of the average wage gap by 22.8 percent. For the occupational income score gap we equally observe a different sign from the TSLS results that are based on the same specification (see column (3) in Table 9). The results suggest that a county with a high share of immigrants between 1870 and 1920 experiences a decrease in the wage gap ratio by 1.3 percent. The result of the unemployment gap aligns with the TSLS results. Above median historical immigration leads to a decrease in the average unemployment gap by 0.4 percentage points. However, the estimated ACLATEs of the economic assimilation outcomes are not significant.

Summarizing, for none of the four outcomes does the instrumental forest find a significant conditional average treatment effect of historical immigration. The results might differ from the TSLS results because of the nonlinearities of controls that are implicitly accounted for in the instrumental forest.

We now assess whether the forest successfully captured heterogeneity by following Athey and Wager (2019) who implement an omnibus test to assess the treatment heterogeneity. For this test, the data set is divided into two subgroups based on whether the out-of-bag predictions are higher or lower than the median CLATE value. Then, a two-sided t-test is performed on the doubly robust ACLATE estimates for the two subgroups. The results can be found in Table 9. For the wage gap outcome in column (2) as well as for the unemployment gap outcome in column (4) there is evidence of detected heterogeneity in the treatment variable. We do not find evidence for a significant difference between the ACLATEs in the two subgroups for columns (1) and (3), indicating that the instrumental forest did not detect strong heterogeneity for the Republican vote share outcome and the occupational income score gap outcome. However, it should be noted that this test is heuristic, since the subgroups are constructed based on the CLATEs

Table 9: Average treatment effects and t-test for assessing treatment heterogeneity

	(1) Republican vote share, 2016	(2) Wage gap, 2016	(3) Occupational income score gap, 2016	(4) Unemployment gap, 2016
ACLATE	-0.081	-0.228	-0.013	-0.004
	(0.117)	(0.143)	(0.094)	(0.012)
	t-value: -0.690	t-value: -1.596	t-value: -0.141	t-value: -0.357
95% CI	(-0.31, 0.148)	(-0.508, 0.052)	(-0.197, 0.171)	(-0.028, 0.02)
ACLATE (high)	-0.014	0.085	-0.001	-0.056***
	(0.042)	(0.143)	(0.042)	(0.015)
	t-value: -0.323	t-value: 0.594	t-value: -0.025	t-value: -3.778
ACLATE (low)	-0.128	-0.557**	-0.021	0.058***
	(0.187)	(0.268)	(0.190)	(0.017)
	t-value: -0.687	t-value: -2.078	t-value: -0.111	t-value: 3.320
95% CI for difference	(-0.261, 0.491)	(0.046, 1.238)	(-0.361, 0.401)	(-0.159, -0.069)
observations	1854	1856	1856	1854
controls	30	26	26	26
controls selected in second forest	7	9	10	10
clustered std. errors	yes	yes	yes	yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Doubly robust ACLATEs. Standard errors in parantheses.

and therefore not independent from the ACLATEs used for the test. Thus, it only gives an insight into the strength of heterogeneity. There still might be heterogeneity in the treatment variable present. We will investigate this in the next section.

5.3.3 Detected heterogeneity

Although there is no evidence of strong treatment heterogeneity based on the omnibus test, there still might be heterogeneity present. We now turn to one of the major advantages of instrumental forests and assess the treatment heterogeneity for each control separately. This allows for considering all controls as potential effect modifiers simultaneously without having to specify subgroups ex ante. For each control, we divide the data set into two subgroups. The first subgroup consists of the observations belonging to the lower 20 percent quantile of that variable. The second subgroup consists of the observations in the 20 percent highest quantile of the distribution. The variable that controls for whether a county was connected to the railway in 1860 is binary and hence we split the data set based on the median value for this control. We then apply a cluster-weighted t-test for assessing the heterogeneity for these two subgroups. The results for the potential effect modifiers are reported in Table 10. The results of the controls for which the subgroups do not differ significantly at a ten percent level can be found in Tables 19 to 26 in Appendices 8.7 to 8.10.

Table 10: ACLATEs per control

Variable	ACLATE low	ACLATE high	95% CI difference	t-value difference	p-value difference
Panel A: Republican vote share, 2016					
hispaniclatinos1870	0.083	-0.216*	(-0.583, -0.015)	-2.058	0.040
	(0.078)	(0.122)			
educfb2016	-0.506	0.109	(-0.126, 1.356)	1.627	0.104
	(0.356)	(0.127)			
Panel B: Wage gap, 2016					
male1870	-0.497***	0.146	(0.183, 1.101)	2.744	0.006
	(0.164)	(0.167)			
hispaniclatinos1870	-0.285*	0.201	(-0.054, 1.026)	1.766	0.077
	(0.164)	(0.221)			
railway1860	0.013	-0.366**	(-0.82, 0.062)	-1.684	0.092
	(0.124)	(0.187)			
educnatives2016	-0.717*	0.054	(0.014, 1.53)	1.996	0.046
	(0.371)	(0.108)			
highdegreenatives2016	-0.693**	0.084	(0.110, 1.442)	2.286	0.022
	(0.325)	(0.099)			
Panel C: Occupational income score gap, 2016					
agenatives2016	0.043	-0.116	(-0.317, -0.001)	-1.972	0.049
	(0.037)	(0.072)			
marriednatives2016	0.031	-0.146*	(-0.359, 0.005)	-1.916	0.055
	(0.042)	(0.082)			
noenglishfb2016	-0.016	-0.250***	(-0.434, -0.034)	-2.292	0.022
	(0.045)	(0.091)			
Panel D: Unemployment gap, 2016					
malenatives2016	0.045	-0.038*	(-0.157, -0.009)	-2.201	0.028
	(0.030)	(0.023)			

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Doubly robust ACLATEs. Standard errors in parantheses.

5.3.4 Republican vote share

For the Republican vote share outcome, the fraction of the population consisting of Hispanics and Latinos in 1870 is found to be a potential effect modifier at the 5 percent level (see panel A in Table 10). The results suggest that counties with above median historical average immigrant shares experience a significant decrease in the support of the Republican vote share in 2016 when they belong to the 20 percent quantile with the highest fraction of Hispanics and Latinos in 1870. For these counties, the long-run effect on the Republican vote share is -0.216, that is, the Republican vote share decreases by 21.6 percentage points when a county's population consisted of a high fraction of Hispanics and Latinos in 1870. The estimated ACLATE for the counties who had comparatively fewer Hispanics and Latinos in their population in 1870 suggests that these counties experience an increase in the Republican vote share in 2016 when they also experienced high levels of historical immigration. The latter result, however, is not significant. These results might suggest that counties where Hispanics were part of the community in 1870 might have been more accepting of the European immigrants who arrived during the Age of Mass Migration due to the cultural diversity in that county. This might have led to an easier transfer of the liberal ideology brought by the European immigrants onto the native population, supporting the finding by Giuliano and Tabellini (2020) that the values of foreign borns are passed on to the natives and that this effect persists over time.

A second potential effect modifier is the average years of schooling of the recent foreign borns, which is marginally significant at the 10 percent level. The result suggests that counties where recent foreign borns are on average at the lower end of the distribution of the number of years of schooling experience a decrease in the Republican vote share by 50.6 percentage points if they experienced high historical immigration. High historical immigration might be an indicator of more established immigrant communities. Especially low-educated foreign borns join these communities, whereas higher-educated foreign borns are less likely to join these communities (Alba and Nee, 1997).

5.3.5 Wage gap

Several potential effect modifiers for the average wage gap between foreign borns and natives in 2016 are identified at the five percent significance level (see panel B in Table 10). The results suggest that historical immigration in counties that had a comparably smaller fraction of males in 1870 decreases the ratio of the average wage of the natives and the average wage of the foreign borns significantly by 49.7 percentage points. In counties with proportionally more women in 1870, the historical immigrants might have been perceived as a welcome workforce, increasing economic productivity without receiving much backlash from the natives. Since immigrants did not threaten the natives economically, the immigrants might have been more accepted and therefore integrated within the community. The subsequent improved assimilation of the immigrants in these counties might have led to a more cultural diverse population with an increased liberal attitude. Hence, foreign borns in these counties are less likely to experience discrimination on the job market.

Another potential effect modifier is a county's fraction of Hispanics and Latinos in 1870. Counties with high historical immigration experience a significant decrease in the ratio between the average wage of the natives and the foreign borns by 28.5 percent if their population consisted of comparably fewer Hispanics and Latinos in 1870. The population of these counties might have been less accepting of the European immigrants due to a lack of ethnic diversity within the population, resulting in an increased segmentation between the natives and the European immigrants. This might have increased the importance of establishing migrant communities over time. Nowadays, immigrants might be able to easily find a job within these well-established communities, resulting in a lower wage gap between the natives and the foreign borns.

The number of years of schooling of a county's native population in 2016 also appears to be an effect modifier. The results suggest that counties with a comparably lower-educated population in 2016 experience a significant decrease in the ratio between the average wage of the natives and the foreign borns by 71.7 percent due to higher historical immigration. This effect is confirmed by the variable that captures a county's share of natives with a college degree or higher. Counties with a low fraction of natives with a college degree or higher experience a decrease in the ratio of the average wages by 69.3 percent when they experienced above-median historical immigration. The high historical immigration might have led to more established communities. Nowadays, these communities receive especially low-educated immigrants. Since these immigrants are likely to find a job within their community, they do not have to compete on the job market with the natives. This potentially leads to a smaller wage gap.

5.3.6 Occupational income score gap

Three potential effect modifiers of the treatment are detected for the occupational income score gap (see panel C in Table 10). However, the estimated ACLATEs in the subsamples with the observations belonging to the lowest and highest end of the distribution concerning the average age of the natives in 2016 are not significant. Significant ACLATEs are detected for the share of married natives in 2016 and the share of foreign borns who do not speak English in 2016.

