Universidad de SanAndrés

Universidad de San Andrés

Departamento de Economía

Licenciatura en Economía

The effect of black population on the electoral performance of the Democratic Party: Evidence from Alabama and Mississippi

Autor: Juan Francisco Millas Caputo

Legajo: 31367

Mentor de Tesis: Tommy E. Murphy

Buenos Aires, Argentina

Septiembre de 2023

Summary

This thesis aims to estimate the effect of the percentage of black population in a county on the Democratic Party's performance in presidential elections in Alabama and Mississippi. The argument highlights the relevance of race when determining the political preference of southern voters in contrast to other explanatory variables, and also assesses its possible endogeneity within a model that ascertains the primary factors that drive the Democrat vote. The identification strategy consists of an IV model under exact identification. The instrument of choice is the percentage of a county's area whose soil belongs to the suborder Uderts, whose validity is sustained using historical and empirical evidence. The main takeaway is that the explanatory power of black population is 10.1% higher when instrumented through Uderts soils compared to an OLS estimation, where an increase of 1 *p.p.* in the black population of a county results in a rise of 0.958 *p.p.* in the votes the Democrats receive in presidential elections. This result is robust to variations in the measures of the dependent and independent variables, the inclusion of controls, and a series of robustness checks.

Motivation

The focus of this study is race, which could be considered among the most divisive issues among Americans, given the long-standing role that racism has played in American politics (Hersch & Nall, 2016). The reason behind this affirmation is the substantial differences in voters' preferences when discriminated by ethnicity. According to the Pew Research Center's (2020) study of electoral results from 1994 up to 2020, all major non-white ethnic groups have leaned towards the Democrats. However, the most striking contrast is seen in individuals identifying as Black or African American, where the percentage of respondents who preferred the Democrats, a whopping 81% in 1994, has remained almost unaltered ever since.

Despite only representing around 19% of the total voter base of the Democrats nationally, black people remain a significant portion of their electorate in the South, where all of the 72 black-majority counties in 2018 are located, according to Schaeffer's (2018) analysis of the Census Bureau population estimates. Figure 1 represents the percentage of votes received by the Democrat Party in the 2020 Election using data from the MIT Election Lab, and Figure 2 represents the percentage of the black population in 2020 using data from the 2020 Census. From visually comparing these two graphics, a crescent-shaped concentration across Mississippi,

Alabama, Georgia, South Carolina, North Carolina and Virginia seems to be present in both of them. This apparent relationship is also corroborated in the data, where the correlation between both variables is equal to 0.726. An interesting question that arises from this analysis is: How much does the presence of black people affect the Democrat vote in the South?

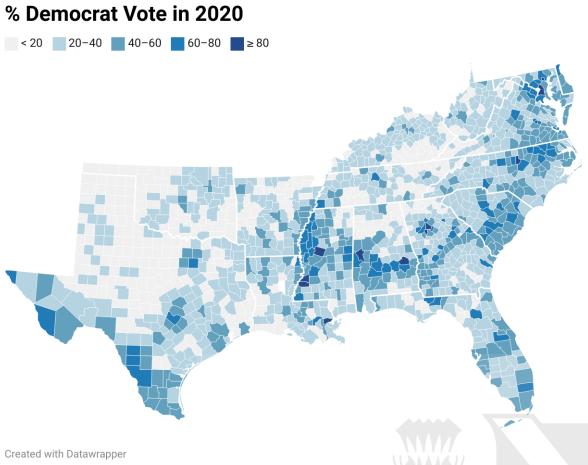

Figure 1: Democrat vote in 2020 in the South

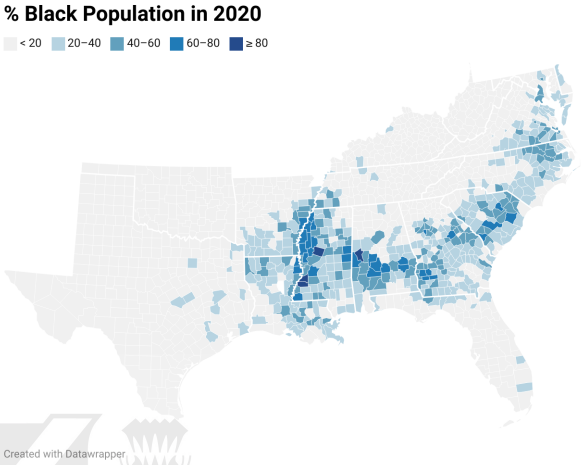

Figure 2: Black population in 2020 in the South

This is what motivates this investigation, which seeks to estimate the effect that the percentage of the black population has on the percentage of votes that the Democrat Party receives at the county level. The first question that could arise from this approach would be if race is actually the explanatory variable behind black people's tendency towards the Democrat Party or if the median black voter's profile substantially differs from those of non-blacks, which might imply that race might not be the only channel that could explain the phenomenon or even be capturing the effect of other group treats. At first, one could attempt to explain the differences in party support by race through turnout, i.e., if black people as a group significantly vote more than non-blacks, then the better Democrat performance is explained by turnout and not race. Using data on voter turnout from the US Census Bureau (2021), it can be observed in Figure 3 that black and white voters' turnout rates followed similar trends since 1964, despite the latter being higher in level. However, since the turn of the century, this gap shrunk significantly, with a difference of about 5% in favor of white voters in the 2020 election. Thus, this relationship between the black population and Democrat performance is most definitely not explained by differences in voter turnout.

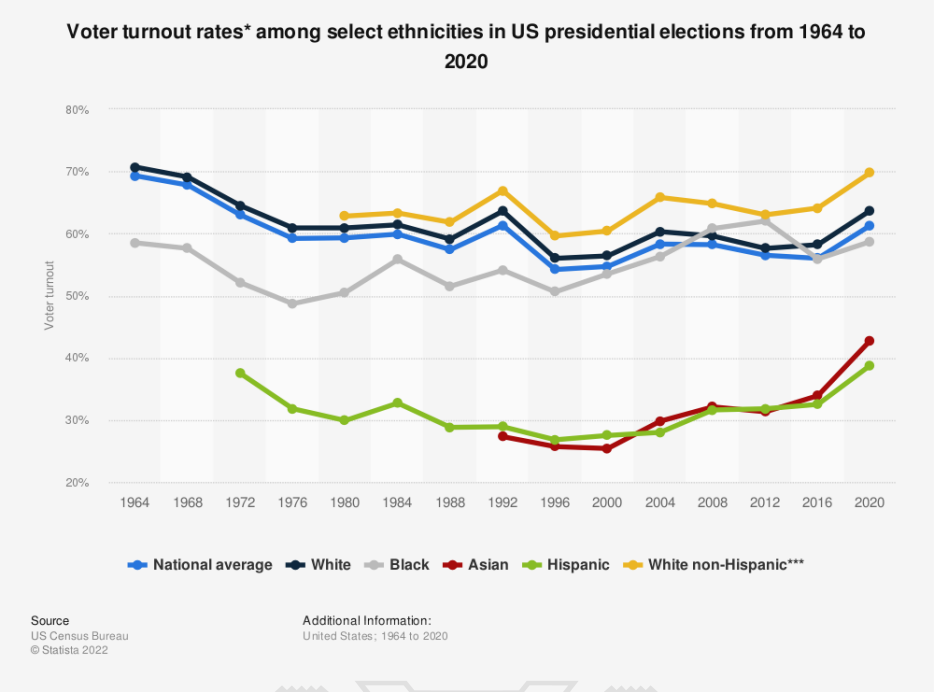

Figure 3: Voter turnout rates by race from 1964 to 2020 - Source: Statista

Another possibility could be that the demographic characteristics of black voters are significantly distinct from those of non-blacks. By inquiring into Moslimani's (2022) comparison of black and non-black eligible voters' demographic traits at the national level, it can be argued that black voters do not significantly differ from non-black voters in gender, age distribution or nativity, but the former does have lower average educational attainment. On the other hand, Li & Fotheringham (2021) measured at county level the relevance of different demographic traits to explain partisanship for presidential elections from 2008 to 2020, finding that median income, income inequality, educational attainment, sex ratio, median age, Hispanic presence, and urbanization rate were all relevant determinants (alongside black presence) of political preference. Therefore, to isolate the possible effect of race on the Democrat vote, a simple OLS regression with demographic controls could potentially identify the effect that pertains to this investigation. However, this approach could lead to the rise of two identification issues: the endogeneity of black presence and reverse causality between black presence and Democrat vote [1].

