Party Lines or Voter Preferences? Explaining Political Realignment

Nicolas Longuet-Marx*

Columbia University
March 19, 2025

Click here for the latest version

Abstract

This paper estimates a political equilibrium model to disentangle demand factors (voters) from supply factors (politicians) in shaping political outcomes, focusing on the recent realignment of blue-collar voters away from left-wing parties. I jointly evaluate the impact of changes in voter preferences and voter demographics (demand side) and party positions and party discipline (supply side) on voters' partisan realignment in U.S. House elections between 2000 and 2020. To measure candidate ideological positioning, I estimate a multimodal text-and-survey model from campaign websites. To estimate voter preferences, I build a new panel of precinct-level election results (N=1.3 million), which allows me to exploit congressional districts' border discontinuities for identification. The paper ultimately identifies parties' stronger polarization on cultural issues compared to economic issues as the main driver of voters' partisan realignment. In contrast, shifts in voter preferences—particularly the increasing preferences of blue-collar voters for progressive economic policies—have mitigated their defection from the Democratic Party. Absent these demand-side changes, voters' partisan realignment would have been even more pronounced. Within specific policy domains, the environment emerges as the topic where parties diverge most in economic versus cultural emphasis: Democrats frame it culturally, while Republicans focus on economic aspects. Simulations reveal that a progressive, economically focused environmental policy would gain greater blue-collar voter support than a culturally focused one.

*I am extremely grateful to Suresh Naidu, Pietro Tebaldi, Francesco Trebbi, Vincent Pons, and Charles Angelucci for their invaluable and supportive guidance throughout this project. For comments that have improved this paper, I thank Elliott Ash, Matthew Backus, Pierre Bodéré, Matilde Bombardini, Nathan Canen, Alessandra Casella, Julia Cagé, Tristan du Puy, Jeffry Frieden, Amory Gethin, Alexis Ghersengorin, Gautam Gowrisankaran, Dan Hopkins, Matias Iaryczower, Karam Kang, Ethan Kaplan, Kei Kawai, Ilyana Kuziemko, Félix Loubaton, John Marshall, Robert Metcalfe, Sergio Montero, Sebastián Otero, Thomas Piketty, Andrea Prat, Carlo Prato, Bernard Salanié, Hassan Sayed, Cailin Slattery, Andrey Simonov, Martin Vaeth, and Ebonya Washington. I am also grateful to seminar participants at Berkeley, Bocconi, Columbia, MIT, Princeton, PSE, Stanford, Yale, ASSA, APSA, and the Monash-Warwick-Zurich Text-as-Data Workshop. I am particularly indebted to Keyon Vafa for sharing and guiding me through the original TBIP code and Jeff Gortmaker for answering questions on PyBLP. I thank Isaiah Colmenero, Micol Galante, Pia Mahajan, Axel Martin, Jacob Posada, Lucas Puttre, and especially Ece Fisgin and Ignacio Ugalde, for excellent research assistance in collecting the precinct-level electoral results. Part of this paper was written while visiting UC Berkeley Haas School of Business—the hospitality of which is gratefully acknowledged. This paper received the IPUMS 2023 Best Student Spatial Work.

1 Introduction

While traditionally favored by the majority of less-educated voters, left-wing parties in Western democracies have seen their base shift significantly toward a more educated electorate (Kitschelt and Rehm, 2019; Gethin et al., 2022; Kuziemko et al., 2023). This trend is widespread, but the pace and intensity of this realignment in the United States stand out as particularly striking. Over the past 20 years, the proportion of voters with a high school diploma or less supporting the Democratic Party has dropped by more than 10 percentage points, while the share of college graduates voting Democratic has increased by a similar margin. This realignment has not only altered voting patterns but also catalyzed profound transformations in political polarization and policymaking, fundamentally reshaping the landscape of American democracy.

The literature is divided between demand-side and supply-side explanations of political realignment. On the one hand, some argue that voters' shifting preferences—such as an increasing focus on cultural issues—are the primary driver behind their realignment (Inglehart, 1997; Enke et al., 2021; Danieli et al., 2022). On the other hand, supply-side explanations suggest that the key factor is parties' changing positions (Rennwald and Evans, 2014; Kuziemko et al., 2023; Choi et al., 2024). No consensus has emerged, largely due to the challenge of jointly assessing changes in voter behavior and party strategies within a single framework.

This paper directly addresses this gap by offering a unified empirical framework that integrates both demand and supply factors to assess changes in political equilibrium outcomes. Specifically, the main empirical goal is to estimate the contributions of changes in voter preferences and voter demographics (demand-side changes) and changes in party leadership positions and party discipline (supply-side changes) to this recent voter realignment episode. The main takeaway is that supply-side factors accounted for all of less-educated voters' political realignment, while demand-side factors pushed in the opposite direction. Absent changes in voter preferences, political realignment by education would have been even more pronounced.

Understanding the drivers of political realignment requires making progress on longstanding questions in political economy: How does the interaction between voters and parties determine political outcomes in equilibrium? How do voters respond to policies offered by their local candidates? And are candidates catering their positions to their constituents, or only following their parties' line? Answering these questions has proven difficult due to a series of independent empirical challenges.

The first challenge concerns the measurement of candidate positions themselves; grappling with the multidimensionality of ideological positioning and comparing these dimensions across multiple elections is particularly difficult (Poole and Rosenthal, 2011; Bateman and Lapinski, 2016). Additionally, previous studies have typically been limited to measures available only for incumbents, as obtaining reliable information on the positions of challengers has often been more complex. Second, to recover the distribution of voter preferences, it is necessary to observe multiple pairs of candidates, each offering their own platform,

along with variation in voter choices for any given choice set (Berry and Haile, 2021). Third, identifying exogenous shifters for both demand and supply variation is essential to circumvent possible endogeneity concerns in the estimation of voters' and candidates' preferences (Berry and Haile, 2024). Finally, individual voting decisions are seldom observed and relying on aggregate voting data makes it difficult to recover individual preferences (King, 2013).

This paper makes progress on these issues in several ways. To begin with, I recover the position of all candidates (incumbents and challengers) in Congressional races on both cultural and economic dimensions by training a multimodal text-and-survey model based on candidate websites.[1] These dimensions are both interpretable and comparable over time thanks to candidate survey questions asked repeatedly. Next, to measure demand, I construct a new panel dataset of precinct-level election results in Congressional races between 2000 and 2020 to capture extremely granular, within-district variation in voter preferences (170,000 precincts per year). The granularity of the election data is key—not only for estimating the rich heterogeneity in voter preferences—but also for exploiting congressional district border discontinuities to identify these preferences. All in all, I estimate a political equilibrium model of U.S. House elections, quantifying the influence of supply-side factors, such as party leadership positions and party discipline, as well as demand-side factors, such as voter ideological preferences, voter demographics, and redistricting.

The key finding is that voters' realignment by education observed in the U.S. between 2000 and 2020 is primarily driven by supply-side factors. This conclusion rests on three main points. First, less-educated voters favor conservative cultural policies but progressive economic policies, with their preference for the latter increasing over time. Second, despite this shift in economic preferences, party polarization has increased twice as much on cultural issues as on economic ones. Taken together, these points indicate that the larger polarization on cultural issues has pushed away less-educated voters from supporting Democratic candidates, even though their preference for progressive economic policies has grown. These changes in voter preferences, however, led less-educated voters to support Democratic candidates more. In other words, absent these demand-side changes, less-educated voters would have shifted even more toward Republican candidates, further deepening political realignment. Below, I outline the empirical framework and intermediate results that support this conclusion.

I begin by documenting descriptive patterns in how voting patterns have changed over my sample period. The precinct-level panel data provides fine-grained variation,[2] allowing me to identify and compare the demographic variables along which this voter realignment has occurred. I show that realignment along educational lines has dominated realignment along any other demographic line, such as race, occupation,

[1] In Section 3, I provide a detailed explanation of the classification of each topic into cultural and economic dimensions. Economic topics typically include taxes, healthcare, welfare, trade policy, and labor relations, while cultural topics encompass crime, gun regulations, reproductive rights, education, LGBTQ rights, affirmative action, and environmental issues.

[2] Each precinct comprises on average 1,000 registered voters, that is, about 50 times smaller than a county and 400 times smaller than a congressional district.

or union affiliation.[3] Regarding candidate positioning, while party polarization is well-documented,[4] combining survey data with natural language processing applied to candidate websites shows that polarization on cultural issues has risen twice as much as on economic ones, with two-thirds of this widening gap driven by the Democratic Party. This analysis applies to both winners and losers.

In the next stage of the analysis, I estimate a political equilibrium model that captures two key aspects of the electoral process. First, candidates compete by selecting positions in the ideological space, aiming to maximize their chances of election while minimizing the ideological distance between their chosen positions and those of their party leadership. The weight given to party leadership reflects the strength of party discipline. Second, heterogeneous voters choose the candidate whose platform and characteristics best align with their preferences.

To recover voter preferences over candidate platforms, I estimate a structural model of voter behavior using fine-grained variation in election results from precinct data combined with individual survey data (Berry et al., 2004). Since candidate positions depend on unobservables affecting demand, I address their endogeneity by matching contiguous precincts across congressional district borders within each state, following the approach used for counties in Spenkuch and Toniatti (2018). Since candidates set strategies at the district level and each individual precinct (about 400 per district) has a negligible impact from the candidate perspective, differences in candidate positions between contiguous precincts across district borders can be treated as random, especially after accounting for time-invariant precinct characteristics. While unobservables cannot be directly tested, I provide evidence that, conditional on fixed effects, candidate positions are uncorrelated with a wide range of precinct-level demographics.

I find that more-educated voters prefer progressive positions on cultural issues, but not on economic ones. In contrast, less-educated voters have increasingly favored progressive economic positions, especially in more recent years. The educational divide on cultural issues has also deepened over time, with education becoming an even stronger predictor of support for progressive cultural policies.

Further, I document a rise in party discipline in both parties. Using estimates of the distribution of ideological preferences in each electoral precinct, I recover supply-side parameters measuring the weights that candidates allocate to the probability of wining versus aligning with the party when choosing the positions they offer to voters. I show that party discipline is stronger among Republicans than Democrats and has dramatically increased over time: the ability of a House candidate to adapt their positions to their constituents has been divided by three, resulting in an uniformization of candidates across districts.

Finally, with demand-side and supply-side estimates in hand, I simulate multiple counterfactual scenarios to assess the contribution of each factor to the overall change in voting behavior. Since candidate positions are an equilibrium outcome, I begin by decomposing the changes in candidate positions, especially the larger polarization on cultural issues, into its multiple drivers. I show that at least 75% of these

[3] The average education gap between Democratic and Republican voters has increased by 2.5 months of schooling per election.
[4] See e.g., McCarty et al. (2016), Gentzkow et al. (2019) for representatives, and Bonica (2013) for candidates.

changes can be attributed to supply-side factors—particularly the polarization of party leadership and increased party discipline. In contrast, demand-side factors explain only 3.5% of the variation in candidate positions.

All these results lead to the paper's key result: when assessing changes in voter choices, I show that virtually all of the explained political realignment is driven by supply-side factors, particularly the rising polarization between the two parties on cultural issues compared to economic issues. In contrast, changes in voter preferences have been pushing in the opposite direction: the growing preference for progressive economic policies among less-educated voters has mitigated the extent of realignment. In other words, while less-educated voters have increasingly supported progressive economic policies, traditionally offered by the Democratic Party, the parties have become more polarized on cultural issues, pushing less-educated voters toward the Republican Party. These findings resonate with theories suggesting that cultural polarization has weakened democracy's ability to address rising inequality (Bonica et al., 2013; Roemer, 1998; Hacker and Pierson, 2020).

As a final step in my analysis, I employ the model to examine how parties' cultural polarization affects voter support for environmental policy. Environmental issues present an interesting case study, not only because they have recently become an important part of political discourse and public opinion (Dunlap et al., 2016; Egan et al., 2022), but also because they carry both economic and cultural significance (Besley and Persson, 2023). Candidates may advocate for environmentally progressive policies that highlight cultural themes which appeal to voters' values, such as climate education, climate justice, ethical consumption or "believe in science" initiatives, or they may emphasize policies with economic implications, such as a "Green New Deal." By projecting candidates' environmental positions on the cultural and economic dimensions, I show that Democratic candidates' environmental positions have a much stronger cultural than economic dimension, while the reverse is true for Republicans.[5] The environment stands out as the topic on which this divergence between the two parties is the most pronounced. Using the empirical model estimated in the paper, I show that current Democratic positions on the environment, which are heavily cultural, deter less-educated voters. In contrast, an equally progressive environmental policy with a stronger economic focus, such as a "Green New Deal," would generate more support from less-educated voters. In a companion paper (Bombardini et al., 2024), we adapt the framework developed in this paper specifically to environmental issues, and precisely examine how demand and supply of environmental policy respond to changes in environmental conditions and employment opportunities in the environmental sector.

[5]In other words, Democratic candidates who have progressive environmental positions tend to be progressive on cultural issues rather than on economic issues, whereas Republican candidates show the opposite pattern: Republicans who are conservative on environmental issues are also conservative on economic issues, not on cultural issues. Appendix J provides examples of website pages on the environment representative of each party's position.

Related Literature This paper makes contributions to several strands of the literature in economics and political science.

First, this paper contributes to the literature estimating structural models of voter preferences, particularly papers that recover preferences from aggregate election results (Coate and Conlin, 2004; Rekkas, 2007; Strömberg, 2008; Gordon and Hartmann, 2013; Sieg and Yoon, 2017; Ujhelyi et al., 2021; Kawai and Sunada, 2022; Iaryczower et al., 2022; Cox, 2023; Cox and Shapiro, 2024; Berry et al., 2024). A unique feature of this paper is the use of election results at a much more granular level than where supply decisions are made. The granularity of the election data is crucial—not only for estimating the rich heterogeneity in voter preferences—but also for exploiting congressional district border discontinuities to address endogeneity in candidate ideology. To the best of my knowledge, this approach offers the first estimates of individual voter preferences over House candidate ideology.

Many articles have also inferred voter preferences from correlations between voters' reported electoral choices in survey data and nationwide party positions obtained from the Comparative Manifesto Project (Adams et al., 2004; Elff, 2009; Evans and Tilley, 2012; Danieli et al., 2022). An important contribution of this paper is to use plausibly exogenous variation in candidate positions to recover voter preferences. Even with micro data, one needs cross-market variation to recover voter preference parameters in order to distinguish changes in product attributes from changes in preferences (Berry and Haile, 2021). The framework developed in this paper leverages the multiplicity of House candidates and granular precinct-level data, exploiting district border discontinuities to address the endogeneity of candidate positions.

A key strength of this paper is the integration of demand estimates in an equilibrium framework which endogenizes candidate strategies. In this context, the estimation of supply parameters contributes new evidence to the literature on political common agency, where politicians must balance the competing demands of their party and constituents. Several papers have explored the impact of party discipline, special interest groups, and constituent preferences on politicians' positioning, with most concluding that party discipline is both strong and increasing over time (Ansolabehere et al., 2001; Lee et al., 2004; Mian et al., 2010; Canen et al., 2020, 2021; Bombardini et al., 2023; Iaryczower et al., 2024; Cox and Shapiro, 2024). My framework contributes to this literature in two key ways. First, it directly measures candidates' sensitivity to voters' actual policy preferences, rather than relying on proxies like Presidential vote share. This approach allows for a direct comparison of how much weight candidates place on constituents' preferences versus party loyalty. Second, it analyzes candidate responsiveness to constituents and party discipline before the election, rather than focusing solely on post-election behaviors.

This article also contributes to the literature on measuring politicians' ideological positioning by developing a framework with four compelling features. First, it relies on information extracted directly from candidates' campaign materials, rather than their actions once elected—such as roll-call votes (Poole and Rosenthal, 1985; Martin and Quinn, 2002), speeches in Congress (Gentzkow et al., 2019; Enke, 2020), or proxies like campaign contributions (Bonica, 2013; Hall and Snyder, 2015) or voters' behavior (Krasa and

Polborn, 2014). Tausanovitch and Warshaw (2017) show these measures are often weakly correlated with each other and may not capture the actual ideological variation presented to voters. Second, by observing both candidates, rather than only elected members, I am able to capture the behavior of both opponents in each election—a critical component for estimating the strategic game of candidate competition. Previous studies using candidate survey data like this paper (Ansolabehere et al., 2001; Shor and McCarty, 2011; Shor and Rogowski, 2018) were constrained by the limited sample size of survey respondents. By incorporating candidate websites alongside surveys, I overcome this limitation. Candidate websites have also been used in the literature to derive unidimensional measures of ideology (Di Tella et al., 2023; Meisels, 2023). Third, this framework allows me to assign candidate positions on multiple political dimensions, which are interpretable and comparable over time due to repeated survey questions. Finally, by combining survey and website data, I estimate candidate ideology without relying on party labels, unlike much of the literature. Party-based measures often misplace extreme candidates, as they reflect a politician's centrality within their party rather than true ideological positioning (Noel, 2014, 2016). Since party dynamics and ideology co-evolve, they also prevent meaningful comparisons over time.

Leveraging panel data of precinct-level results, this paper provides new insights into the documentation of political realignment in the U.S. by utilizing official, granular election results. While previous studies (e.g., Kitschelt and Rehm (2019); Gethin et al. (2022); Kuziemko et al. (2023) among others) have relied on survey data to document the evolution of voting behaviors, I complement these *stated preferences* approaches by providing evidence from *revealed preferences* based on actual voting behaviors. Using precinct-level data also prevents biases in sample selection and allows me to bring more nuance to the extent of political realignment, along many dimensions and across various geographical scales.

Lastly, this paper contributes to the literature looking at the demand and supply of environmental and climate policy. It provides new evidence on candidate strategic positioning on environmental issues, offering insights into the factors shaping politicians' supply of environmental policies (Fredriksson et al., 2011; Gagliarducci et al., 2019; Gazmararian and Milner, 2021; Kaplan et al., 2024). It also highlights how these strategic differences in candidate positioning influence voter support for environmental policies, contributing to the literature on the determinants of policy preferences for such policies (Egan and Mullin, 2017; Besley and Persson, 2023; Dechezleprêtre et al., 2022; Drews and Van den Bergh, 2016). In a companion paper (Bombardini et al., 2024), we adapt the framework developed in this paper specifically to environmental issues, and examine how demand and supply respond to changes in environmental conditions and employment opportunities in the environmental sector.

The remaining of this paper is organized as follows: Section 2 outlines the conceptual framework of the paper, Section 3 presents the data, the methods to measure candidate ideology, and provides descriptive statistics, Section 4 describes the estimation of demand-side parameters, Section 5 describes the estimation of the supply-side parameters, Section 6 decomposes changes in candidate positions and voting behavior between demand-side and supply-side factors, Section 7 studies positioning and voting on environmental

issues, and Section 8 concludes and discusses the results.

2 Empirical Model

2.1 Setup

I consider a static model of local political competition where two candidates compete to maximize their probability of being elected while complying with party discipline. Citizens engage in sincere voting based on candidates' ideology and voting is compulsory.[6]

The setup is characterized by two types of agents. For a given election, indexed by t, there is a set of voters $\mathcal{I}_t$, each indexed by i with observable characteristics $\mathbf{w_{it}}$ (e.g., education, race, age). Voters choose between two candidates, D and R, each indexed by j, who choose their k-dimensional political platforms, $\mathbf{x_{jt}} \in \mathbb{R}^k$, for $j = D, R$.

When voting for candidate j, voter i's utility is $u_{it}(\mathbf{x_{jt}}; \mathbf{w_{it}})$. The probability that i votes for the Democratic candidate in election t is then

$$s_{iDt}\left(\mathbf{x_{Dt}}, \mathbf{x_{Rt}}; \mathbf{w_{it}}\right) \quad = \quad \Pr\left(u_{it}(\mathbf{x_{Dt}}; \mathbf{w_{it}}) \geq u_{it}(\mathbf{x_{Rt}}; \mathbf{w_{it}})\right). \tag{1}$$

The total vote share obtained by the Democratic candidate can be written as:

$$s_{Dt} = \int s_{iDt}\left(\mathbf{x_{Dt}}, \mathbf{x_{Rt}}; \mathbf{w_{it}}\right) dF_t(\mathbf{w_{it}}), \tag{2}$$

where $F_t(\cdot)$ is the distribution of voters' observable characteristics in election t. Since voting is compulsory, the Republican vote share is $s_{Rt} = 1 - s_{Dt}$.

The vector $\mathbf{x_{jt}} \in \mathbb{R}^k$ is chosen by the local candidate ($j = \{D, R\}$) trading off the probability of being elected on the one hand, and comply with the national leadership of their party on the other. Formally, candidates seek to be elected as they derive a rent $Q > 0$ from being in office, and maximize

$$\Pi_{jt}(\mathbf{x_{jt}}) = \underbrace{P_j(\mathbf{x_{jt}}, \mathbf{x_{-jt}})}_{\text{probability of winning}} Q - \underbrace{\lambda_{jt}(\mathbf{x_{jt}}, \mathbf{N_{jt}})}_{\substack{\text{cost of deviation from} \\ \text{national party platform}}}, \tag{3}$$

where the probability of winning $P_j(\mathbf{x_{jt}}, \mathbf{x_{-jt}})$ depends on the total vote share s_{Dt} and a random shock, and the function $\lambda_{jt}(\cdot)$ captures candidates' cost of deviating from the national party position $\mathbf{N_{jt}} \in \mathbb{R}^k$.

[6] While voter turnout decisions are undoubtedly important, this paper focuses on the choice between Democratic and Republican candidates, as this has been the primary emphasis of most of the existing literature. Note that, from a candidate's perspective, what matters is not the absolute number of votes but the relative vote share compared to the opponent, which differs from most economic markets. Appendix Section G provides estimates from an alternative demand model that includes an endogenous turnout decision. The estimates from the specification with turnout are similar to those under compulsory voting, though of smaller magnitude.

The lower the cost of deviating, the greater is candidates' ability to adapt to their local conditions.

The need to comply with the position of the national leadership is not modeled explicitly, yet it is justified by a variety of political and institutional factors. First, party leaders reward candidates who follow the party line, both before and after the election. Second, it is costly for local candidates to come up with their own political strategy. I treat the evolution of the national leadership platforms as exogenous. Exogeneity here is intended as relative to the determinants of the local elections I study, and can be justified by the limited impact of each congressional district with respect to federal concerns.

A Nash equilibrium in election t is characterized by a collection of positions $(\mathbf{x_{Dt}}, \mathbf{x_{Rt}})$ and vote shares (s_{Dt}, s_{Rt}) such that (i) $\mathbf{x_{jt}}$ maximizes $\Pi_{jt}(\mathbf{x_{jt}})$ and (ii) $s_{jt} = \int s_{ijt}(\mathbf{x_{Dt}}, \mathbf{x_{Rt}}; \mathbf{w_{it}})\, dF_t(\mathbf{w_{it}})$, for $j \in \{D, R\}$.

Such duopoly models with spatial competition in multiple dimensions do not necessarily admit a unique equilibrium (Hotelling, 1929; Caplin and Nalebuff, 1991). In Appendix F, I show that with sufficiently high party discipline, an equilibrium always exists and is unique. In particular, I can verify that with the estimated parameters, the model always admits a unique equilibrium.

2.2 Empirical Goal

The main objective of this paper is to decompose the change in individual vote between two periods into changes in $s_{ijt}(\cdot)$ and $F_t(\mathbf{w_{it}})$ on the demand side and changes in $\mathbf{N_t} = (\mathbf{N_t^D}, \mathbf{N_t^R})$ and $\lambda_{jt}(\cdot)$ on the supply side. Importantly, I evaluate the impact of changes in leadership positions exclusively through the resulting adjustments to $(\mathbf{x_{Dt}}, \mathbf{x_{Rt}})$ made by local candidates. This assumes away any direct influence that leadership changes might exert on local vote shares. Formally, the utility in (1) does not depend on $\mathbf{N_{jt}}$ directly.

To make these decompositions, I proceed in five steps. First, I employ granular precinct-level data to obtain precise measures of the observed vote shares S_{jt}. Second, I obtain measures of $\mathbf{x_{jt}}$ and $\mathbf{N_t}$ using an ideal point model of candidate ideology. Third, I leverage variation in $\mathbf{x_{jt}}$ between adjacent precincts in bordering districts to identify and estimate $s_{ijt}(\cdot)$ conditional on t and voter characteristics $\mathbf{w_{it}}$. Fourth, I recover the supply-side cost functions $\lambda_{jt}(\cdot)$ underlying candidates' choice of $\mathbf{x_t}$ given demand. Lastly, the parameters recovered in the first four steps allow me to conduct a number of counterfactual exercises. For example, I recover alternative candidate positions and voting behavior, if the party leadership positions $\mathbf{N_t} = (\mathbf{N_t^D}, \mathbf{N_t^R})$ had not changed but all other factors changed, or if voter preferences $s_{ijt}(\cdot)$ had not changed but all other factors had.

I will start by discussing the measurement of S_{jt}, $\mathcal{I}_t$, $\mathbf{x_t}$, and $\mathbf{N_t}$, and document the time series variation whose decomposition is my main object of interest. I will then introduce an econometric specification for the estimation of the primitives $s_{ijt}(\cdot)$ and $\lambda_{jt}(\cdot)$, and discuss exclusion restrictions that are sufficient for identification.

3 Data and Descriptive Statistics

3.1 Panel of Precinct-level Election results

I collect precinct-level electoral results from 2000 to 2020. Precincts are the smallest geographical unit at which election results are available in the U.S. Each precinct has an average population of 1,000 registered voters. There are about 175,000 precincts in the U.S., with on average 400 precincts per congressional district, which allows me to obtain considerable variation within congressional districts. Despite these election results being made public by county or state officials, there has been so far no unified dataset that contains the precinct-level Congressional results for the entire period that includes the geographical boundaries of each precinct. I therefore combined data from more than 30 different sources, such as Secretaries of State and county officials, in order to obtain results for the largest possible number of states over the period. The list of sources is described in the Appendix section dedicated to the description of the electoral data (see Appendix C).

Figure 1 shows the distribution of precincts in the U.S. for the 2020 election. Using precinct-level data instead of county-level data offers three advantages. First, precincts are much smaller and more demographically homogeneous than counties, reducing concerns about ecological fallacy and facilitating the estimation of individual preferences. Second, unlike counties, precincts are designed to have roughly equal populations everywhere, which allows for significant variation in both urban and rural areas. In contrast, using county-level voting data provides very scarce within-congressional district variation in urban areas. For instance, there are 16 congressional districts in Los Angeles county alone and Manhattan spans 3 distinct districts. In comparison, these counties have 4,312 and 1,266 precincts, respectively. The third advantage of precinct-level over county-level data is that the finer granularity offers a stronger foundation for identification assumptions, as I will describe in Section 4.

Since precinct boundaries are changing over time, building a panel of electoral precincts is challenging. I therefore apportion all votes to the 2010 block-group level, which have approximately the same population as electoral precincts.[7] This offers the benefit of creating a consistent panel dataset with geographies that can be compared across elections. This is also the most granular level at which census demographic variables are available. Here and throughout, I will use the term *precinct* instead of *block-group* as the unit of observation, with a minor abuse of terminology. Appendix C describes the strategy implemented to apportion the electoral results to the block group level by using the spatial overlaps of precincts and census blocks (35 blocks per block-groups). In order to maintain a balanced panel of states across the

[7]Note that while block-group populations are relatively consistent across states (averaging 1,400 people, with a range from 1,150 to 1,800), precinct sizes vary significantly. For example, New Jersey and Washington have roughly comparable populations, but Washington has 4,794 electoral precincts, whereas New Jersey has only 746. As a result, in New Jersey, a single precinct corresponds to an average of eight block-groups, whereas in Washington, each precinct corresponds to one block-group, though they do not spatially align. Note, however, that since the identification strategy relies solely on within-state comparisons, this variation does not pose a threat to identification. Additionally, all results are clustered by congressional district and election, ensuring that the precision of the estimates is not overstated.