The result that historical immigration leads to a significant decrease in the ratio for the occupational income score gap by 14.6 percent in counties where the fraction of married natives is higher seems rather controversial. According to Giuliano and Tabellini (2020), intermarriage between natives and foreign borns might enhance the transmission of liberal values. However, high historical immigration supposedly leads to an increased establishment of migrant communities and migrants might be more likely to marry within their communities. Hence, high historical immigration is expected to decrease the intermarriage rate between natives and foreign borns, causing an increase in segmentation between the natives and foreign borns on the labor market. We cannot investigate this theory further as we do not have data on the intermarriage between natives and foreign borns. Another possible explanation is that being married is a proxy for another, unobserved variable. Altogether, the effect of this variable needs further investigation in order to be properly interpreted.

Counties where fewer foreign borns speak English experience a significant decrease in the occupational income score gap ratio by 25 percent when they experienced high historical immigration as opposed to counties where a higher proportion of foreign borns speak English nowadays. This might again be due to the establishment of communities due to high historical immigration. It might be that foreign borns with a low level of English proficiency are well-received by these established communities and can find similar occupations as the natives within these communities. Immigrants with a low level of English proficiency who settle in counties with low historical immigration and therefore with less established migrant communities have to compete on the same job market as the native population, making economic assimilation more difficult due to for instance discrimination.

5.3.7 Unemployment gap

In counties with relatively more males in 2016, historical immigration leads to a decrease in the unemployment gap by 3.8 percentage points (see panel D in Table 10). The establishment of migrant communities in these counties might help male foreign borns in finding a job within those communities, as the economic assimilation between male foreign borns and natives is more challenging on the natives' job market than the economic assimilation between women (Butcher and DiNardo, 2002; Alba and Nee, 1997).

5.3.8 Summary

Altogether, the results suggest that the effect of historical immigration has persisted through two channels. First, we find evidence that supports the theory by Sequeira et al. (2020) which states that the liberal ideology of the European immigrants was passed on to the natives during the Age of Mass Migration. This led to a shift in the natives' ideology. Secondly, the historical immigration led to the establishment of migrant communities which allow migrants to find similar employments as the natives within those communities. However, this might also imply that there is more segregation between the natives and foreign borns nowadays.

5.4 Robustness checks for the instrumental forest

This section contains several robustness checks for the instrumental forest. We start with a sensitivity analysis for the robustness of our model specification in Section 5.4.1. Then, we investigate the effect of tuning the parameters on the results, as the estimates of the instrumental forest are sensitive to the chosen tuning parameters (Section 5.4.2). Finally, we check whether the overlap assumption is met in Section 5.4.3, and study how using a binary treatment influences the results in Section 5.4.4.

5.4.1 Robustness of the model specification

The large standard errors in Table 9 demonstrate that when allowing for a more flexible specification regarding the nonlinearities in the data, the estimated average conditional local treatment effects for all four outcomes are insignificant. The standard errors measure the uncertainty of the model and depend on the specification of the model. Uncertainty of the causal effect is especially important for correct policy making. To linger concerns on the implementation of a double forest instead of a single forest as well as the inclusion of the 2016 controls which might be influenced by the treatment, we follow Athey and Imbens

(2015) and investigate the robustness of our model to the specification. We compare the robustness of the double forest including the full set of controls to the robustness of a single forest as well as to the robustness when excluding the 2016 controls from the model. The approach allows for checking whether these different model specifications lead to different point estimates.

For this purpose, we first split the data sample on each control based on their respective median values. Then, we estimate the average conditional local treatment effect for each control on the two subsamples separately. Subsequently, we take the mean over the estimated ACLATEs and their corresponding standard errors. In case of the Republican vote share outcome, this results in 30 choices of the controls, and in case of the economic assimilation outcomes this results in 26 splits.

Table 11: Sensitivity analysis for the model specification

	(1) Republican vote share, 2016	(2) Wage gap, 2016	(3) Occupational income score gap, 2016	(4) Unemployment gap, 2016
Panel A: Double forest				
mean ACLATE over splits	-0.068	-0.210	-0.004	-0.003
mean se over splits	(0.127)	(0.163)	(0.102)	(0.0166)
$\hat{\sigma}_\tau$	0.088	0.138	0.047	0.008
percent of standard error $\hat{\theta}_B$	0.750	0.965	0.506	0.659
Panel B: Single forest				
mean ACLATE over splits	0.010	-0.201	0.012	-0.009
mean se over splits	(0.092)	(0.183)	(0.087)	(0.016)
$\hat{\sigma}_\tau$	0.070	0.143	0.043	0.010
Percent of standard error $\hat{\theta}_B$	0.891	0.847	0.547	0.843
Panel C: Double forest - only 1870 controls				
mean ACLATE over splits	0.018	0.471	0.034	-0.00008
mean se over splits	(0.413)	(0.597)	(0.096)	(0.015)
$\hat{\sigma}_\tau$	0.160	0.453	0.044	0.008
percent of standard error $\hat{\theta}_B$	0.433	0.832	0.477	0.720
Panel D: Double forest - no tuning				
mean ACLATE over splits	0.056	-0.075	0.025	-0.007
mean se over splits	(0.124)	(0.151)	(0.100)	(0.017)
$\hat{\sigma}_\tau$	0.104	0.102	0.051	0.011
percent of standard error $\hat{\theta}_B$	1.011	0.807	0.559	0.985

We can then calculate a robustness measure based on the standard deviation as follows:

$$\hat{\sigma}_\theta = \sqrt{\frac{1}{\#(P)} \sum_{p \in P} \left(\hat{\theta}_p(x) - \hat{\theta}(x) \right)^2} \tag{10}$$

where P is the number of splits, $\hat{\theta}(x)$ the estimated ACLATE as calculated in Table 9, and $\hat{\theta}_p(x)$ the mean of the two estimated ACLATEs for each sample split.

The percent of standard error $\hat{\theta}_B$ is calculated by dividing the estimated standard error by the robustness measure. Panel A in Table 11 shows that $\hat{\sigma}_\tau$ is about 75.0 percent, 96.5 percent, 50.6 percent, and 65.9 percent of the standard errors for the Republican vote share, the wage gap, the occupational income score gap, and the unemployment gap respectively, when including all controls in the model specification. Although the single forest appears to be more robust, the double forest performs relatively well. For the occupational income score gap outcome, the robustness is rather low in both cases. Reassuringly, the double forest appears to be more robust when including all controls instead of only the subset of 1870 controls for the Republican vote share, the wage gap, and the occupational income score (see Panel A and Panel C in Table 11).

5.4.2 Tuning parameters

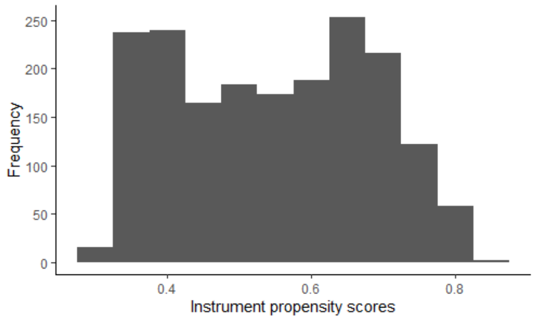
Instrumental forests are very sensitive to tuning, even more than causal forests (Tibshirani et al., 2020). Hence, we investigate the effect of using the default settings of all tunable parameters and manually set the parameter *mtry* which determines the number of controls at each split to $\lfloor K/3 \rfloor$, where K is the total number of controls included in the forest. When comparing Table 9 with Table 27 in Appendix 8.11, we note that we now get nonsensical results for the estimated ACLATEs in the subsamples with below and above median estimated treatment effects for the occupational income score gap outcome. The signs of the estimated ACLATEs for the Republican vote share and the occupational income score gap outcome are flipped for the two subsamples. More interestingly, for the unemployment gap outcome it now holds that the ACLATE for the higher subsample of the estimated individual treatment effects is higher than the estimated ACLATE for the lower subsample of the individual treatment effects. Thus, for the unemployment gap outcome, the default values seem to yield more reasonable results. Moreover, the ACLATE (high) and ACLATE (low) are no longer significant. This is probably due to increasing the tuned value of the sample fraction of each tree from 0.05 and 0.10 for the first and second forest respectively to the default value of 0.5. Using only a very small subsample of the full sample for each tree results in a decorrelation of the predictions of the individual trees in the forest. This decreases the variance, but simultaneously might increase the bias. In Table 9, this bias-variance tradeoff probably results in biased estimates for the two subsamples, causing the doubly robust estimator in the high subsample to be lower than the doubly robust ACLATE in the low subsample.

5.4.3 Violation of the overlap assumption

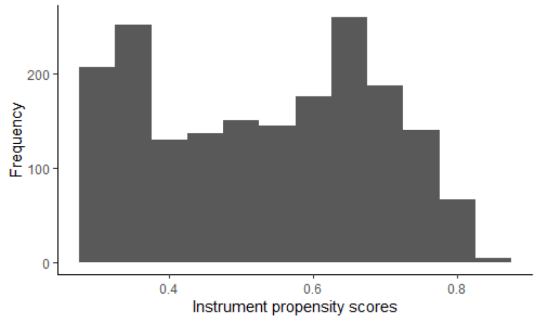
One of the assumptions underlying the model is the overlap assumption which is also referred to as the matching assumption. This assumption states that there should be sufficient overlap between the treated and the untreated cases in the sample, that is, for each county with high historical immigration there is a county with low historical immigration with similar observed values for the controls. This assumption translates to the propensity scores being uniformly distributed on the interval $[0, 1]$. In case of the instrumental forest, propensity scores for the treatment as well as for the instrument are estimated. For all four outcomes, the instrument propensity scores are bounded away from one and relatively well-distributed on the interval $[0, 1]$ (see Figure 2). The instrument propensities are also bounded away from zero for the Republican vote share outcome. The treatment propensities in Figure 3 are not bounded away from zero and one. Skewed propensity scores might lead to nonsensical results and an inflation of the estimated ACLATEs. Overlap-weighting as proposed by Li et al. (2018) might solve this issue, since it circumvents the division by propensity scores close to zero. However, this method is only established for causal forests and has not been established yet for instrumental forests. Therefore, we investigate the effect of overlap-weighting in case of the causal forest. Reassuringly, the overlap-weighting does yield similar results as the non-weighted case, as can be seen from a comparison between Table 28 and Table 30 in Appendices 8.12 and 8.13 respectively. As expected, the overlap-weighting seems to slightly deflate the estimated effects. Since the treatment propensities for the instrumental forest are the same as the propensities in the causal forest, we can generalize these results. We conclude that skewed propensities are unlikely a problem in our setting with the instrumental forest.