A notable example of the former can be found in psychological traits at individual and communal levels (such as openness, agreeableness, extroversion, emotional stability and conscientiousness) that affect both political preference and residence election (Gerber et al., 2011).

[1]Previous investigations have attempted to explain electoral preference using counties' demographic traits, yet they have not contemplated these issues (Leighler & Nagler, 2013; Tyson & Manian, 2016; Schaffner et al., 2018; Budiman et al., 2020; Gramlich, 2020; Hanson & Chen, 2020).

The latter concern would be that not only does black presence affect the Democrat vote, but also that the relationship goes the other way around. The evidence for this claim is present in Liu et al. (2019), where the authors conclude that migration patterns reinforce partisan sorting (i.e., Democrats tend to move to Democrat-majority counties).

Therefore, the issue with this approach is that the effect of black presence cannot be properly identified by an OLS model. Nevertheless, by presenting a soil-related variable that acts as an instrument for the presence of the black population, it will be argued that an IV strategy can identify the effect in the states of Alabama and Mississippi.

Background

Before quantifying the effect that black presence has on the Democrat vote, it is important to first understand the basis of the positive association between black people and Democrat performance, which seems to go beyond electoral or demographic differences. White & Laird (2020) argue that this alignment started with the pro-diversity labor and race laws that were being implemented in Northern cities by the mid-20^{th} century during the Democrat-backed New Deal (Schickler, 2016) and was later reinforced with the party's support of the 1964 Civil Rights Act (Rigueur, 2015), which prohibited discrimination based on race, color, religion, sex, or origin. Yet, the crucial part of their argument is their theoretical framework to understand the persistence of the phenomenon, which models a process they denominate *racialized social constraint*. They contend that, given the heightened group consciousness that the period of party identification consolidation had, Democrat support became a norm within the community and was constantly reinforced by in-group accountability. This led to " (...) [a] constrain [on] black political behavior by making salient the possible social benefits of conformity and the likely social consequences for defection" (p. 39).

Moreover, it is also relevant to the investigation to comprehend why, apart from consistently voting for the Democrat Party, the Southern black population concentrated in the crescent-shaped area that was depicted in Figure 2. In order to do that, it is necessary to rely on the historical colonization process of the South, which not only determined the population distribution but also limited its modification throughout time. This historical revision, apart from helping to understand the distribution of the black population, also provides even more evidence towards the hypothesis that black presence cannot be considered an exogenous variable

within a Democrat vote model.

It can hardly be argued that the racial distribution across U.S. counties is exogenous[2], given the ethnic heterogeneity of the colonization process and the later establishment of new communities from the late 18^{th} century to the mid-20^{th} century. This diversity is founded on the legal restrictions that were imposed on specific ethnic groups' economic, political, and personal rights throughout that period, which created geographically-based ethnic patterns that somewhat persisted over time. Considering their introduction to the U.S. was initially to be enslaved, Black people were granted highly restricted rights until the mid-20^{th} century, which, according to Lynch (2022), profoundly severed their agency by equating them to property and, thus, constraining them to their masters' properties.

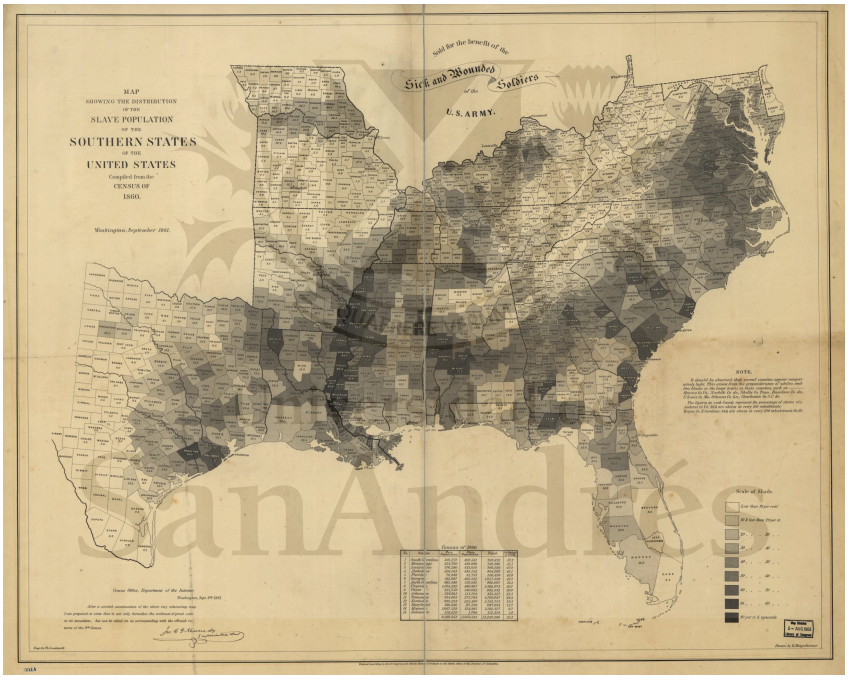

Figure 3: Slave population in the South in 1860

The result of this process was a sharp shift in the racial composition of some counties, where the percentage of the black population compared to the total population rapidly rose in a magnitude that can be appreciated in Figure 3 by Hergesheimer (1861). There were two underlying factors that determined the concentration of this process within the American South: a political one because Southern states allowed slavery to continue up until the Civil War and a geographical one since the South's climate and soil were suitable for the growth of labor-

[2]While race expression is usually considered exogenous, Nix & Qiang (2015) and Villareal & Bailey (2020) present evidence against that assumption based on opportunistic behavior.

intensive cash crops (such as cotton and tobacco) that were reliant on slave workforce. By visually comparing the distributions in Figures 2 and 3, it is not implausible to think that this initial pattern of enslaved black people persisted throughout time and probably determined the current distribution of their descendants.

Thus, soils' cotton suitability had a significant role in determining the settling patterns of Alabama and Mississippi and, consequently, the racial distribution of their counties. After the Creek War of 1813-14 ended, which derived in the Treaty of Fort Jackson that granted the U.S. 93.000 km^2 of Creek tribal land (H.R. 4015, 56^{th} Congress, 1814), both territories were incorporated as states. Their initial colonization by settlers from neighboring seaboard states had the expansion of the cotton frontier as its aim, a process which received the name of *Alabama Fever*, during a period where international cotton prices were at an all-time high given the tremendous demand from British textile factories (Keith, 2011). From their previous experience in Georgia and South Carolina, these agricultural settlers had learned that cotton sowing was the most suitable in soils that were dense, dark, and rich (Giesen, 2004). The later production output of this expansion can be appreciated in Figure 4 by Gannet & Fletcher (1883).