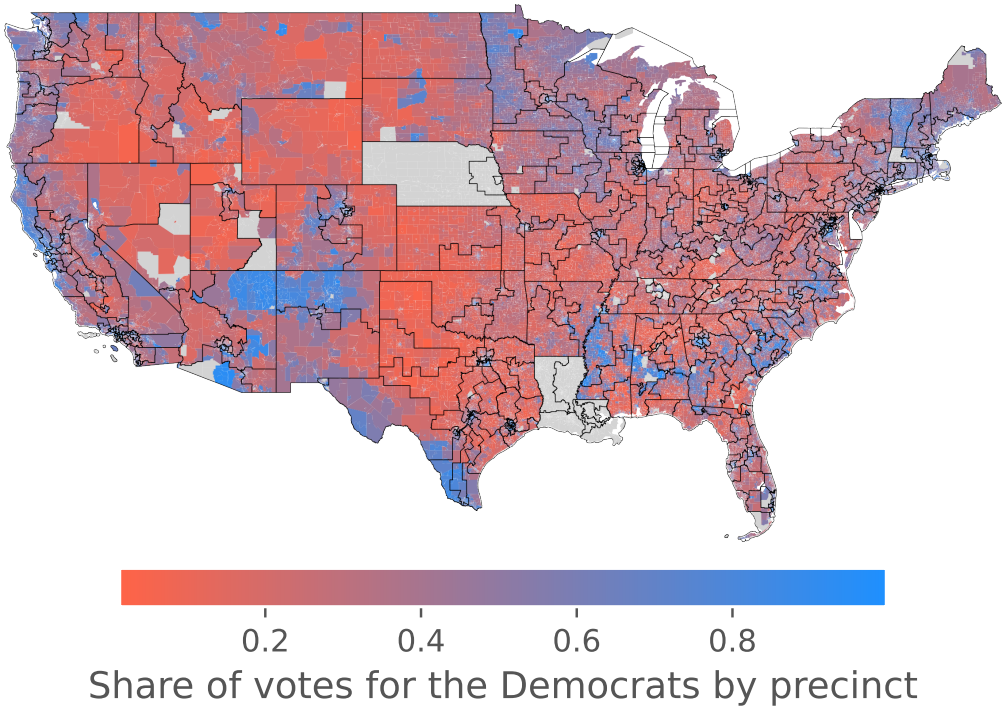
two periods of analysis, I include only those states that are available in both the pre-2010 and post-2010 periods. Table A.2 in the Appendix presents the sample of states included in the analysis.

I then combine election results with census demographics at the block-group level from NHGIS IPUMS (Manson et al., 2021) using the decennial census and the American Community Survey (ACS). In order to estimate individual-level preferences and not only aggregate preferences, I need the full distribution of demographics along three dimensions: education, race, and age. This requires having both the average, the variance, and the covariance of each of these demographic variables at the block-group level. Since only marginal distributions are reported in the decennial census and the ACS, I recover the joint distribution at the block-group level from a multi-scale model combining block-group level and tract-level demographic counts with PUMA-level individual data from the ACS individual files. Appendix D details the strategy to recover the full distribution of demographics in each block-group.

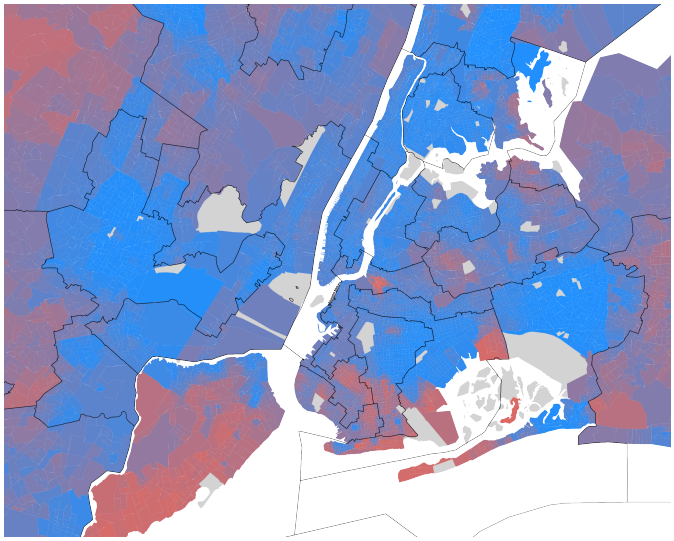
In the remainder of this paper, I will refer to white and non-white voters with a specific definition. The census asks both about race and ethnicity, and people can identify as belonging to multiple races. In order to have demographic groups with sufficiently large sizes in most precincts, I classify both "Hispanic white" and "Hispanic non-white" as "non-white" and individuals reporting multiple races as "non-white". The group of "white" voters is therefore constituted of "Non-Hispanic whites", with a single race. Under this definition, the average share of whites is 75% in 2000 and 68% in 2020.

In addition to census demographics, I also collected the share of unionized workers using localized contract-level data from the Federal Mediation and Conciliation Service (FMCS), which gives establishment-level data on unionization. I geocode each establishment to obtain its census block, I then apportion employment from the workplace to the voting place using LEHD Origin-Destination Employment Statistics (LODES) data which give yearly block-to-block job flows. This strategy allows me to obtain a block-group-level estimate of voters that are part of a labor union. I also collected the share of voters that belong to any church, and specifically that belong to an evangelical church, using a combination of data from the Census of Religious Bodies and data from Axle.

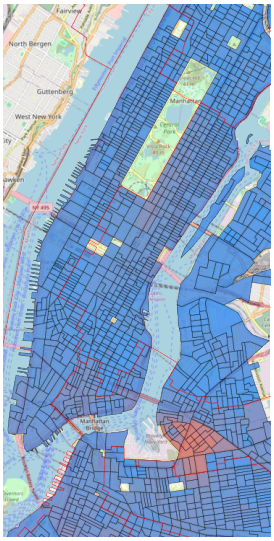
Lastly, in order to have additional individual-level demographic heterogeneity, I use individual vote-choice survey data from the Cooperative Election Survey (CES), The American National Election Survey (ANES), the General Social Survey (GSS), and from Gallup, obtained from Kuziemko et al. (2023).



(a) Continental United States



(b) New York City, NY



(c) Manhattan, NY

Figure 1: Precinct-level democratic vote share in 2020

Notes: The figure shows the 2020 distribution of House Democratic vote share in each precinct. Congressional district borders are shown in black (panels (a) and (b)) and in red (panel (c)). Each congressional district contains on average 400 precincts. Precinct results have been spatially interpolated at the block-group level as described in Appendix C.

3.2 Evidence of Political Realignment

I start by providing new evidence of political realignment using this novel panel of precinct-level electoral results. Most of the existing evidence on realignment has been documented through survey data (Kitschelt and Rehm, 2019; Gethin et al., 2022; Kuziemko et al., 2023). Using aggregate election results presents three main advantages. First, it allows me to study the revealed rather than stated preferences of voters, which may more accurately reflect their voting choices, as survey respondents might misreport their votes. Second, it helps reduce potential sample bias resulting from the partial coverage of surveys, which may not reach the entire voter population. Lastly, it gives more power to study precise phenomena and detect non-linearities. Naturally, relying on aggregate results also has some drawbacks. The main issue is to recover individual voting behavior from aggregate election results, referred to as the ecological inference problem (King, 2013). I discuss these challenges and the strategy proposed to overcome them in Section 4.

While much of the literature on political realignment has focused on education, I begin by examining the variables along which this realignment has been most pronounced. Specifically, I analyze 11 variables capturing precinct-level demographic composition, occupational distribution, and affiliation with religious or labor unions. To compare the extent of realignment across these dimensions, I run a single linear regression of precinct-level Democratic vote shares on demographic variables interacted with a time trend. Each demographic variable is normalized to have a mean of zero and a standard deviation of one to ensure comparability of coefficients across dimensions. The regression specification is as follows:

$$S_{Dpt} = \sum_{w} \beta^{w} \times w_{pt} \times t + \sum_{w} \gamma^{w} \times w_{pt} + \mu_t + \epsilon_{pt}, \tag{4}$$

where S_{Dpt} is the House Democratic vote share in precinct p at election t, each w_{pt} captures a demographic dimension of precinct p, μ_t are election fixed effects. Figure 2 reports both the estimates from the joint regression including all demographic variables, Appendix Figure A.3 reports the unconditional estimates from regressing one coefficient at a time. Positive values ($\beta^{w} > 0$) indicate a realignment toward the Democratic Party while negative values indicate a realignment toward the Republican Party. The figure shows that the effect of education dominates all other demographic variables in absolute terms, with more educated areas shifting toward the Democrats. Realignment along education is twice as strong as along income lines, and is even more pronounced among white voters. Other variables also play an important role, particularly the share of white residents and precinct urban density, with predominantly white and more rural areas shifting toward the Republicans.

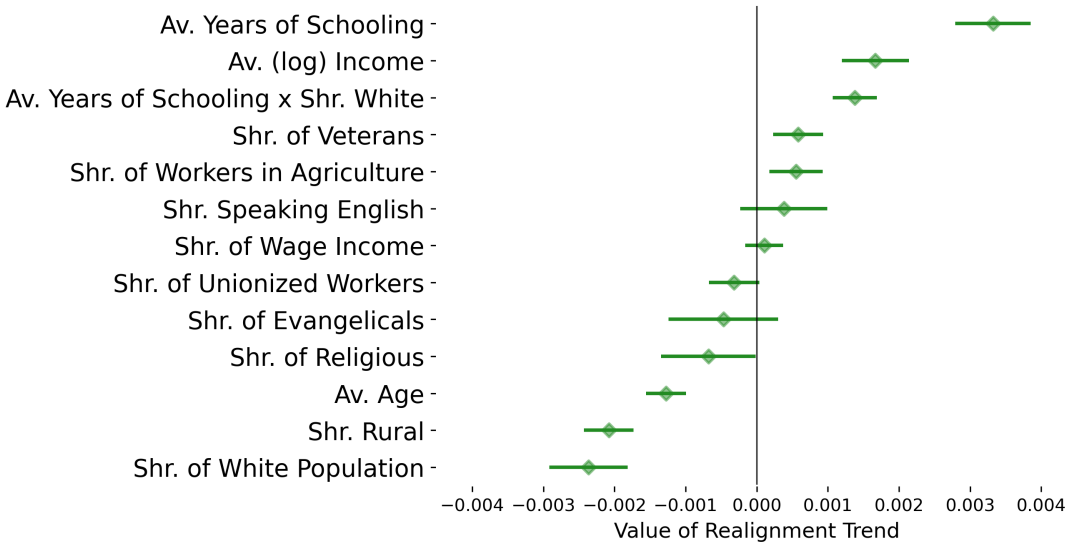


Figure 2: Observed Trends in Political Realignment: Education as the Key Driver

Notes: The figure shows, for each demographic variable w_{pt}, the coefficients β^w from the following linear regression: $S_{Dpt} = \sum_w \beta^w w_{pt} \times t + \sum_w \gamma^w w_{pt} + \mu_t + \epsilon_{pt}$ where S_{Dpt} is the House Democratic vote share in precinct p at time t, and μ_t are election fixed-effects. Positive coefficients indicate realignment toward the Democratic Party while negative coefficients indicate realignment toward the Republican Party. The bars around each marker show the 95% confidence intervals with standard errors clustered two ways, at the precinct level and at the congressional-district-by-year level. Appendix Figure A.3 shows the results from a regression of each demographic variable separately. Appendix Figure A.4 shows the results from regressions holding demographics at their 2000 level and controlling for precinct fixed effects.

I now turn to measures of political realignment focusing on educational lines. Figure 3 shows, for each election, the linear relationship between precinct-level education and Democratic voting. In each successive election, the average gap in education between Democratic and Republican voters has increased by 2.5 months of schooling. Importantly, the rise in this correlation has preceded the election of Donald Trump in 2016, with no specific jump in the educational gradient in that year.

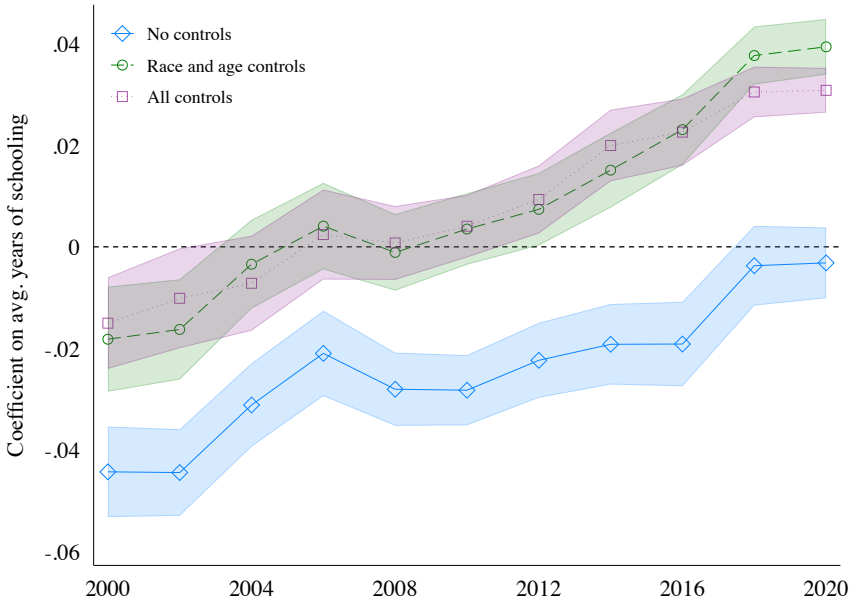


Figure 3: Increasing correlation between precinct-level education and Democratic vote shares.

Notes: The figure shows the coefficient β_t^{edu} from the following regression: $S_{Dpt} = \beta_t^{edu} edu_{pt} + \mu_t + \epsilon_{pt}$. The coefficients in blue with diamond markers are unconditional while the coefficients in green with circle markers are from a regression controlling for the precinct-level share of white and the precinct-level average age by year. The coefficients in purple with square markers additionally control for average income, share of unionized workers, share of religious people, share of veterans share of veterans, share of workers in agriculture, share of the population that lives in rural areas, separately for each year. Standard errors are clustered two ways, at the precinct level and at the congressional-district-by-year level

Precinct-level election results allow me to delve further into this relationship and investigate potential non-linearities. Figure A.5 shows the average Democratic vote share for each 5% quantile of the education distribution. The figure shows a U-shape relationship between education and Democratic vote shares for both the early and later years, with most changes over time occurring at the tails of the distribution.

Figure A.6 shows the evolution of the education gradient both within and between congressional districts. The relationship within district is strongly negative, with more educated precincts inside each congressional district voting more for the Republican candidates, with little change over time. This negative relationship contrasts with the pattern observed between districts, where more educated districts have been consistently voting more for the Democratic Party, echoing findings from Gelman (2009) for income. Overall, the precinct-level U-shape relationship can be thought as the sum of these two opposite relationships, between and within districts.[8]

[8]For each year, the total covariance between education and Democratic vote share $\mathbb{C}(S_{Dpt}, edu_{pt})$ can be decomposed into the within district covariance $\mathbb{E}\big[\mathbb{C}(S_{Dpt}, edu_{pt}|d(p))\big]$ and the between district covariance $\mathbb{C}\big[\mathbb{E}(S_{Dpt}|d(p)), \mathbb{E}(edu_{pt}|d(p))\big]$. For instance, in 2020, $\mathbb{C}(S_{Dpt}, edu_{pt}) = 0.005$, $\mathbb{E}\big[\mathbb{C}(S_{Dpt}, edu_{pt}|d(p))\big] = -0.012$, and $\mathbb{C}\big[\mathbb{E}(S_{Dpt}|d(p)), \mathbb{E}(edu_{pt}|d(p))\big] = 0.017$.

3.3 Candidate ideological positioning

In order to recover candidate ideologies on multiple dimensions, I combine candidate survey data from VoteSmart with text data from candidate websites obtained from the United States Elections Web Archive, maintained by the Library of Congress.[9] Appendix Figures A.1 and A.2 provide illustrative examples of the raw data.

Project Votesmart is a non-partisan organization that has been sending surveys to candidates since the 1990s, covering a wide range of questions about candidates' political stances on various topics. Response rates began to decline in the late 2000s, prompting Project Votesmart to conduct internal research using candidate statements, press releases, and interest group ratings to impute answers for candidates who have not responded. When candidates' positions on an issue remain unclear from these sources, the response was left unknown.

Because some questions vary across election cycles, I match questions between cycles only when their framing is nearly identical. I include questions that were asked in at least two election cycles and exclude those focused on specific politicians or relationships with specific countries. I have a total of 132 distinct questions from 2000 to 2020. Each question belongs to one of the 15 topics: abortion, crime, education, environment, gun regulation, campaign finance, immigration, international relations and security, diversity questions, employment, trade, taxes, health care, social security, and welfare. I then classify each of these topics into two main categories: cultural and economic issues. I categorize employment, trade, taxes, health care, social security, and welfare as economic issues, while classifying all other topics as cultural. Although some other topics, like immigration and the environment, also hold economic relevance, I show in Appendix Section E that candidates' positions on these topics tend to align more closely with candidates' cultural than economic ideal points. In Section 7, I relax this binary classification and examine candidates' relative cultural versus economic positioning on environmental issues.

In order to obtain an ideal point for each candidate, I apply a Bayesian item response model, similar to Clinton et al. (2004), Jessee (2009), and Shor and Rogowski (2018), detailed in Appendix Section E. The model estimates each candidate's underlying ideology, along with a difficulty parameter for each survey question that characterizes its position in the ideological space and a discrimination parameter that indicates how polarizing the question is. The model is estimated by Marginal Maximum Likelihood, separately for cultural and economic issues. I assign standard normal priors to the ideal points. By definition, all candidates answered the set of questions only partially since not all questions were asked every year. I also obtain the standard errors of each ideal point.

To complement these survey-based ideal points for candidates who did not respond or provided only partial answers, I also incorporate data from candidate websites. Using the United States Elections Web Archive, I scrape each candidate's website on the day before the general election. I then process and transform the text into embedding vectors to extract valuable information (Dai et al., 2015), as detailed

[9] Websites are only available starting in 2002.

in Appendix Section E.

I then train a machine learning regressor using the text features with an Extreme Gradient Boosting algorithm (Chen et al., 2015). I obtain a mean squared error (MSE) of 0.17 for the economic prediction and 0.21 for the cultural dimension (19% and 21% of a standard deviation, respectively).

Finally, I combine the information from survey and websites using the relative uncertainty of both measures. Candidates without any survey answer (29%) are assigned their website ideal points only (x_{jk}^{website}), similarly candidates for whom I do not have the website (18%) are assigned their survey ideal point only. For all those with both survey and website ideal points (46%), I take a weighted average of the two measures using their relative uncertainty as the weighting factor:

$$x_{jk} = \omega_{jk} x_{jk}^{\text{survey}} + (1 - \omega_{jk}) x_{jk}^{\text{website}}, \tag{5}$$

with $\omega_{jk} = \frac{MSE(x_k^{\text{website}})}{se(x_{jk}^{\text{survey}})^2 + MSE(x_k^{\text{website}})}$.

Appendix Figure A.22 shows the correlation between the survey-based x_{jk}^{survey} and website-based ideal points (x_{jk}^{website}) on each topic, which is above 90% overall and above 70% within party.

I do not observe the ideology of about 9% of the sample of candidates which are therefore excluded from the analysis.

Appendix K describes an alternative estimation of politicians' ideology based on an unsupervised probabilistic topic model, adapted from Vafa et al. (2020). The estimated dimensions of the two methods have a correlation of about 0.5 (see Appendix Figure A.34), both between and within party.

Figure 4 shows the distribution of candidate positions on the cultural and economic dimensions. The figure shows the positions of notable candidates in 2020, along with the positions of party leadership, which are an average of the ideological positions of key party figures: the Speaker of the House, the Majority and Minority Leaders, the Whips, and the Chairs of the Caucus and the Conference.

The distribution of Democratic and Republican candidates are clearly distinct from each other. However, contrary to measures such as DW-Nominate, the distribution of candidates from the two parties overlap substantively, with the most conservative Democratic candidates located to the right of the most progressive Republican candidates. Canen et al. (2021) found similar polarization patterns in Congress, once accounting for party discipline from party leaders. Importantly, while the overall Spearman (rank) correlation between the cultural and economic dimensions is 0.88, the within-party correlations are much smaller: 0.49 among Democratic candidates and 0.46 among Republican candidates.[10] These relatively low correlations suggest that candidates' cultural and economic positions vary independently, with some candidates being significantly more progressive on one dimension than the other.

Appendix Figure A.8 compares my measures with commonly used measures of ideology, such as DW-Nominate (Poole and Rosenthal, 1985), two-dimensional positions from Canen et al. (2021), which are

[10] The overall covariance between cultural and economic positions is $\mathbb{C}(x_{j,\text{cult}}, x_{j,\text{econ}}) = 0.623$, which can be decomposed into: $\mathbb{E}\big[\mathbb{C}(x_{j,\text{cult}}, x_{j,\text{econ}}|p(j))\big] = 0.074$ and $\mathbb{C}\big[\mathbb{E}(x_{j,\text{cult}}|p(j)), \mathbb{E}(x_{j,\text{econ}}|p(j))\big] = 0.549$, where $p(j)$ denotes the party of each candidate.

both available for election-winners only, and the contribution-based measure of ideology from Bonica (2014) which are also available for election-losers. Both the cultural and economic dimensions are highly correlated with other measures.

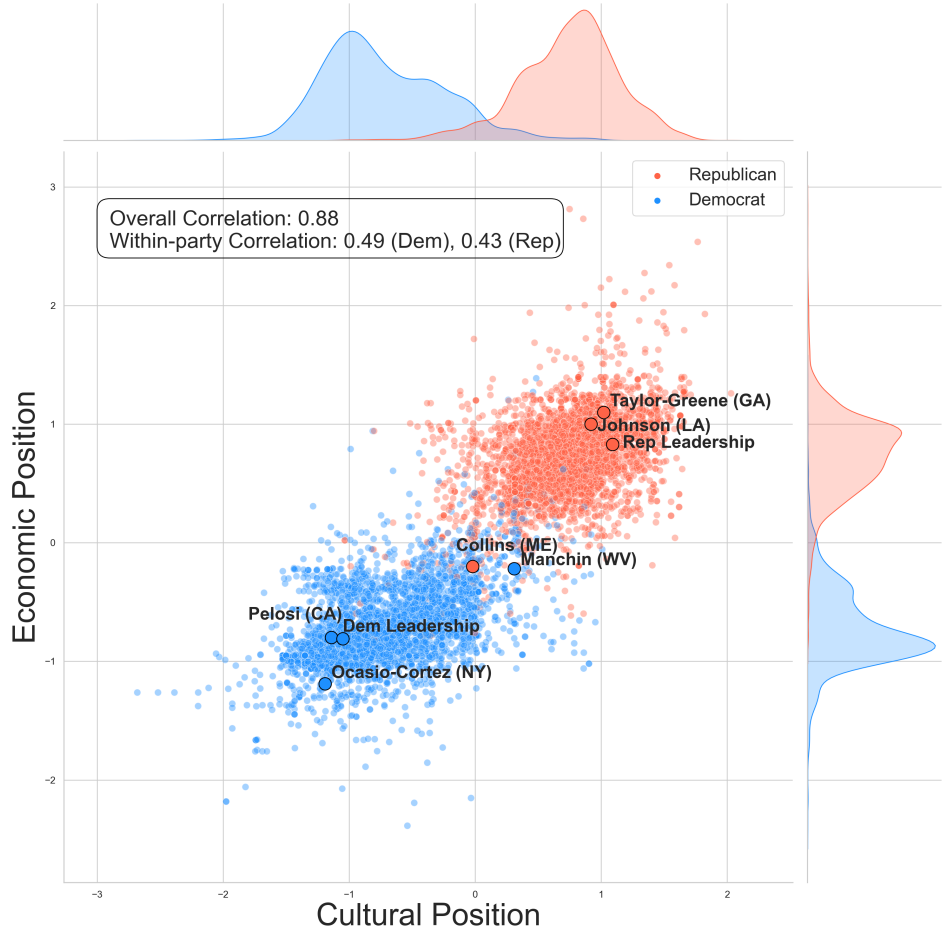


Figure 4: Distribution of candidate ideal points

Notes: Each dot shows one House candidate's estimated two-dimensional ideal point from the multimodal text-and-survey model on cultural and economic issues for each election. Notable candidates' positions as well as the party leadership position in 2020 are added to the graph. The marginal densities are plotted on each axis. The graph excludes third parties and independent candidates. The correlation between the two dimensions is 0.88 in the whole sample and the within-party correlations are 0.49 for the Democratic Party and 0.43 for the Republican Party. Appendix Figure A.9 shows the distribution of the difference in positions between Democratic and Republican candidates in each Congressional district.

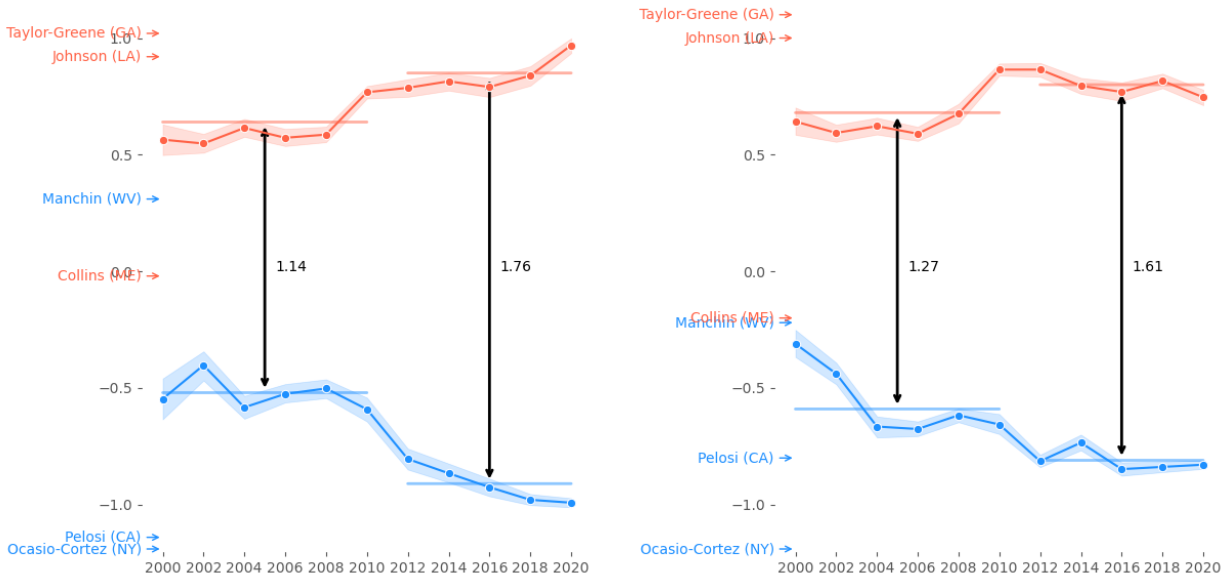
3.4 Evidence of Polarization

Figure 5 shows the evolution of candidate ideal points on both cultural and economic issues over time. The average distance between the two parties has increased in both dimensions, with a much more dramatic divergence on cultural issues compared to economic issues. The average distance between a Democratic and a Republican candidate on cultural issues has doubled between 2000 and 2020, while it has risen by around 50% on economic issues. Figure reports the average differences between the two parties between

from 2000 to 2010 and from 2012 to 2020.

The overall dynamic of polarization align with similar findings from other types of data (rollcall votes, speeches in Congress, campaign contributions, etc.). To the best of my knowledge, this paper is the first one to document a clear contrast in polarization, with cultural issues showing significantly more divergence than economic issues. Appendix Figure A.11 shows the same relationship separately for election-winners and election-losers.

Notably, the rise in political polarization between the two main parties has almost mechanically led to an increase in the overall correlation between candidates' cultural and economic ideal points. However, when looking within party, the correlation between candidate positions across the two dimensions has decreased over time. In 2000-2010, candidates' cultural and economic positions had a 0.46 Spearman correlation coefficient for Democratic candidates and a 0.51 coefficient for Republican candidates. This rank correlation went down to 0.31 for Democratic candidates and 0.26 for Republican candidates in 2012-2020. This decline in the correlation indicates that as the two parties moved further apart, candidates began to differentiate themselves more across the two dimensions.



(a) Cultural Positions

(b) Economic Positions

Figure 5: Larger Polarization on Cultural vs. Economic Issues

Notes: The figure shows the evolution of the average position of candidates in each party on each dimension, with the 95% confidence interval around the mean. Appendix Figure A.11 shows the same relationship separately for election-winners and election-losers. The average distance between the two parties on cultural issues increased from 1.14 in the first period (2000-2010) to 1.76 in the second period (2012-2020). On economic issues, the average distance increased from 1.27 to 1.61. Notable candidates' positions in 2020 on both dimensions are added to the graph.