5.4.4 Binary treatment variable

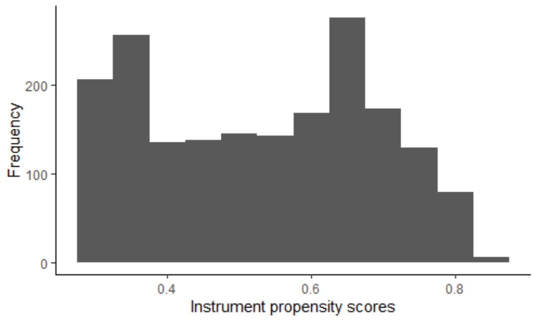
Another possible reason for the difference in results of the two-stage least squares regression and the instrumental forest might be the use of a binary treatment instead of a continuous treatment. The use of a binary instrument might imply a loss of information. This does not appear to be a concern in case of the two-stage least squares regression, as the robustness check using a binary treatment results in a significant effect that is similar in sign (see Tables 13 to 16 in Appendices 8.2 to 8.5). We equally want to linger this concern for the instrumental forest. Since the instrumental forest does not allow for estimation of the doubly robust ACLATE, we again turn to the causal forest to investigate this issue. From a comparison between Table 28 and Table 32 in Appendices 8.12 and 8.14 respectively, we observe that the use of a binary treatment variable does



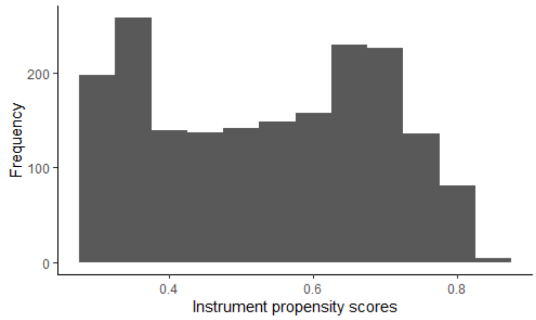
(a) Estimated instrument propensity scores for the Republican vote share outcome



(b) Estimated instrument propensity scores for the wage gap outcome

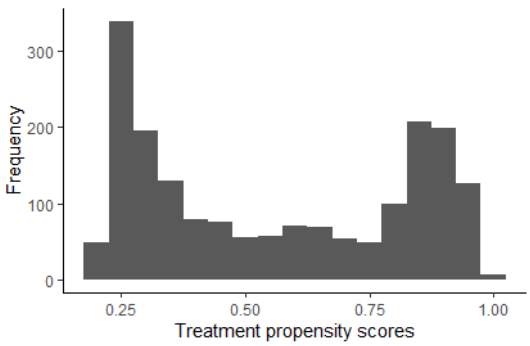


(c) Estimated instrument propensity scores for the occupational income score gap outcome

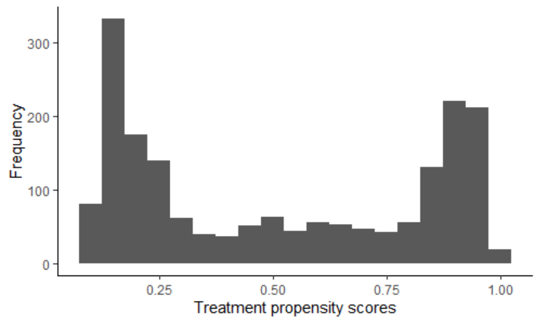


(d) Estimated instrument propensity scores for the unemployment score gap outcome

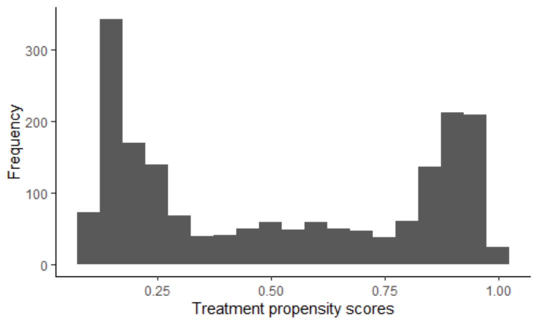
Figure 2: Histograms of the estimated instrument propensities for the four outcome variables



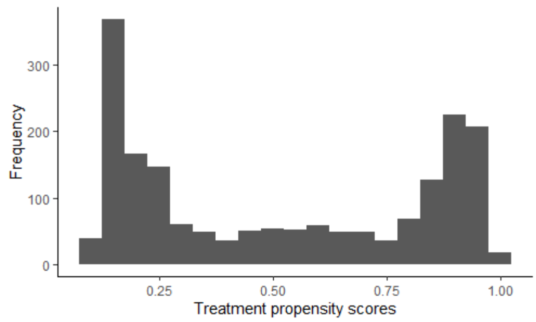
(a) Estimated treatment propensity scores for the Republican vote share outcome



(b) Estimated treatment propensity scores for the wage gap outcome

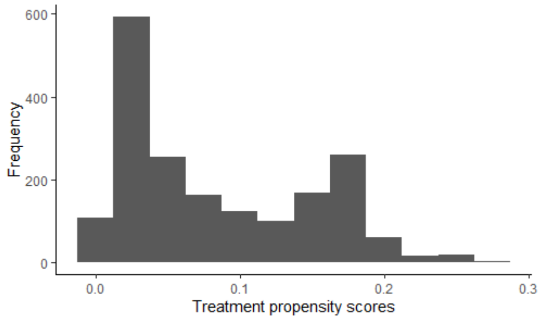


(c) Estimated treatment propensity scores for the occupational income score gap outcome

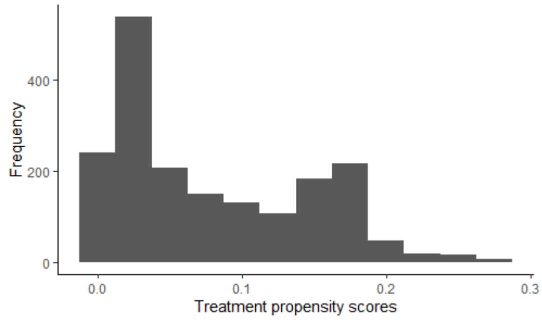


(d) Estimated treatment propensity scores for the unemployment score gap outcome

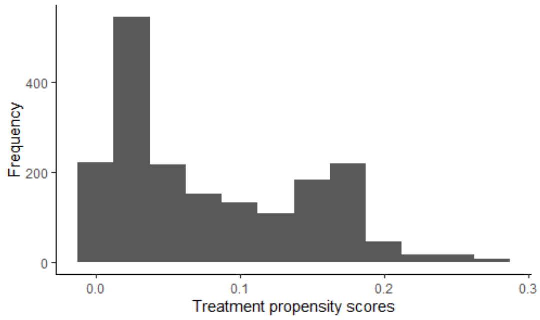
Figure 3: Histograms of the estimated treatment propensities for the four outcome variables



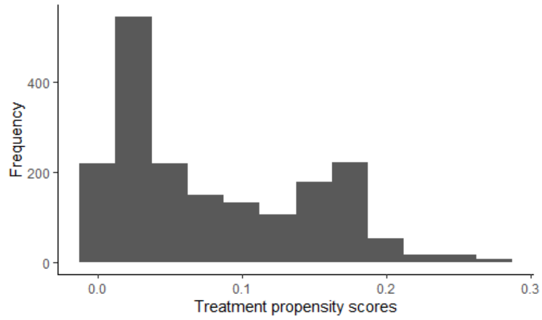
(a) Estimated treatment propensity scores for the Republican vote share outcome



(b) Estimated treatment propensity scores for the wage gap outcome



(c) Estimated treatment propensity scores for the occupational income score gap outcome



(d) Estimated treatment propensity scores for the unemployment score gap outcome

Figure 4: Histograms of the estimated treatment propensities for the four outcome variables in case of a causal forest with continuous treatment

change the results. This is most likely due to the propensity scores that are skewed towards zero when a continuous treatment is used (see Figure 4). Therefore, a more fair comparison would be the comparison between the overlap-weighted results. However, overlap-weighting is not implemented for a continuous treatment variable which impedes a proper investigation of this matter.

6 Conclusion

In this paper, we investigate the long-run effect of the historical immigration during the Age of Mass migration on the support of the Republican party in 2016 as well as its effect on the economic assimilation of recent foreign-borns. Since the estimates of the OLS regression might not be reliable due to possible endogeneity in the treatment variable, we implement a TSLS analysis where we instrument with a "leave-out" version of the shift-share instrument. We compare the signs and significance of the estimated causal effects with the estimated ACLATES of the instrumental forest. The latter method offers the advantage of identifying potential effect modifiers from the full set of controls, giving an insight into the underlying mechanisms of the ACLATEs.