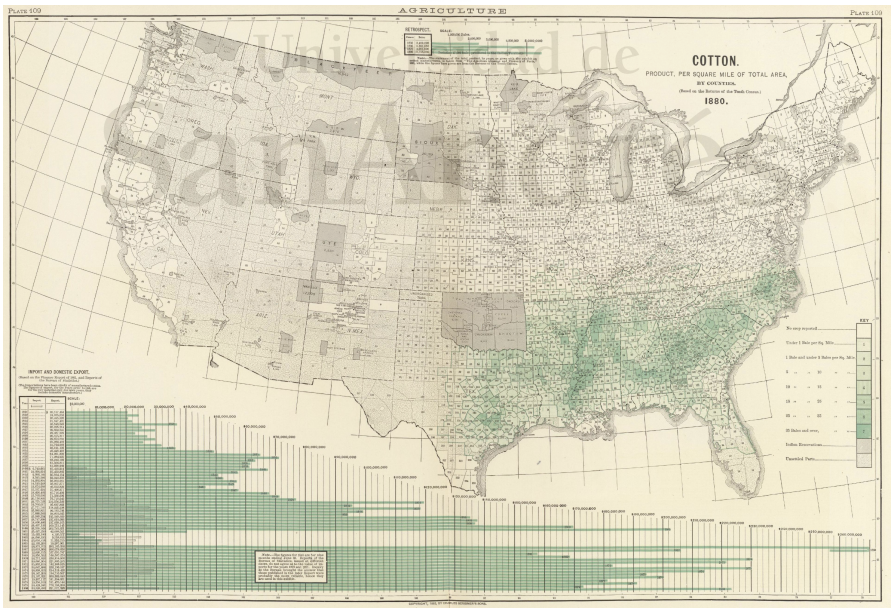

Figure 4: Cotton production per square mile in 1880

Given the technology of the early 19^{th} century, cotton harvesting was a labor-intensive process. Consequently, these new plantation owners markedly increased their demand for the enslaved workforce after settling, going from 400.000 enslaved individuals in 1790 to 4.000.000

in 1860 across the nation (Dattel, 2006). Despite slave importation from abroad being banned from 1808, the domestic trade of enslaved people emerged as a mechanism that directed them toward the new cotton frontier (Evans, 1961). African Americans represented the largest share of enslaved individuals, with almost nine out of ten being bonded (Corbett et al., 2014).

After the 1863 Emancipation Proclamation and the Union's victory over the Confederacy in the 1865 Civil War, slavery was effectively abolished in the U.S. However, since most formerly enslaved people could not afford to migrate to Northern states, they started to engage in new forms of economic dependency, such as tenant farming and sharecropping, with their former white owners (Bode, 2020). In the former, landowners leased their lands in exchange for rent; in the latter, formerly enslaved people provided their workforce in exchange for a percentage of crop sales. Both systems required machinery and materials investments which further deepened formerly enslaved people's dependency on their former owners since they were their moneylenders (Walbert, 2010). Yet, their popularity rapidly grew, and by 1890, three out of four black Southerners engaged in either of these practices.

During the 20^{th} century, there were multiple episodes of mass internal migration of black Southerners towards urban centers in the North and West, in a process called the Great Migration. According to Collins (2021), black people went from representing almost 90% of the total population in southern states in 1910 to less than half by 1970. Still, black people continued to be part of the racial majority in those counties that were black-majority before the Great Migration

Steven Dutch, a geologist from the University of Wisconsin-Green Bay, was the first to establish a correlation between party preference and a geological factor that had not been considered in previous studies. In *Geology and Election 2000* (2002), he highlights how the county-level percentage of Democrat votes in Mississippi, Alabama, Georgia, and South Carolina in the 2000 presidential election considerably resembles the approximate extent of Cretaceous rock unit deposits in these states. This formation was a result of the partial flooding of these lands by the Gulf of Mexico during the Cretaceous Period (∼85 million years ago), which created wedges of marine sediments along it that were later re-exposed to the elements when the sea withdrew back to the current coastline at the end of the Mesozoic Era (∼60 million years ago) (Schwimmer & Frazier, 2013).

Figures 5, 6, and 7, which are all from Dutch (2002), illustrate the electoral performance

of the Democrats by county in the 2000 presidential election, the approximate localization of Cretaceous rock unit deposits (in shades of green), and the percentage of black population by county, respectively. After running a correlation between the first two, Dutch observed a strong and positive relationship between them in Mississippi and Alabama and a relatively weak relationship in Georgia and South Carolina.

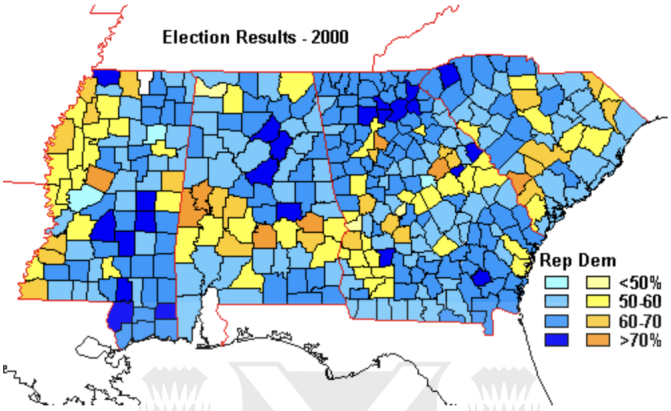

Figure 5: Election results in 2000 by county

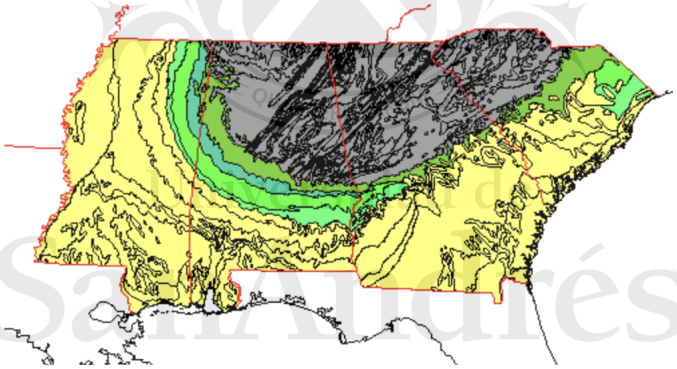

Figure 6: Cretaceous rock sediments approximate distribution

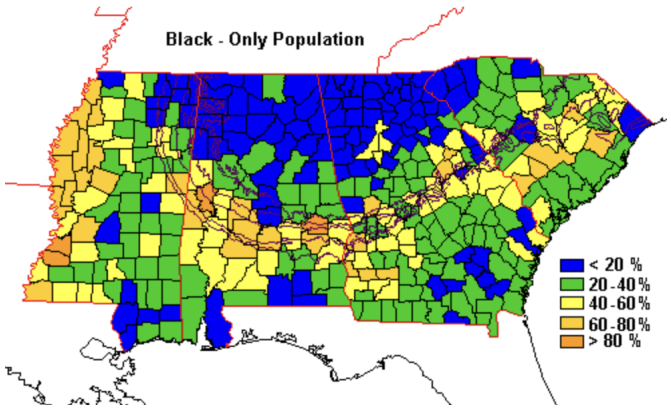

Figure 7: Black population in 2000 by county

The correlation between these three variables does not enlighten the argument of this inves-

tigation by itself. Up to now, the historical area covered by these sediments has been scarcely mentioned in the literature, since its only highlight was its rich marine fossil variety (Renger, 1935). However, the interesting fact about these Cretaceous rocks is that the re-exposure of the seabed (known as Selma Chalk) provoked its weathering for millions of years and its later decomposition into a particular type of soil of the order Vertisols or, more specifically, of the suborder Uderts (Tullos, 2004). Uderts can be characterized as dark, clayish, and alkaline soils prone to cracking. Even so, their most relevant characteristic for the investigation is that they are particularly suitable for cotton and rice farming (USDA, 2006; SARE, 2020). Thus, it is not surprising that most cotton plantations during the *Alabama Fever* were established where these soils were present, and that was the case for the region which would be later known as the Blackland Prairie or Black Belt, which is primarily composed of Uderts soils and its extension can be seen in Figure 8 by SARE (2020).