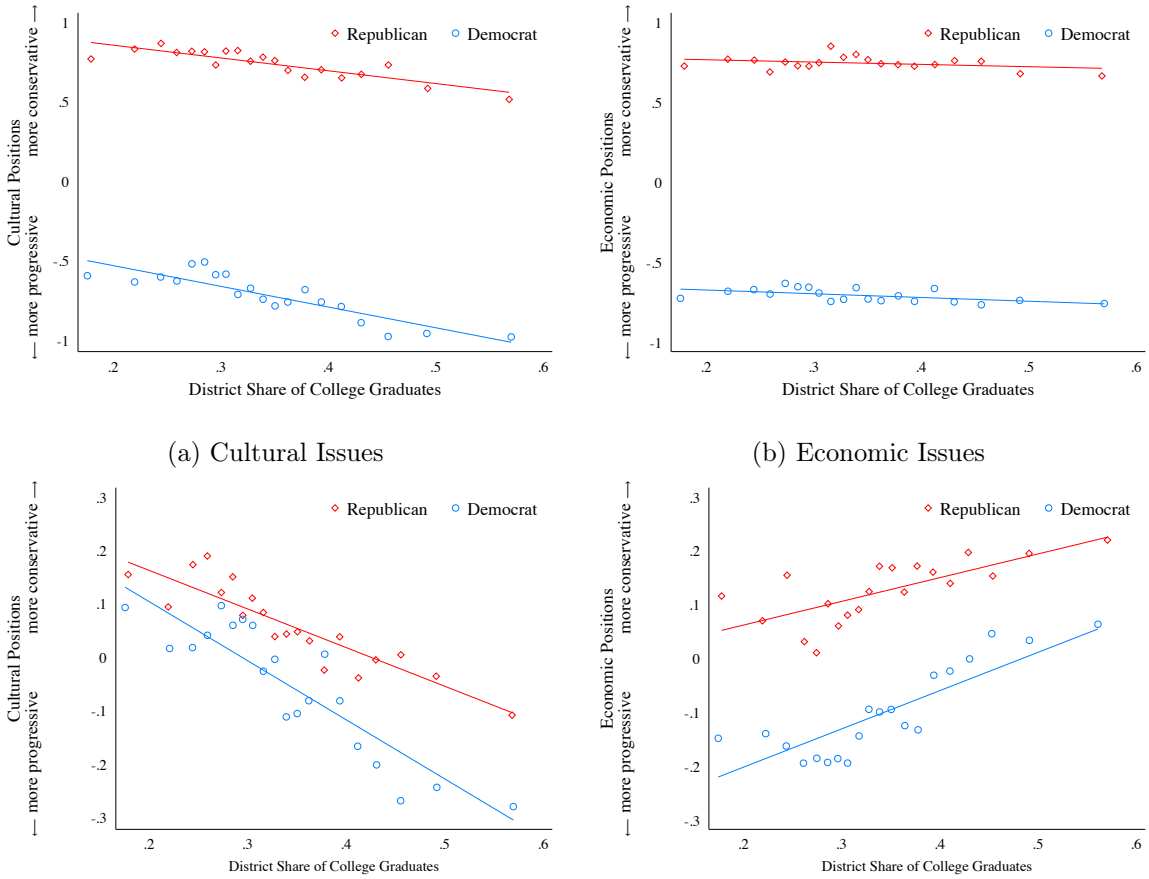
3.5 Determinants of candidate positioning

This section examines the extent to which the demographic composition of congressional districts correlates with candidates' ideal points. Based on classic spatial models, such as Downs (1957), and the Median Voter Theorem (Black, 1958), one would expect candidates to adapt their positions to align with their constituents' ideology, resulting in significant ideological variation across districts.

Figure 6 illustrates these ideological differences by focusing on the educational composition of districts. For each 5% quantile of the education distribution, it shows the average positions of Democratic and Republican candidates. The top two panels, which display unconditional relationships, reveal substantial differences in candidate positions based on district education levels. On cultural issues, Democratic candidates in the most educated districts tend to be one standard deviation more progressive than those in the least educated districts. Republican candidates exhibit a similar, though slightly less pronounced, pattern. By contrast, panel (b) shows that there is almost no difference in candidate positions on economic issues across education levels in these unconditional graphs.

The two bottom panels display the same relationships, but controlling for candidates' position on the other dimension. The negative relationship between candidates' cultural positions and district education holds true for both Democratic and Republican candidates. The economic dimension exhibits the opposite pattern: both Democratic and Republican candidates tend to offer more conservative economic positions in more-educated districts, once accounting for their cultural positions. Appendix Figure A.10 presents similar results for other demographic variables, showing that candidates offer more conservative cultural positions to districts with a larger share of white voters, with more Evangelicals, with fewer unionized workers, and more veterans. Candidates also tend to offer more progressive economic positions to districts with more unionized workers.

All these patterns provide suggestive evidence that candidates might be adjusting their positions to the demand of their constituents. Section 5 precisely estimates the weight that candidates allocate to their constituents in their choice of positions.



(a) Cultural Issues

(b) Economic Issues

(c) Cultural Issues, controlling for Economic

(d) Economic Issues, controlling for Cultural

Figure 6: Candidate positions and congressional district composition

Notes: Each panel shows the relationship between candidate positions and the district share of college graduates. Each dot represents 5% of the education distribution and shows the average position of candidates, separately for Democrats and Republicans, for the districts in this quantile. Panels (a) and (b) present unconditional candidate positions, with cultural issues on the left and economic issues on the right. Panels (c) and (d) show the same relationship but controlling for the position on the other dimension. For example, on panel (a), Democratic candidates running in the most educated districts (55% of college graduates) tend to be 0.5 points more progressive (1 standard deviation) than Democratic candidates running in the least educated districts (17% of college graduates). Figure A.10 shows the same relationship for other demographic variables.

4 Estimation of voter preferences

4.1 Model

The model of vote choice features individual voters with heterogeneous preferences over candidate characteristics. Since voting is compulsory (cf. footnote 6) and there are only two candidates ($j \in \{D, R\}$), voters make a binary choice which I model as resulting from maximizing a random indirect utility with heterogeneous preferences. I give the following functional form to the utility of voter i, who resides in precinct $p(i)$, itself located in district $d(i)$, when choosing the Democratic candidate in election t:

$$u_{it} = \sum_k \beta_{ikt} x_{d(i)kt} + \alpha_{it} + \xi_{p(i)t} + \epsilon_{it}, \tag{6}$$

where $x_{d(i)kt} = x_{Dd(i)kt} - x_{Rd(i)kt}$ is the difference in positions on each dimension k between the two candidates, α_{it} is voter i's utility of voting for the Democratic rather than the Republican candidate, independently of their positions, $\xi_{p(i)t}$ is a precinct-level taste shock in favor of the Democratic candidate, and ϵ_{it} is an individual-level taste shock, which I assume follows a type-I extreme value distribution.

To capture the heterogeneity in voter preferences over ideology and over partisan preferences, I decompose voter preference parameters as:

$$\alpha_{it} = \alpha_t + \mathbf{w}'_{\mathbf{it}} \boldsymbol{\alpha}^{\mathbf{w}}_{\boldsymbol{t}} + \sigma^{\alpha}_t \nu_{0it}, \tag{7}$$

$$\beta_{ikt} = \beta_{kt} + \mathbf{w}'_{\mathbf{it}} \boldsymbol{\beta}^{\mathbf{w}}_{\boldsymbol{kt}} + \sigma^{\beta}_{kt} \nu_{kit}, \tag{8}$$

for each ideological dimension k. $\mathbf{w_{it}}$ is a vector of observed voter characteristics (years of schooling, race, age, and interactions), $\boldsymbol{\alpha}^{\mathbf{w}}_{\boldsymbol{t}}$ and $\boldsymbol{\beta}^{\mathbf{w}}_{\boldsymbol{kt}}$ are vectors of dimension $|\mathbf{w}|$ capturing how voters' partisan and ideological preferences vary with demographics, $\boldsymbol{\nu_{it}} \sim P_\nu = \mathcal{N}(0, I_{K+1})$ adds some non-demographic based individual heterogeneity in partisan and ideological preferences, which is captured by the parameters σ^{α}_t and σ^{β}_{kt}. The distribution of demographics in each precinct p and election t is denoted by $F_t(\cdot; p(i))$.

The share of votes received by the Democratic candidate in each precinct p is obtained by integrating over both demographic characteristics and unobserved heterogeneity:

$$s_{Dpt} = \int\int \frac{\exp\left(\sum_k \beta_{ikt} x_{d(i)kt} + \alpha_{it} + \xi_{p(i)t}\right)}{1 + \exp\left(\sum_k \beta_{ikt} x_{d(i)kt} + \alpha_{it} + \xi_{p(i)t}\right)} dP_\nu dF_t(\mathbf{w}; p(i)). \tag{9}$$

I follow Berry et al. (2004) (BLP) to set up a GMM estimator combining aggregate and individual moments.[11][12]

4.2 Identification

I discuss first identification of voter preferences over candidates' endogenous characteristics (ideology) using a subset of electoral precincts located around the congressional district border. Then, I discuss how to recover the remaining parameters of the demand side of the model on the entire sample.

[11] I implement this estimation using the Python package provided by Conlon and Gortmaker (2023).

[12] As an alternative specification, shown in the appendix, I assume voter homogeneity within precincts $(\mathbf{w_{it}} = \mathbf{w_{p(i)t}})$, following Berry (1994). Under this assumption, I take the log-odds ratio of the Democratic vote share to express it as a function of candidate characteristics, enabling parameter estimation by Ordinary Least Squares, conditional on the fixed effects. The two specifications provide similar results.

First Step The first step recovers the parameters β_{ikt} which capture voters' heterogeneous preferences over candidate positions. Since candidate positions are likely correlated with voters' taste shocks, using naive identification strategies would result in biased demand estimates. To address this endogeneity concern, I propose an identification strategy that controls for unobservables affecting voters by exploiting the fine-grained structure of election data and the hierarchical nature of political competition. Candidates make decisions at a more aggregate level, targeting constituencies larger than individual precincts. With each district containing on average 400 precincts, candidates are unlikely to respond to taste shocks at the precinct level and are more likely to base their position choices on district-level trends. This strategy uses spatial discontinuity in candidate positioning, which should not correlate with underlying voter discontinuities. By grouping precincts located on either side of a congressional district border, I create sets of adjacent precincts that plausibly share the same taste shock.

I can then account for, at each election, any unobserved taste shock that is common to both sides of the district border, I denote the average taste for Democratic candidates in group $g(p)$ associated with precinct p at election t by $\xi_{g(p)t}$. Additionally, I control for precinct fixed effects, which handle any unobserved elements specific to each precinct that remain constant over time, denoted by ξ_p.[13] The precinct-level unobserved shock can be expressed as: $\xi_{pt} = \xi_{g(p)t} + \xi_p + \widetilde{\xi_{pt}}$. The identifying assumption is given by:

$$\mathbb{E}[\Delta x_{d(p)kt} \cdot \widetilde{\xi_{pt}}] \quad = \quad 0, \tag{10}$$

where $\Delta x_{d(p)kt}$ represents the double difference in positioning, relative to the precinct average position over the period $\overline{x}_{d(p)k}$: $\Delta x_{d(p)kt} = \left[x_{d(p)kt} - \overline{x}_{d(p)k}\right] - \left[x_{d(p')kt} - \overline{x}_{d(p')k}\right]$.

This assumption follows the reasoning by which differences in candidate positions across neighboring districts are not correlated with temporary, unobserved differences between adjacent precincts. This assumption is supported by the fact that each precinct has only a negligible impact on candidates' vote shares, with an average of 400 precincts per district, and by the absence of discontinuities at district borders in non-election-related factors. Additionally, I only use district borders within the same state, ensuring no systematic differences in legislation or upper-level offices on either side of the district boundary. Congressional districts are obviously not drawn at random and state legislators might be aware of some precinct-level unobserved taste shocks when deciding which precincts to include in each district. For the identifying assumption to hold, *temporary* deviations between precincts on both sides of the border should not lead to changes in candidate positions. Table 1 shows that while in an OLS specification, precinct demographics are strongly correlated with candidate positions, this correlation disappears after conditioning on precinct-group-by-election and precinct fixed effects. The table regresses candidate difference in positions on precinct-level demographics, using only the sample of border precincts. One out

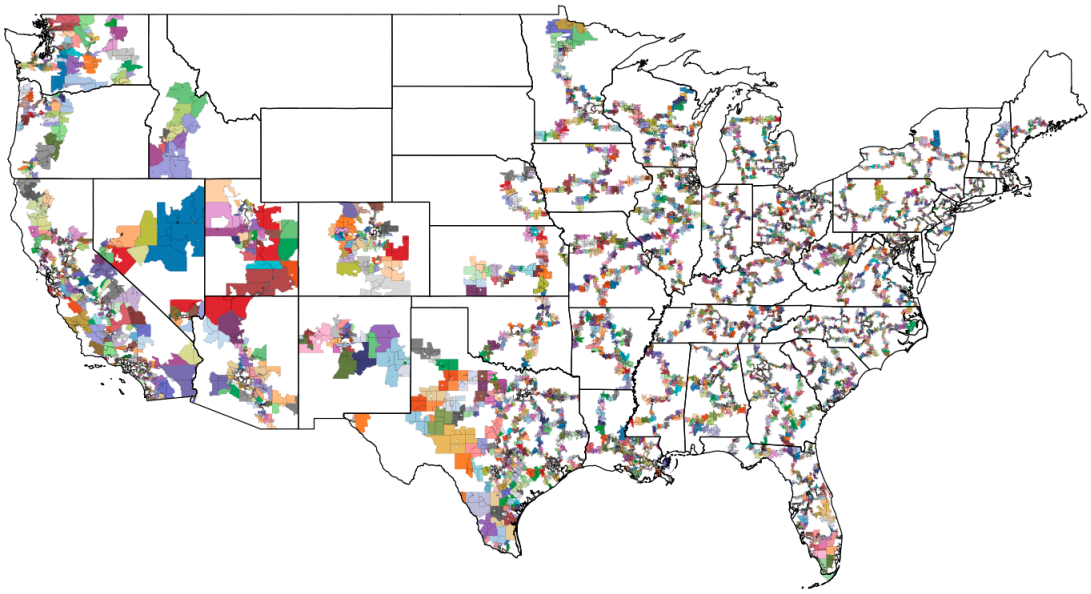
[13] Note that I estimate two separate models, one for 2000-2010 and one for 2012-2020, each with its own set of precinct fixed effects.

of 22 coefficients is significant at the 5% level, as would be expected. No coefficient is significant at the 1% level.

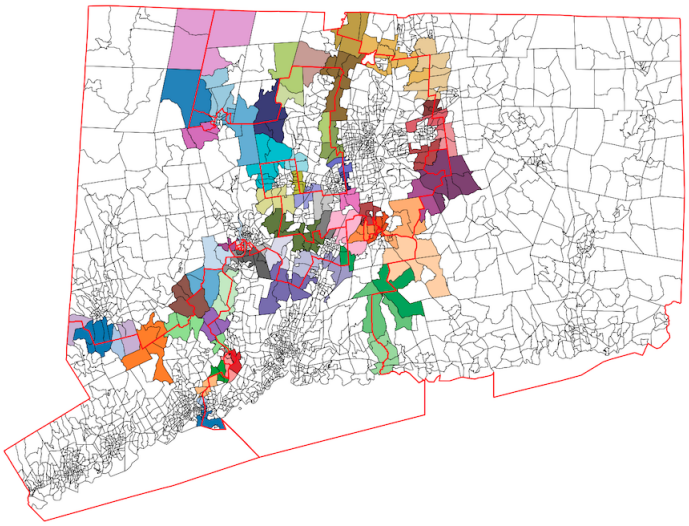
Figure 7 illustrates the distribution of precincts located at district borders for the 2018 elections and depicts the precinct-group-by-election fixed effects. The process of creating precinct-groups is not trivial, as each precinct might be contiguous to several others. Instead of simple pairs, I construct precinct-groups as follows: (1) At each election, for each precinct $p \in d_1$ located on the border between congressional districts d_1 and d_2, I identify the closest precinct $p' \in d_2$, using the distance between their population-weighted centroids. (2) If p is also p''s closest precinct in district d_1, then p and p' form a pair. (3) If, however, p' is closer to another precinct $p'' \in d_1$ than to p, then p, p', and p'' form a group together. (4) I continue extending these chains until they close, meaning until the last added precinct already has its closest precinct included in the group. This approach ensures that the matching process is independent of the order in which it is performed. Figure A.12 in appendix offers a visual representation of the methodology.

The median number of precincts in each of these precinct-groups is 4 and 22% of the country's precincts are adjacent to a congressional district border. This strategy, which leverages spatial discontinuities, has been used in past studies to estimate the impacts of minimum wage (Card and Krueger, 1994; Dube et al., 2010), school valuations (Black, 1999), and advertising effects (Spenkuch and Toniatti, 2018).[14]

[14] Dube et al. (2010) and Spenkuch and Toniatti (2018) create a dataset at the county-pair level where each county has as many observations as it has contiguous neighbors. When the treatment effect is assumed to be homogeneous and standard errors are clustered at the state-pair level (or district-pair level in my case), the multiplication of the number of observations does not affect the estimates and their standard errors but in my case, the treatment effect is highly heterogeneous and precincts are 60 times smaller than counties. It is therefore more reasonable to keep the number of observations constant and create groups of precincts, rather than pairs.



(a) Continental U.S.



(b) Connecticut

Figure 7: Precinct-groups distribution for the 2018 elections (116^{th} Congress).

Notes: The first panel shows all the precincts (block-groups) in the U.S. that sit on a Congressional district border. I use only borders within the same state and therefore exclude at-large districts. The second panel zooms on the distribution of precinct groups in Connecticut. Appendix section C details the construction of the precinct groups. Each color shows a different group of precincts.

The moments used for the identification of the parameters in the first step, for each political dimension $k = \{cultural, economic\}$, for each demographic characteristic w_{pt} (years of schooling, race, age, and interactions), are:

$$
\begin{aligned}
\mathbb{E}[\Delta x_{d(p)kt} \cdot \Delta\xi_{pt}] &= 0 \\
\mathbb{E}[\Delta\big(x_{d(p)kt} \cdot w_{pt}\big) \cdot \Delta\xi_{pt}] &= 0 \\
\mathbb{E}[\Delta w_{pt} \cdot \Delta\xi_{pt}] &= 0 \\
\mathbb{E}[\Delta x^2_{d(p)kt} \cdot \Delta\xi_{pt}] &= 0 \\
\mathbb{E}[\Delta\big(x^2_{d(p)kt} \cdot w_{pt}\big) \cdot \Delta\xi_{pt}] &= 0
\end{aligned}
\tag{11}
$$

Figure A.13 in appendix presents Monte Carlo simulations of the demand specification, demonstrating that the coefficients are properly identified.

Appendix G demonstrates robustness of the results to using an alternative identification strategy using a different source of variation. I estimate the same parameters with congressional-district-by-election fixed effects instead of precinct-group-by-election fixed effects. With precinct-group-by-election fixed effects, I compare arguably similar precincts that were facing a choice between different pairs of candidates. In contrast, with district-by-election fixed effects, I compare different precincts facing the same pair of candidates. This alternative specification gives very consistent estimates of voter preferences.

Second Step By design, the first step removes any identifying variation for characteristics that lack a discontinuity at the district border, such as precinct demographics ($w_{pt} \approx w_{g(p)t}$). This setup leaves voters' party preferences (α_{it}) largely unidentified in the first step.[15]

Given that a candidate's party affiliation is considered exogenous, it is not necessary to limit the analysis to precincts at the district border to identify party preferences. Consequently, in the second step, I utilize the entire sample, not just precincts sitting on the district border. This second step incorporates the estimates from the first step regarding voter preferences over candidate ideology and uses the following moments to recover preferences over partisan affiliations:

$$
\mathbb{E}[w_{pt} \cdot \xi_{pt}] = 0 \tag{12}
$$

for each demographic characteristic w_{pt} (years of schooling, race, age, and interactions). Essentially, this second step recovers the demographic heterogeneity in party preferences, accounting for the ideological preferences estimated in the first step.

4.3 Incorporation of micro-level variation

The precinct-level election results provide aggregate variation in vote shares. As precincts are relatively homogeneous demographically (1,000 registered voters on average), one can hope to recover demographic heterogeneity in vote choice based on aggregate vote shares at the precinct-level. There is, however,

[15] Note that I still estimate voter preferences over party in the first step to account for small differences in voter demographics that might subsist within precinct groups.

	DemCult - RepCult		DemEcon - RepEcon	
	(1)	(2)	(3)	(4)
Av. Edu	-0.023**	-0.006	-0.005	-0.007
	(0.009)	(0.005)	(0.008)	(0.005)
Shr. White	0.084***	0.001	0.017	0.006
	(0.013)	(0.022)	(0.012)	(0.020)
Av. (log) Income	-0.056***	0.001	-0.002	0.002
	(0.010)	(0.005)	(0.008)	(0.005)
Av. Age	-0.065***	0.013**	-0.019***	-0.001
	(0.006)	(0.005)	(0.005)	(0.005)
Pop	-0.012***	-0.002	-0.009***	-0.005
	(0.003)	(0.003)	(0.003)	(0.004)
Shr. Unionized	0.032***	0.003	0.007	-0.006
	(0.009)	(0.007)	(0.007)	(0.006)
Shr. Farmers	0.006	-0.002	-0.002	-0.000
	(0.006)	(0.002)	(0.006)	(0.002)
Shr. Wage Income	-0.005	-0.001	0.001	-0.001
	(0.004)	(0.004)	(0.003)	(0.003)
Shr. Speaking English	-0.027**	0.007	-0.027**	0.002
	(0.013)	(0.009)	(0.011)	(0.008)
Shr. Rural	-0.018***	-0.010	-0.004	0.016
	(0.007)	(0.015)	(0.006)	(0.014)
Shr. Veterans	0.079***	-0.000	0.049***	-0.000
	(0.007)	(0.002)	(0.006)	(0.002)
Precinct-group x Election FE		X		X
Precinct FE		X		X
F-statistic	30.257	0.999	7.188	0.547
Observations	226,509	226,509	226,509	226,509

Table 1: No systematic correlation between candidate positions and precinct demographics, conditional on the fixed effects

Notes: This table shows the coefficients from a regression of candidate positions on precinct-level demographic variables using only precincts located at the congressional district borders, columns (1) and (3) are simple OLS regressions and columns (2) and (4) are conditional on precinct fixed effects and precinct-group-by-election fixed effects. Standard errors clustered both by congressional district by year and by precinct are reported in parentheses. The F-statistic for the joint significance of the coefficients is reported in the table notes.

a longstanding debate in social science about the ability of inferring individual-level preferences from aggregate election results (King et al., 2004). The Monte Carlo simulations in appendix show that, with very granular voting data and rich heterogeneity in voter preferences, one can recover consistent estimates of the true individual-level demand parameters. In addition to aggregate variation, one could also use individual-level variation to help estimate the demographic hetereogeneity parameters. I therefore use large sample (N=400,000) survey data from ANES, CES, Gallup, and GSS from Kuziemko et al. (2023)

containing voter demographics, congressional district and House candidate choice.

As in Petrin (2002), Berry et al. (2004), and Conlon and Gortmaker (2023), I construct new moments $g_M(\theta)$ from this individual-level data, which are functions of the model parameters. Micro-moments include the interaction between demographics of interest (education, white dummy, age, and interactions) and each individual's partisan choice, as well as the ideological position of their chosen candidate. I then stack these moments with aggregate moments from precinct-level data $g_A(\theta)$, giving a vector of moments: $g(\theta) = \begin{pmatrix} g_A(\theta) \\ g_M(\theta) \end{pmatrix}$. The full list of moments is shown in Appendix G.[16]

4.4 Demand Estimates

Table 2 reports the results of equation 9, corresponding estimates without within-precinct heterogeneity are also reported in Appendix Table A.6. The two tables give very similar estimates. The main parameters of interest are β_{cult} and β_{econ}, with particular focus on the heterogeneity of these dimensions across educational lines. For instance, $\beta_{edu,W,cult}$ captures the change in cultural preferences associated with an additional year of schooling for white voters. The coefficient β_{cult} represents the average preferences for cultural issues across all voters. I begin by briefly describing the estimated coefficients and their evolution over time, before comparing the coefficients for the group of less-educated voters to those of more-educated voters. Here and throughout, less-educated voters are defined as those with education below the yearly national median (13.65 years of schooling in 2020), while more-educated voters have education at or above the median.

Overall, in both time periods, the ideological coefficients for white voters are generally larger in magnitude and more precise than those for non-white voters, suggesting that ideology plays a relatively greater role in shaping the choices of white voters than those of non-white voters.

The impact of education on white voters' preferences for cultural and economic policies diverges: higher levels of education are associated with more progressive preferences on cultural issues but more conservative preferences on economic issues. The coefficients are larger for cultural issues, indicating a stronger educational gradient in that dimension. Over time, the coefficients on both dimensions increase, with a more pronounced rise in the economic dimension. Less-educated non-white voters also show a similar shift toward more progressive economic positions.

Appendix Figure A.14 illustrates how Democratic candidate vote shares respond to unilateral shifts in Democratic candidates' positions, separately for more- and less-educated voters. Positive slopes indicate a preference for more conservative policies. The figure shows that less-educated voters increasingly favor progressive economic positions, as reflected in the steeper negative slopes in 2012-2020 compared to 2000-2010, with the opposite trend for more-educated voters. By contrast, the change in slope on cultural issues is less pronounced, indicating a smaller shift in preferences along this dimension over time.

[16] Note that the GMM weighting matrix is block-diagonal as the covariance between aggregate and micro-moments is assumed to be zero.

The parameter α_{it} captures factors that influence voters' utility of voting for a Democratic candidate and do not vary across districts. It reflects preferences for characteristics such as pure partisanship and the ideology of higher-level candidates (e.g., Senate and Presidential), who may exert a *coattail effect* on House candidate choices (Calvert and Ferejohn, 1983). I treat α_{it} as a confounder for ideology that requires control, without assigning a structural interpretation to it. Over time, α_{it} has shifted significantly, with educated voters becoming more likely, and white voters less likely, to vote for Democratic candidates.

The demand-side ideological estimates can be visualized using voters' indifference curves. For each voter, there exists a continuum of candidate positions that leave them indifferent between the Democratic and Republican candidates. These positions form their indifference curve, represented by:

$$x_{\text{cult}} = -\frac{\beta_{i,\text{econ}}}{\beta_{i,\text{cult}}} x_{\text{econ}}, \tag{13}$$

where x_{cult} (resp. x_{econ}) is the difference in positions between the Democratic and the Republican candidate on cultural (resp. economic) issues. The set of positions for which voter i votes for the Democratic candidate depends on the sign of $\beta_{i,\text{cult}}$. If $\beta_{i,\text{cult}} > 0$, for any positions such that $x_{\text{cult}} \geq -\frac{\beta_{i,\text{econ}}}{\beta_{i,\text{cult}}} x_{\text{econ}}$, i will vote for the Democratic candidate, and the opposite if $\beta_{i,\text{cult}} < 0$. Figure 8 shows the average indifference curve of less-educated and more-educated voters, separately for 2000 to 2010 and 2012 to 2020. The graph also includes the difference in the positions of party leadership, represented by $N^D_{2000} - N^R_{2000}$ in 2000. It shows that in 2000, the Democratic leadership was 1.1 ideological units more progressive than the Republican leadership on both the cultural and economic dimensions. By 2020, this gap had increased, with a particularly pronounced shift on cultural issues to 2.2 and a smaller increase on economic issues to 1.6.

Since non-ideological effects are normalized, I represent indifference curves for voters who are indifferent between the initial positions of the two parties, with their indifference curve intersecting at the leadership positions in 2000, allowing for a clearer comparison of how shifts in leadership positions impact voter preferences over time. Voters who are initially indifferent between the two parties can be thought of as swing voters.