Using TSLS, we find that historical immigration negatively affects the support of the Republican party nowadays and increases the wage gap between the natives and foreign-borns. However, when allowing for a more flexible specification with the instrumental forest, we do not find a significant average effect of historical immigration on the voting behavior or economic assimilation of immigrants. Nevertheless, we do find significant ACLATEs for subgroups of the population. The identified effect modifiers suggest two mechanisms underlying the effect of historical immigration. Firstly, we find evidence supporting the finding of Giuliano and Tabellini (2020) that liberal values of the European immigrants are passed on to the natives. Thus, the assimilation of the foreign-borns appears to be a two-sided process. Secondly, the treatment effect modifiers suggest that immigrants establish migrant communities which persist over time. This supports the observation that migrants tend to settle in established ethnic communities which offer some economic advantages, enhancing economic assimilation (Bartel, 1989; Portes and Zhou, 1993). It should be noted that this simultaneously might lead to an increase in the segmentation between the native population and the foreign-borns, as stated by Alba and Nee (1997).

The findings in this paper demonstrate the advantages of an instrumental forest compared to the TSLS method. Potential effect modifiers were identified in a data-driven fashion. Some of these effect modifiers might otherwise not have been selected ex ante as a subgroup. Moreover, the difference in significance between the treatment effects estimated by the TSLS regressions and the effects estimated by the instrumental forests highlights the importance of accounting for nonlinearities in the data.

The findings have some policy indications. The results suggest an increased segmentation between the native population and the immigrants due to the settlement into ethnic communities of the latter. Although immigrants might assimilate economically within their communities, the segmentation might threaten the social cohesion of the population (Gathmann and Keller, 2018). Moreover, the segmentation impedes the two-sided transmission of values between the foreign-borns and natives, thereby threatening the "melting-pot" society (Giuliano and Tabellini, 2020). Therefore, policies should be adjusted such that the population becomes more inclusive of immigrants and subsequently less segmented.

7 Limitations and further research

The data, methods, and results in this paper have some limitations and offer starting points for further research. Firstly, it could be of interest to further extend the created data set. It would be of added value to include the characteristics of the immigrants who arrived during the Age of Mass Migration in the set of controls, as these might be potential treatment effect modifiers. This data was not available for our study. Moreover, the results might change when including data on the unregistered, recent immigrants.

Finally, we aggregated the data from IPUMS without considering the possibility of the data containing only subgroups of the population. It might be of interest to investigate sample selection and the necessity of weighting the data in the aggregation process.

Regarding the methods implemented in this paper, the instrumental forest has several shortcomings. It currently only allows for estimation of the doubly robust ACLATE in case of a binary treatment variable. Moreover, it only allows for the specification of one endogenous variable, and hence it is not possible to simultaneously instrument for the possibly endogenous control for the average immigrant share between 1970 and 2010.

There further are some possible improvements regarding the propensity scores. Although the results of the causal forest implied that overlap-weighting was not necessary in our application, extending the method to allow for overlap-weighting might improve the results. Instead of overlap-weighting, trimming the propensity scores such that they are bounded away from zero and one might also be of interest in future research. Furthermore, we estimated the propensity scores using regression forests. Other methods for the estimation of the propensity scores might be investigated as they might yield improved results.

Finally, it should be noted that the long-run effect of the immigration during the Age of Mass Migration cannot be generalized. The estimated effects only hold for the specific sample and instrument used in this paper. In addition, the composition of sending countries is very different nowadays compared to the past. While the majority of immigrants between 1870 and 1920 came from Europe, recent immigrants are mainly from Latin America and Asia. This emphasizes the importance of extending the data set with migrant characteristics for future research.

References

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

Abramitzky, R. and Boustan, L. (2017). Immigration in american economic history. *Journal of economic literature*, 55(4):1311–45.

Acemoglu, D., Johnson, S., and Robinson, J. A. (2005). Institutions as a fundamental cause of long-run growth. *Handbook of economic growth*, 1:385–472.

Ager, P. and Brückner, M. (2013). Cultural diversity and economic growth: Evidence from the us during the age of mass migration. *European Economic Review*, 64:76–97.

Ager, P. and Hansen, C. W. (2017). Closing heaven's door: Evidence from the 1920s us immigration quota acts. *Available at SSRN 3059439*.

Akee, R., Copeland, W., Costello, E. J., and Simeonova, E. (2018). How does household income affect child personality traits and behaviors? *American Economic Review*, 108(3):775–827.

Alba, R. and Nee, V. (1997). Rethinking assimilation theory for a new era of immigration. *International migration review*, 31(4):826–874.

Angrist, J. D. and Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Athey, S. and Imbens, G. (2015). A measure of robustness to misspecification. *American Economic Review*, 105(5):476–80.

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

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

Bartel, A. P. (1989). Where do the new us immigrants live? *Journal of Labor Economics*, 7(4):371–391.

Borjas, G. J. (2015). The slowdown in the economic assimilation of immigrants: Aging and cohort effects revisited again. *Journal of Human Capital*, 9(4):483–517.

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

Butcher, K. F. and DiNardo, J. (2002). The immigrant and native-born wage distributions: evidence from united states censuses. *ILR Review*, 56(1):97–121.

Card, D. (2001). Immigrant inflows, native outflows, and the local labor market impacts of higher immigration. *Journal of Labor Economics*, 19(1):22–64.

Chernozhukov, V., Escanciano, J. C., Ichimura, H., Newey, W. K., and Robins, J. M. (2016). Locally robust semiparametric estimation. *arXiv preprint arXiv:1608.00033*.

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

Droller, F. (2018). Migration, population composition and long run economic development: Evidence from settlements in the pampas. *The Economic Journal*, 128(614):2321–2352.

Duncan, G. J., Magnuson, K., and Votruba-Drzal, E. (2014). Boosting family income to promote child development. *The Future of Children*, pages 99–120.

Dustmann, C., Vasiljeva, K., and Piil Damm, A. (2019). Refugee migration and electoral outcomes. *The Review of Economic Studies*, 86(5):2035–2091.

Gathmann, C. and Keller, N. (2018). Access to citizenship and the economic assimilation of immigrants. *The Economic Journal*, 128(616):3141–3181.

Giuliano, P. and Tabellini, M. (2020). The seeds of ideology: Historical immigration and political preferences in the united states. Technical report, National Bureau of Economic Research.

Grand, C. l. and Szulkin, R. (2002). Permanent disadvantage or gradual integration: explaining the immigrant–native earnings gap in sweden. *Labour*, 16(1):37–64.

Halla, M., Wagner, A. F., and Zweimüller, J. (2012). Immigration and voting for the extreme right. Technical report, Working paper.

Halla, M., Wagner, A. F., and Zweimüller, J. (2017). Immigration and voting for the far right. *Journal of the European Economic Association*, 15(6):1341–1385.

Heckman, J. J. and Vytlacil, E. (2005). Structural equations, treatment effects, and econometric policy evaluation 1. *Econometrica*, 73(3):669–738.

Johannemann, J., Hadad, V., Athey, S., and Wager, S. (2019). Sufficient representations for categorical variables. *arXiv preprint arXiv:1908.09874*.

Li, F., Morgan, K. L., and Zaslavsky, A. M. (2018). Balancing covariates via propensity score weighting. *Journal of the American Statistical Association*, 113(521):390–400.

Manson, S., Schroeder, J., Van Riper, D., and Ruggles, S. (2019). Ipums national historical geographic information system: Version 14.0 [database]. minneapolis, mn: Ipums. 2019.

Maurin, E. (2002). The impact of parental income on early schooling transitions: A re-examination using data over three generations. *Journal of public Economics*, 85(3):301–332.

Mayda, A. M., Peri, G., and Steingress, W. (2016). Immigration to the us: A problem for the republicans or the democrats? Technical report, National Bureau of Economic Research.

Mayda, A. M., Peri, G., and Steingress, W. (2018). The political impact of immigration: Evidence from the united states. Technical report, National Bureau of Economic Research.

MIT Election Data and Science Lab (2018). County Presidential Election Returns 2000-2016.

Portes, A. and Zhou, M. (1993). The new second generation: Segmented assimilation and its variants. *The annals of the American academy of political and social science*, 530(1):74–96.

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American statistical Association*, 89(427):846–866.

Rocha, R., Ferraz, C., and Soares, R. R. (2017). Human capital persistence and development. *American Economic Journal: Applied Economics*, 9(4):105–36.

Rodríguez-Pose, A. and Von Berlepsch, V. (2014). When migrants rule: the legacy of mass migration on economic development in the united states. *Annals of the Association of American Geographers*, 104(3):628–651.

Sequeira, S., Nunn, N., and Qian, N. (2020). Immigrants and the making of america. *The Review of Economic Studies*, 87(1):382–419.

Smith, W. C. and Fernandez, F. (2015). Education and wage gaps: A comparative study of immigrant and native employees in the united states and canada. *Program for the International Assessment of Adult Competencies*.

Tabellini, G. (2010). Culture and institutions: economic development in the regions of europe. *Journal of the European Economic association*, 8(4):677–716.

Tabellini, M. (2020). Gifts of the immigrants, woes of the natives: Lessons from the age of mass migration. *The Review of Economic Studies*, 87(1):454–486.

Tibshirani, J., Athey, S., Friedberg, R., Hadad, V., Miner, L., Wager, S., and Wright, M. (2020). *grf: Generalized Random Forests (Beta)*. URL https://github.com/grf-labs/grf. R package version 1.2.0.

Villarreal, A. and Tamborini, C. R. (2018). Immigrants' economic assimilation: Evidence from longitudinal earnings records. *American Sociological Review*, 83(4):686–715.

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

Zúñiga, V. and Hernández-León, R. (2005). *New destinations: Mexican immigration in the United States*. Russell Sage Foundation.