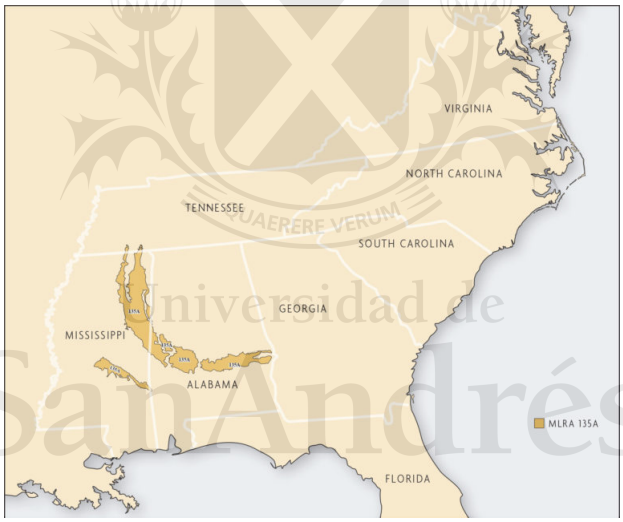

Figure 8: Blackland Prairie extension

In summary, the presence of Uderts determined soils' cotton suitability in Alabama and Mississippi and, since this was known by settlers, cotton plantations were established there. Given cotton's labor intensity, enslaved black people were located there and, after the abolition of slavery, freed black people remained within the Black Belt given their economic restrictions. Thus, it appears as if the presence of Uderts could be used as an instrument for black presence in a county. Figure 9 depicts a graphical simplification of the proposed framework, whose validity will be further discussed in the Methodology section.

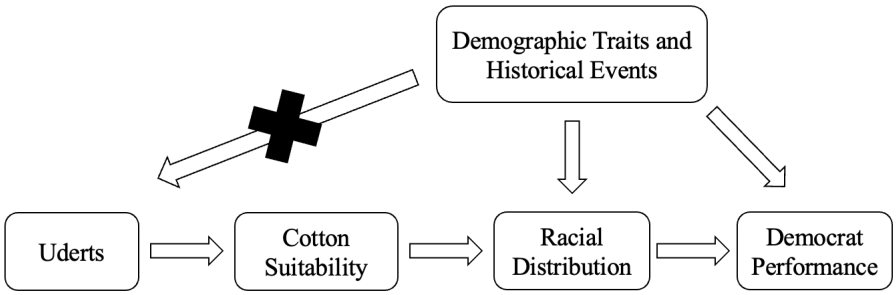

Figure 9: Proposed framework

Data

As previously mentioned, county-level data from Alabama and Mississippi will be used. Administratively, they are subdivided into 67 and 82 counties, respectively; and 26 (Autaga, Barbour, Bibb, Bullock, Butler, Choctaw, Conecuh, Dallas, Elmore, Franklin, Greene, Hale, Lawrence, Lowndes, Macon, Madison, Marengo, Montgomery, Morgan, Perry, Pickens, Pike, Russell, Sumter, Washington, and Wilcox) and 8 (Chickasaw, Clay, Kemper, Monroe, Noxubee, Oktibbeha, Pontotoc, and Tate) of them have at least 0.1% of their total county land covered by a Udert, also respectively. Figure 10 shows the percentage of Uderts by county as a percentage of the total county area. This data was created by adding all the percentages of soils within the county area corresponding to the Uderts suborder, as reported in the Web Soil Survey by the United States Department of Agriculture (USDA).

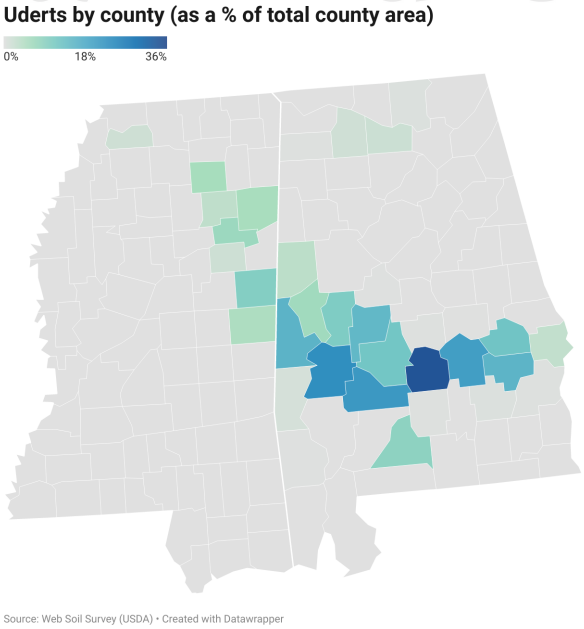

Figure 10: Percentage of Uderts by county

Data regarding the election results for each county from 2012 to 2020 was extracted from the MIT Election Data and Science Lab. It contains the number of votes each candidate received, whether they were Democrats, Republicans, or from other minor parties, in every presidential election celebrated within that period.

For demographic data, Table S0601, denominated "Selected characteristics of the total and native populations in the United States", from the 2012, 2016, and 2020 5-year estimates American Community Survey for each county in the two states of interest was used. These surveys contain a wide range of variables. However, only the following will be kept: the percentage of the population identifying as black, the percentage of the population identifying as Hispanic, the median income, the percentage of people in the top income decile and in the bottom five deciles, the median age, the percentage of individuals who completed a higher degree or a bachelor's degree and the percentage of the population identifying as female. A variable for the percentage of the population living in urban areas is also included, which was obtained from the 2010 Census.

The choice of variables that will serve as controls in this model was based on the results obtained by Li & Fotheringham (2021) for the counties in Alabama and Mississippi. Additionally, based on Kelly (2020), both the latitude and longitude of each county are included in order to control for any spatial autocorrelation issue. Table 1 shows the mean differences between counties with and without Uderts for each variable discriminating by state.

As expected, the percentage of Democrat vote is significantly higher in counties within the Black Belt than outside of it in both states. However, Black Belt counties in Mississippi do not significantly have a greater presence of black people. A possible reason for this might be the fact that Mississippi, as opposed to Alabama, has two sub-regions within its borders where, through different mechanisms and historical circumstances, there is an exceptionally high density of black population: the Blackland Prairie to the east and the Yazoo-Mississippi Delta to the west. All other control variables, except educational attainment and urbanization rate, have significant differences in at least one of the states.

Table 1: Two-sample t tests by Uderts presence

	Alabama	Mississippi
Democrat Vote (%)	-24.78***	-5.48**
	(2.51)	(3.76)
Black (%)	-28.55***	-2.25
	(2.88)	(3.93)
Hispanic (%)	0.85**	-0.17
	(0.46)	(0.42)
Log Median Income	0.11***	0.03
	(0.03)	(0.03)
Median Age	1.29***	1.62*
	(0.44)	(0.95)
Higher Ed (%)	0.45	-1.05
	(1.06)	(2.08)
Female (%)	-0.33	-0.69**
	(0.27)	(0.31)
Urban (%)	4.12	2.39
	(3.86)	(4.21)
Top10/Bottom50 Index	1.49*	1.91**
	(1.04)	(0.89)

Notes: Standard deviations in parenthesis.

Methodology

The linear OLS model with state fixed effects that can be proposed in order to identify the effect that black population has on the Democrat vote at county level for the two states would be:

$$DemVote_{is} = \alpha_s + \beta_1 Black_{is} + \beta_2 Hispanic_{is} + \beta_3 ln(MedianIncome_{is}) + \beta_4 MedianAge_{is} +$$
$$\beta_5 HigherEd_{is} + \beta_6 Urban_{is} + \beta_7 Female_{is} + \beta_8 Inequality_{is} + \beta_9 Lat_i + \beta_{10} Lon_i + u_{is}$$
$$s = 1, 2$$
$$i = 1, ..., 149$$

Nonetheless, given the arguments presented previously, $\hat{\beta}_1$ or the estimated effect of black population on Democrat vote could be a biased estimation. Thus, the identification strategy consists of the proposal of an IV model under exact identification, where the percentage of a

county's area covered in Uderts soils (z_{is}) would be the instrument for the percentage of black population (x_{is}).