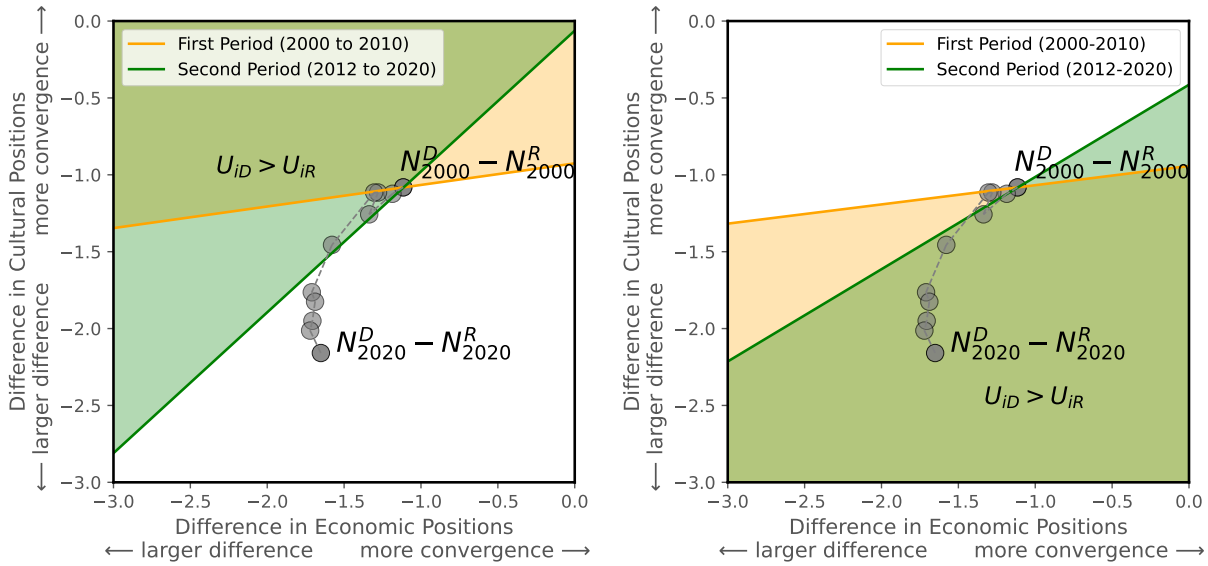
All indifference curves are upward slopping, indicating that $\beta_{i,\text{cult}}$ and $\beta_{i,\text{econ}}$ are of opposite signs. The shaded areas indicate the set of positions where each voter type would choose the Democratic candidate. Less-educated voters tend to vote for the Democratic candidate when candidates are positioned in the upper left corner, where there is low differentiation on cultural issues but a strong differentiation on economic issues, with the Democratic candidate being much more progressive. Conversely, more-educated voters support the Democratic candidate when candidates are located in the lower right corner, where they are similar on economic issues but highly differentiated on cultural issues.

Over time, the slope for both less-educated and more-educated voters has increased, reflecting the growing educational gradient on economic issues. More formally, the marginal rate of substitution between the economic and cultural dimensions needed to keep voters indifferent has decreased. In 2012-2020,

Democratic candidates needed to shift less to the left on economic issues to offset a leftward move on cultural issues than they had to in 2000-2010.

There are two main takeaways from Figure 8. First, the shift in leadership positions has caused initially indifferent less-educated voters to favor the Republican Party over the Democratic Party. Similarly, this shift has led more-educated voters who were once indifferent to now prefer the Democrats. To put it differently, the change in party positions has caused less-educated swing voters to increasingly favor the Republicans, while it has caused more-educated swing voters to increasingly favor the Democrats. The second takeaway is that changes in voter preferences have mitigated the impact of these leadership shifts. For less-educated voters, the distance between the new leadership positions and less-educated voters' initial indifference curve is much larger than the distance to their new indifference curve. As a result, changes in their preferences have brought them closer to the Democrats, preventing an even greater shift toward the Republicans. In contrast, for more-educated voters, their ideological shifts have moved them further away from the Democrats, not closer, mitigating their alignment with the Democratic Party.

The next section examines the extent to which individual candidates can deviate from the leadership's position and instead align more closely with their voters' preferences. This flexibility could mitigate the overall impact of the phenomenon described above, potentially reducing the degree of voter realignment driven by leadership shifts.



(a) Less-educated voters

(b) More-educated voters

Figure 8: Voters' indifference curves

Notes: The figure shows the average indifference curves of less-educated (panel (a)) and more-educated voters (panel (b)), separately for 2000-2010 and 2012-2020. The indifference curve represents the combination of candidate positions that leave voters indifferent between the Democratic and Republican candidate. For each voter i, the indifference curve is given by $x_{\text{cult}} = -\frac{\beta_{i,\text{econ}}}{\beta_{i,\text{cult}}} x_{\text{econ}}$, where x_{cult} (resp. x_{econ}) is the difference in positions between the Democratic and the Republican on cultural (resp. economic) issues. Non-ideological components of the indifference curve are omitted. The shaded areas show the combinations of positions for which voters choose to vote for the Democratic candidate. Less-educated voters choose the Democratic candidate for any pairs of positions located to the Upper-Left of their indifference curve. More-educated voters choose the Democratic candidate for any pairs of positions located to the Lower-Right of the indifference curve. $N^D_t - N^R_t$ represents the difference in positions between the two leaderships in each year t. The indifference curves are evaluated at the initial leadership position $N^D_{2000} - N^R_{2000}$ and therefore consider voters who were initially indifferent between the leaderships of the two parties.

Coefficients	Parameters	Estimates	Standard Errors
First Period: 2000-2010			
CultDem - CultRep	β_{cult}	0.0110	(0.0259)
EconDem - EconRep	β_{econ}	-0.0397	(0.0342)
Yrs. Schooling × Non-white	$\alpha_{edu,NW}$	-0.1576	(0.0210)
Yrs. Schooling × White	$\alpha_{edu,W}$	0.0249	(0.0163)
White	α_W	-2.2885	(0.1196)
Age	α_{age}	0.0255	(0.0025)
(CultDem - CultRep) × Yrs. Schooling × Non-white	$\beta_{edu,NW,\text{cult}}$	-0.0042	(0.0069)
(CultDem - CultRep) × Yrs. Schooling × White	$\beta_{edu,W,\text{cult}}$	-0.0357	(0.0051)
(CultDem - CultRep) × White	$\beta_{W,\text{cult}}$	0.0414	(0.0447)
(CultDem - CultRep) × Age	$\beta_{age,\text{cult}}$	-0.0007	(0.0005)
(EconDem - EconRep) × Yrs. Schooling × Non-white	$\beta_{edu,NW,\text{econ}}$	-0.0268	(0.0072)
(EconDem - EconRep) × Yrs. Schooling × White	$\beta_{edu,W,\text{econ}}$	0.011	(0.0061)
(EconDem - EconRep) × White	$\beta_{W,\text{econ}}$	0.1352	(0.0266)
(EconDem - EconRep) × Age	$\beta_{age,\text{econ}}$	-0.0016	(0.0035)
Unobserved partisan heterogeneity	σ^{α}	0.0562	(0.1437)
Unobserved cultural heterogeneity	$\sigma^{\beta}_{\text{cult}}$	-0.0107	(0.1277)
Unobserved economic heterogeneity	$\sigma^{\beta}_{\text{econ}}$	0.0118	(0.8791)
Second Period: 2012-2020			
CultDem - CultRep	β_{cult}	0.0438	(0.0135)
EconDem - EconRep	β_{econ}	-0.0685	(0.0132)
Yrs. Schooling × Non-white	$\alpha_{edu,NW}$	-0.0502	(0.0140)
Yrs. Schooling × White	$\alpha_{edu,W}$	0.2288	(0.0160)
White	α_W	-3.0364	(0.0779)
Age	α_{age}	0.0245	(0.0024)
(CultDem - CultRep) × Yrs. Schooling × Non-white	$\beta_{edu,NW,\text{cult}}$	0.0112	(0.0149)
(CultDem - CultRep) × Yrs. Schooling × White	$\beta_{edu,W,\text{cult}}$	-0.0444	(0.0039)
(CultDem - CultRep) × White	$\beta_{W,\text{cult}}$	-0.0623	(0.4851)
(CultDem - CultRep) × Age	$\beta_{age,\text{cult}}$	0.0003	(0.0027)
(EconDem - EconRep) × Yrs. Schooling × Non-white	$\beta_{edu,NW,\text{econ}}$	0.0249	(0.0053)
(EconDem - EconRep) × Yrs. Schooling × White	$\beta_{edu,W,\text{econ}}$	0.0314	(0.0041)
(EconDem - EconRep) × White	$\beta_{W,\text{econ}}$	-0.084	(0.0751)
(EconDem - EconRep) × Age	$\beta_{age,\text{econ}}$	0.0023	(0.0046)
Unobserved partisan heterogeneity	σ^{α}	0.0847	(0.1541)
Unobserved cultural heterogeneity	$\sigma^{\beta}_{\text{cult}}$	0.1171	(0.1924)
Unobserved economic heterogeneity	$\sigma^{\beta}_{\text{econ}}$	0.0336	(0.0751)

Table 2: Estimated demand-side parameters with demographic heterogeneity

Notes: The table shows the estimated coefficients from Equation 9, estimated by BLP with precinct and precinct-group-by-election fixed effects. Standard errors are clustered by congressional district by election. Standard errors from the second step (α estimates) are not yet bootstraped and only incorporate the error from the second BLP estimation. Results from a simpler specification (Berry, 1994) are reported in Appendix Table A.6. Similar results, using an alternative set of fixed effects: congressional-district-by-election fixed effects and precinct fixed effects, are reported in Appendix G. Appendix Table A.4 provides descriptive statistics on voters' demographics.

5 Estimating candidates' choice of ideology

5.1 Model

After estimating the demand-side parameters, which capture the distribution of voters' ideological preferences across congressional districts, I then model candidates' positioning decisions. Specifically, I estimate a standard candidate competition model (e.g., Wittman (1983)) in which candidates balance the need to cater to their constituents' preferences with the pressure to adhere to their party's platform. This framework allows me to identify the factors influencing candidates' strategic decisions in offering policy positions that may deviate from or align with the party line.

I define each candidate's objective function as Π_{jt}, where candidates seek to maximize their probability of being elected, given that they derive a positive rent $Q > 0$ from holding office, while simultaneously minimizing their deviation from the national party platform. This framework captures the trade-off candidates face between aligning with their local voters' preferences and adhering to the national party line. I write

$$\Pi_{jt}(\mathbf{x_{jt}}) = \underbrace{P_j(\mathbf{x_{jt}}, \mathbf{x_{-jt}})}_{\text{probability of winning}} Q - \underbrace{\lambda_{jt}||\mathbf{x_{jt}} - \mathbf{N_{jt}}||^2}_{\substack{\text{cost of deviation from} \\ \text{national party platform}}} + \eta_{jt}, \tag{14}$$

where the parameter λ_{jt} captures candidates' cost of deviating from the party line.

There is an aggregate shock ζ_{jt} that introduces uncertainty into candidates' probability of winning (Persson and Tabellini, 2002). Assuming ζ_{jt} follows a uniform distribution:[17] $\zeta_{jt} \sim U(-\frac{1}{\phi}, \frac{1}{\phi})$, I can write

$$\begin{aligned} P_j(\mathbf{x_{jt}}, \mathbf{x_{-jt}}) &= \mathbb{P}(s_j(\mathbf{x_{jt}}, \mathbf{x_{-jt}}) + \zeta_{jt} \geq 0.5) \\ &= 1 - \mathbb{P}(\zeta_{jt} \leq 0.5 - s_j(\mathbf{x_{jt}}, \mathbf{x_{-jt}})) \\ &= \frac{\phi}{2} s_j(\mathbf{x_{jt}}, \mathbf{x_{-jt}}) + \frac{1}{2} - \frac{\phi}{4}. \end{aligned}$$

Taking the first order conditions with respect to each ideal point dimension gives the following equilibrium conditions:

$$(x_{jkt} - N_{jkt}) = \frac{1}{\widetilde{\lambda_{jt}}} \frac{\partial s_{jt}(\mathbf{x_{jt}}, \mathbf{x_{-jt}})}{\partial x_{jkt}} + \eta_{jkt}, \tag{15}$$

on each dimension k, and where $\frac{\partial s_{jt}(\mathbf{x_{jt}}, \mathbf{x_{-jt}})}{\partial x_{jkt}}$ is the derivative of the demand function with respect to candidates' position on dimension k, recovered from the demand estimation and $\widetilde{\lambda_{jt}} = \frac{4\lambda_{jt}}{\phi Q}$. I only recover the strength of party discipline relative to candidates' rent if elected (Q) and the support of the aggregate

[17] While not an uncommon assumption in political economy (see e.g., Acemoglu et al. (2013)), uniform shocks lead to a linear relationship between vote shares and candidates' objective function. Supplemental Appendix H therefore shows that relative estimates of λ under an alternative distributional assumption for the shock, using a Logistic distribution, give very similar results.

shock (ϕ). Intuitively, the larger the support of the shock and the greater the rent, the stronger the actual level of party discipline will be for a given parameter $\widetilde{\lambda}$.

I use two different measures of leadership positions (N_{jt}): (i) the average across party leaders only (Majority/Minority leader, Speakers, Whips, Caucus and Conference Secretaries and Chairs), (ii) a simple average over all candidates.

5.2 Identification

Equation (15) cannot be estimated via OLS due to simultaneity, originating from the demand-side derivative depending on the unobservable η_{jkt}. For instance, a candidate with a personal preference for progressive cultural positions (i.e., $\eta_{\text{cultural}} < 0$) may adopt a more progressive stance (lower x_{cultural}), which in turn affects the demand derivative they face, generating an endogeneity problem.

To resolve the endogeneity issue, I instrument the predicted vote share derivative with congressional district demographics. These instruments are relevant since district demographics directly influence voter preferences, which in turn affect the slope of the demand function. They are exogenous under the assumption that candidates' idiosyncratic ideological shocks are uncorrelated with district demographic characteristics. The core identifying assumption is that district demographics influence candidates' policy positions only through their effect on voter demand.

The parameters are estimated by GMM imposing the following moment conditions:

$$\mathbb{E}[w_{jt} \cdot \eta_{jkt}] = 0 \tag{16}$$

where w_{jt} are the district-level demographics and η_{jkt} the candidate-specific shock on each topic k.

Figure A.15 in appendix presents Monte Carlo simulations demonstrating the proper identification of the supply-side parameters. Appendix Figure A.16 illustrates the intuition. I recover the strength of party discipline from the slope of the function that links candidate positions to the demand derivative in their congressional district. A flatter curve indicates a higher degree of party discipline, as it reflects situations where candidates have limited flexibility to adjust their positions in response to local conditions. Conversely, a steeper slope suggests candidates have more leeway to deviate from the party line based on district-specific preferences.

5.3 Supply Estimates

Table 3 presents the estimated parameters for each period and party, where a higher λ reflects a greater electoral cost for candidates deviating from the party line. I report results from three different specifications: the first uses a simple average to define the party leadership position, the second employs a weighted average of party leaders, and the third also uses a weighted average but excludes party leaders from the estimation. All three specifications produce similar estimates.

The parameter λ from 2000 to 2010 is nearly three times larger for the Republican Party than for the Democratic Party, indicating stronger party discipline among Republican candidates. Additionally, I find that party discipline has increased over time, with λ in 2012 to 2020 being almost three times larger than in the earlier period for both parties. This trend is consistent with previous findings, such as those documented by Canen et al. (2020) and Canen et al. (2021) in Congress, and through campaign contributions from party PACs (Cox and Shapiro, 2024). Notably, Cox and Shapiro (2024) also report that the Republican leadership imposes a higher penalty on policy deviations than the Democratic leadership.

For a specific candidate in a district, increased party discipline translates into electoral losses if they fail to sufficiently align their positions with local voter preferences. In practice, this trend also leads to a greater uniformity of candidates within each party across the country, as stronger party discipline pressures candidates to adhere more closely to the national platform, regardless of district-specific conditions.

	λ		
Dem 2000-2010	0.0267	0.0267	0.0147
	[0.0231, 0.0316]	[0.0231, 0.0316]	[0.0127, 0.0167]
Dem 2012-2020	0.1068	0.1075	0.0660
	[0.0845, 0.1450]	[0.0849, 0.1464]	[0.0412, 0.0909]
Rep 2000-2010	0.0900	0.0898	0.0507
	[0.0631, 0.1565]	[0.0630, 0.1564]	[0.0312, 0.0702]
Rep 2012-2020	0.3092	0.3024	0.1267
	[0.1569, 0.7534]	[0.1548, 0.4452]	[0.0105, 0.2423]
Leadership Measure	Simple Average	Party leaders only	Party leaders only
Sample	All	All	Excluding Party Leaders

Table 3: Estimated supply parameters

Notes: The table shows the supply coefficients from equation 16 estimated using an optimal GMM. The first and second column use the entire sample of House Representatives while the last column excludes party leaders. The first column uses a simple average across all candidates as a measure of party leadership position while the second and third column uses an average across party leaders. Standard errors are not yet bootstrapped and do not incorporate the error from the demand estimation.

It is important to note that the interpretation of λ as a measure of party discipline is broad and not directly micro-founded. I do not observe the specific constraints that the party leadership imposes on candidates to align their positions with the national platform. Instead, I interpret any systematic deviation of a candidate's position toward the national party line as a reflection of party discipline. This could include various mechanisms, such as increased pressure through career concerns, changes in primary voter preferences or structure, or shifts in the selection process favoring candidates more ideologically aligned with party leadership. Overall, the objective of this supply model is to capture the trade-off between local and national influences, where party discipline represents the strength of national, or federal, constraints relative to local preferences.

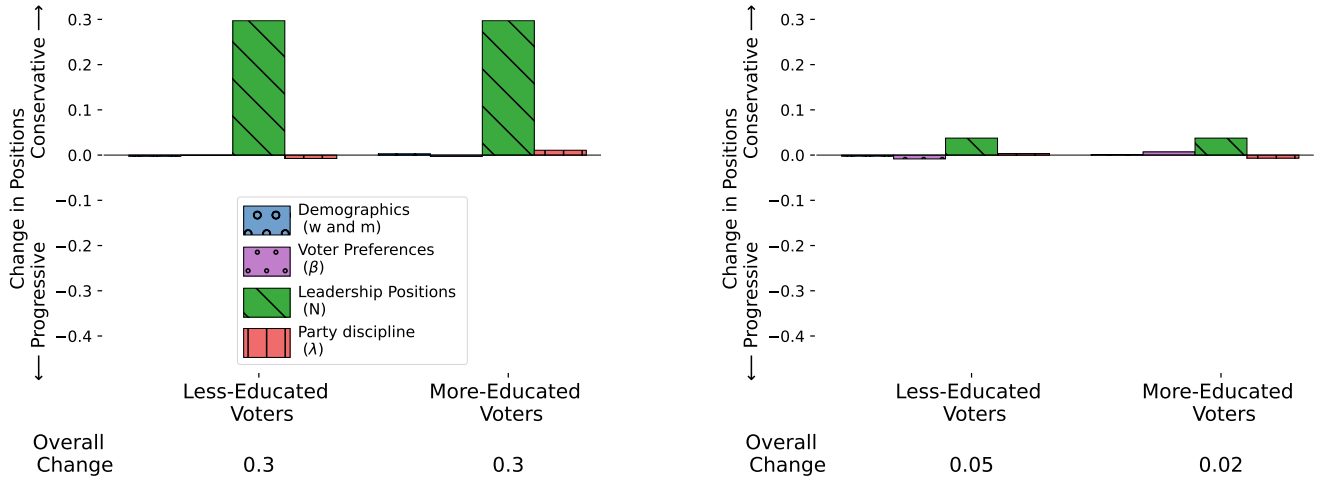
6 Political Realignment: Demand vs. Supply Factors

The previous two sections have estimated the factors influencing voters' choices based on candidate positions and candidates' positioning decisions in response to district composition and party dynamics. With these estimates in hand, I can simulate counterfactual equilibrium outcomes to explore the contribution that each demand-side and supply-side factor has had to the observed changes in candidate positions and in voting. Each counterfactual scenario is characterized by a set of parameters that give a unique equilibrium defined by $(\mathbf{x}^*, \mathbf{s}^*) = f\left(\mathbf{m_t}, F_t\left(\mathbf{w}; p\right), \boldsymbol{\beta_t}, \mathbf{N_t}, \boldsymbol{\lambda_t} \middle| \boldsymbol{\alpha_t}, \boldsymbol{\xi_t}, \boldsymbol{\eta_t}\right)$ where $\mathbf{x}^*, \mathbf{s}^*$ are vectors of candidate positions and vote probabilities, $\mathbf{m_t} = (m_{p,d})_{\forall p,d}$ is a redistricting scenario, that is a map of precincts to districts, $F_t(\mathbf{w}; p)$ is the distribution of demographics in each precinct p, $\boldsymbol{\beta_t}$ captures voters' ideological preferences, $\mathbf{N_t}$ denotes national party leaders' positions, $\boldsymbol{\lambda_t}$ measures the strength of party discipline, $\boldsymbol{\alpha_t}$ accounts for voters' partisan preferences, $\boldsymbol{\xi_t}$ is a vector of precinct-level taste shocks, and $\boldsymbol{\eta_t}$ is a vector of candidate-level ideological shocks.

Under each scenario, I determine new equilibrium candidate positions as the solution to a fixed point algorithm, where each candidate plays their best response to the other's strategy. The fixed point is solved using an imitation game algorithm (McLennan and Tourky, 2005; Batista et al., 2024). These counterfactual candidate positions are then used as inputs in the demand system to generate counterfactual electoral outcomes. I assess the impact of each factor by comparing the realized outcomes under both its initial and final values. Since the effect of each factor depends on the values of the others, I calculate the average contribution for each factor across all possible permutations of the remaining factors, similar to a Shapley value decomposition (Shapley et al., 1953), following the approach of Guriev et al. (2023). Appendix I details the computation of each factor's contribution.

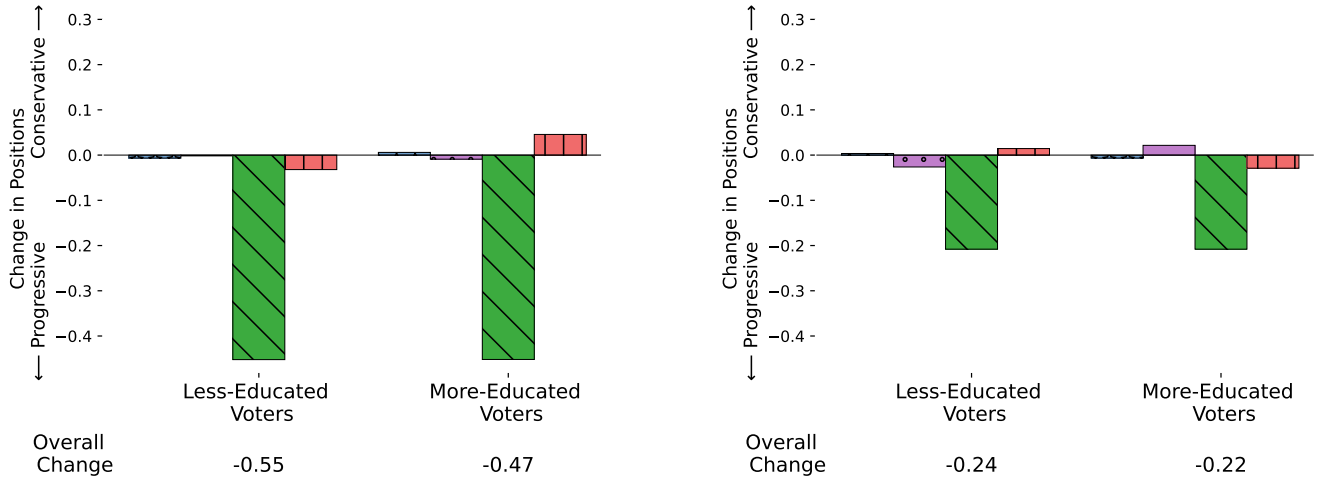
Changes in candidate positions Figure 9 illustrates the contributions of each factor to changes in candidate positions, separately for less-educated and more-educated voters. It compares the 2020 election with the average of the baseline period (2000 to 2010). For both parties and both topics, most of the change is driven by shifts in national party leaders' positions. The rise in party discipline either amplifies or mitigates these shifts, depending on voters' education. For example, increased party discipline has made Democratic candidates more progressive on cultural issues for less-educated voters, while moderating this progressiveness for more-educated voters. Party discipline has a stronger moderating effect on Democratic candidates' cultural positions in highly educated precincts than in less-educated ones. This effect arises because highly educated precincts are often concentrated in districts with similar demographics, as illustrated in Appendix Figure A.17. This concentration is largely due to urban areas forming districts that are homogeneously culturally progressive, consistent with the findings of Rodden (2019). The rise in party discipline plays an opposite role when it comes to economic positions. It has prevented Democratic candidates from adopting more progressive stances in less-educated precincts and from becoming more conservative in more-educated precincts. Looking at demand-side factors, shifts in voter preferences

explain only a small part of the variation in candidate positions. These shifts tend to push candidates toward more conservative economic positions in more-educated precincts and more progressive ones in less-educated precincts. Redistricting and changes in voter demographics have a negligible impact.



(a) Republican Cultural Positions

(b) Republican Economic Positions



(c) Democratic Cultural Positions

(d) Democratic Economic Positions

Figure 9: Equilibrium Contributions of each Factor Changes in Candidate Positions, by Education

Notes: The figure shows the relative contributions of each factor to the change in candidate positions between 2020 and the average of the baseline period (2000 to 2010), separately for less-educated and more-educated voters. Panel (a) shows the change in Republican candidates' cultural positions. Panel (b) shows the change in Republican candidates' economic positions. Panel (c) shows the change in Democratic candidates' cultural positions. Panel (d) shows the change in Democratic candidates' economic positions. Below each bar, the total changes in positions for this type of voters for that period, including unexplained elements, are reported.

Changes in voting behaviors Next, I assess the relative impact of supply-side and demand-side factors on voting behavior, splitting voters by education. Panel (a) of Figure 10 presents the contributions

of demand-side and supply-side factors, separately for less-educated and more-educated voters. Beginning with voter ideological preferences, the change in their preferences has resulted in Democratic candidates gaining 1.5 percentage points among less-educated voters, while losing 0.5 percentage points among more-educated voters. This shift can largely be attributed to the growing preference for progressive economic positions among less-educated voters and the decline in such preferences among more-educated voters. Importantly, these shifts contrast with the broader trend of political realignment, where less-educated voters have moved away from the Democratic Party. In other words, changes in voter preferences have moderated the overall realignment; without them, the shift would have been even more pronounced. Changes in voter demographics and redistricting have only a minor impact. Overall, demographic shifts benefited Democratic candidates everywhere, among both less-educated and more-educated voters. This advantage is largely due to the population becoming more educated and more racially diverse, particularly among less-educated voters.

In contrast, supply-side factors have driven less-educated voters toward the Republican Party and more-educated voters toward the Democratic Party. The shift in leadership positions on cultural issues has caused Democratic candidates to lose 1 percentage point of support among less-educated voters while gaining almost 1.5 percentage points from more-educated voters. This effect has been only partially offset by changes in economic positions.

The underlying dynamic can be summarized as follows: while parties were equally polarized on economic and cultural dimensions in the early 2000s, they have since diverged more sharply on cultural issues. Less-educated voters, who tend to prefer progressive economic policies but conservative cultural ones, now derive less utility from supporting Democratic candidates and are increasingly turning to the Republicans. Conversely, more-educated voters now find greater utility in supporting Democratic candidates as cultural issues have gained relative prominence.

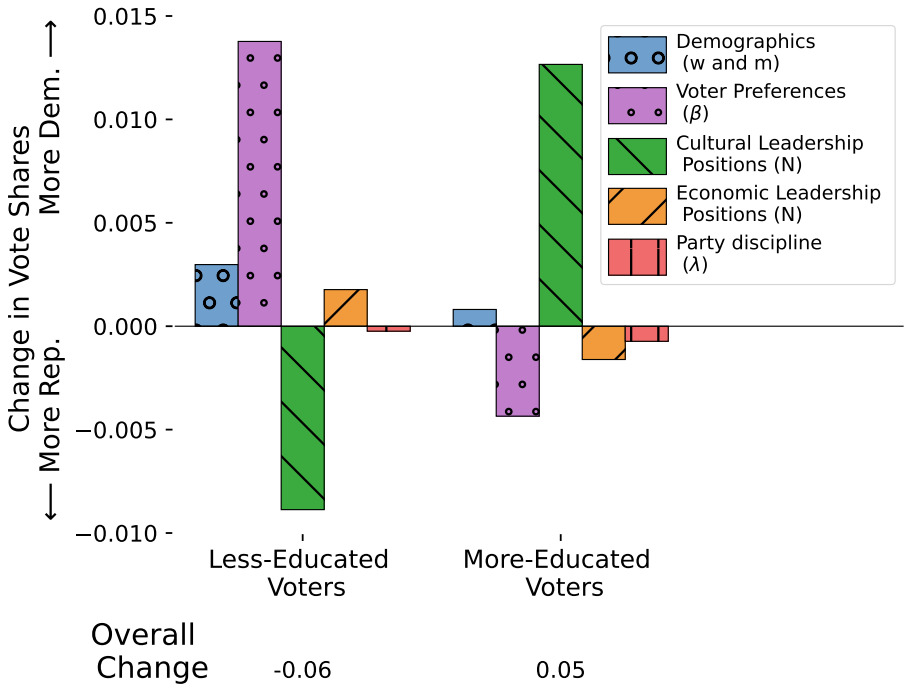


Figure 10: Equilibrium Contributions of each Factor to Changes in Voting Patterns.