8 Appendix

8.1 Variable descriptions and corresponding sources

Table 12: Variable names, descriptions, and sources

Variable	Description (Source)
Outcomes	
repvoteshare2016	Republican vote share in presidential elections of 2016 (Data and Lab, 2018)
logwagenatfb	Wage gap between natives and foreign borns in 2016 (IPUMS)
logoccnatfb	Unemployment gap between natives and foreign borns in 2016 (IPUMS)
unempnatfb	Occupational income score gap between natives and foreign borns in 2016 (IPUMS)
Immigrant variables	
avgimmshare1870_1920	Average immigrant share for 1870 - 1920 (NHGIS)
avgimmshare1930_1950	Average immigrant share for 1930 - 1950 (NHGIS)
avgimmshare1970_2010	Average immigrant share for 1970 - 2010 (NHGIS)
County controls 1870	
male1870	Share of male population in 1870 (NHGIS)
hispaniclatinos1870	Share of Hispanics and Latinos in 1870 (NHGIS)
africanamericans1870	Share of African Americans in 1870 (NHGIS)
illiteracy1870	Illiteracy rate in 1870 (NHGIS)
urban1870	Share of urban population in 1870 (NHGIS)
manufacturingwages1870	Average of manufacturing wages in 1870 (NHGIS)
agriculturalwages1870	Average of agricultural wages in 1870 (NHGIS)
manufacturingemployed1870	Share of population employed in manufacturing in 1870 (NHGIS)
personalestate1870	Average personal estate in 1870 (NHGIS)
railway1860	Railway connection in 1860 (NHGIS)
Demographics 2016	
malenatives2016	Share of males of native population in 2016 (IPUMS)
malefb2016	Share of males of foreign-born population in 2016 (IPUMS)
agenatives2016	Average age of native population in 2016 (IPUMS)
agefb2016	Average age of foreign-born population in 2016 (IPUMS)
marriednatives2016	Share of married natives in 2016 (IPUMS)
marriedfb2016	Share of married foreign borns in 2016 (IPUMS)
educnatives2016	Average years of schooling of native population in 2016 (IPUMS)
educfb2016	Average years of schooling of foreign-born population in 2016 (IPUMS)
highdegreenatives2016	Share of natives with a college degree or higher in 2016 (IPUMS)
highdegreefb2016	Share of foreign borns with a college degree or higher in 2016 (IPUMS)
noenglishnatives2016	Share of natives who do not speak English in 2016 (IPUMS)
noenglishfb2016	Share of foreign borns who do not speak English in 2016 (IPUMS)
Geographical controls	
longitude2000	Longitude value of a county's center in 2000 (United States Census Bureau)
latitude2000	Latitude value of a county's center in 2000 (United States Census Bureau)
urban2010	Average share of urban population in 2010 (NHGIS)

8.2 TSLS checks for the Republican vote share outcome

Table 13: TSLS checks for the Republican vote share outcome

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
avgimmshare1870_1920	0.076^{*}	-0.200^{***}	-0.124^{***}			-0.123^{***}	-0.075^{***}		
	(0.044)	(0.039)	(0.016)			(0.040)	(0.014)		
avgimmshare1900_1920				-0.049^{*}				-0.034^{**}	
				(0.029)				(0.015)	
bin_avgimmshare1870_1920					-0.625^{***}				-0.426^{**}
					(0.204)				(0.202)
county controls 1870	no	yes	yes	yes	yes	yes	yes	yes	yes
demographics 2016	no	no	no	no	no	yes	yes	yes	yes
geographical controls	no	yes	yes	yes	yes	yes	yes	yes	yes
avgimmshare1970_2010	no	no	no	no	no	yes	yes	yes	yes
state fixed effects	yes	yes	yes	yes	yes	yes	yes	yes	yes
clustered standard errors	yes	yes	no	yes	yes	yes	no	yes	yes
weighted by total votes 2016	yes	no	yes	yes	yes	no	yes	yes	yes
N	2131	1860	1860	1860	1860	1854	1854	1854	1854
R^2	-0.481	0.153	0.429	0.427	-0.481	0.434	0.663	0.684	0.247
AIC	-1374.5	-2764.9	-2940.1	-2933.7	-1168.5	-3480.8	-3869.2	-3992.4	-2380.2
BIC	-1368.8	-2693.0	-2868.3	-2861.8	-1096.6	-3304.0	-3686.9	-3815.6	-2203.4
Kleibergen Paap F-statistic	9.99	30.18		10.74	11.95	6.05		8.49	3.50
SW F-statistic avgimmshare1870_1920			182.64			12.51	169.63		
SW F-statistic avgimmshare1900_1920								18.24	
SW F-statistic bin_avgimmshare1870_1920									7.02
SW F-statistic avgimmshare1970_2010						13.07	269.67	36.77	9.70

Standard errors in parentheses. The dependent variable is the Republican vote share in 2016. Column (1) contains the specification without any controls. Columns (2) to (5) contain the historical controls. Columns (6) to (9) additionally contain the recent controls. The treatment variable in columns (4) and (8) is the average immigrant share between 1900 and 1920. The treatment variable in columns (5) and (9) is a binary variable which equals one when the average immigrant share between 1870 and 1920 is higher than the median value of this variable.

Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

8.3 TSLS checks for the wage gap outcome

Table 14: TSLS checks for the wage gap outcome

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
avgimmshare1870_1920	0.177***	0.214*	0.030			0.112	0.081		
	(0.015)	(0.129)	(0.048)			(0.112)	(0.067)		
avgimmshare1900_1920				0.263				0.098	
				(0.419)				(0.070)	
bin_avgimmshare1870_1920					0.127				0.407**
					(0.113)				(0.178)
county controls 1870	no	yes	yes	yes	yes	yes	yes	yes	yes
demographics 2016	no	no	no	no	no	yes	yes	yes	yes
geographical controls	no	yes	yes	yes	yes	yes	yes	yes	yes
avgimmshare1970_2010	no	no	no	no	no	yes	yes	yes	yes
state fixed effects	yes	yes	yes	yes	yes	yes	yes	yes	yes
clustered standard errors	yes	yes	no	yes	yes	yes	no	yes	yes
weighted by persons 2016	yes	no	yes	yes	yes	no	yes	yes	yes
N	2127	1857	1857	1857	1857	1856	1856	1856	1856
R^2	-0.333	-0.067	0.001	-0.433	-0.019	0.125	0.132	0.119	-0.050
AIC	488.0	1448.3	-115.3	553.9	-78.74	1109.7	-345.4	-318.0	7.930
BIC	493.6	1520.2	-43.48	625.7	-6.889	1270.0	-185.1	-157.7	168.2
Kleibergen Paap F-statistic	123.75	30.29		0.47	7.52	13.02		0.96	2.42
SW F-statistic avgimmshare1870_1920			90.01			29.91	53.82		
SW F-statistic avgimmshare1900_1920								2.53	
SW F-statistic bin_avgimmshare1870_1920									5.16
SW F-statistic avgimmshare1970_2010						27.55	82.80	8.71	4.74

Standard errors in parentheses. The dependent variable is the Republican vote share in 2016. Column (1) contains the specification without any controls. Columns (2) to (5) contain the historical controls. Columns (6) to (9) additionally contain the recent controls. The treatment variable in columns (4) and (8) is the average immigrant share between 1900 and 1920. The treatment variable in columns (5) and (9) is a binary variable which equals one when the average immigrant share between 1870 and 1920 is higher than the median value of this variable.
Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

8.4 TSLS checks for the occupational income score gap outcome

Table 15: TSLS checks for the occupational income score gap outcome

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
avgimmshare1870_1920	-0.096***	0.043	0.002			0.058	0.002		
	(0.008)	(0.034)	(0.013)			(0.041)	(0.018)		
avgimmshare1900_1920				0.092				0.042	
				(0.136)				(0.029)	
bin_avgimmshare1870_1920					0.007				0.009
					(0.045)				(0.062)
county controls 1870	no	yes	yes	yes	yes	yes	yes	yes	yes
demographics 2016	no	no	no	no	no	yes	yes	yes	yes
geographical controls	no	yes	yes	yes	yes	yes	yes	yes	yes
avgimmshare1970_2010	no	no	no	no	no	yes	yes	yes	yes
state fixed effects	yes	yes	yes	yes	yes	yes	yes	yes	yes
clustered standard errors	yes	yes	no	yes	yes	yes	no	yes	yes
weighted by persons 2016	yes	no	yes	yes	yes	no	yes	yes	yes
N	2127	1858	1858	1858	1858	1856	1856	1856	1856
R^2	-1.229	-0.019	0.012	-0.716	0.011	0.058	0.116	0.030	0.116
AIC	-4007.8	-3569.1	-5101.8	-4076.7	-5099.8	-3690.7	-5272.5	-5099.7	-5271.5
BIC	-4002.2	-3497.2	-5029.9	-4004.9	-5028.0	-3530.4	-5112.3	-4939.4	-5111.2
Kleibergen Paap F-statistic	122.55	30.94		0.47	7.35	13.02		0.96	2.38
SW F-statistic avgimmshare1870_1920			90.05			29.91	53.59		
SW F-statistic avgimmshare1900_1920								2.53	
SW F-statistic bin_avgimmshare1870_1920									5.05
SW F-statistic avgimmshare1970_2010						27.55	82.62	8.64	4.68

Standard errors in parentheses. The dependent variable is the Republican vote share in 2016. Column (1) contains the specification without any controls. Columns (2) to (5) contain the historical controls. Columns (6) to (9) additionally contain the recent controls. The treatment variable in columns (4) and (8) is the average immigrant share between 1900 and 1920. The treatment variable in columns (5) and (9) is a binary variable which equals one when the average immigrant share between 1870 and 1920 is higher than the median value of this variable.
Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