Assuming that the conditions of random sample and finite and positive definite variance matrix are fulfilled, it is also necessary to corroborate the validity of our instrument to uphold the unbiasedness of the estimation of the causal relationship. First, in order to fulfill the rank condition ($E(z_{is}x'_{is}) = \Sigma_{zx}$ finite and invertible matrix), it is required that the instrument correlates with the variable to be instrumented. By running a simple correlation matrix between $Uderts_{is}$ and $Black_{is}$, we can corroborate that they have a positive correlation of 0.6849, thus satisfying the condition. Second, to fulfill the orthogonality condition ($E(z_{is}u_{is}) = 0$), it is required that the instrument does not correlate with any other determinant of the dependent variable that is contained in the error term. We cannot test this condition empirically, but we can attempt to argue that the variables omitted from the model specification do not significantly correlate to the presence of Uderts soils.

The first variable that comes to mind is agricultural productivity. If counties in the Blackland Prairie initially had a higher cotton yield than those outside of it and, as Brown et al. (2023) establish, counties with higher agricultural productivity tend to be more Republican-leaning, then excluding it from our model would result in a negative bias on our estimation of β_1. However, the agricultural practices carried through the 19^{th} century in cotton plantations, primarily the absence of crop rotation and persistent infestations of boll weevil (a beetle that feeds on cotton buds), severely undermined the productivity of Blackland Prairie counties (Gibson, 1941; Smith, 2008; Yahn, 2009). This productivity loss persisted over time, with a recent comparative study by Zhang et al. (2017) finding no statistical differences in total agricultural productivity between Black Belt counties and adjacent counties. Hence, it would not seem plausible to believe that agricultural productivity explains the current Democrat vote, at least in Alabama and Mississippi.

The second variable that can be considered is the wealth of initial settlers. If the settlement patterns were determined by the economic resources of the prospective settler (i.e., wealthier settlers were able to acquire the most productive plots when auctioned, such as those within the Black Belt) and, as Kantrowitz (2020) and Greenberg (2021) propose, wealthy settlers ideologically aligned with conservative postures towards racial issues; then those counties where they

established might have been more prone to institutions that undermined political participation of non-whites. If these institutions persisted up to our times, then excluding settler's wealth from our model would also result in a negative bias on our estimation.

By first enquiring into the county-aggregated self-reported personal wealth in current U.S. dollars and the white population by county, gathered from the 1860 Census by IPUMS for 109 out of the 149 counties; a per capita net worth variable that should work as a proxy to the settler's wealth was constructed. By comparing the kernel density estimations discriminating by the presence of Uderts, individuals who settled in the Black Belt seem wealthier than their counterparts who relocated to other counties.

In this case, it may be possible that in Black Belt counties, where the average settler net worth was 43% higher than outside, the political institutions that emerged after the Emancipation Proclamation suppressed black voters to a greater extent and reinforced this income gap between white settlers and enslaved blacks. Alternatively, this mechanism could be possible if we consider it from Acemoğlu's et al. (2005) framework of *de facto* and *de iure* political power as determinants of political and economic institutions. Multiple channels were used to institutionalize voter suppression before the Voting Rights Act of 1965, such as poll taxes, literacy tests, or arbitrary registration practices (Kousser, 1974). Within the proposed model, this possible channel of long-run wealth inequality can be approximated by using Chancel & Picketty's (2021) ratio between the percentage of people in the top income decile and the percentage in the bottom half, which acts as an alternative for a Gini index given the lack of individual income data at the county level in the ACS.

Notable literature that probes into these institutional channels is Engerman et al. (2002), where the authors' theoretical framework establishes that initial factor endowments (such as geographical characteristics) affect the quality of institutions, which in turn affects economic development. However, as opposed to our line of research, they argue that the importance of factor endowments lies in their indirect effects (i.e., their effects on institutions) rather than on themselves *per se*. Even with this, they highlight the U.S. South's case as a particular one because it does not fit under their theory as well as others (like the U.S. North or the Caribbean). According to them, the factor endowments of the U.S. South (i.e., soils suitable for cash crops) and the persistence of slavery should have led to the development of institutions that resembled those of similarly endowed regions (e.g., Latin America). Even so, they were later constrained

because some of their aspects were determined at a national level, where the "good" institution of the U.S. North prevailed due to the Civil War. As a result, the institutional channel to explain economic differences (or, in our case, party identification) is more complex than it seemed, thus, making the consideration of the direct effects of factor endowments an attractive alternative.

Having argued in favor of our instrument's validity, we can further corroborate its strength by performing various tests. In the following section, we will perform a Hausman-Durbin-Wu test to determine whether treating the black population of a county as an endogenous variable (and, thus, using an IV approach) is efficient. If we obtain a p-value < 0.05, the test indicates an inconsistent estimation of the coefficient of interest via OLS and a consistent and efficient estimation via IV. We will also be using two post-estimation tools (*estat endogenous* and *weakivtest*) to help us determine if our instrumented variable is exogenous (in which case, using OLS would be more efficient) and if our instrument has a weak correlation with our suspected endogenous variable, respectively. In the former, Wooldridge's (1995) robust score test and robust regression-based test are reported since we are assuming heteroskedasticity. If the test result is significant, we should treat the instrumented variable as exogenous. In the latter, Montiel Olea & Pflueger's (2013) test for weak instruments is reported, and, as a rule of thumb, if the F-statistic is lower than 10, then the instrument is considered weak.

Results

For our approach through an IV model to be logical, it is necessary to check whether a county's black population is endogenous within our model, given our dataset. So, we will first estimate an OLS and an IV model for 2020 without any control variable, whose results are shown in columns 1 and 3 of Table 2.

By performing a Durbin-Wu-Hausman test on our estimated coefficients for both models, we obtain a p-value of 0.0195, thus concluding that an OLS approach would render inconsistent estimations and make the decision to use an IV model plausible. Under this estimation without any control variable, the IV coefficient is 4.7% higher than the OLS coefficient, possibly indicating that our coefficient of interest is underestimated or negatively biased when approximated through OLS. Nonetheless, these coefficients are not representative since we are omitting many variables that most definitely affect the Democrat vote and are correlated to the black population.

On the other hand, by estimating that same coefficient using these two models but adding the set of control variables, the obtained results are depicted in columns 2 and 4 of Table 3. As observed, only the percentage of the population that achieved an educational level equal or higher to a high school degree and the percentage of the population identifying as Hispanic are the control variables that have a statistically significant coefficient in both models, being the two positive. On the other hand, the median income, the median age, the percentage of the urban population, and the percentage of the population identifying as female do not have a statistically significant effect in either model.

Now, it can be observed that the IV coefficient is 10.1% higher than the OLS one, indicating that the estimation is robust to adding a relevant set of controls to the model and greater in magnitude than the OLS coefficient compared to the specification without controls. The validity of this estimation can be further checked by performing an endogeneity test on the instrumented variable, which rendered a p-value of 0.0062 and, therefore, confirmed the hypothesis and the previous result of the Durbin-Wu-Hausman test that the percentage of the black population in a county suffers from endogeneity within the model we have proposed.