Notes: The figure shows the relative contributions of demand-side and supply-side factors to changes in voting behaviors. Panel (a) studies changes in votes, by education. Panel (b) studies the correlation between education and Democratic vote shares between 2020 and the average of the baseline period (2000 to 2010). For each voter, I compare their probability of voting for the House Democratic candidate under each counterfactual scenario. Over this period, overall Democratic vote shares among less-educated voters decreased by 6 percentage points, while they increased by 5 percentage points among more-educated voters.

Lastly, Appendix Figure A.18 examines the contributions of demand-side and supply-side factors to the change in the correlation between education and Democratic voting at the individual level. This correlation increased by 25 points between the 2000-2010 average and 2020. Without the shift in voter preferences, the increase would have been almost 5 points higher. Similarly, without demographic changes, the correlation would have risen by an additional point. In contrast, without the shift in leadership's cultural positions, the correlation would have been about 4 points lower. Changes in economic positions have moderated the impact of cultural positions by driving down the correlation. Party discipline had a net effect close to zero, as it led to losses among both more-educated voters and less-educated voters.

As mentioned earlier, the framework developed in this paper focuses on district-level ideological variation and does not fully explain the variation in voter choices. A significant portion is captured by the parameter α_{it}, which reflects general voter preferences for Democratic candidates, independent of ideology. These preferences are likely influenced by shifts in leadership positions as well as the positions of higher-level candidates (coattail effect (Calvert and Ferejohn, 1983)). Although the impact of presidential, senatorial, and party leaders' positions on House candidate vote shares cannot be causally identified

within this framework, it is reasonable to assume similar voter preferences extend to these upper-level candidates, suggesting that the results presented here may represent a lower bound of the overall effect of leadership's ideological shifts.

It is important to note that the primary driver of realignment is not merely party polarization but the divergence in polarization across different issue dimensions. Specifically, the greater polarization on cultural issues compared to economic ones has been the decisive factor. Had polarization on economic issues increased alongside cultural issues, the observed realignment would have been less pronounced. The next section explores the political implications of candidates polarizing more on cultural than on economic issues particularly for policies that span both economic and cultural dimensions, such as environmental policy.

7 Environmental Issues: Cultural or Economic?

As a final step in my analysis, I examine how parties' cultural polarization affects voter support for environmental policy. My model allows me to assess the impact of shifting the relative polarization on cultural and economic dimensions within a given topic. Here, I focus on environmental policy where—in addition to the prominence in recent political discourses (Egan et al., 2022)—the topic carries both cultural and economic significance (Besley and Persson, 2023). For example, candidates may advocate for environmentally progressive policies that highlight cultural themes which appeal to voters' values, such as climate education, climate justice, ethical consumption or "believe in science" initiatives, or they may emphasize policies with economic implications, such as a "Green New Deal." The same duality applies to conservative positions on these issues. In Appendix J, I detail that such duality is particularly relevant in the context of environmental policy with significant heterogeneity between the two parties. While the Democratic Party's weighting on environmental issues is much more cultural, it is more economic for the Republicans.[18]

Therefore, a question one may ask is whether there would be a broader support for environmental policy if parties, through their national platforms ($\mathbf{N_{Dt}}, \mathbf{N_{Rt}}$ above) were to change their relative weighting on cultural and economic issues. To investigate this, I modify the relative cultural-economic weighting of the Democratic leadership position on the environment, holding their overall level of progressiveness constant. I then re-compute equilibrium positioning of each candidate, and the resulting voting results for each counterfactual scenario. I detail the procedure in Appendix J.

A natural counterfactual scenario is to assess voter support for Democratic candidates had the Demo-

[18] In other words, Democratic candidates who are progressive on environmental issues tend to be progressive on cultural issues rather than economic ones. The reverse is true in the Republican Party, Republican candidates who are progressive on environmental issues tend to be progressive on economic issues rather than cultural ones. Appendix J provides examples of website pages on the environment located near the centroid of each party, illustrating the stronger economic emphasis of Republican candidates.

cratic leadership adopted the same cultural weighting on environmental policy as the Republicans.[19] In this scenario, the Democratic Party would retain its level of progressiveness on environmental issues but shift the emphasis from cultural themes (e.g., "the science behind climate disruptions is virtually undisputable," Michael Mason's website, 2016) to economic ones (e.g., "providing public jobs for energy transition," Rishi Kumar's website, 2020).[20] Consequently, each Democratic candidate would have incentives to select $\mathbf{x_{jt}}$ closer to $\mathbf{N_{Dt}}$, meaning they would be strongly progressive on the economic dimension while more moderate on the cultural dimension. Note that Republican candidates would also adjust their positions in response to this shift in Democratic leadership. Voters would then choose the candidate that maximizes their utility, based on their individual preferences.

Figure 11 presents the corresponding results, comparing vote shares obtained under the current Democratic weighting, which leans more cultural, with those obtained under the Republican weighting, which is more economic, and positioning on the environment that would be fully cultural or fully economic. If the Democratic leadership were to adopt the Republican weighting (stronger economic emphasis), it would increase Democratic candidates' vote share among less-educated voters by 1 point but reduces support from more-educated voters by nearly 1.5 points. Strengthening further the economic dimension to switch to a fully economic environmental policy leads to an even larger shift of support toward less-educated voters, while cultural policies have an opposite effect. Overall, candidates offering environmental policies framed as more economic—such as a "Green New Deal"—tend to attract more support from less-educated voters but less from more-educated ones. Interestingly, environmental positions with a stronger cultural dimension receive broader overall support, as more-educated voters are more sensitive to policy changes, although these policies tend to deter less-educated voters. In a companion paper (Bombardini et al., 2024), we adapt the framework developed in this paper specifically to environmental issues, and examine how the demand and supply respond to changes in environmental conditions and employment opportunities in the environmental sector.

[19] For each counterfactual scenario, I isolate support for environmental policies, disregarding candidate positioning and voter reactions to other issues.

[20] Appendix J provides examples of website pages on the environment that are either heavily cultural or economic.

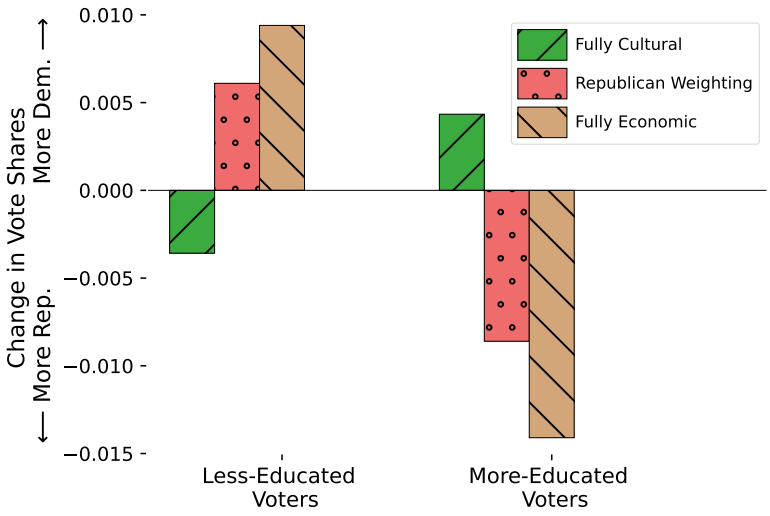


Figure 11: Economically Oriented Progressive Environmental Positions Attract More Support from Less-Educated Voters

Notes: The figure shows the difference in vote shares between the baseline Democratic environmental policy, with a strong cultural weighting, and alternative cultural-economic weighting. For instance, the red bars show the change in vote shares obtained by shifting the emphasis of environmental policies toward more economic policies, adopting the same weighting parameter than the Republican Party. Appendix J provides technical details.

8 Conclusion

Separating supply and demand is a major challenge in the study of economic markets, and the political arena is no exception. It is arguably the "market" where demand and supply are most intertwined. Without clear costs on the supply side, the policies offered by politicians could be seen as purely endogenous to voter preferences. As a result, the observed equilibrium outcomes, such as electoral results, cannot be simply understood as voter choices among political proposals. Moreover, many determinants of voter choices, like candidate positions, are unobserved or difficult to measure. These challenges have hindered scholars from fully estimating a political equilibrium model.

To better understand recent political shifts, such as increased polarization and realignment, it is crucial to disentangle the contributing demand-side and supply-side factors. Evidence shows that parties have moved further apart (Hare and Poole, 2014; Gentzkow et al., 2019), while voters have shifted their allegiances from those of twenty years ago (Kitschelt and Rehm, 2019; Gethin et al., 2022). This paper connects these two facts and tries to remedy the measurement and endogeneity issues aforementioned by building and estimating a multidimensional political equilibrium model of joint candidates' choice of positions and voters' choice of candidates.

I show that less-educated voters prefer more conservative cultural policies but more progressive economic policies, and increasingly so over time. In parallel, political parties have increasingly polarized on cultural issues rather than on economic ones, and rising party discipline has constrained local candidates from adapting to their local conditions. By combining these phenomena, I demonstrate through counterfactual scenarios that most of the shifts in candidate positions and changes in voter choices can be attributed to supply-side factors. In contrast, concurrent changes in voter preferences among less-educated voters have favored the Democratic Party, suggesting that political realignment would have been even more pronounced without these changes in preferences.

Finally, I employ the empirical framework developed in this paper to assess counterfactual levels of support for environmental policies. I show that more economically-oriented environmental policies would draw more support from less-educated voters than the equivalent culturally-oriented policies. However, Democratic candidates' positions on the environment in the last two decades have been much more cultural than economic, contradictory to Republicans, which has hindered less-educated voters' support for such policies.

References

Acemoglu, D., J. A. Robinson, and R. J. Santos (2013). The monopoly of violence: Evidence from colombia. *Journal of the European Economic Association 11*(suppl_1), 5–44.

Adams, J., M. Clark, L. Ezrow, and G. Glasgow (2004). Understanding change and stability in party ideologies: do parties respond to public opinion or to past election results? *British Journal of Political Science*, 589–610.

Ansolabehere, S., P. Ban, and M. Morse (2018). Precinct-Level Election Data, 2014.

Ansolabehere, S., M. Palmer, and A. Lee (2014). Precinct-Level Election Data, 2002-2012.

Ansolabehere, S., J. M. J. Snyder, and C. I. Stewart (2001). Candidate positioning in us house elections. *American Journal of Political Science*, 136–159.

Baltz, S., A. Agadjanian, D. Chin, J. Curiel, K. DeLuca, J. Dunham, J. Miranda, C. H. Phillips, A. Uhlman, C. Wimpy, et al. (2022). American election results at the precinct level. *Scientific Data 9*(1), 651.

Bateman, D. A. and J. Lapinski (2016). Ideal points and american political development: Beyond dw-nominate. *Studies in American Political Development 30*(2), 147–171.

Batista, Q., C. Coleman, Y. Furusawa, S. Hu, S. Lunagariya, S. Lyon, M. McKay, D. Oyama, T. J. Sargent, Z. Shi, et al. (2024). Quantecon. py: A community based python library for quantitative economics. *Journal of Open Source Software 9*(93), 5585.

Berry, S., C. Cox, and P. Haile (2024). Voting in two-party elections: An exploration using multi-level data. *Working Paper*.

Berry, S., J. Levinsohn, and A. Pakes (2004). Differentiated products demand systems from a combination of micro and macro data: The new car market. *Journal of Political Economy 112*(1), 68–105.

Berry, S. T. (1994). Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics*, 242–262.

Berry, S. T. and P. A. Haile (2021). Foundations of demand estimation. *4*(1), 1–62.

Berry, S. T. and P. A. Haile (2024). Nonparametric identification of differentiated products demand using micro data. *Econometrica 92*(4), 1135–1162.

Besley, T. and T. Persson (2023). The political economics of green transitions. *The Quarterly Journal of Economics 138*(3), 1863–1906.

Black, D. (1958). The theory of committees and elections.

Black, S. E. (1999). Do better schools matter? parental valuation of elementary education. *The quarterly journal of economics 114*(2), 577–599.

Bombardini, M., F. Finan, N. Longuet-Marx, S. Naidu, and F. Trebbi (2024). Climate politics in the united states. *Working Paper*.

Bombardini, M., B. Li, and F. Trebbi (2023). Did us politicians expect the china shock? *American Economic Review 113*(1), 174–209.

Bonica, A. (2013). Ideology and interests in the political marketplace. *American Journal of Political Science 57*(2), 294–311.

Bonica, A. (2014). Mapping the ideological marketplace. *American Journal of Political Science 58*(2), 367–386.

Bonica, A., N. McCarty, K. T. Poole, and H. Rosenthal (2013). Why hasn't democracy slowed rising inequality? *Journal of Economic Perspectives 27*(3), 103–124.

Bouton, L., J. Cagé, E. Dewitte, and V. Pons (2022). Small campaign donors. Technical report, National Bureau of Economic Research.

Calvert, R. L. and J. A. Ferejohn (1983). Coattail voting in recent presidential elections. *American Political Science Review 77*(2), 407–419.

Canen, N., C. Kendall, and F. Trebbi (2020). Unbundling polarization. *Econometrica 88*(3), 1197–1233.

Canen, N. J., C. Kendall, and F. Trebbi (2021). Political parties as drivers of us polarization: 1927-2018. Technical report, National Bureau of Economic Research.

Caplin, A. and B. Nalebuff (1991). Aggregation and imperfect competition: On the existence of equilibrium. *Econometrica*, 25–59.

Card, D. and A. B. Krueger (1994). Minimum wages and employment: A case study of the fast food industry in new jersey and pennsylvania. *American Economic Review 84*(4), 772–793.

Chen, T., T. He, M. Benesty, V. Khotilovich, Y. Tang, H. Cho, K. Chen, R. Mitchell, I. Cano, T. Zhou, et al. (2015). Xgboost: extreme gradient boosting. *R package version 0.4-2 1*(4), 1–4.

Choi, J., I. Kuziemko, E. Washington, and G. Wright (2024). Local economic and political effects of trade deals: Evidence from nafta. *American Economic Review 114*(6), 1540–1575.

Clinton, J., S. Jackman, and D. Rivers (2004). The statistical analysis of roll call data. *American Political Science Review 98*(2), 355–370.

Coate, S. and M. Conlin (2004). A group rule—utilitarian approach to voter turnout: Theory and evidence. *American Economic Review 94*(5), 1476–1504.

Conlon, C. and J. Gortmaker (2023). Incorporating micro data into differentiated products demand estimation with pyblp. Technical report, National Bureau of Economic Research.

Cox, C. (2023). The equilibrium effects of campaign finance deregulation on us elections. *Available at SSRN 3794817*.

Cox, C. and I. Shapiro (2024). Party discipline in elections and latent policy ideals. *Available at SSRN 4098078*.

Dai, A. M., C. Olah, and Q. V. Le (2015). Document embedding with paragraph vectors. *arXiv preprint arXiv:1507.07998*.

Danieli, O., N. Gidron, S. Kikuchi, and R. Levy (2022). Decomposing the rise of the populist radical right. *Available at SSRN 4255937*.

Dechezleprêtre, A., A. Fabre, T. Kruse, B. Planterose, A. S. Chico, and S. Stantcheva (2022). Fighting climate change: International attitudes toward climate policies. Technical report, National Bureau of Economic Research.

Di Tella, R., R. Kotti, C. Le Pennec, and V. Pons (2023). Keep your enemies closer: Strategic platform adjustments during us and french elections. Technical report, National Bureau of Economic Research.

Downs, A. (1957). An economic theory of political action in a democracy. *Journal of Political Economy 65*(2), 135–150.

Drews, S. and J. C. Van den Bergh (2016). What explains public support for climate policies? a review of empirical and experimental studies. *Climate policy 16*(7), 855–876.

Dube, A., T. W. Lester, and M. Reich (2010). Minimum wage effects across state borders: Estimates using contiguous counties. *The Review of Economics and Statistics 92*(4), 945–964.

Dunlap, R. E., A. M. McCright, and J. H. Yarosh (2016). The political divide on climate change: Partisan polarization widens in the us. *Environment: Science and Policy for Sustainable Development 58*(5), 4–23.

Egan, P. J., D. M. Konisky, and M. Mullin (2022). Ascendant public opinion: the rising influence of climate change on americans' attitudes about the environment. *Public Opinion Quarterly 86*(1), 134–148.

Egan, P. J. and M. Mullin (2017). Climate change: Us public opinion. *Annual Review of Political Science 20*(1), 209–227.

Elff, M. (2009). Social divisions, party positions, and electoral behaviour. *Electoral Studies 28*(2), 297–308.

Enke, B. (2020). Moral values and voting. *Journal of Political Economy 128*(10), 3679–3729.

Enke, B., M. Polborn, and A. A. Wu (2021). Morals as luxury goods and political polarization.

Evans, G. and J. Tilley (2012). How parties shape class politics: Explaining the decline of the class basis of party support. *British Journal of Political Science 42*(1), 137–161.

Fredriksson, P. G., L. Wang, and K. A. Mamun (2011). Are politicians office or policy motivated? the case of us governors' environmental policies. *Journal of Environmental Economics and Management 62*(2), 241–253.

Gagliarducci, S., M. D. Paserman, and E. Patacchini (2019). Hurricanes, climate change policies and electoral accountability. Technical report, National Bureau of Economic Research.

Gazmararian, A. F. and H. V. Milner (2021). Political cleavages and changing exposure to global warming. *Comparative Political Studies*, 00104140241283014.

Gelman, A. (2009). *Red state, blue state, rich state, poor state*. Princeton University Press.

Gentzkow, M., J. M. Shapiro, and M. Taddy (2019). Measuring group differences in high-dimensional choices: method and application to congressional speech. *Econometrica 87*(4), 1307–1340.

Gethin, A., C. Martínez-Toledano, and T. Piketty (2022). Brahmin left versus merchant right: Changing political cleavages in 21 western democracies, 1948–2020. *The Quarterly Journal of Economics 137*(1), 1–48.

Gordon, B. R. and W. R. Hartmann (2013). Advertising effects in presidential elections. *Marketing Science 32*(1), 19–35.

Guriev, S., E. Henry, T. Marquis, and E. Zhuravskaya (2023). Curtailing false news, amplifying truth. *Amplifying Truth (October 29, 2023)*.

Hacker, J. S. and P. Pierson (2020). *Let them eat tweets: How the right rules in an age of extreme inequality*. Liveright Publishing.

Hall, A. B. and J. M. J. Snyder (2015). Candidate ideology and electoral success. *Manuscript, Stanford University*.

Hare, C. and K. T. Poole (2014). The polarization of contemporary american politics. *Polity 46*(3), 411–429.

Hotelling, H. (1929). Stability in competition. *Economic Joumal 39*, 41–57.

Iaryczower, M., G. Lopez-Moctezuma, and A. Meirowitz (2024). Career concerns and the dynamics of electoral accountability. *American Journal of Political Science 68*(2), 696–713.

Iaryczower, M., S. Montero, and G. Kim (2022). Representation failure. Technical report, National Bureau of Economic Research.

Inglehart, R. (1997). *Modernization and postmodernization in 43 societies*. Princeton University Press.

Jessee, S. A. (2009). Spatial voting in the 2004 presidential election. *American Political Science Review 103*(1), 59–81.

Kaplan, E., J. L. Spenkuch, and H. Yuan (2024). Pandering in the shadows: How natural disasters affect special interest politics.

Kawai, K. and T. Sunada (2022). Estimating candidate valence. Technical report, National Bureau of Economic Research.

King, G. (2013). *A solution to the ecological inference problem*. Princeton University Press.

King, G., M. A. Tanner, and O. Rosen (2004). *Ecological inference: New methodological strategies*. Cambridge University Press.

Kitschelt, H. P. and P. Rehm (2019). Secular partisan realignment in the united states: The socioeconomic reconfiguration of white partisan support since the new deal era. *Politics & Society 47*(3), 425–479.

Krasa, S. and M. Polborn (2014). Policy divergence and voter polarization in a structural model of elections. *The Journal of Law and Economics 57*(1), 31–76.

Kuziemko, I., N. Longuet-Marx, and S. Naidu (2023). "compensate the losers?" economic policy and partisan realignment in the us.

Lee, D. S., E. Moretti, and M. J. Butler (2004). Do voters affect or elect policies? evidence from the us house. *The Quarterly Journal of Economics 119*(3), 807–859.

Manson, S., J. Schroeder, D. Van Riper, T. Kugler, and S. Ruggles (2021). IPUMS National Historical Geographic Information System: Version 16.0 [dataset].

Martin, A. D. and K. M. Quinn (2002). Dynamic ideal point estimation via markov chain monte carlo for the us supreme court, 1953–1999. *Political analysis 10*(2), 134–153.

McCarty, N., K. T. Poole, and H. Rosenthal (2016). *Polarized America: The dance of ideology and unequal riches*. mit Press.

McLennan, A. and R. Tourky (2005). From imitation games to kakutani. *Manuscript, available at http://www. econ. umn. edu/mclennan/Papers/papers. html 37*, 41.

Meisels, M. (2023). Positioning in congressional primary campaigns. *Unpublished manuscript. URL: https://www. mellissameisels. com/files/MMPCPC. pd f*.

Mian, A., A. Sufi, and F. Trebbi (2010). The political economy of the us mortgage default crisis. *American Economic Review 100*(5), 1967–1998.

Noel, H. (2014). Separating ideology from party in roll call data. *Unpublished Manuscript, Department of Government, Georgetown University*.

Noel, H. (2016). Ideological factions in the republican and democratic parties. *The ANNALS of the American Academy of Political and Social Science 667*(1), 166–188.

Ozdaglar, A. (2013). *Strategic Form Games and Nash Equilibrium*. Springer.

Persson, T. and G. Tabellini (2002). *Political economics: explaining economic policy*. MIT press.

Petrin, A. (2002). Quantifying the benefits of new products: The case of the minivan. *Journal of Political Economy 110*(4), 705–729.

Poole, K. T. and H. Rosenthal (1985). A spatial model for legislative roll call analysis. *American Journal of Political Science*, 357–384.

Poole, K. T. and H. L. Rosenthal (2011). *Ideology and congress*, Volume 1. Transaction Publishers.

Rekkas, M. (2007). The impact of campaign spending on votes in multiparty elections. *The Review of Economics and Statistics 89*(3), 573–585.

Rennwald, L. and G. Evans (2014). When supply creates demand: Social democratic party strategies and the evolution of class voting. *West European Politics 37*(5), 1108–1135.

Rodden, J. A. (2019). *Why cities lose: The deep roots of the urban-rural political divide*. Basic Books.

Roemer, J. E. (1998). Why the poor do not expropriate the rich: an old argument in new garb. *Journal of Public Economics 70*(3), 399–424.

Rosen, J. B. (1965). Existence and uniqueness of equilibrium points for concave n-person games. *Econometrica*, 520–534.

Sanchez, R. (2021). Girth: G. item response theory. https://github.com/eribean/girth. Computer software.

Shapley, L. S. et al. (1953). A value for n-person games.

Shor, B. and N. McCarty (2011). The ideological mapping of american legislatures. *American Political Science Review 105*(3), 530–551.

Shor, B. and J. C. Rogowski (2018). Ideology and the us congressional vote. *Political Science Research and Methods 6*(2), 323–341.

Sieg, H. and C. Yoon (2017). Estimating dynamic games of electoral competition to evaluate term limits in us gubernatorial elections. *American Economic Review 107*(7), 1824–1857.

Spenkuch, J. L. and D. Toniatti (2018). Political advertising and election results. *The Quarterly Journal of Economics 133*(4), 1981–2036.

Strömberg, D. (2008). How the electoral college influences campaigns and policy: the probability of being florida. *American Economic Review 98*(3), 769–807.

Tausanovitch, C. and C. Warshaw (2017). Estimating candidates' political orientation in a polarized congress. *Political Analysis 25*(2), 167–187.

Ujhelyi, G., S. Chatterjee, and A. Szabo (2021). None of the above: Protest voting in the world's largest democracy. *Journal of the European Economic Association 19*(3), 1936–1979.

Vafa, K., S. Naidu, and D. M. Blei (2020). Text-based ideal points. *arXiv preprint arXiv:2005.04232*.

Wittman, D. (1983). Candidate motivation: A synthesis of alternative theories. *American Political science review 77*(1), 142–157.