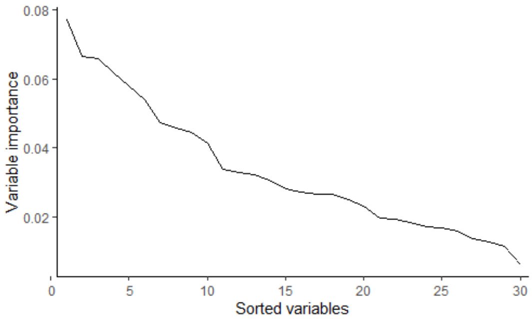
8.5 TSLS checks for the unemployment gap outcome

Table 16: TSLS checks for the unemployment gap outcome

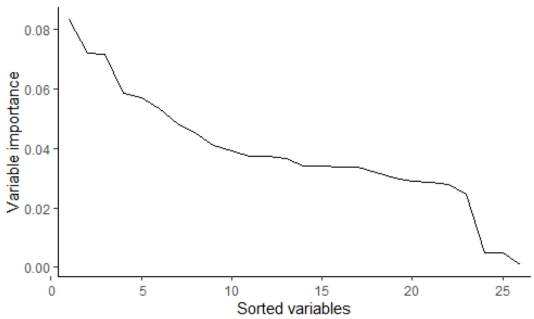
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
avgimmshare1870_1920	-0.002	-0.022***	-0.009			-0.022*	-0.009		
	(0.001)	(0.008)	(0.006)			(0.013)	(0.009)		
avgimmshare1900_1920				-0.012				-0.007	
				(0.018)				(0.009)	
bin_avgimmshare1870_1920					-0.040**				-0.046
					(0.019)				(0.039)
county controls 1870	no	yes	yes	yes	yes	yes	yes	yes	yes
demographics 2016	no	no	no	no	no	yes	yes	yes	yes
geographical controls	no	yes	yes	yes	yes	yes	yes	yes	yes
avgimmshare1970_2010	no	no	no	no	no	yes	yes	yes	yes
state fixed effects	yes	yes	yes	yes	yes	yes	yes	yes	yes
clustered standard errors	yes	yes	no	yes	yes	yes	no	yes	yes
weighted by persons 2016	yes	no	yes	yes	yes	no	yes	yes	yes
N	2124	1856	1856	1856	1856	1854	1854	1854	1854
R^2	-0.001	-0.059	-0.015	-0.035	-0.096	-0.032	0.011	0.018	-0.111
AIC	-8907.0	-6516.5	-7878.3	-7842.6	-7734.8	-6525.4	-7885.4	-7899.3	-7670.9
BIC	-8901.3	-6444.7	-7806.4	-7770.8	-7663.0	-6365.1	-7725.2	-7739.1	-7510.6
Kleibergen Paap F-statistic	121.05	13.54		0.46	5.91	6.16		0.95	2.08
SW F-statistic avgimmshare1870_1920			94.80			12.82	56.47		
SW F-statistic avgimmshare1900_1920								2.49	
SW F-statistic bin_avgimmshare1870_1920									4.21
SW F-statistic avgimmshare1970_2010						13.26	88.13	8.48	4.41

Standard errors in parentheses. The dependent variable is the Republican vote share in 2016. Column (1) contains the specification without any controls. Columns (2) to (5) contain the historical controls. Columns (6) to (9) additionally contain the recent controls. The treatment variable in columns (4) and (8) is the average immigrant share between 1900 and 1920. The treatment variable in columns (5) and (9) is a binary variable which equals one when the average immigrant share between 1870 and 1920 is higher than the median value of this variable.
Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

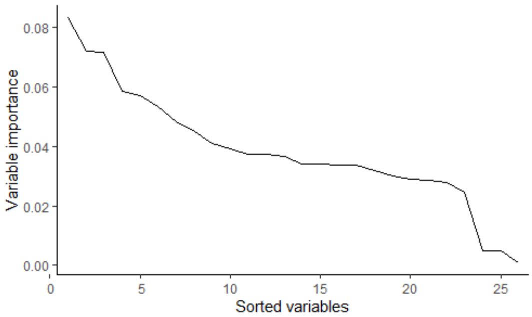
8.6 Variable importances for the instrumental forest



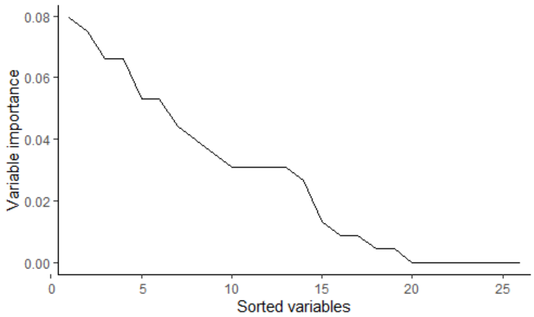
(a) Sorted variable importances in first forest for the Republican vote share outcome



(b) Sorted variable importances in first forest for the wage gap outcome



(c) Sorted variable importances in first forest for the occupational income score gap outcome



(d) Sorted variable importances in first forest for the unemployment score gap outcome

Figure 5: Sorted variable importances in first forests

Table 17: Variable importances of the variables in the first forest

Variable	(1) Republican vote share, 2016	(2) Wage gap, 2016	(3) Occupational income score gap, 2016	(4) Unemployment gap, 2016
avgimmshare1970_2010	6.60%	3.03%	2.90%	0.88%
male1870	7.72%	2.78%	3.41%	0.00%
hispaniclatinos1870	6.18%	2.64%	3.93%	0.00%
africanamericans1870	6.66%	4.29%	3.40%	0.88%
illiteracy1870	3.38%	2.70%	3.74%	0.00%
urban1870	1.95%	0.37%	0.49%	0.00%
manufacturingwages1870	4.44%	1.97%	3.65%	2.65%
agriculturalwages1870	3.28%	2.06%	3.02%	5.30%
manufacturingemployed1870	2.64%	2.24%	2.87%	0.44%
personalestate1870	1.57%	2.14%	5.34%	3.09%
railway1860	0.62%	0.05%	0.10%	0.00%
agenatives2016	3.08%	2.22%	4.10%	4.42%
agefb2016	1.96%	9.53%	4.49%	7.95%
malefb2016	1.73%	5.13%	3.21%	7.51%
malenatives2016	4.72%	2.68%	2.80%	3.09%
marriednatives2016	3.24%	3.19%	3.37%	6.63%
marriedfb2016	2.81%	6.65%	2.47%	3.98%
educfb2016	2.32%	3.54%	5.86%	3.53%
educnatives2016	2.73%	5.30%	8.34%	5.30%
highdegreenatives2016	1.67%	5.68%	7.15%	3.09%
highdegreefb2016	1.83%	12.18%	7.20%	3.09%
noenglishnatives2016	1.15%	1.03%	0.51%	0.00%
noenglishfb2016	5.39%	4.24%	3.37%	6.63%
incwagenatives2016	2.51%			
incwagefb2016	2.67%			
unempnatives2016	1.26%			
unempfb2016	1.37%			
longitude2000	4.13%	5.46%	5.72%	0.44%
latitude2000	4.59%	6.50%	4.82%	0.00%
urban2010	5.80%	2.38%	3.74%	1.33%

Table 18: Variable importances of the variables in the second forest

Variable	(1) Republican vote share, 2016	(2) Wage gap, 2016	(3) Occupational income score gap, 2016	(4) Unemployment gap, 2016
avgimmshare1970_2010	13.53%			
male1870	17.86%			
hispaniclatinos1870	11.15%			
africanamericans1870	16.43%	9.60%		
agriculturalwages1870				5.72%
personalestate1870				9.39%
malenatives2016	13.87%			18.09%
malefb2016		9.19%		
agenatives2016				4.21%
agefb2016		7.78%		16.95%
marriednatives2016				5.84%
educnatives2016		12.66%	25.36%	6.06%
educfb2016				11.18%
highdegreenatives2016		12.75%	20.47%	
highdegreefb2016		14.19%	21.08%	6.82%
noenglishfb2016	13.91%	9.61%		
urban2010	13.26%			3.55%
longitude2000		11.31%	20.16%	
latitude2000		12.94%	12.93%	

8.7 ACLATEs without significant difference for the Republican vote share

Table 19: Average treatment effects and insignificant t-tests for assessing treatment heterogeneity for the Republican vote share - Part 1

Variable	ACLATE low	ACLATE high	95% CI difference	t-value difference	p-value difference
avgimmshare1970_2010	-0.008 (0.075)	-0.140 (0.180)	(-0.515, 0.249)	-0.680	0.496
male1870	-0.056 (0.104)	-0.529 (0.469)	(-1.414, 0.470)	-0.983	0.326
africanamericans1870	0.044 (0.130)	0.170 (0.118)	(-0.219, 0.469)	0.714	0.475
illiteracy1870	-0.035 (0.114)	0.054 (0.128)	(-0.247, 0.427)	0.522	0.602
urban1870	0.027 (0.082)	-0.153 (0.164)	(-0.540, 0.180)	-0.978	0.328
manufacturingwages1870	-0.169 (0.197)	0.117 (0.078)	(-0.130, 0.702)	1.348	0.178
agriculturalwages1870	0.009 (0.179)	-0.072 (0.151)	(-0.540, 0.378)	-0.344	0.731
manufacturingemployed1870	-0.066 (0.147)	0.008 (0.093)	(-0.267, 0.415)	0.423	0.672
personalestate1870	0.046 (0.056)	-0.024 (0.094)	(-0.285, 0.145)	-0.635	0.525
railway1860	-0.287 (0.260)	0.076 (0.052)	(-0.156, 0.882)	1.369	0.171
agenatives2016	-0.447 (0.350)	0.042 (0.081)	(-0.215, 1.191)	1.36	0.174
agefb2016	0.040 (0.110)	-0.133 (0.104)	(-0.468, 0.124)	-1.139	0.255
malefb2016	-0.088 (0.101)	-0.071 (0.114)	(-0.282, 0.316)	0.113	0.910
malenatives2016	0.124 (0.155)	-0.051 (0.07)	(-0.509, 0.159)	-1.027	0.304
marriednatives2016	-0.481 (0.362)	-0.150 (0.172)	(-0.453, 1.117)	0.828	0.408
marriedfb2016	0.003 (0.137)	0.009 (0.126)	(-0.358, 0.372)	0.037	0.970
educnatives2016	-0.153 (0.144)	0.017 (0.128)	(-0.207, 0.547)	0.884	0.377