A test for weak instruments can also be run. Its result, as was previously mentioned, indicates weakness of the instrument when the F-statistic is below 10. The resulting F-statistic of the test ran on the IV estimation is 47.38, consequently discarding the possibility of the percentage of Uderts soils being a weak instrument of the black population in a county within our model. This result can also be appreciated in the first stage estimation of our IV model, which is depicted in Table 3, where the percentage of Uderts soils is a solid explanatory variable in both model specifications, only slightly decreasing in magnitude when controls are added. In this table, it can also be observed that the percentage of the black population is negatively affected by the percentage of the Hispanic population and by the median income of the county, having an average effect on the independent variable of -1.557 and -0.769% when they increase by 1 *p.p.* and 1%, respectively. On the other hand, the percentage of the urban population has a positive average effect of 0.246 when it increases by 1 *p.p.*.

These results are relevant because, as previously shown, race is a crucial factor in determining the electoral decision of a person. Nonetheless, it has also been proved that in the case of the two states considered, the effect of the racial composition of a county on the result for the

Democratic Party can be underestimated if it is naively estimated through an OLS model. This bias can be exemplified in the case of Lowndes, the county in Alabama with the highest presence of Uderts, where, under the estimation of an OLS model, the black population explains 88.7% of the variability. In contrast, under the estimation of an IV model, that same variable explains 97.8% of the variability. If this same calculation was repeated for each county, it would be seen that, as expected, the explanatory power of the variability in the black population is 10.1% higher when estimated through an IV model than through an OLS model.

Table 2: Estimations

	OLS		IV	
	No Controls	Controls	No Controls	Controls
Black (%)	0.853***	0.870***	0.893***	0.958***
	(0.015)	(0.023)	(0.028)	(0.034)
Hispanic (%)		0.450***		0.608***
		(0.112)		(0.125)
Log Median Income		-0.918		6.785
		(4.075)		(4.548)
Median Age		-0.023		0.050
		(0.145)		(0.156)
Higher Ed (%)		0.382***		0.389***
		(0.091)		(0.103)
Urban (%)		0.017		-0.003
		(0.016)		(0.017)
Female (%)		0.105		0.024
		(0.147)		(0.121)
Top10/Bottom50 Index		-0.088		-0.112
		(0.108)		(0.107)
Observations	149	149	149	149
State FE	Yes	Yes	Yes	Yes
Lat/Lng Controls	No	Yes	No	Yes

Notes: Standard errors in parenthesis.

Table 3: First-stage estimations

	No Controls	Controls
Uderts (%)	1.746***	1.459***
	(0.270)	(0.212)
Hispanic (%)		-1.557***
		(0.370)
Log Median Income		-76.890***
		(14.500)
Median Age		-0.569
		(0.544)
Higher Ed (%)		-0.133
		(0.416)
Urban (%)		0.246***
		(0.057)
Female (%)		0.472
		(0.515)
Top10/Bottom50 Index		0.306
		(0.427)
Observations	149	149
State FE	Yes	Yes
Lat/Lng Controls	No	Yes

Notes: Robust standard errors in parenthesis.

Robustness checks

It is also imperative that to uphold the representativeness of the results that have been found in the previous section; a series of robustness checks can be carried out as a means to corroborate that the coefficients are actually representing the relationship that was established and not just a product of chance or fit of the data used.

To achieve this, a plurality of approaches is proposed. First, as a way to demonstrate that the coefficients that were found are similar to ones that can be estimated using other electoral performances of the Democratic Party, the same IV model was estimated using the results for the 2012 and 2016 elections, as it can be seen in columns 1 to 2 of Table 4. Despite some of the control variables' coefficients varying in magnitude or significance, it can be observed that the one related to the black population firmly upholds in both senses, with inter-year estimations of the coefficient having little variability. It could be attempted to replicate our estimations using data from previous elections. Unfortunately, the American Community Survey data available

on the United States Census Bureau website only covers the 2010-2020 period. Thus, further research on the estimations' robustness could be carried out if new county-level demographic data previous to 2010 is presented.

Second, to verify whether the coefficients' magnitude and significance vary using an alternative measure for Uderts soils, the IV model for the three electoral results was re-estimated using the approximate extension by county of the Blackland Prairie calculated by Barone (2005) using the General Land Office Surveys from the 1830s (county-level coverage is depicted in Figure 10), as it can be seen in columns 4 to 6 of Table 4. By comparing the resulting coefficients to the ones obtained from using the measures from the Web Soil Survey, no significant differences between the coefficients for the black population using any of the mensuration for Uderts are observed. Notwithstanding the coefficients from the alternative instrument being slightly larger in magnitude, the selected measurement for Uderts as an instrument is more adequate because the strength of prairie extension as in instrument, also calculated using Montien Olea & Pflueger's (2013) method, is significantly lower (10.55) than that of the chosen instrument (47.38).

Table 4: Robustness checks

	Uderts			Prairie		
	2012	2016	2020	2012	2016	2020
Black (%)	0.921***	0.958***	0.958***	0.938***	1.012***	0.998***
	(0.035)	(0.029)	(0.034)	(0.068)	(0.074)	(0.069)
Hispanic (%)	0.470***	0.625***	0.608***	0.552***	0.733***	0.680***
	(0.153)	(0.122)	(0.125)	(0.184)	(0.186)	(0.163)
Log Median Income	6.219	1.312	6.785	10.415	5.628	10.288
	(4.370)	(3.632)	(4.548)	(6.366)	(6.148)	(6.736)
Median Age	-0.005	0.007	0.050	0.056	0.089	0.083
	(0.140)	(0.135)	(0.156)	(0.154)	(0.174)	(0.169)
Higher Ed (%)	0.246***	0.285***	0.389***	0.256***	0.294***	0.392***
	(0.080)	(0.093)	(0.103)	(0.089)	(0.109)	(0.113)
Urban (%)	-0.003	-0.016	-0.003	-0.012	-0.026	-0.012
	(0.017)	(0.016)	(0.017)	(0.020)	(0.021)	(0.022)
Female (%)	0.075	-0.005	0.024	0.032	-0.046	-0.013
	(0.145)	(0.124)	(0.121)	(0.164)	(0.139)	(0.127)
Top10/Bottom50 Index	-0.401***	-0.085	-0.112	-0.438***	-0.104	-0.123
	(0.129)	(0.113)	(0.107)	(0.134)	(0.123)	(0.113)
Observations	149	149	149	149	149	149
State FE	Yes	Yes	Yes	Yes	Yes	Yes
Lat/Lng Controls	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Standard errors in parenthesis.

Counties	Hectares	Percentage of County Area	Number of Patches
Alabama			
Dallas	2,300	1.0	9
Hale	4,600	2.7	9
Lowndes	11,730	6.2	14
Marengo	4,150	1.6	17
Montgomery	29,900	14.0	10
Perry	2,910	1.5	4
Pickens	3,800	1.6	6
Sumter	13,230	5.5	53
Wilcox	440	0.2	2
Total	73,060		
Mississippi			
Chickasaw	6,690	5.1	14
Clay	8,730	8.0	11
Kemper	2,320	1.2	23
Lee	1,880	1.6	17
Lowndes	15,370	11.4	51
Monroe	16,500	8.2	11
Noxubee	15,880	8.6	104
Oktibbeha	4,120	3.4	27
Pontotoc	140	0.1	5
Total	71,630		

Figure 11: Areas of prairie in the GLO surveys in the 1830s in Alabama and Mississippi - Source: Barone (2005)

Third, a leave-one-out cross-validation test (LOOCV) to test the coefficients' dependence on specific observations and the performance of the model across different subsets of our data can be performed. Barron's (2014) *loocv* command will be used, which performs the LOOCV test on the IV model with controls and reports the root mean squared errors (RMSE), the mean absolute errors (MAE), and the pseudo-R2. After running it for each year, an average RMSE of approximately 3.83 was obtained. This value is similar to the RMSE that would be obtained from simply regressing the model using all the observations, implying that the model is, indeed, consistent in its estimations no matter which subset of the data it is using and not overly dependent on specific observations, hence providing more evidence in favor of the robustness of the estimations.