Appendix

Contents

A Notation **52**

B Additional Tables and Figures not shown in the main paper **53**

C Electoral results and precinct boundaries **68**

C.1 Data sources 68

C.2 Panel of electoral precincts 69

C.3 Absentee Ballots 71

D Estimating the Full Distribution of Precinct-Level Demographics **74**

E Multimodal Text-and-Survey Ideal Point Model **76**

F Theory Appendix **80**

G Additional Demand Results **83**

G.1 List of Moments used for Demand Estimation 83

G.2 log-log specification: Homogeneous voters 83

G.3 Alternative identification strategy 84

G.4 Additional candidate observable characteristics 87

G.5 Voters' turnout decision 89

H Alternative supply specification **91**

I Details on Realignment Counterfactuals **93**

J Details on Counterfactual Support for Environmental Policies **94**

K Alternative candidate ideology model **99**

A Notation

Symbol	Description
p	index of precinct (block-group)
t	index of election year
j	index of candidate
k	index of policy dimension
i	index of voter
$g(p)$	precinct-group
$d(p)$	district of precinct p
x_{jkt}	position of candidate j on topic k in election t
x_{kt}	difference in position between the Democrat and the Republican on topic k at election t
ξ_{pt}	precinct p's unobserved taste shock for the Democratic candidate at election t
ξ_p	precinct p's unobserved constant taste shock for Democratic candidates
$\xi_{g(p)t}$	Precinct-group $g(p)$ unobserved taste shock for Democratic candidates at election t
ϵ_{ijt}	Voter i's unobserved taste shock for candidate j at election t
β_{ikt}	Voter i's preference over candidate's positions on dimension k at election t
α_{it}	Voter i's "non-ideological" partisanship at election t
η_{jkt}	Candidate j's unobserved preference for dimension k
$\mathbf{w_{it}}$ (resp. $\mathbf{w_{pt}}$)	Demographics of voter i (resp. precinct p) at election t
$F_t(\mathbf{w}, p)$	Distribution of demographics in precinct p at election t
$s_{ijt}(.)$	Voter i' preferences in election t
$s_{jpt}(.)$	Candidate j's vote share in precinct p at election t
ν_{it}	Unobserved voter i heterogeneity in preferences at time t
Π_{jt}	Candidate j objective function at time t
λ_{jt}	Parameter of party discipline at election t in candidate j's party
N_{jkt}	Position of party P's leadership on dimensions k at time t
$\mathbf{m}$	Mapping of precincts p to districts d: $\mathbf{m} = (m_{pd})_{\forall p,d}$

B Additional Tables and Figures not shown in the main paper

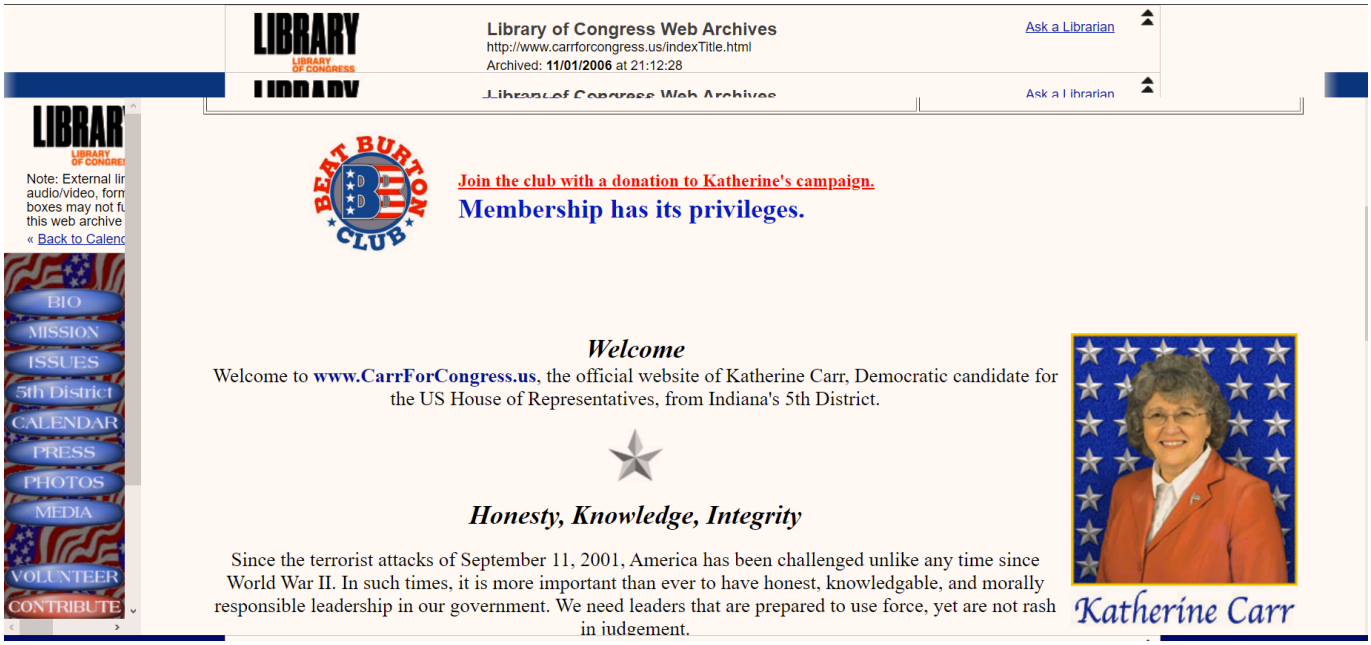


Figure A.1: Example of a website for a U.S. House candidate in Indiana in 2006

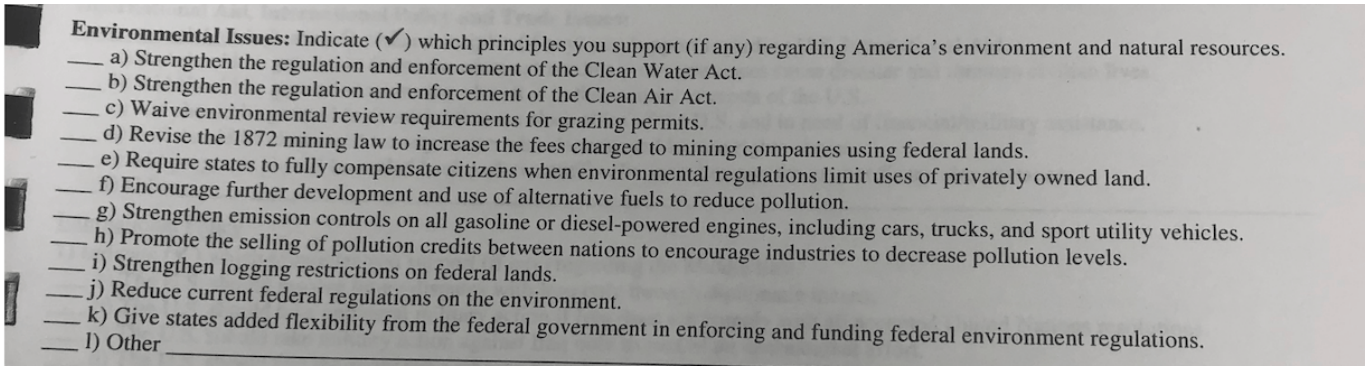


Figure A.2: Example of candidate survey (environment section in 2002)

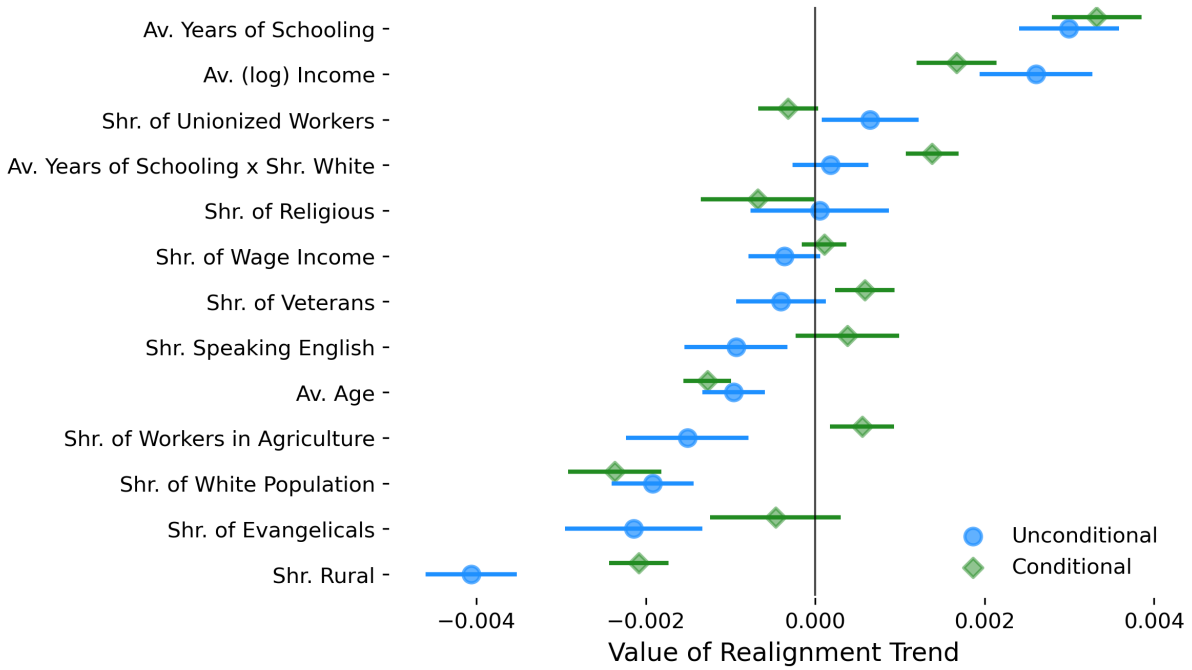


Figure A.3: Observed Trends in Political Realignment - Unconditional Estimatess

Notes: The figure shows the impact of z-score of precinct-level demographic variables on Democratic vote shares. The unconditional regression (shown with a round blue marker) tests for the trend separately for each variable, showing β_2^w from the following linear regression: $S_{Dpt} = \sum_w \beta^w \times w_{pt} \times t + \sum_w \gamma^w \times w_{pt} + \mu_t + \epsilon_{p,t}$ where $S_{D,p,t}$ is the share of vote obtained by Democratic candidates in precinct p at time t, w_{pt} is the normalized value of the demographic in that precinct at time t, and μ_t are election fixed-effects. Conditional coefficients are as in figure 2, unconditional coefficients show the results from a regression of each demographic variable separately.

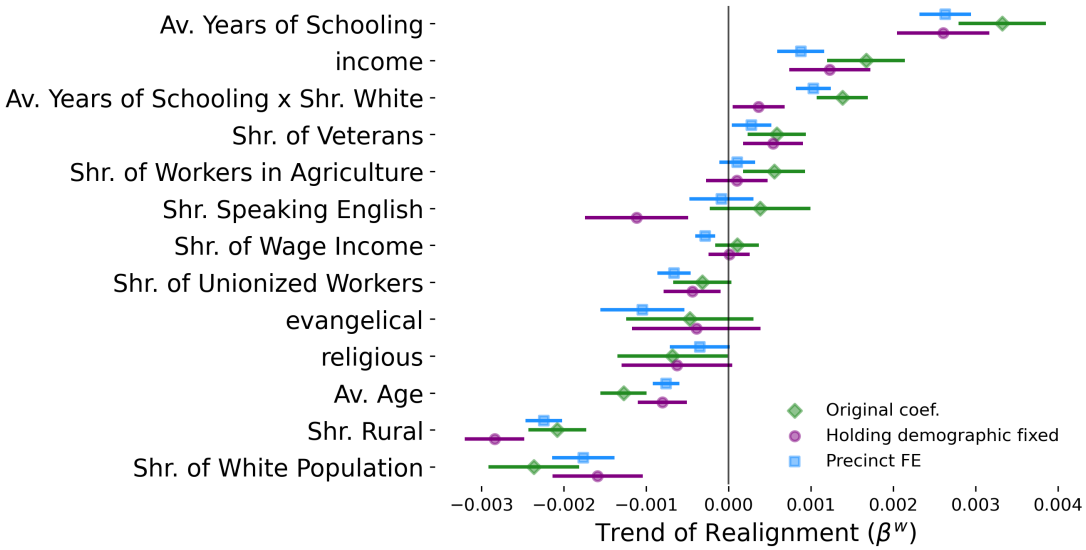


Figure A.4: Robustness of Observed Trends in Political Realignment

Notes: As in Figure 2, the figure shows, for each demographic variable w_{pt}, the coefficients β^w from the following linear regression: $S_{Dpt} = \sum_w \beta^w w_{pt} \times t + \sum_w \gamma^w w_{pt} + \mu_t + \epsilon_{pt}$ where S_{Dpt} is the House Democratic vote share in precinct p at time t, and μ_t are election fixed-effects. Positive coefficients indicate realignment toward the Democratic Party while negative coefficient indicate realignment toward the Republican Party. The bars around each marker show the 95% confidence intervals with standard errors clustered two ways, at the precinct level and at the congressional district by year. The green diamond markers show the original coefficients, as in Figure 2, the purple circle markers show the coefficients from a specification using only the demographics in 2000, therefore accounting for any sorting. The blue square markers show the coefficients from a specification with precinct fixed effects, therefore using only changes over time.

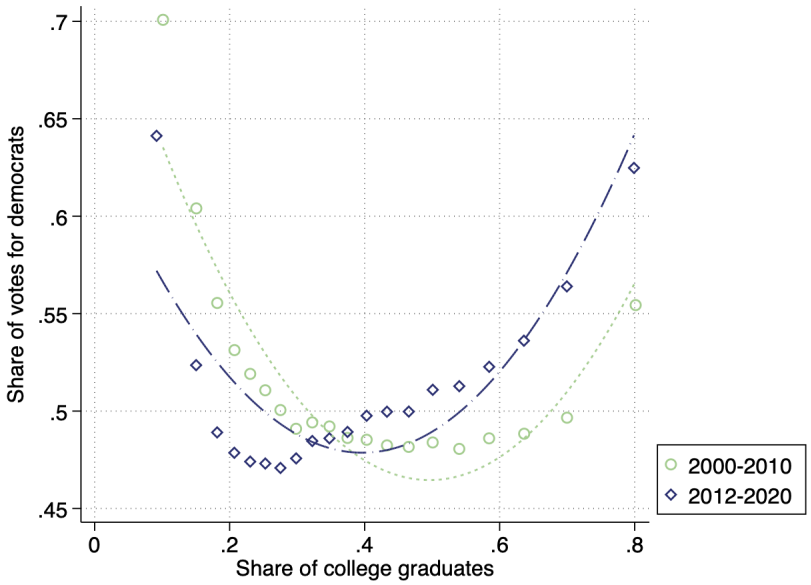
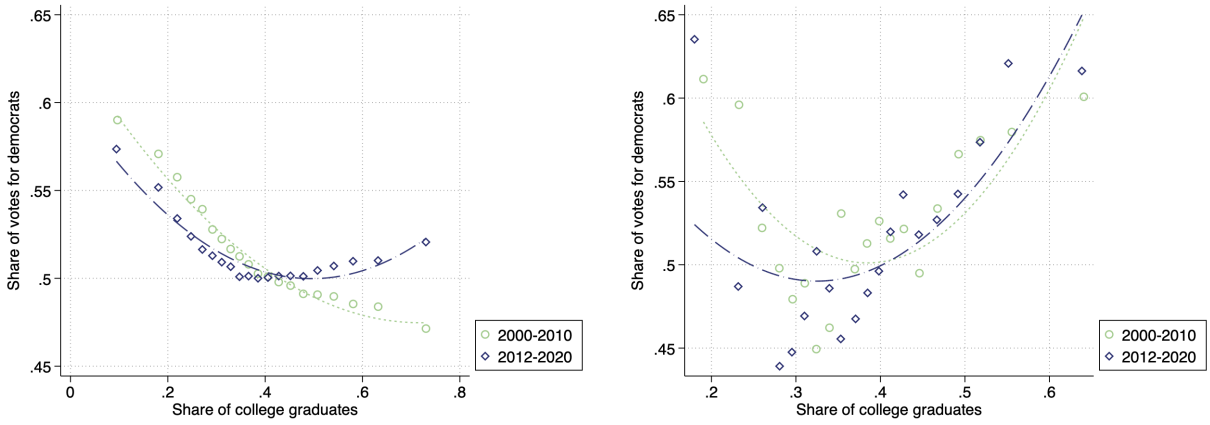


Figure A.5: Relationship between education and Democratic vote.

Notes: The figure shows the 5% quantiles fo the distribution of education and Democratic voters. Each dot represents 5% of the population. The curve is a quadratic fit of the data. Equal weights is given to each year within each period. Appendix Figure A.7 shows the same relationship separately for each election.



(a) Within-district

(b) Between-district

Figure A.6: Evolution of the within and between congressional district education gradient

Notes: The two Figures show the quantiles of the distribution of education and Democratic votes. Each dot represents 5% of the population. The curve is a quadratic fit of the data. Equal weights is given to each year within each period. The first panel shows the relationship within-district, conditional on congressional district by election fixed effects. The second panel shows the relationship between districts.

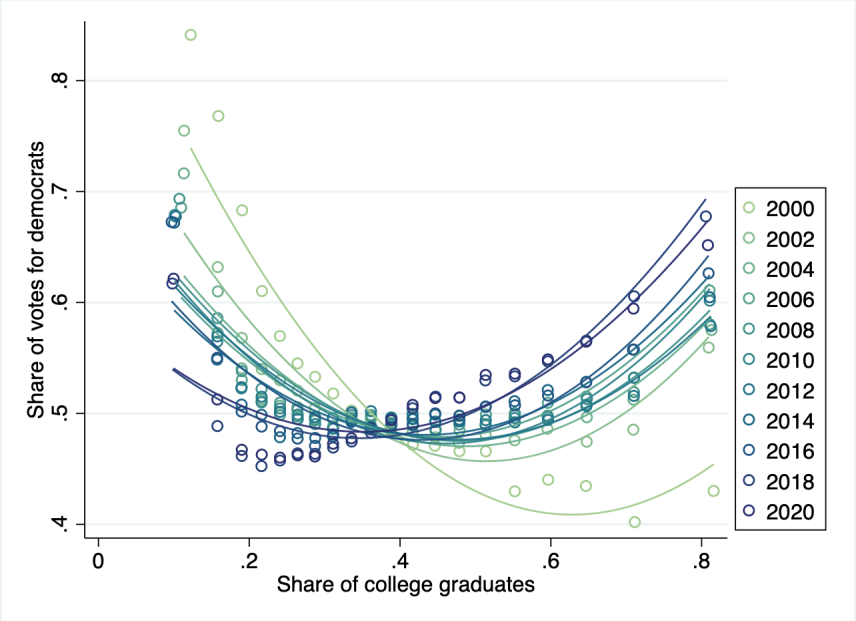
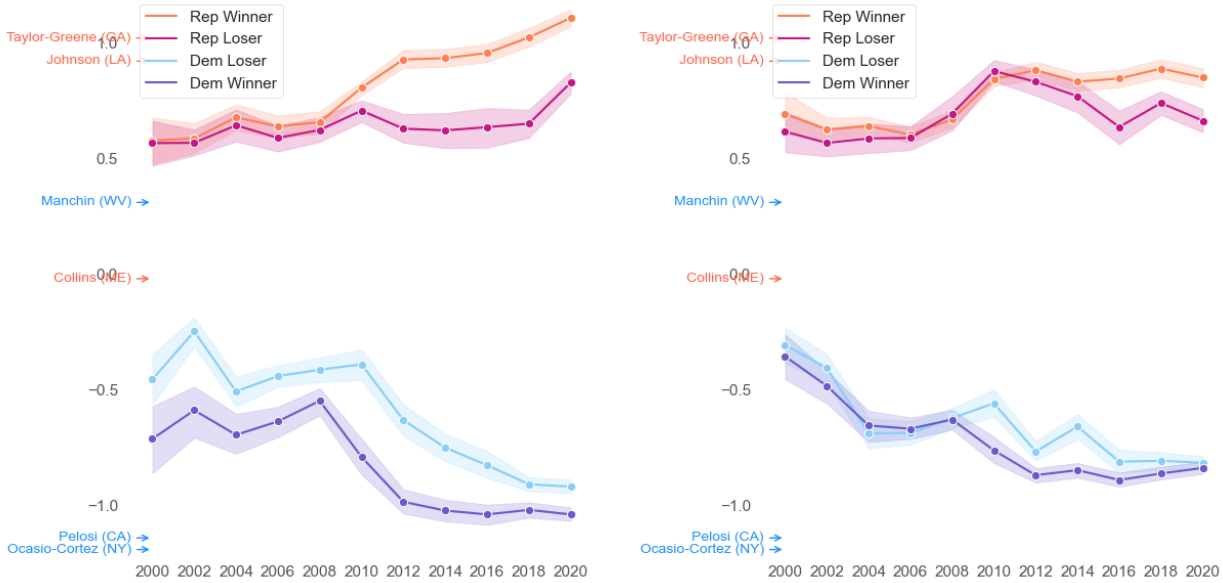


Figure A.7: Relationship between education and Democratic vote, separately for each elections.

Notes: As in Figure A.5, this figure shows the relationship between precinct education and Democratic vote share, separately for each year.

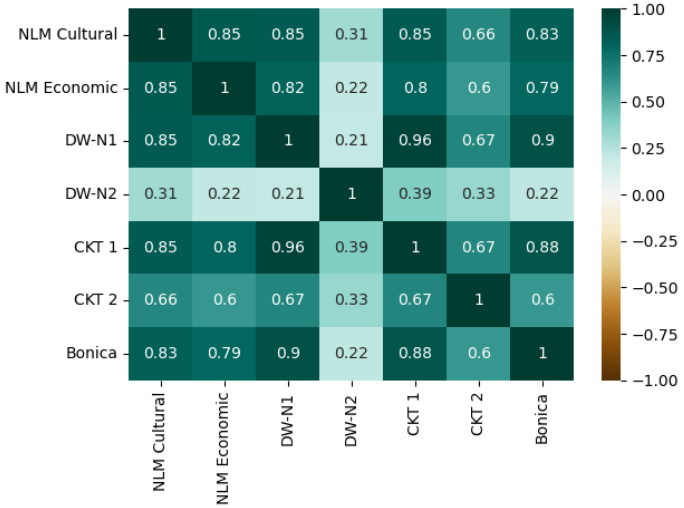


(a) Cultural Positions

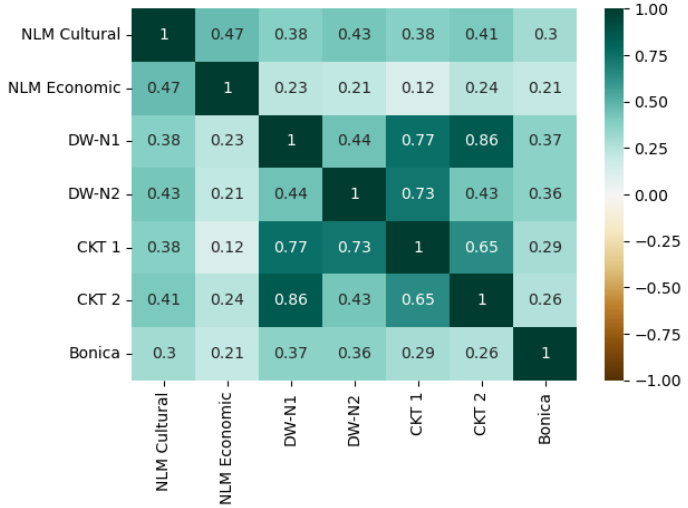
(b) Economic Positions

Figure A.11: Evolution of candidate positions on economic and cultural questions, separately for election winners and losers

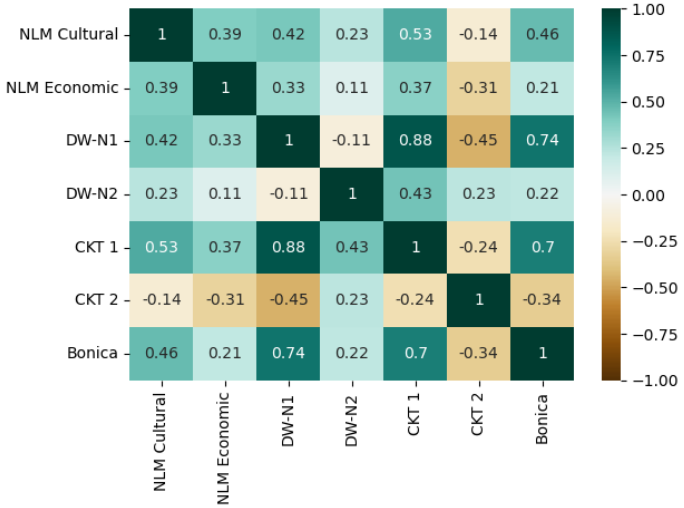
Notes: As in Figure 5, this figure shows the evolution of the average position of candidates in each party on each dimension, separately for Congressional candidates who have won and those who have lost the election.



(a) All Sample



(b) Democrats only



(c) Republicans only

Figure A.8: Comparison between own measures and common measure of ideology

Notes: The first panel shows the pairwise Spearman (rank) correlation between measures computed for this paper (denoted as NLM) and other commonly used measures of ideology. The correlations with DW-Nominate first and second dimensions (DW-N1 and DW-N2) are only for House and Senate winners. The pairwise correlations with Canen, Kendall, and Trebbi (2021) use only Senate election winners. The pairwise correlations with (Bonica, 2014) uses all candidates. The second and third panel show the within-party correlations for Democrats and Republicans, respectively.



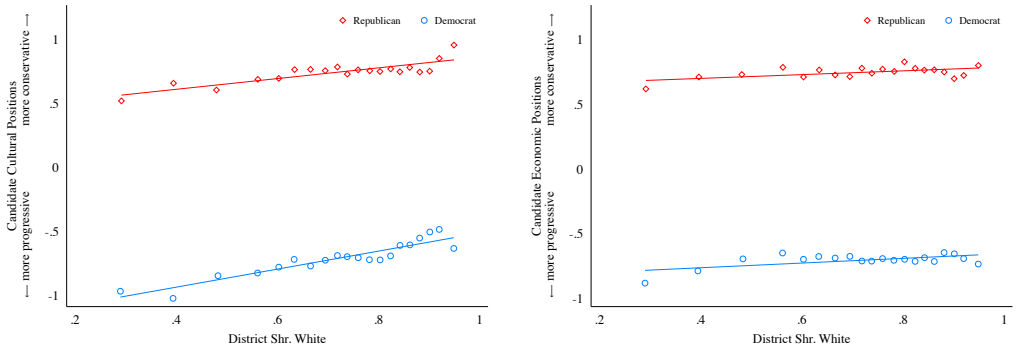
Figure A.9: Distribution of Candidates' Difference in Ideology

Notes: The figure shows the distribution of ideological differences between the Democratic and the Republican candidate in each Congressional race from Figure 4. Ideological differences in districts with notable candidates are displayed on the graph for the 2020 elections. The overall correlation between the difference on cultural issues and on economic issues is 0.46.

Parameter	MNL (Homogenous voters)				BLP (Heterogenous voters)			
	Bias	MSE	Coverage Proba	Power	Bias	MSE	Coverage Proba	Power
x_1	0.158	0.087	0.933	0.933	0.032	0.089	0.981	0.894
x_2	0.103	0.024	0.837	1.000	0.010	0.019	0.942	1.000
$x_1 \times$ yrs schooling	0.320	0.253	0.875	0.394	0.007	0.218	0.971	0.929
$x_1 \times$ yrs schooling $\times$ race	0.113	0.019	0.779	1.000	0.005	0.008	0.962	1.000
$x_1 \times$ race	0.141	0.034	0.740	1.000	0.003	0.027	0.952	1.000
$x_1 \times$ age	0.202	0.098	0.885	0.913	0.004	0.076	0.962	0.962
$x_2 \times$ yrs schooling	0.130	0.035	0.894	1.000	0.011	0.026	0.971	1.000
$x_2 \times$ yrs schooling $\times$ race	0.067	0.020	0.933	1.000	0.006	0.019	0.952	1.000
$x_2 \times$ race	0.496	0.261	0.010	0.990	0.004	0.031	0.913	1.000
$x_2 \times$ age	0.142	0.126	0.904	0.788	0.006	0.139	0.923	0.817
yrs schooling	0.191	0.097	0.885	0.865	0.034	0.086	0.942	0.885
yrs schooling $\times$ race	0.105	0.025	0.817	1.000	0.002	0.017	0.952	1.000
race	0.458	0.258	0.433	0.740	0.043	0.102	0.923	0.942
age	0.134	0.160	0.952	0.625	0.058	0.190	0.962	0.712

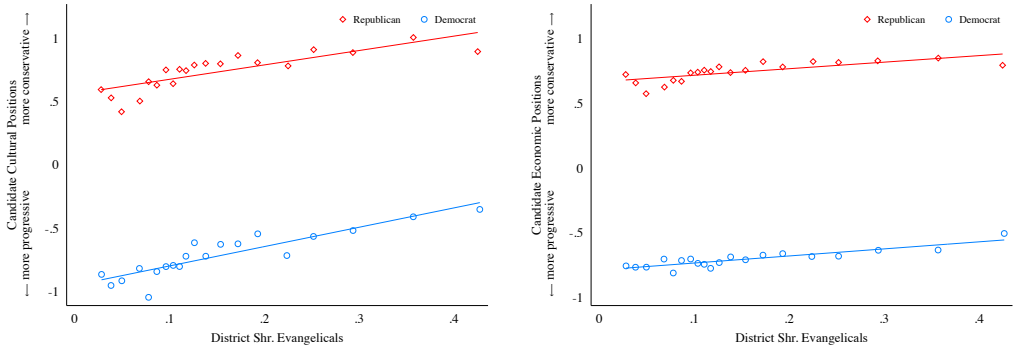
Table A.1: Statistics on simulated parameters

Notes: The table reports statistics of simulations of parameter identifications. For both the model with homogeneous voters: Multinomial Logit (MNL) as in Berry (1994) and the model with heterogeneous voters (BLP), it gives the bias of the estimates, the mean squared error (MSE), the coverage probability, and the power.



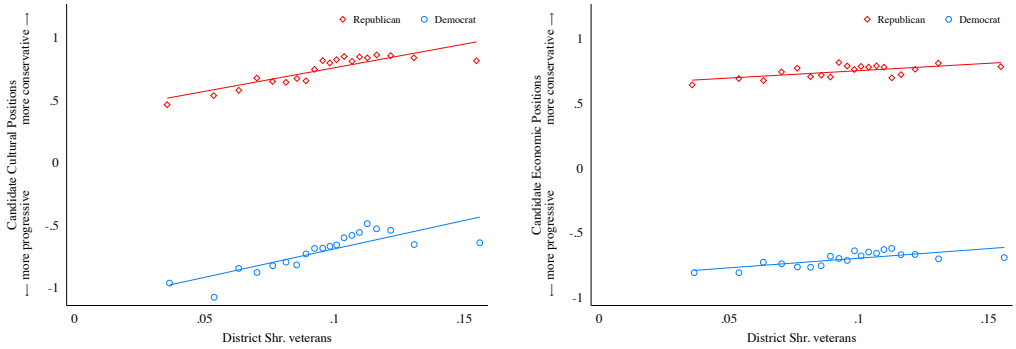
(a) Share of White and Cultural Issues

(b) Share of White and Economic Issues



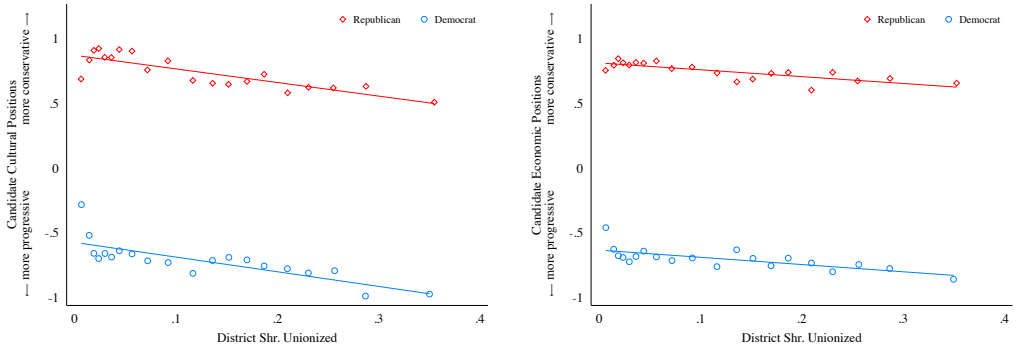
(c) Share of Evangelicals and Cultural Issues

(d) Share of Evangelicals and Economic Issues



(e) Share of Veterans and Cultural Issues

(f) Share of Veterans and Economic Issues



(g) Share of Unionized and Cultural Issues

(h) Share of Unionized and Economic Issues

Figure A.10: Candidate positions and congressional district composition

Notes: As in Figure 6, each panel shows the relationship between candidate positions and district demographics. Each dot represents 5% of the distribution and shows the average position of candidates, separately for Democrats and Republicans.