Table 20: Average treatment effects and insignificant t-tests for assessing treatment heterogeneity for the Republican vote share - Part 2

Variable	ACLATE low	ACLATE high	95% CI difference	t-value difference	p-value difference
highdegreenatives2016	-0.139	0.041	(-0.179, 0.539)	0.981	0.327
	(0.146)	(0.111)			
highdegreefb2016	-0.336	-0.030	(-0.245, 0.857)	1.087	0.277
	(0.269)	(0.080)			
noenglishnatives2016	0.015	-0.081	(-0.354, 0.162)	-0.726	0.468
	(0.061)	(0.117)			
noenglishfb2016	0.053	-0.181	(-0.582, 0.114)	-1.317	0.188
	(0.071)	(0.163)			
incwagenatives2016	-0.197	-0.420	(-1.028, 0.582)	-0.544	0.587
	(0.143)	(0.385)			
incwagefb2016	-0.292	0.091	(-0.169, 0.933)	1.359	0.174
	(0.258)	(0.112)			
unempnatives2016	0.043	-0.292	(-0.875, 0.203)	-1.222	0.222
	(0.100)	(0.256)			
unempfb2016	-0.090	0.067	(-0.252, 0.566)	0.754	0.451
	(0.184)	(0.099)			
longitude2000	-0.198	0.108	(-0.363, 0.975)	0.897	0.370
	(0.268)	(0.212)			
latitude2000	0.134	0.053	(-0.454, 0.292)	-0.425	0.671
	(0.154)	(0.112)			
urban2010	-0.090	-0.137	(-0.385, 0.293)	-0.268	0.788
	(0.081)	(0.153)			

8.8 ACLATEs without significant difference for the wage gap

Table 21: Average treatment effects and insignificant t-tests for assessing treatment heterogeneity for the wage gap - Part 1

Variable	ACLATE low	ACLATE high	95% CI difference	t-value difference	p-value difference
avgimmshare1970_2010	-0.015 (0.154)	-0.280 (0.219)	(-0.789, 0.259)	-0.989	0.323
africanamericans1870	-0.021 (0.339)	-0.163 (0.112)	(-0.842, 0.558)	-0.397	0.691
illiteracy1870	0.110 (0.144)	-0.107 (0.136)	(-0.605, 0.169)	-1.103	0.270
urban1870	-0.250 (0.141)	-0.213 (0.205)	(-0.451, 0.525)	0.147	0.883
manufacturingwages1870	-0.148 (0.113)	-0.543 (0.263)	(-0.957, 0.167)	-1.379	0.168
agriculturalwages1870	-0.165 (0.346)	-0.475 (0.267)	(-1.168, 0.546)	-0.710	0.478
manufacturingemployed1870	-0.204 (0.186)	-0.317 (0.231)	(-0.695, 0.469)	-0.381	0.704
personalestate1870	-0.265 (0.194)	-0.379 (0.259)	(-0.749, 0.521)	-0.351	0.726
agenatives2016	0.049 (0.127)	-0.619 (0.402)	(-1.495, 0.157)	-1.586	0.113
agefb2016	-0.285 (0.236)	-0.148 (0.183)	(-0.448, 0.724)	0.461	0.645
malefb2016	-0.277 (0.226)	-0.497 (0.242)	(-0.870, 0.430)	-0.663	0.508
malenatives2016	-0.301 (0.329)	0.055 (0.191)	(-0.389, 1.103)	0.938	0.348
marriednatives2016	-0.026 (0.157)	-0.128 (0.191)	(-0.588, 0.382)	-0.415	0.678
marriedfb2016	-0.559 (0.275)	-0.476 (0.267)	(-0.669, 0.833)	0.215	0.830
educfb2016	-0.139 (0.226)	-0.455 (0.281)	(-1.023, 0.391)	-0.877	0.381
highdegreefb2016	-0.415 (0.336)	-0.496 (0.265)	(-0.921, 0.759)	-0.190	0.850

Table 22: Average treatment effects and insignificant t-tests for assessing treatment heterogeneity for the wage gap - Part 2

Variable	ACLATE low	ACLATE high	95% CI difference	t-value difference	p-value difference
noenglishnatives2016	-0.326	-0.228	(-0.314, 0.510)	0.468	0.640
	(0.154)	(0.143)			
noenglishfb2016	-0.357	-0.542	(-0.993, 0.625)	-0.446	0.655
	(0.292)	(0.292)			
longitude2000	-0.520	-0.021	(-0.25, 1.248)	1.305	0.192
	(0.254)	(0.286)			
latitude2000	-0.072	-0.256	(-0.909, 0.541)	-0.497	0.619
	(0.114)	(0.352)			
urban2010	-0.050	-0.183	(-0.832, 0.566)	-0.372	0.710
	(0.319)	(0.160)			

8.9 ACLATEs without significant difference for the occupational income score gap

Table 23: Average treatment effects and insignificant t-tests for assessing treatment heterogeneity for the occupational income score gap - Part 1

Variable	ACLATE low	ACLATE high	95% CI difference	t-value difference	p-value difference
avgimmshare1970_2010	-0.048	0.057	(-0.217, 0.429)	0.640	0.522
	(0.047)	(0.158)			
male1870	-0.098	-0.083	(-0.163, 0.193)	0.166	0.869
	(0.059)	(0.070)			
hispaniclatinos1870	-0.138	0.238	(-0.193, 0.945)	1.297	0.195
	(0.071)	(0.281)			
africanamericans1870	-0.030	-0.047	(-0.250, 0.214)	-0.148	0.882
	(0.113)	(0.036)			
illiteracy1870	-0.147	-0.040	(-0.105, 0.319)	0.994	0.320
	(0.100)	(0.042)			
urban1870	-0.064	0.012	(-0.248, 0.400)	0.457	0.648
	(0.038)	(0.161)			
manufacturingwages1870	-0.035	-0.194	(-0.372, 0.054)	-1.467	0.142
	(0.028)	(0.105)			
agriculturalwages1870	-0.021	0.004	(-0.344, 0.394)	0.135	0.893
	(0.060)	(0.178)			
manufacturingemployed1870	-0.069	-0.017	(-0.293, 0.395)	0.292	0.770
	(0.058)	(0.166)			
personalestate1870	-0.038	-0.217	(-0.439, 0.081)	-1.352	0.176
	(0.057)	(0.120)			
railway1860	-0.040	-0.001	(-0.221, 0.299)	0.292	0.770
	(0.050)	(0.123)			
agefb2016	-0.157	-0.029	(-0.066, 0.322)	1.29	0.197
	(0.078)	(0.062)			
malefb2016	0.016	-0.130	(-0.392, 0.100)	-1.163	0.245
	(0.079)	(0.098)			
malenatives2016	-0.038	0.068	(-0.066, 0.276)	1.207	0.228
	(0.055)	(0.067)			
marriedfb2016	-0.104	-0.080	(-0.137, 0.185)	0.294	0.769
	(0.054)	(0.062)			
educfb2016	-0.149	0.005	(-0.05, 0.356)	1.481	0.139
	(0.062)	(0.082)			
educnatives2016	0.010	0.051	(-0.782, 0.864)	0.099	0.922
	(0.418)	(0.040)			

Table 24: Average treatment effects and insignificant t-tests for assessing treatment heterogeneity for the occupational income score gap - Part 2

Variable	ACLATE low	ACLATE high	95% CI difference	t-value difference	p-value difference
highdegreenatives2016	0.085 (0.401)	0.034 (0.033)	(-0.839, 0.737)	-0.128	0.898
highdegreefb2016	-0.156 (0.077)	-0.036 (0.133)	(-0.181, 0.421)	0.778	0.437
noenglishnatives2016	-0.096 (0.054)	-0.013 (0.094)	(-0.129, 0.295)	0.764	0.445
longitude2000	-0.201 (0.108)	-0.113 (0.083)	(-0.178, 0.354)	0.649	0.516
latitude2000	-0.071 (0.047)	0.062 (0.310)	(-0.481, 0.747)	0.426	0.670
urban2010	-0.073 (0.047)	0.059 (0.141)	(-0.159, 0.421)	0.888	0.374

8.10 ACLATEs without significant difference for the unemployment gap

Table 25: Average treatment effects and insignificant t-tests for assessing treatment heterogeneity for the unemployment gap - Part 1