Fourth, a placebo test can be performed to test the validity of the Uderts' presence as an instrument of the black population. If the instrument is valid, replacing the endogenous variable with an unrelated variable when estimating the IV model should render statistically non-significant coefficients because the instrument has no relationship to the unrelated variable and no other omitted mechanism is driving the effect that the presence of Uderts has on black population.

In order to achieve that, the variable *Hispanic* is used, which, as aforementioned, contains the percentage of the population in a county that identifies as Hispanic. This variable has two desirable characteristics in a placebo: it is somewhat correlated to the endogenous variable (-0.384), yet it is weakly correlated to the instrument (-0.13). This means that although there is a relationship between the residence election of Hispanics and African Americans, the former has not chosen the Black Belt as often as the latter. As seen in Table 5, none of the coefficients for the placebo are statistically significant, which indicates that the hypothesis that the presence of Uderts is solely affecting Democrat vote through black presence is robust.

Table 5: Placebo test

	IV		
	2012	2016	2020
Hispanic (%)	2.043	6.245	5.667
	(1.955)	(6.285)	(4.816)
Black (%)	0.909***	1.167***	1.109***
	(0.063)	(0.302)	(0.209)
Log Median Income	0.695	-7.258	-7.882
	(4.880)	(10.526)	(12.287)
Median Age	0.397	1.729	1.611
	(0.566)	(2.022)	(1.513)
Higher Ed (%)	0.444*	1.099	1.097*
	(0.269)	(0.947)	(0.661)
Urban (%)	-0.028	-0.132	-0.114
	(0.039)	(0.137)	(0.116)
Female (%)	0.415	1.143	0.834
	(0.399)	(1.312)	(0.802)
Top10/Bottom50 Index	-0.474**	-0.413	-0.237
	(0.210)	(0.581)	(0.319)
Observations	149	149	149
State FE	Yes	Yes	Yes
Lat/Lng Controls	Yes	Yes	Yes

Notes: Standard errors in parenthesis.

Final remarks

Throughout this investigation, it has been attempted to prove two main aspects. First, that the race an individual identifies with is a relevant explanatory variable for their political preference within the U.S. two-party system. However, it was argued that is not possible to assess the magnitude of its effect through an OLS approach because the current racial composition of a county's population has been affected by multiple historical events and is also subject to reverse causality. And second, by using the percentage of county area covered in Uderts, the

effect of black population on the Democrat vote for the states of Alabama and Mississippi was appropriately identified, given the geological particularities that shaped these two states' landscapes.

Going back to the motivation of this study, it is believed that this work is of importance because, considering the interest that political parties have on correctly identifying their key target voters, partisan spending on federal elections nationwide has more than quadrupled from 2000 to 2020. In the case of the Democratic Party, it reached 9.6 billion dollars in 2020 (Center for Responsive Politics, 2022), making the efficient allocation of these funds increasingly relevant.

These results have highlighted that a naive approach to estimating the black population's relevance in the Democrat vote underestimates their political importance. Since the Republican Party has progressively targeted this demographic group in their advertisements in the South (Caspani, 2018; Weisman, 2022), Democrats should prioritize relocating advertising resources toward this group. This issue becomes even more urgent given the recent literature that pointed towards a greater reliance on ideological affiliation rather than demographic traits to determine an individual's political choice.

Further research on the direct effects of factor endowments (not just geographical characteristics) in different U.S. states would help, as the results of this investigation have, to correctly identify the effects that demographic traits have on the political preferences of voters, given the immense diversity that can be found in the demographics of the U.S. population across its territory.

References

Acemoglu, D., Johnson, S., & Robinson, J. A. (2005). Institutions as a Fundamental Cause of Long-Run Growth. In Aghion, P., & Durlauf, S. D. (Eds.), Handbook of Economic Growth (pp. 385–472). North Holland. Retrieved from https://doi.org/10.1016/s1574-0684(05)01006-3

Barone, J. A. (2005). Historical Presence and Distribution of Prairies in the Black Belt of Mississippi and Alabama. Castanea, 70(3), 170–183. Retrieved from http://www.jstor.org/stable/4034326.

Barron, M. (2014). LOOCV: Stata module to perform Leave-One-Out Cross-Validation. Retrieved from https://rb.gy/vmdgf.

Bode, F. A. (2003). Tenant Farming. In New Georgia Encyclopedia. Retrieved from https://shorturl.at/epO26.

Brown, J. R., Cantoni, E., Enos, R. D., Pons, V., & Sartre, E. (2023). The increase in partisan segregation in the United States. NICEP Working Paper. Retrieved from https://shorturl.at/oryL3.

Budiman, A., Noe-Bustamante, L., & Lopez, M. L. (2020). Naturalized Citizens Make Up Record One-in-Ten U.S. Eligible Voters in 2020. Pew Research Center. Retrieved from https://shorturl.at/uHV49.

Caspani, M. (2018, October 19). Black 'lynching' ad illustrates race overtones, anger in U.S. campaign ads. Thompson Reuters. Retrieved from https://shorturl.at/bcgxB.

Center for Responsive Politics (2022). Total Cost of Election. Retrieved from https://rb.gy/t6sw4.

Chancel, L., & Piketty, T. (2021). Global Income Inequality, 1820–2020: the Persistence and Mutation of Extreme Inequality. Journal of the European Economic Association, 19(6), 3025–3062. Retrieved from https://doi.org/10.1093/jeea/jvab047.

Collins, W. J. (2021). The Great Migration of Black Americans from the US South: A guide and interpretation. Explorations in Economic History, 80, 101382. Retrieved from https://doi.org/10.1016/j.eeh.2020.101382.

Corbett, P. S., Janssen, V., Lund, J. M., Pfannestiel, T., Waskiewicz, S., & Vickery, P. (2014). In U.S. History. OpenStax. Retrieved from https://shorturl.at/cHNQ1.

Dattel, E. R. (2006). Cotton in a Global Economy: Mississippi (1800-1860). Mississippi Historical Society. Retrieved from https://shorturl.at/wyCJ6.

Dutch, S. (2002). Geography and Election 2000. Retrieved from https://shorturl.at/bxDIW.

Engerman, S. L., Sokoloff, K. L., Urquiola, M., & Acemoglu, D. (2002). Factor Endowments, Inequality, and Paths of Development among New World Economies [with Comments]. Economía, 3(1), 41–109. Retrieved from http://www.jstor.org/stable/20065432.

Evans, R. (1961). Some Economic Aspects of the Domestic Slave Trade, 1830-1860. Southern Economic Journal, 27(4), 329–337. Retrieved from https://doi.org/10.2307/1055531.

Gannett, H. & Hewes, F. W. Cotton. Product, per square mile of total area, by counties. Based on the returns of the tenth census. 1880. Digital Public Library of America. Retrieved from https://shorturl.at/epzMX.

Gerber, A. S., Huber, G. A., Doherty, D., & Dowling, C. M. (2011). The big five personality traits in the political arena. Annual Review of Political Science, 14(1), 265–287. Retrieved from https://doi.org/10.1146/annurev-polisci-051010-111659.

Gibson, J. S. (1941). The Alabama Black Belt: Its Geographic Status. Economic Geography, 17(1), 1–23. Retrieved from https://doi.org/10.2307/141741.

Giesen, J. C. (2004). Cotton. In New Georgia Encyclopedia. Retrieved from https://shorturl.at/acgsK.

Gramlich, J. (2020). What the 2020 electorate looks like by party, race and ethnicity, age, education and religion. Pew Research Center. Retrieved from https://shorturl.at/cdeuL.

Greenberg, A. S. (2021). US Expansionism during the Nineteenth Century. In P. Fibiger Bang, C. A. Bayly, & W. Scheidel (Eds.), The Oxford World History of

Empire: Volume Two (pp. 1011–1034) Oxford University Press. Retrieved from https://doi.org/10.1093/oso/9780197532768.003.0037.