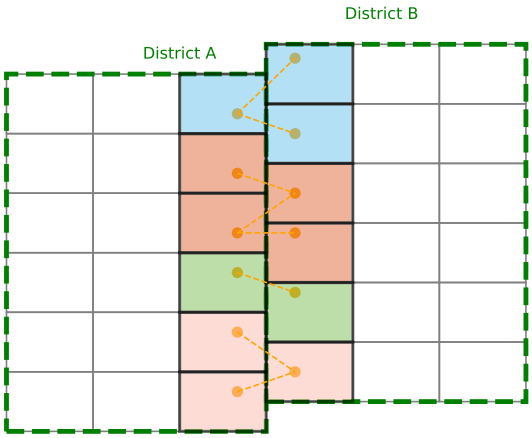


Figure A.12: Construction of precinct *pairs*

Notes: This figure describes the method adopted to match contiguous precincts with each other. Each square represents a precinct. Each precinct's population-weighted centroid is represented by a dot in the precinct. Each precinct is matched to the precinct to the closest precinct on the other side of the border. Each color represents a precinct group.

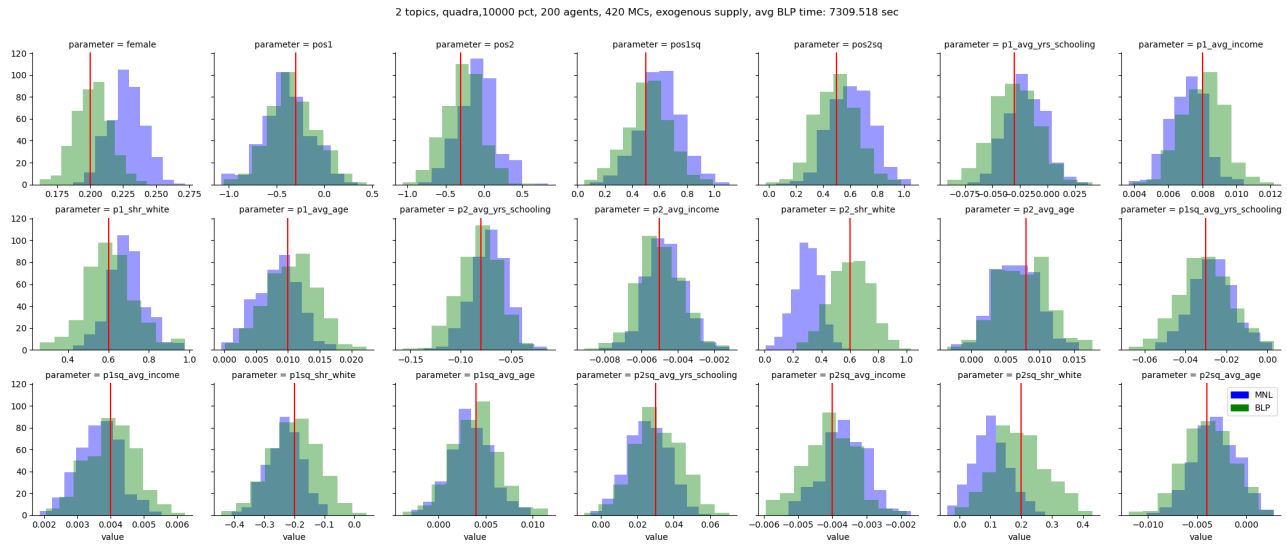
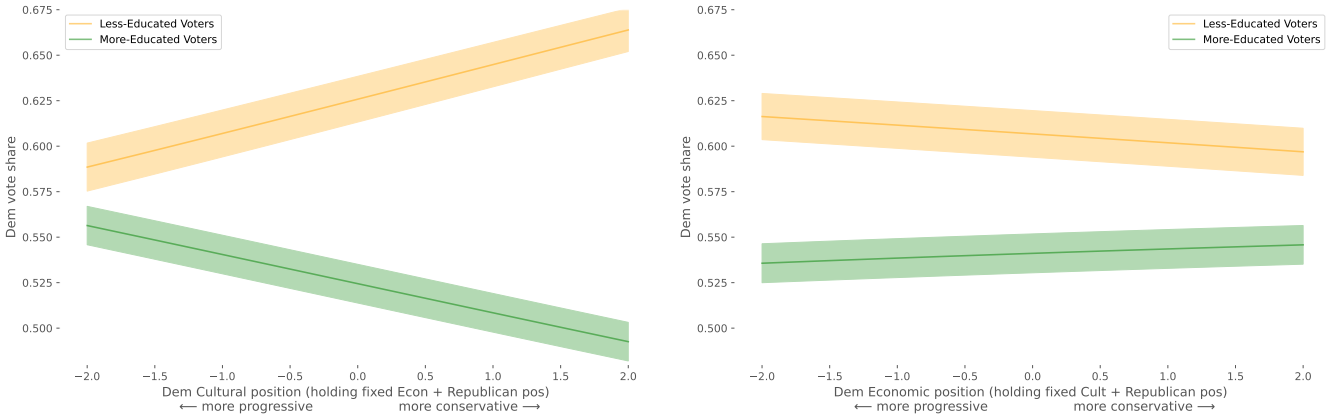


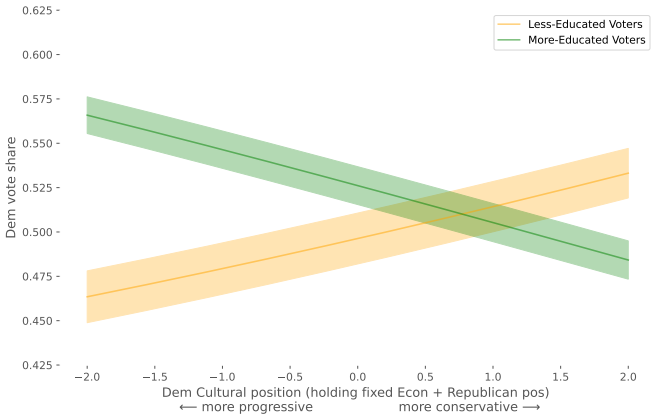
Figure A.13: Distribution of demand-side parameters from 420 Monte Carlo simulations using 20,000 precincts by election

Notes: Each histogram shows the distribution of parameters, in blue using the multinomial logit (MNL), Berry (1994) and in green using within-precinct heterogeneity (BLP). The red lines show the true parameters.

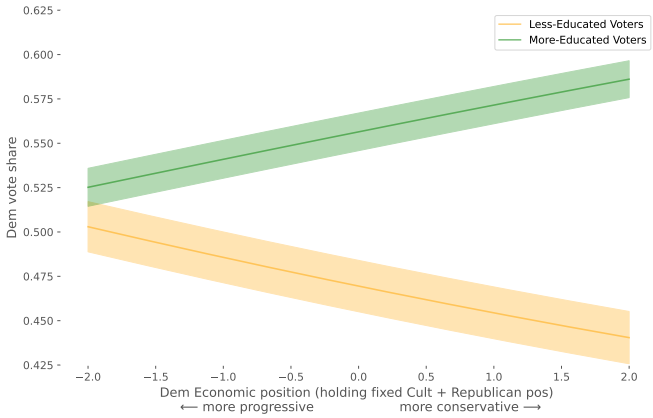


(a) Cultural issues (2000-2010)

(b) Economic issues (2000-2010)



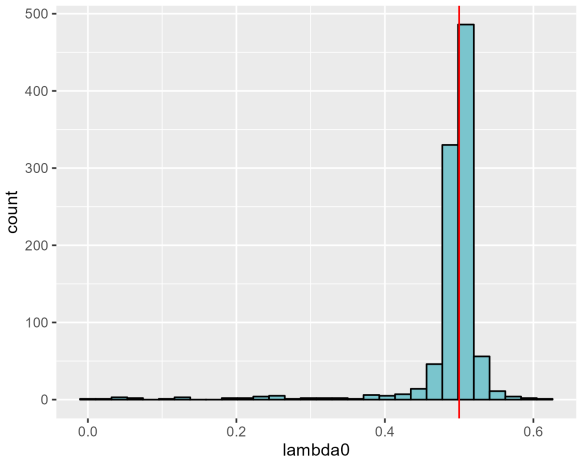
(c) Cultural issues (2012-2020)



(d) Economic issues (2012-2020)

Figure A.14: Deviations from candidate positions and vote shares

Notes: The figure presents predicted Democratic vote shares as a function of candidate positioning on cultural issues (left panels) and economic issues (right panels), incorporating precinct-group-by-election and precinct fixed effects. The top panels show the initial preferences in 2000-2010 and the bottom panel shows preferences in 2012-2020. Negative values represent more progressive candidate stances. A negative slope in each panel indicates a preference for more progressive positions among the respective group. Each panel isolates unilateral changes by holding constant the position on the alternate topic and the stance of the Republican candidate. 95% confidence intervals are included.



(a) λ_0

Figure A.15: Distribution of supply-side parameters from 420 Monte Carlo simulations using 435 districts.

Notes: The figure shows the distribution of parameters λ_{jt}, estimated by GMM. The red line shows the true parameters.

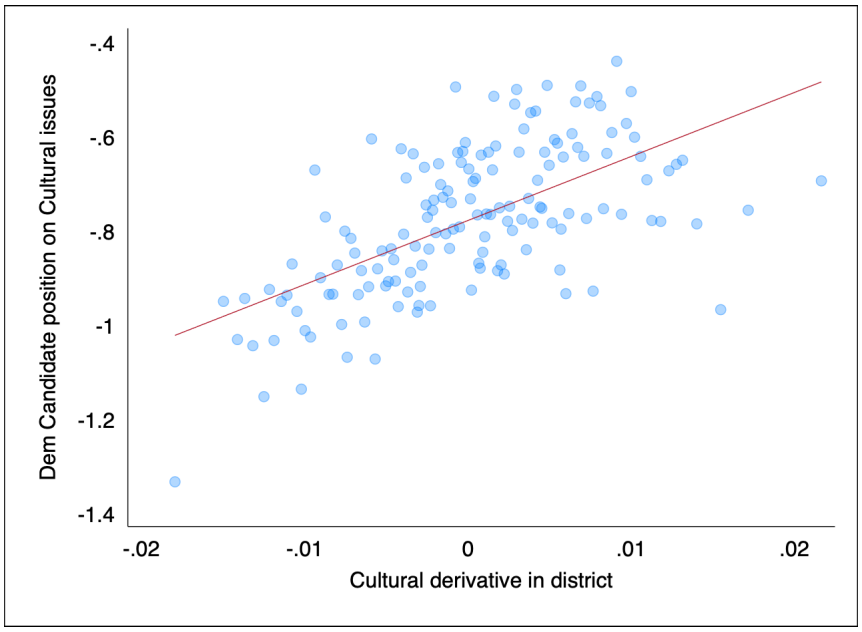


Figure A.16: Demand derivative and candidate ideological positions.

Notes: This figure illustrates the relationship between district-level demand derivatives for cultural issues and the positions adopted by Democratic candidates, highlighting the underlying supply estimation. The figure indicates that Democratic candidates adopting more progressive cultural stances tend to compete in districts where a marginal shift toward progressive positions would yield higher vote shares.

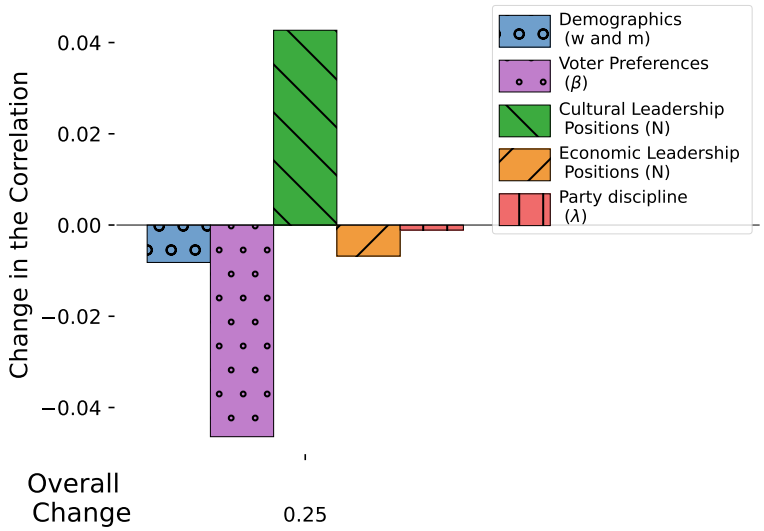
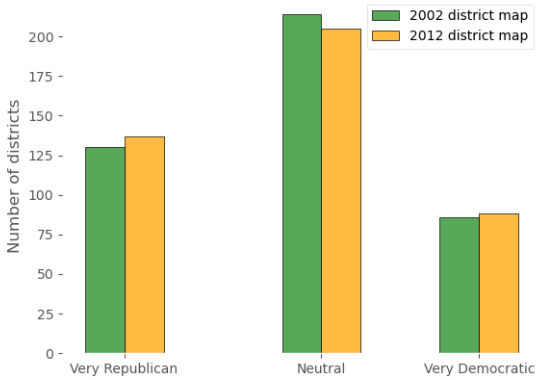
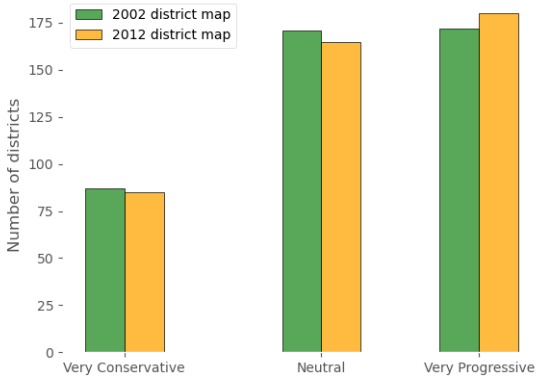


Figure A.18: Equilibrium Contributions of each Factor to Changes in the Correlation between Education and Democratic Voting.

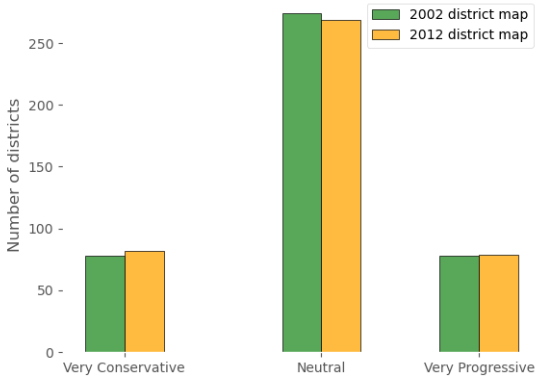
Notes: The figure displays the relative contributions of demand-side and supply-side factors to the change in the correlation between education and Democratic vote shares from 2000–2010 to 2020. For each voter, I compare their probability of voting for the House Democratic candidate under each counterfactual scenario. Over that period, the correlation increased by 25 percentage points.



(a) Average Democratic Partisanship



(b) Average Cultural Preferences



(c) Average Economic Preferences

Figure A.17: Number of politically biased districts following the 2010 Congressional redistricting.

Notes: The Figure shows the number of congressional district that can be considered as very politically biased, either in terms of partisanship (a) or ideology (b and c). For each graph, I compute the district level average partisanship or ideological coefficient with the 2002 or 2012 district maps. Districts are classified as neutral if they fall between the 25^{th} and 75^{th} percentiles of the precinct distribution. For example, following the 2010 redistricting, there are 7 fewer districts that can be considered as neutral in terms of partisanship.

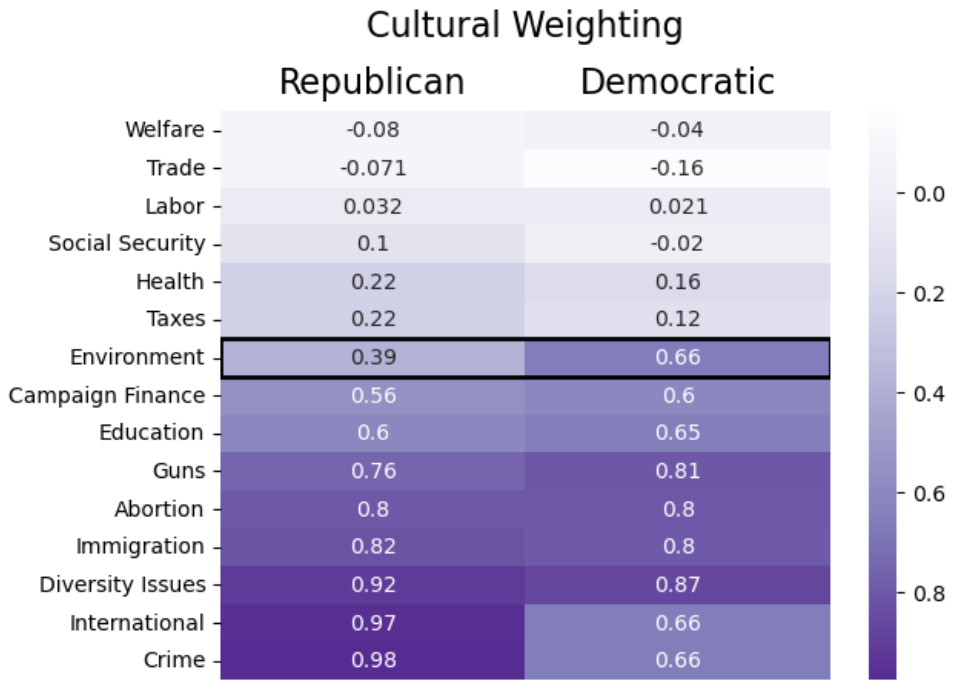


Figure A.19: Party Cultural Weights on each Topic

Notes: For each of the 15 topics, I re-estimate the ideal point model to get a position specifically on that topic and a cultural and economic dimensions, excluding that topic. I then project the topic ideal points on the cultural and economic dimensions by regressing the topic-specific ideal point on the economic and cultural ideal points: $x_{j,k} = \gamma_0 + \gamma_{k,cult} x_{j,cult} + \gamma_{k,econ} x_{j,econ} + \zeta_{jk}$, separately for each party. I plot the relative weighting coefficients $\rho_k = \frac{\gamma_{k,cult}}{\gamma_{k,cult} + \gamma_{k,econ}}$. Appendix Figure A.20 shows the same results graphically.

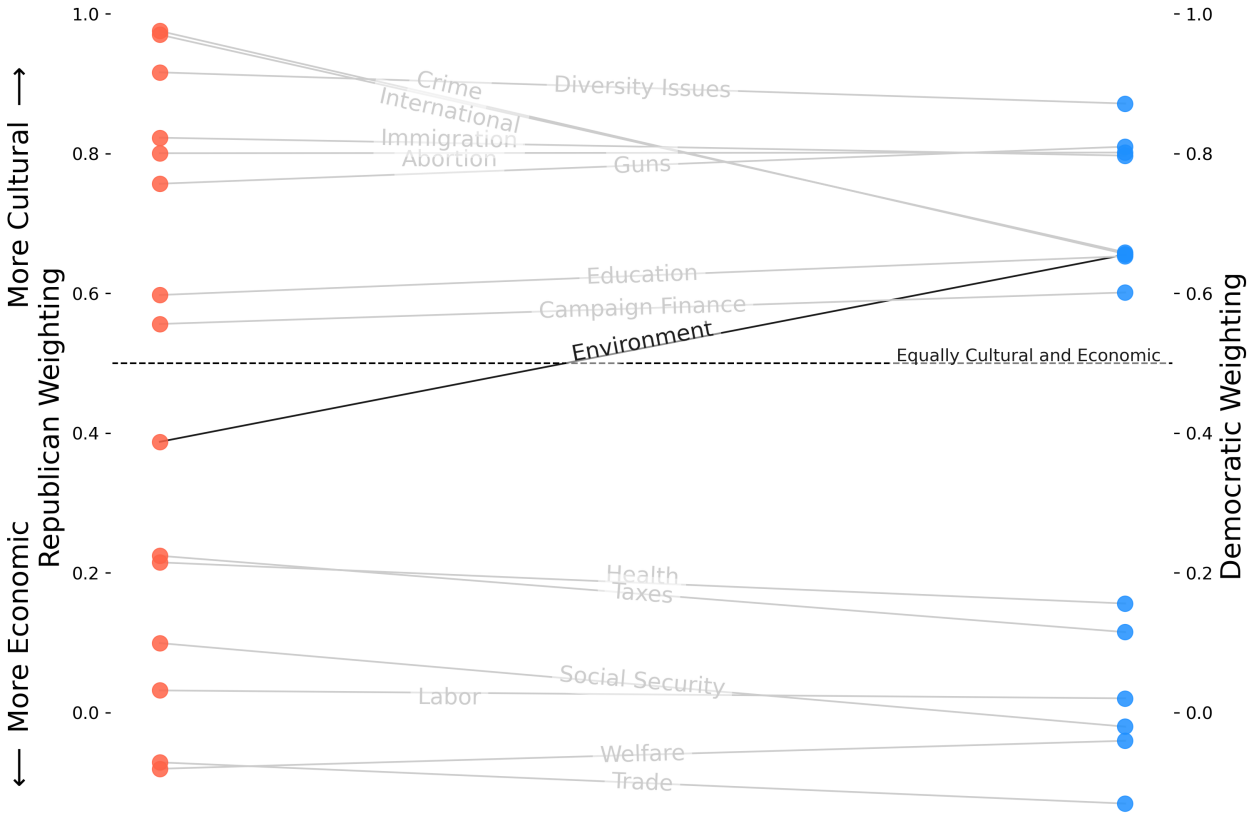


Figure A.20: Party Cultural Weights on each Topic

Notes: As in Figure A.19, for each of the 15 topics, I re-estimate the ideal point model to get a position specifically on that topic and a cultural and economic dimensions, excluding that topic. I then project the topic ideal points on the cultural and economic dimensions by regressing the topic-specific ideal point on the economic and cultural ideal points: $x_{j,k} = \gamma_0 + \gamma_{k,cult} x_{j,cult} + \gamma_{k,econ} x_{j,econ} + \zeta_{jk}$, separately for each party. I plot the relative weighting coefficients $\rho_k = \frac{\gamma_{k,cult}}{\gamma_{k,cult} + \gamma_{k,econ}}$.

C Electoral results and precinct boundaries

This section describes the methods and sources used to build the panel of electoral precincts. Since boundaries of electoral precincts change over time, I have interpolated all the precinct results to the census block-group level, which have approximately the same population (1,875 for precincts vs. 1,375 for block-groups). Table A.2 shows the distribution of state-years included in the sample. Table A.3 shows the distribution of state-years collected and cleaned for the project but not included in the sample, either because of important unbalanceness or states having a single congressional district.

I first describe the sources used for the electoral results and the precinct shapefiles. I then explain the interpolation strategy to obtain results at the block-group level. Third, I explain how I treat absentee ballots. Finally, I describe the mechanisms implemented to check the quality of the data.

C.1 Data sources

For each election year and each state, I started by collecting all the data that was already available from previous data collection initiatives. Specifically, I collected the data available from the following five initiatives: the Harvard Election Data Archive (Ansolabehere et al., 2014, 2018), data from the MIT Election Data and Science Lab (Baltz et al., 2022), data from the Voting and Election Science Team at the University of Florida, data from the *OpenElections* initiative, and data from the *Redistricting Data Hub*. I also collected the U.S. Census Voting Districts (VTD) boundary shapefiles for 2000, 2010, and 2020.

If the results and/or the boundaries were missing from these sources or if the matching rate was too low I collected the results and the boundaries myself from Secretaries of State and county election officials. I provide below the list of alternative sources when election results or precinct shapefiles were not provided by any of the above-listed initiatives.

The list of state-years included in the final dataset is provided in Appendix Table A.2.

Since most of the precinct-level election results and precinct shapefiles do not come from the same sources, I match precinct-level results to their shapefiles using fuzzy-string matching within county, keeping the most likely pairs that have a normalized Levenshtein similarity above 0.7. Also, since I do not have shapefiles for every single year in the period for each state, I match the precinct-level election results with their closest shapefiles from the same decennial period. Since precinct boundaries can change while keeping the same name, I conduct some tests on the number of votes between the election data and the precinct data. I exclude all precincts that have a difference in the total number of votes between election for the same offices of more than 25%.

- For Arizona, data for 2010 and 2014 come from the Arizona Secretary of State.
- For Arkansas, precinct shapefiles come from the Arkansas GIS Office.

- For California, all the election results and precinct boundaries come from the Statewide Database at the University of California, Berkeley.
- For Colorado, 2000 and 2002 election results from the Colorado Secretary of State.
- For Connecticut, precinct shapefiles for 2016 come from the Metric Geometry and Gerrymandering Group
- For Florida, election results from 2012 onward come from the state-level Division of Elections.
- For Georgia, 2012 election results come from the Secretary of State.
- For Kentucky, results come from the State Board of Elections.
- For Massachusetts, precinct shapefiles come from the Metric Geometry and Gerrymandering Group (MGGG).
- For Michigan, results from 2000 to 2014 come from the Department of State.
- For Minnesota, election results were obtained from the Office of the Secretary of State and precinct boundaries from the Legislative Coordinating Commission of the Minnesota legislature.
- For New Mexico, 2006 and 2010, I digitized the electoral results from PDF reports from each county.
- For North Carolina, electoral results and precinct shapefiles were obtained from the North Carolina State Board of Elections.
- For Ohio, electoral results for 2004, 2006, and 2008 come from the Department of State.
- For Virginia, precinct boundaries come from Erika Lopresti's Github
- For Washington State, King County results for 2012 and 2014 come from the county election commission.
- For Wisconsin, all the results and precinct boundaries come directly from the State Legislature.

I systematically used the congressional district information included in the electoral results, if the district was not mentioned for each precinct, I assign them using the shapefile of congressional districts compared with the precinct boundaries.

C.2 Panel of electoral precincts

To assign votes to each block group I implement the following strategy, illustrated on Figure A.21.

First, I use geospatial analysis to compare the precinct shape with each census block and create spatial crosswalks. Second, I compute the total precinct population that belongs to the block group by using the block-level population. When a block is divided into several precincts, I assign only the share of the population that is covered by the precinct.

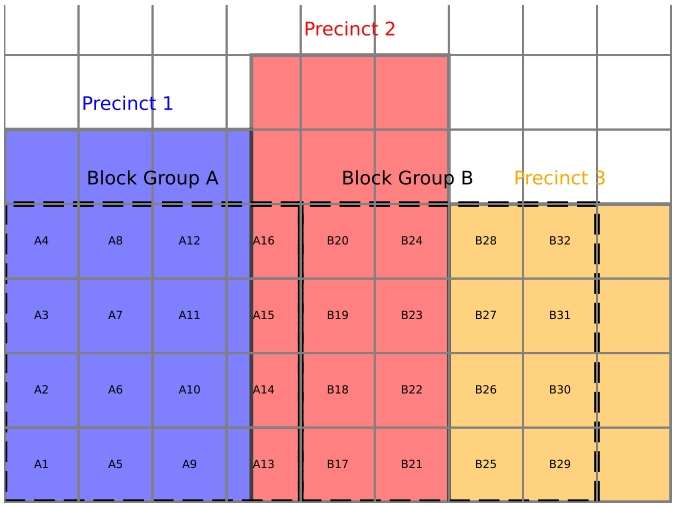


Figure A.21: Precinct votes allocation

Notes: This figure describes the strategy to allocate votes from the precinct to the block-group level. Each square of the grid shows a census block. Block-groups are shown in green and precincts are shown in blue, red, and yellow.

The total number of votes for candidate j in block group bg can be written as:

$$Votes_{j,bg} = \sum_{p:\{p \cap bg\} \neq \emptyset} \sum_{b \in \{p \cap bg\}} Votes_{j,p} \cdot w_{b,p} \cdot \frac{Pop_b}{Pop_p}$$

where $w_{b,p}$ is the share of block's b area that falls within precinct p's boundaries, Pop_b and Pop_p the population of the block and precinct population, respectively.