Variable	ACLATE low	ACLATE high	95% CI difference	t-value difference	p-value difference
avgimmshare1970_2010	-0.001 (0.020)	0.004 (0.022)	(-0.053, 0.063)	0.165	0.869
male1870	-0.032 (0.027)	0.012 (0.039)	(-0.048, 0.136)	0.926	0.355
hispaniclatinos1870	0.003 (0.017)	-0.025 (0.029)	(-0.094, 0.038)	-0.827	0.408
africanamericans1870	0.021 (0.021)	-0.003 (0.037)	(-0.108, 0.060)	-0.571	0.568
illiteracy1870	0.006 (0.020)	-0.023 (0.030)	(-0.100, 0.042)	-0.792	0.428
urban1870	0.010 (0.015)	-0.015 (0.017)	(-0.070, 0.020)	-1.067	0.286
manufacturingwages1870	0.000 (0.029)	0.013 (0.021)	(-0.056, 0.084)	0.386	0.700
agriculturalwages1870	-0.021 (0.014)	0.004 (0.022)	(-0.026, 0.078)	0.975	0.329
manufacturingemployed1870	-0.034 (0.024)	-0.014 (0.016)	(-0.036, 0.076)	0.708	0.479
personalestate1870	-0.026 (0.023)	0.012 (0.026)	(-0.031, 0.107)	1.091	0.275
railway1860	0.022 (0.023)	-0.018 (0.014)	(-0.093, 0.013)	-1.456	0.146
agenatives2016	0.019 (0.018)	0.017 (0.033)	(-0.078, 0.072)	-0.076	0.939
agefb2016	0.008 (0.032)	0.005 (0.031)	(-0.090, 0.084)	-0.078	0.938
malefb2016	0.014 (0.024)	0.000 (0.024)	(-0.082, 0.052)	-0.436	0.663
marriednatives2016	0.013 (0.017)	-0.037 (0.029)	(-0.118, 0.016)	-1.497	0.134
marriedfb2016	-0.020 (0.035)	0.017 (0.021)	(-0.044, 0.118)	0.903	0.367
educfb2016	0.003 (0.033)	0.053 (0.038)	(-0.049, 0.149)	0.988	0.323

Table 26: Average treatment effects and insignificant t-tests for assessing treatment heterogeneity for the unemployment gap - Part 2

Variable	ACLATE low	ACLATE high	95% CI difference	t-value difference	p-value difference
educnatives2016	-0.009 -(0.050)	-0.004 (0.016)	(-0.099, 0.109)	0.092	0.927
highdegreenatives2016	-0.016 (0.052)	-0.001 (0.019)	(-0.094, 0.124)	0.270	0.787
highdegreefb2016	0.013 (0.028)	0.057 (0.037)	(-0.048, 0.134)	0.928	0.353
noenglishnatives2016	0.003 (0.014)	-0.004 (0.012)	(-0.044, 0.03)	-0.371	0.711
noenglishfb2016	0.006 (0.026)	-0.010 (0.029)	(-0.092, 0.06)	-0.424	0.672
longitude2000	0.009 (0.023)	-0.013 (0.038)	(-0.109, 0.065)	-0.492	0.623
latitude2000	0.010 (0.040)	-0.033 (0.024)	(-0.135, 0.047)	-0.94	0.347
urban2010	0.017 (0.026)	0.007 (0.018)	(-0.072, 0.052)	-0.323	0.747

8.11 Results instrumental forest without tuning

Table 27: Average treatment effects and t-test for assessing treatment heterogeneity for the double forest without tuning

	(1) Republican vote share, 2016	(2) Wage gap, 2016	(3) Occupational income score gap, 2016	(4) Unemployment gap, 2016
ACLATE	0.044	-0.084	0.018	-0.008
	(0.103)	(0.126)	(0.091)	(0.011)
	t-value: 0.428	t-value: -0.667	t-value: 0.194	t-value: -0.699
95% CI	(-0.157, 0.245)	(-0.332, 0.164)	(-0.16, 0.196)	(-0.030, 0.014)
ACLATE (high)	0.155	0.025	-0.026	0.001
	(0.117)	(0.116)	(0.047)	(0.019)
	t-value: 1.331	t-value: 0.214	t-value: -0.551	t-value: 0.045
ACLATE (low)	-0.032	-0.223	0.084	-0.016
	(0.168)	(0.227)	(0.200)	(0.020)
	t-value: -0.188	t-value: -0.982	t-value: 0.419	t-value: -0.784
95% CI for difference	(-0.214, 0.588)	(-0.251, 0.747)	(-0.513, 0.293)	(-0.038, 0.072)
observations	1854	1856	1856	1854
controls	30	26	26	26
controls selected in second forest	7	9	10	10
clustered std. errors	yes	yes	yes	yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Doubly robust ACLATEs. Standard errors in parantheses.

8.12 Results causal forest with binary treatment

Table 28: Average treatment effects and t-test for assessing treatment heterogeneity for the double causal forest with binary treatment

	(1) Republican vote share, 2016	(2) Wage gap, 2016	(3) Occupational income score gap, 2016	(4) Unemployment gap, 2016
ACLATE	-0.040**	-0.036**	0.001	-0.001
	(0.020)	(0.015)	(0.007)	(0.004)
	t-value: -2.002	t-value: -2.337	t-value: 0.106	t-value: -0.345
95% CI	(-0.080, 0.000)	(-0.066, -0.006)	(-0.012, 0.014)	(-0.008, 0.006)
ACLATE (high)	-0.030	0.006	-0.002	-0.001
	(0.029)	(0.021)	(0.009)	(0.004)
	t-value: -1.040	t-value: 0.295	t-value: -0.223	t-value: -0.312
ACLATE (low)	-0.052**	-0.083***	0.001	0.000
	(0.026)	(0.021)	(0.007)	(0.004)
	t-value: -1.975	t-value: -4.001	t-value: 0.159	t-value: -0.018
95% CI for difference	(-0.054, 0.098)	(0.032, 0.148)	(-0.025, 0.019)	(-0.012, 0.01)
observations	1854	1856	1856	1854
controls	30	26	26	26
controls selected in second forest	6	5	8	5
clustered std. errors	yes	yes	yes	yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Doubly robust ACLATEs. Standard errors in parantheses.

Table 29: Parameter tuning

Parameter	(1) Republican vote share, 2016	(2) Wage gap, 2016	(3) Occupational income score gap, 2016	(4) Unemployment gap, 2016
sample.fraction	0.46	0.42	0.38	0.42
mtry	6.00	3.00	4.00	3.00
min.node.size	1.00	1.00	4.00	51.00
honesty.fraction	0.63	0.64	0.66	0.57
honesty.prune.leaves	0.00	1.00	0.00	1.00
alpha	0.10	0.01	0.10	0.10
imbalance.penalty	0.60	0.27	1.00	0.049

8.13 Results causal forest with overlap-weighting

Table 30: Average treatment effects and t-test for assessing treatment heterogeneity for double causal forest with overlap-weighting

	(1) Republican vote share, 2016	(2) Wage gap, 2016	(3) Occupational income score gap, 2016	(4) Unemployment gap, 2016
ACLATE	-0.023	-0.030**	-0.003	0.001
	(0.025)	(0.015)	(0.006)	(0.004)
	t-value: -0.945	t-value: -1.943	t-value: -0.489	t-value: 0.15
95% CI	(-0.071, 0.025)	(-0.06, 0)	(-0.014, 0.008)	(-0.006, 0.008)
ACLATE (high)	-0.015	0.002	-0.01	-0.001
	(0.024)	(0.023)	(0.01)	(0.004)
	t-value: -0.625	t-value: 0.094	t-value: -1.004	t-value: -0.322
ACLATE (low)	-0.021	-0.063***	0.003	0.003
	(0.039)	(0.023)	(0.006)	(0.003)
	t-value: -0.538	t-value: -2.767	t-value: 0.609	t-value: 1.152
95% CI for difference	(-0.083, 0.095)	(0.001, 0.129)	(-0.035, 0.009)	(-0.015, 0.005)
observations	1854	1856	1856	1854
controls	30	26	26	26
controls selected in second forest	6	5	8	5
clustered std. errors	yes	yes	yes	yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Doubly robust ACLATEs. Standard errors in parantheses.

Table 31: Parameter tuning

Parameter	(1) Republican vote share, 2016	(2) Wage gap, 2016	(3) Occupational income score gap, 2016	(4) Unemployment gap, 2016
sample.fraction	0.46	0.42	0.38	0.42
mtry	6.00	3.00	4.00	3.00
min.node.size	1.00	1.00	4.00	51.00
honesty.fraction	0.63	0.64	0.66	0.57
honesty.prune.leaves	0.00	1.00	0.00	1.00
alpha	0.10	0.01	0.10	0.10
imbalance.penalty	0.60	0.27	1.00	0.049

8.14 Results causal forest with continuous treatment

Table 32: Average treatment effects and t-test for assessing treatment heterogeneity for double causal forest with continuous treatment

	(1) Republican vote share, 2016	(2) Wage gap, 2016	(3) Occupational income score gap, 2016	(4) Unemployment gap, 2016
ACLATE	-0.191	0.101	-0.024	0.039
	(0.129)	(0.187)	(0.049)	(0.031)
	t-value: -1.485	t-value: 0.539	t-value: -0.488	t-value: 1.271
95% CI	(-0.444, 0.062)	(-0.266, 0.468)	(-0.12, 0.072)	(-0.022, 0.1)
ACLATE (high)	-0.232	-0.053	-0.042	-0.003
	(0.195)	(0.177)	(0.062)	(0.046)
	t-value: -1.189	t-value: -0.298	t-value: -0.682	t-value: -0.063
ACLATE (low)	-0.08	0.093	-0.004	0.07*
	(0.132)	(0.22)	(0.082)	(0.037)
	t-value: -0.608	t-value: 0.421	t-value: -0.051	t-value: 1.904
95% CI for difference	(-0.613, 0.309)	(-0.698, 0.408)	(-0.24, 0.164)	(-0.188, 0.042)
observations	1854	1856	1856	1854
controls	30	26	26	26
controls selected in second forest	5	5	10	6
clustered std. errors	yes	yes	yes	yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Doubly robust ACLATEs. Standard errors in parantheses.

Table 33: Parameter tuning

Parameter	(1) Republican vote share, 2016	(2) Wage gap, 2016	(3) Occupational income score gap, 2016	(4) Unemployment gap, 2016
sample.fraction	0.22	0.40	0.49	0.18
mtry	1.00	3.00	6.00	4.00
min.node.size	1.00	2.00	2.00	1.00
honesty.fraction	0.51	0.56	0.77	0.54
honesty.prune.leaves	0.00	0.00	1.00	1.00
alpha	0.07	0.03	0.17	0.23
imbalance.penalty	1.21	1.21	0.19	1.93