Hanson, P., & Chen, Y. (2020). The Demographic Profiles of Democrats and Republicans. Data Analysis and Social Inquiry Lab. Grinnell College. Retrieved from https://t.ly/9L0oF.

Hergesheimer, E. (1861). Map showing the distribution of the slave population of the southern states of the United States. Retrieved from https://www.loc.gov/item/99447026/.

Hersh, E. D., & Nall, C. (2016). The Primacy of Race in the Geography of Income-Based Voting: New Evidence from Public Voting Records. American Journal of Political Science, 60(2), 289–303. Retrieved from http://www.jstor.org/stable/24877622.

H.R. Doc. 7 Stat. 120, 13th Cong., 2nd Sess. (1814). Retrieved from https://shorturl.at/acCNX.

Kantrowitz, S. (2020). White Supremacy, Settler Colonialism, and the Two Citizenships of the Fourteenth Amendment. Journal of the Civil War Era, 10(1), 29–53. Retrieved from https://www.jstor.org/stable/26888071.

Keith, L. (2011). Alabama Fever. In Encyclopedia of Alabama. Retrieved from https://shorturl.at/acsvI.

Kelly, M. (2020). Understanding Persistence. CEPR Discussion Paper, No. 15246.

Kousser, J. M. (1974). The Shaping of Southern Politics: Suffrage Restriction and the Establishment of the One-Party South, 1880-1910. Yale Historical Publications Miscellany. Volume 102. Yale University Press.

Leighley, J. E., & Nagler, J. (2013b). Who votes now?: demographics, issues, inequality, and turnout in the United States.

Li, Z., & Fotheringham, A. S. (2021). The spatial and temporal dynamics of voter preference determinants in four U.S. presidential elections (2008–2020). Transactions in Gis, 26(3), 1609–1628. Retrieved from https://doi.org/10.1111/tgis.12880.

Liu, X., Andris, C., & Desmarais, B. A. (2019). Migration and political polarization

in the U.S.: An analysis of the county-level migration network. PloS one, 14(11), e0225405. https://doi.org/10.1371/journal.pone.0225405.

Lynch, H. (2022). African Americans. In Encyclopedia Britannica. Retrieved from https://shorturl.at/bczH6.

Manson, S., Schroeder, J., Van Riper, D., Kugler, T, & Ruggles, S. IPUMS National Historical Geographic Information System: Version 17.0 [Dataset]. Minneapolis, MN: IPUMS. 2022. Retrieved from http://doi.org/10.18128/D050.V17.0.

MIT Election Data and Science Lab (2018). County Presidential Election Returns 2000-2020. Harvard Dataverse, V11. Retrieved from https://doi.org/10.7910/DVN/VOQCHQ.

Montiel Olea, J. L. & Pflueger, C. (2013) A Robust Test for Weak Instruments, Journal of Business & Economic Statistics, 31:3, 358-369, DOI: 10.1080/00401706.2013.806694.

Moslimani, M. (2022). Key facts about Black eligible voters in 2022. Pew Research Center. Retrieved from https://shorturl.at/gqMNV.

National Geographic Society. (2022). The Plantation System. Retrieved from https://shorturl.at/JX028.

Nix, E., & Qian, N. (2015). The Fluidity of Race: 'Passing' in the United States, 1880–1940. NBER Working Paper 20828. Retrieved from https://shorturl.at/fhxz0.

Pew Research Center. (2020). The Changing Racial and Ethnic Composition of the U.S. Electorate. Retrieved from https://shorturl.at/anDHM.

Renger, J. J. (1935). Excavation of Cretaceous Reptiles in Alabama. The Scientific Monthly, 41(6), 560–565. http://www.jstor.org/stable/16029.

Rigueur, L. W. (2015). The Loneliness of the Black Republican: Pragmatic Politics and the Pursuit of Power (Repr. ed.). Princeton, NJ: Princeton University Press.

Schaeffer, K. (2019). In a rising number of U.S. counties, Hispanic and black Americans are the majority. Pew Research Center. Retrieved from https://shorturl.at/kBJZ5.

Schaffner, B. F., MacWilliams, M. C., & Nteta, T. M. (2018). Understanding white polarization in the 2016 vote for President: The sobering role of racism and sexism. Political

Science Quarterly, 133(1), 9–34. Retrieved from https://doi.org/10.1002/polq.12737.

Schickler, E. (2016). Racial Realignment: The Transformation of American Liberalism, 1932–1965. Princeton University Press.

Schwimmer, D. R. & Frazier, W. J. (2013). Cretaceous Period in Alabama. In Encyclopedia of Alabama. Retrieved from https://shorturl.at/inuK1.

Smith, R. (2008). Boll weevil in Alabama. In Encyclopedia of Alabama. Retrieved from https://shorturl.at/cijqv.

United States Department of Agriculture (USDA). Web Soil Survey [Dataset]. Retrieved from https://shorturl.at/cmnU0.

Sustainable Agriculture Research and Education (SARE). (2020). Conservation Tillage Systems in the Southeast: Production, Profitability and Stewardship. SARE Handbook Series 15. Retrieved from https://shorturl.at/dvIL7.

Tullos, A. (2004). The Black Belt. Southern Spaces. Retrieved from https://shorturl.at/lDEQ2.

Tyson, A. & Maniam, S. (2016). Behind Trump's victory: Divisions by race, gender, education. Pew Research Center. Retrieved from https://shorturl.at/mptCZ.

U.S. Census Bureau. (2010). 2010 Census Urban and Rural Classification and Urban Area Criteria, Percent Urban and Rural in 2010 by State and County. Retrieved from https://shorturl.at/cdI68.

U.S. Census Bureau. (2020). 2020 American Community Survey 5-year estimates, Table S0601. Retrieved from https://shorturl.at/gjtUV.

U.S. Census Bureau. (2020). 2020 Decennial Census, Table P1. Retrieved from https://rb.gy/49xpx.

United States Department of Agriculture (USDA). (2006). Land Resource Regions and Major Land Resource Areas of the United States, the Caribbean, and the Pacific Basin. U.S. Department of Agriculture Handbook 296. Retrieved from https://shorturl.at/qsFTW.

Villarreal, A., & Bailey, S. R. (2020). The Endogeneity of Race: Black Racial Identifica-

tion and Men's Earnings in Mexico. Social forces; a scientific medium of social study and interpretation, 98(4), 1744–1772. Retrieved from https://doi.org/10.1093/sf/soz096.

Walbert, D. (2010). Sharecropping and Tenant Farming. In Encyclopedia of North Carolina. Retrieved from https://shorturl.at/fDKP9.

Weisman, J. (2022, October 25). With Ads, Imagery and Words, Republicans Inject Race Into Campaigns. The New York Times. Retrieved from https://shorturl.at/egUX4.

White, I. K., & Laird, C. N. (2020). Steadfast Democrats. Princeton University Press. Retrieved from https://doi.org/10.1515/9780691201962.

Wooldridge, J. M. (1995). Score diagnostics for linear models estimated by two stage least squares. In Advances in Econometrics and Quantitative Economics: Essays in Honor of Professor C. R. Rao, ed. G. S. Maddala, P. C. B. Phillips, and T. N. Srinivasan, 66–87. Oxford: Blackwell.

Yahn, M. (2019). Soil Erosion and Conservation in Alabama. In Encyclopedia of Alabama. Retrieved from https://shorturl.at/tPTY8.

Zhang, D., Xie, J., & Affuso, E. (2017). An efficiency and productivity analysis of the agricultural sector in Alabama. International Journal of Applied Economics, 14(2), 19-36. Retrieved from https://shorturl.at/zF126.