For example, taking the situation described in Appendix Figure A.21; where there are three precincts. In order to allocate the votes to block group A and block group B, I first compute the share that intersects with each block within these block groups. Blocks are shown as grey squares of the grid. Assuming that the population of each block is equal: 80% of the population in precinct 1 belongs to block group A (block A1 to A12 and one third of A13, A14, A15, and A16 ($w_{b,p} = \frac{1}{3}$)). Similarly, 16.7% of precinct 2 belongs to block group A, while 50% belongs to block group B. 66% of precinct 3's population belongs to block group B. The total votes of each candidates in block group A, is therefore obtained by taking 80% of the votes from precinct 1 and 16.7% of the votes from precinct 2.

This strategy makes the following underlying assumption: (1) the population is uniformly distributed within each block, (2) the share of votes for each candidate is uniformly distributed across precincts. Importantly, I do not need to assume that the population is uniformly distributed within precincts or within block groups across blocks.

C.3 Absentee Ballots

Each state and sometimes each county reports absentee ballots differently. There are three main types of reporting of absentee votes. First, some counties directly assign absentee votes to each candidate results in each precinct, without distinction with the election-day votes. Second, some counties record absentee ballots separately for each precinct, providing a separate estimate of number of votes for each candidate in each precinct. Third, some counties report absentee votes at a more aggregate level than the precinct level. The rule applied in the election results has been to assign absentee ballots to specific precincts when possible and only when they were reported at a more disaggregated level than the county. For example, Michigan counties usually reports absentee ballots at the ward level, which is more aggreagated level than the precinct (about 3 precincts per ward on average). I therefore allocate the absentee votes to each precinct proportionally. If the Democratic candidate obtained 40% of their in-person votes from precinct A in ward 1, I allocate 40% of the absentee ballots from ward 1 to precinct A.

State	2000	2002	2004	2006	2008	2010	2012	2014	2016	2018	2020
Alabama	x	x	x	x	x	x	x	x	x	x	x
Arizona		x	x	x	x	x	x	x	x	x	x
Arkansas	x	x	x	x	x	x	x	x	x	x	x
California	x	x	x	x	x	x	x	x	x	x	x
Colorado			x	x	x	x	x	x	x	x	x
Connecticut	x	x	x	x	x	x	x	x	x	x	x
Florida				x	x	x	x	x	x	x	x
Idaho	x	x	x	x	x	x	x	x	x	x	x
Illinois							x	x	x	x	x
Iowa	x	x	x	x	x	x	x	x	x	x	x
Kansas	x	x	x	x	x	x	x		x	x	x
Maine	x	x	x	x	x	x	x	x	x	x	x
Maryland		x	x	x	x	x	x	x	x	x	x
Massachusetts	x	x	x	x	x	x	x	x	x	x	x
Michigan	x	x	x	x	x	x	x	x	x	x	x
Minnesota	x	x	x	x	x	x	x	x		x	x
Mississippi			x	x	x	x	x		x	x	x
Missouri	x	x	x	x	x	x		x	x	x	x
Nebraska		x		x		x		x		x	x
New Hampshire	x	x	x	x	x	x	x	x		x	x
New Jersey	x	x	x	x	x				x	x	x
New Mexico	x		x	x	x	x	x	x	x	x	x
New York				x	x	x	x	x	x	x	x
North Carolina	x	x	x	x	x	x	x	x	x	x	x
Oklahoma	x	x	x	x	x	x		x		x	x
Ohio				x	x	x		x	x	x	x
Oregon	x	x	x	x	x	x		x	x	x	x
Pennsylvania	x	x	x	x	x	x	x	x	x	x	x
Rhode Island	x	x	x	x	x		x	x	x	x	x
South Carolina			x	x	x	x	x	x		x	x
Tennessee		x	x	x	x	x	x	x	x	x	x
Texas	x	x	x	x	x	x	x	x	x	x	x
Virginia	x	x	x	x	x	x	x	x	x	x	x
Washington				x	x	x	x	x	x	x	x
Wisconsin		x	x	x	x	x	x	x	x	x	x

Table A.2: States with Electoral Results Included in the Sample

State	2000	2002	2004	2006	2008	2010	2012	2014	2016	2018	2020
Alaska		x	x	x	x	x	x	x	x	x	x
Delaware		x	x	x	x	x	x		x	x	x
Georgia							x	x	x	x	x
Hawaii				x	x	x	x	x	x	x	x
Indiana							x	x	x	x	x
Kentucky									x	x	x
Louisiana	x	x	x	x	x	x	x	x	x	x	x
Montana										x	x
Nebraska					x					x	x
Nevada					x	x				x	x
North Dakota			x	x	x	x	x	x	x	x	x
South Dakota				x	x	x	x	x	x	x	x
Utah									x	x	x
Vermont		x	x	x	x	x	x	x	x	x	x
West Virginia										x	x
Wyoming	x	x	x	x	x	x	x	x	x	x	x

Table A.3: Electoral Results for States Not Included in the Sample

Notes: States are excluded from the sample either because their coverage is too low (e.g., Indiana or Michigan), because they are At Large seats (e.g., Delaware or Wyoming) or because they do not belong to the continental U.S. (Alaska and Hawaii).

D Estimating the Full Distribution of Precinct-Level Demographics

In order to recover the distribution of voter preferences in each precinct, I need the joint distribution of voter demographics, however, only the marginal distributions are available from U.S. Census data. I therefore implement a multi-level demographic model to recover the joint distribution. I consider two types of demographics: binary (white vs. non white) and continuous (education, income and age). The objective is to recover the joint distribution of the three continuous variables for both white and non-white voters.

Census data from IPUMS (Manson et al., 2021) provide count variables that give the total number of people in some education, income, and, age brackets, by race. This data allows me to obtain the average and the variance of each of these variables by race. Note that the estimated variance is probably under-estimated since the census data only provides brackets. There are two additional challenges. First, while IPUMS provides the marginal distribution of education by race, income by race, and age by race from block-group-level data, the data does not contain the joint distribution along these three dimensions. I therefore use PUMA-level individual data from the ACS from IPUMS to obtain correlations between education, income, and age for each year. Second, since the marginal distribution by race for education and income for 2010 and 2020 are only available at the tract level, I use tract-level marginal distribution by race and block-group level distribution for the whole population to extend each block group's 2000 marginal distribution by race.

The overall approach is described below in more detail:

In each block-group, for each decennial year, I need to estimate four matrices:

$$\boldsymbol{\mu^{white}} = (\mu^{white}_{edu}, \mu^{white}_{income}, \mu^{white}_{age}) \tag{17}$$

$$\boldsymbol{\mu^{NW}} = (\mu^{NW}_{edu}, \mu^{NW}_{income}, \mu^{NW}_{age}) \tag{18}$$

$$\boldsymbol{\sigma^{white}} = \begin{pmatrix} \sigma^{white}_{edu} & \sigma^{white}_{edu,income} & \sigma^{white}_{edu,race} \\ \sigma^{white}_{edu,income} & \sigma^{white}_{income,income} & \sigma^{white}_{income,age} \\ \sigma^{white}_{edu,age} & \sigma^{white}_{income,age} & \sigma^{white}_{age,age} \end{pmatrix} \tag{19}$$

$$\boldsymbol{\sigma^{NW}} = \begin{pmatrix} \sigma^{NW}_{edu} & \sigma^{NW}_{edu,income} & \sigma^{NW}_{edu,race} \\ \sigma^{NW}_{edu,income} & \sigma^{NW}_{income,income} & \sigma^{NW}_{income,age} \\ \sigma^{NW}_{edu,age} & \sigma^{NW}_{income,age} & \sigma^{NW}_{age,age} \end{pmatrix} \tag{20}$$

1. I use census individual level data to get the covariance of the continuous demographics by white/non-white by PUMA (about 1,000 census block-groups by PUMA). This gives me the non-diagonal terms of $\boldsymbol{\sigma^{W}}$ and $\boldsymbol{\sigma^{NW}}$
2. $\boldsymbol{\mu^{white}}$ and $\boldsymbol{\mu^{NW}}$ are fully observed observed in 2000 at the block-group level but only partially in

2010 and 2020, for which μ_{edu}^{white} and μ_{edu}^{NW} are only reported at the tract-level.

3. For μ_{edu}^{white} and μ_{edu}^{NW} in 2010 and 2020, I estimate them by using a combination of the overall education distribution in the block-group (BG) in these years, the initial distribution of education by race in 2000 and the distribution of education by race in 2010 at the tract-level. I recover $\mu_{edu,2010,BG}^{white}$ by making the following assumption:

$$\frac{\frac{\mu_{edu,t,BG}^{white}}{\mu_{edu,t,BG}^{all}}}{\frac{\mu_{edu,2000,\text{tract}}^{white}}{\mu_{edu,2000,\text{tract}}^{all}}} = \frac{\frac{\mu_{edu,2010,BG}^{white}}{\mu_{edu,2010,BG}^{all}}}{\frac{\mu_{edu,2010,\text{tract}}^{white}}{\mu_{edu,2010,\text{tract}}^{all}}} \tag{21}$$

Everything is observed except $\mu_{edu,2010,BG}^{white}$, the underlying hypothesis is that the ratio of education of white to non white in the block group versus in the tract has stayed constant between 2000 and 2010. I recover the distributions for non-white and for 2020 in the same way.

4. The diagonal terms of $\sigma^{\mathbf{NW}}$ and $\sigma^{\mathbf{white}}$ are only observed in 2000. In the other years I just observe the diagonal terms of an aggregated matrix for all races. I therefore use the same as strategy as above:

$$\frac{\frac{\sigma_{edu,2000,BG}^{white}}{\sigma_{edu,2000,BG}^{all}}}{\frac{\sigma_{edu,2000,\text{tract}}^{white}}{\sigma_{edu,2000,\text{tract}}^{all}}} = \frac{\frac{\sigma_{edu,2010,BG}^{white}}{\sigma_{edu,2010,BG}^{all}}}{\frac{\sigma_{edu,2010,\text{tract}}^{white}}{\sigma_{edu,2010,\text{tract}}^{all}}} \tag{22}$$

I recover the distributions for non-white, for 2020, and for age and income in the same way.

	Average	Overall std. deviation	Within precinct std. deviation	Within district std. deviation
Years of schooling	13.45	3.56	2.87	3.21
Age	48.91	17.98	12.87	16.32
Race (white dummy)	0.74	0.44	0.27	0.37

Table A.4: Statistics of demographics distribution

Notes: The table shows descriptive statistics on the distribution of education, age, and race and ethnicity at the individual level.

E Multimodal Text-and-Survey Ideal Point Model

For each election, I want to recover the underlying position x_j of candidate j. For simplicity, I write the ideal point with $k = 1$. For each candidate j, I observe the patterns of responses to the Votesmart survey questions $\mathbf{y_j} = (y_{j1}, y_{j2}, ..., y_{jQ})$ with Q the number of questions answered by that candidate at this election. Note that each question re-ordered so that $y_{jq} = 1$ corresponds to a conservative position.

I can write the probability that candidate j would have adopted this specific response pattern as:

$$L(x_j|\mathbf{y}_j) = \prod_{q=1}^{Q} \Pr(y_{jq}|x_j, a_q, b_q)$$

with $\Pr(y_{jq}|x_j)$ the probability of responding y_{jq} to question q.

$$\Pr(y_{jq} = 1|x_j) = \frac{1}{1 + e^{-a_q(x_j - b_q)}}$$

where y_{jq} denotes the response of candidate j to question q, x_j is the underlying position of candidate j, a_q is the discrimination (polarization) parameter for question q, b_q is the political orientation parameter for question q. In the specific context of political questions, a_q can be understood as the level of polarization of the question. An apolitical question where responses would be essentially uncorrelated with the underlying political dimension would get a very low a_q. b_q captures the political location of the question; it gives the threshold above which I expect a candidate to answer *yes* to the question. For high b_q, only very conservative candidates are expected to answer *yes*, for low b_q, most candidates are expected to answer *yes*. Note that I do not use a quadratic utility model as in Shor and Rogowski (2018) since most question ask more about a direction (e.g., increase social security contributions) than about a specific location of a policy

The model is estimated using Marginal Maximum Likelihood: I maximize the marginal likelihood of the observed data, integrating over the distribution of the latent trait:

$$\mathcal{L} = \prod_{j=1}^{N} \int_{-\infty}^{+\infty} L(x_j|\mathbf{y}_j) f(x_j) dx_j \tag{23}$$

where N is the number of candidates and $f(x_j)$ is the probability density function of the latent trait in the population that I assume is a Normal distribution with mean 0 and variance 1. The abilities are obtained by Expected A Posteriori (EAP) Estimation. I also obtain bootstrapped standard errors for each parameter after the estimation.

I estimate one ideal point model per issue (cultural, economic, environment, and cultural without environment) using Girth (Sanchez, 2021) on Python. The survey includes two main types of questions: questions asking whether a given candidate supports a policy (binary answer, such as: "Strengthening the regulation and enforcement of the Clean Water Act") or questions asking about the desired level of

spending or taxes on a specific dimension, such as "Do you think inheritance taxes should be greatly increased/slightly increased/maintain status/slightly decreased/eliminated". For each of these ordered questions, I build a set of new dummy variables equals to one if the answer is lower than each threshold. This gives me a set of binary questions for each election cycle. Appendix Table A.5 summarizes the number of raw question by sub-topic and their classification in issues. I exclude questions that mention a specific country which are likely to be context specific (e.g., "Should the United States increase financial support to Afghanistan?", "Should the U.S. support the creation of a Palestinian state). I also exclude four questions that were misinterpreted by candidates: e.g., "Abortions should be legal when the pregnancy resulted from incest or rape.", or "Aid should be granted to countries when it is in the security interests of the U. S." was interpreted by some candidates as "only in these cases" as it leads to some internal contradictions in candidates' answers. I also exclude the question: "Should trade agreements include provisions to address environmental concerns and to protect workers' rights?" since it could be classified as either "trade" or "environment" and therefore either cultural or economic. I also exclude the question: "Prohibit the dilation and extraction procedure.", asked in 2008 and 2010, which gives very inconsistent responses by the same candidates from year to the next.

Subtopic	Main Topic	Number of distinct raw questions
Abortion	Cultural	7
Crime	Cultural	11
Education	Cultural	7
Environment	Cultural[a]	17
Labor and Employment	Economic	9
Gun regulations	Cultural	8
International Trade	Economic	6
Campaign Finance	Cultural	4
Immigration	Cultural	8
Diversity Questions	Cultural	3
Health Care	Economic	9
Taxes and Spending	Economic	19
Security and International Policy	Cultural	7
Social Security	Economic	13
Welfare	Economic	4

Table A.5: Summary of topic classifications

[a]For the analysis in section 7, I re-compute the cultural ideal points, excluding environment from the classification.

Once I have obtained an ideal point and a standard error for each candidate that has answered the survey, I train and apply a machine learning regressor (Extreme Gradient Boosting) using features extracted from their website. I scrape the website in the following way: I collect all the text available on the first page, all the text available on any page listed on the first page and iterate once again by taking all the text available on any page listed on these second pages. I only collect text that is on the

website (i.e., I do not scrape external websites even if they are referenced on the candidate's website). Starting from the raw text data scraped from candidate websites, I clean the text in regular ways by removing all words that are a consequence of the data being displayed on a website (e.g., "contact me", "send an email", "access photos", etc.), removing names of candidates, state names and region names. I then construct unigrams, bigrams and trigrams (sequences of 1, 2 or 3 words) and I remove words that are too infrequent (used in less than 0.05% of the documents), this gives me a large matrix of occurences of tokens for each candidate. In addition, I also compute document embeddings (Dai et al., 2015) which give a vector representation of each document.

The mean squared error (MSE) for the economic prediction is 0.17 (R-squared of the prediction is at 0.91) and 0.21 for the cultural dimension (R-squared of 0.89). For candidates for whom I only have the website (26%) of the sample, I only assign the website ideal point, for those with only the survey (18%), I only assign the survey ideal points and for all those with both survey and website ideal points (46%), I take a weighted average of the two measures using the following formula:

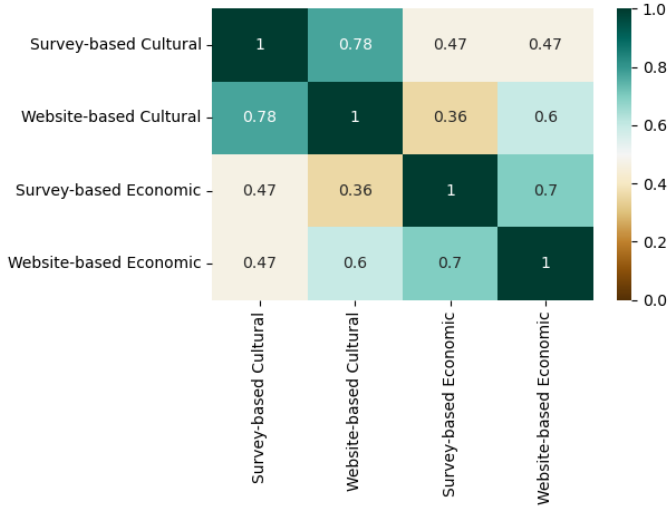
$$x_j = w_j x_j^{survey} + (1 - w_j)\widehat{x_j^{website}} \tag{24}$$

with $w_j = \frac{MSE(x^{website})}{se(x_j^{survey})^2 + MSE(x^{website})}$.

Figure A.22 shows the correlation between the survey-based and website-based measures.



(a) All Sample



(b) Democrats only



(c) Republicans only

Figure A.22: Comparison between survey-based and website-based measures

Notes: The first panel shows the pairwise correlation between survey-based and website-based measures. The first panel look at the overall correlation. The second and third panel show the within-party correlations for Democrats and Republicans, respectively.

F Theory Appendix

This section studies the existence and uniqueness of an equilibrium in the framework estimated in this paper where two candidates compete in a two-dimensional space. I remind the reader that such an equilibrium, in each congressional district, is defined by a collection of positions $(\mathbf{x_D}, \mathbf{x_R})$ and vote shares (s_D, s_R) such that (i) $\mathbf{x_D}$ maximizes $\Pi_j(\mathbf{x_D}, \mathbf{x_R})$ when $\mathbf{x_R}$ is fixed, (ii) $\mathbf{x_R}$ maximizes $\Pi_j(\mathbf{x_D}, \mathbf{x_R})$ when $\mathbf{x_D}$ is fixed, and (iii) $s_j = \int s_{ijt}(\mathbf{x_D}, \mathbf{x_R}; \mathbf{w_i})\, dF_t(\mathbf{w_i})$ for $j \in \{D, R\}$, where Π_j are candidates' objective functions defined in section 2, omitting the election subscripts t for clarity.

As explained in section 5, I assume that there is a shock to candidates' probability of wining that follows a uniform distribution $\zeta_{jt} \sim U[-\frac{1}{\phi}, \frac{1}{\phi}]$, I therefore re-write the probability of winning as:

$$\begin{aligned} P_j(\mathbf{x_j}, \mathbf{x_{-j}}) &= \mathbb{P}(s_j(\mathbf{x_j}, \mathbf{x_{-j}}) + \zeta_{jt} \geq 0.5) \\ &= 1 - \mathbb{P}(\zeta_{jt} \leq 0.5 - s_j(\mathbf{x_j}, \mathbf{x_{-j}})) \\ &= \frac{\phi}{2} s_j(\mathbf{x_j}, \mathbf{x_{-j}}) + \frac{1}{2} - \frac{\phi}{4} \end{aligned}$$

which allows the express candidates' objective as a function of their vote share.

- **Existence:**

  I follow Ozdaglar (2013) to demonstrate the existence of the equilibrium. I denote by $S := [-10, 10] \times [-10, 10]$ candidates' strategy space, which is convex and compact. Π_j is continuous in candidate $-j$'s strategy. In order to apply Theorem 2 of *op. cit.*, I need to show that Π_j is concave in each candidate's strategy. This is equivalent to showing that the following inequalities hold:

$$\begin{aligned} \frac{\partial^2 s_j(\mathbf{x})}{\partial x_{1,D}^2} &\leq 2\lambda_D \\ \frac{\partial^2 s_j(\mathbf{x})}{\partial x_{2,D}^2} &\leq 2\lambda_D \\ \frac{\partial^2 s_j(\mathbf{x})}{\partial x_{1,R}^2} &\leq 2\lambda_R \\ \frac{\partial^2 s_j(\mathbf{x})}{\partial x_{2,R}^2} &\leq 2\lambda_R \end{aligned}$$

  where λ_j has been re-normalized by $\frac{2\lambda_j}{\phi Q}$. The threshold for which these conditions are true depends on the exact structure of voter preferences. Using the demand-side and supply-side parameters estimates in the paper, whose estimation does not rely on the existence of an equilibrium, I get that

$$\min(\widehat{\lambda}_j) = 0.025 \quad \text{and} \quad \max(\frac{\widehat{\partial^2 s_j}(\mathbf{x})}{\partial x_{k,j}^2}) = 0.001,$$

for $j \in \{D, R\}$ and $k \in \{\text{cultural}, \text{economic}\}$, across all elections.

The previous inqualities hold, then implying that Π_j is concave in each candidate's strategy which concludes the proof of the existence of the equilibrium.

- **Uniqueness:**

  I follow Rosen (1965) and Ozdaglar (2013) to study the uniqueness of the equilibrium. Rosen shows that if the objective functions are diagonally strictly concave then the game has a unique pure-strategy Nash Equilibrium. Rosen also shows that if the symmetric matrix $U(\mathbf{x}) + U^T(\mathbf{x})$ is negative definite for all $x \in S$, then the objective functions are diagonally strictly concave, where $U(\mathbf{x})$ is the Jacobian of the gradient vector of the objective functions.

  In the framework of this paper, the gradient of the objective functions can be written as:

$$\begin{aligned}
\nabla\Pi(\mathbf{x}) &= \big(\nabla\Pi_D(\mathbf{x}), \nabla\Pi_R(\mathbf{x})\big)^T \\
&= \big(\frac{\partial\Pi_D(\mathbf{x})}{\partial x_{1,D}}, \frac{\partial\Pi_D(\mathbf{x})}{\partial x_{2,D}}, \frac{\partial\Pi_R(\mathbf{x})}{\partial x_{1,R}}, \frac{\partial\Pi_{re}(\mathbf{x})}{\partial x_{2,R}}\big)^T \\
&= \big(\frac{\partial s_D(\mathbf{x})}{\partial x_{1,D}} - 2\lambda_D(x_{1,D} - N_{1,D}), \frac{\partial s_D(\mathbf{x})}{\partial x_{2,D}} - 2\lambda_D(x_{2,D} - N_{2,D}), \\
&\quad -\frac{\partial s_D(\mathbf{x})}{\partial x_{1,R}} - 2\lambda_R(x_{1,R} - N_{1,R}), -\frac{\partial s_D(\mathbf{x})}{\partial x_{2,R}} - 2\lambda_R(x_{2,R} - N_{2,R})\big)^T
\end{aligned}$$

  The Jacobian of the gradient $U(\mathbf{x})$ can be written as:

$$U(\mathbf{x}) = \begin{bmatrix}
\frac{\partial^2\Pi_D(\mathbf{x})}{\partial x_{1,D}^2} & \frac{\partial^2\Pi_D(\mathbf{x})}{\partial x_{1,D}\partial x_{2,D}} & \frac{\partial^2\Pi_D(\mathbf{x})}{\partial x_{1,D}\partial x_{1,R}} & \frac{\partial^2\Pi_D(\mathbf{x})}{\partial x_{1,D}\partial x_{2,R}} \\
\frac{\partial^2\Pi_D(\mathbf{x})}{\partial x_{1,D}\partial x_{2,D}} & \frac{\partial^2\Pi_D(\mathbf{x})}{\partial^2 x_{2,D}} & \frac{\partial^2\Pi_D(\mathbf{x})}{\partial x_{2,D}\partial x_{1,R}} & \frac{\partial^2\Pi_D(\mathbf{x})}{\partial x_{2,D}\partial x_{2,R}} \\
\frac{\partial^2\Pi_R(\mathbf{x})}{\partial x_{1,D}\partial x_{1,R}} & \frac{\partial^2\Pi_R(\mathbf{x})}{\partial x_{2,D}\partial x_{1,R}} & \frac{\partial^2\Pi_R(\mathbf{x})}{\partial x_{1,R}^2} & \frac{\partial^2\Pi_R(\mathbf{x})}{\partial x_{1,R}\partial x_{2,R}} \\
\frac{\partial^2\Pi_R(\mathbf{x})}{\partial x_{1,D}\partial x_{2,R}} & \frac{\partial^2\Pi_R(\mathbf{x})}{\partial x_{2,D}\partial x_{2,R}} & \frac{\partial^2\Pi_R(\mathbf{x})}{\partial x_{1,R}\partial x_{2,R}} & \frac{\partial^2\Pi_R(\mathbf{x})}{\partial x_{2,R}^2}
\end{bmatrix}$$

  and

$$U(\mathbf{x}) + U^T(\mathbf{x}) = \begin{bmatrix}
2(A - 2\lambda_D) & 2B & 0 & 0 \\
2B & 2(C - 2\lambda_D) & 0 & 0 \\
0 & 0 & -2(A + 2\lambda_R) & -2B \\
0 & 0 & -2B & -2(C + 2\lambda_R)
\end{bmatrix}$$

where

$$\begin{aligned} A &= \frac{\partial^2 s_D(\mathbf{x})}{\partial x_{1,D}^2} \\ B &= \frac{\partial^2 s_D(\mathbf{x})}{\partial x_{1,D}\partial x_{2,D}} \\ C &= \frac{\partial^2 s_D(\mathbf{x})}{\partial x_{2,D}^2} \end{aligned}$$

I denote by Δ_1, Δ_2, Δ_3, and Δ_4 the leading principal minors. $U(\mathbf{x}) + U^T(\mathbf{x})$ is negative definite if $\Delta_1 < 0$, $\Delta_2 > 0$, $\Delta_3 < 0$, and $\Delta_4 > 0$. The leading principal minors can be expressed as:

$$\begin{aligned} \Delta_1 &= A - 2\lambda_D \\ \Delta_2 &= (A - 2\lambda_D)(C - 2\lambda_D) - B^2 \\ \Delta_3 &= -2(A - 2\lambda_D)(C - 2\lambda_D)(A + 2\lambda_R) - (A + 2\lambda_R)\big[(A - 2\lambda_D)(C - 2\lambda_D) - B^2\big] \\ \Delta_4 &= 8(A - 2\lambda_D)(C - 2\lambda_D)(A + 2\lambda_R)(C + 2\lambda_R) + \\ &\quad 2(A - 2\lambda_D)(C - 2\lambda_D)\big[(A + 2\lambda_R)(C + 2\lambda_D) - B^2\big] + \\ &\quad 2(A + 2\lambda_R)(C + 2\lambda_R)\big[(A - 2\lambda_D)(C - 2\lambda_D) - B^2\big] \end{aligned}$$

In the framework estimated in the paper, I get the following minimum and maximum values across all elections:

$$\max(\Delta_1) = -0.0396, \quad \min(\Delta_2) = 0.0015 \ \ , \max(\Delta_3) = -0.0092, \ \text{ and } \ \min(\Delta_4) = 0.0018,$$

which implies that $U(\mathbf{x}) + U^T(\mathbf{x})$ is negative definite, leading to the uniqueness of the equilibrium. It should be noted that the logic behind these conditions is pretty intuitive. Indeed, it imposes that the potential convexity in the vote share is compensated by the concavity of the party discipline component.

G Additional Demand Results

This section first provides an exhaustive list of the moments used for estimation and then explores the robustness of the demand results in four dimensions: (1) Using a simple specification following (Berry, 1994), (2) using an alternative identification strategy, (3) adding additional candidate observable characteristics, and (4) recovering voters' endogenous turnout decision.

G.1 List of Moments used for Demand Estimation

I use both aggregate moments from precinct-level election results and micro-moments from survey data.

The vector of aggregate moments is denoted by $g_A(\theta)$ while the vector of micro moments is denoted $g_M(\theta)$.

I write for each topic $k = \{cultural, economic\}$, for each demographic characteristic $w_{pt} = \{education_{pt}, race_{pt}, education_{pt} \times race_{pt}, age_{pt}\}$

$$g(\theta) = \begin{pmatrix} \left.\begin{array}{c} \mathbb{E}[x_{d(p)kt} \cdot \xi_{pt} | \xi_{g(p)t}, \xi_p] \\ \mathbb{E}[\left(x_{d(p)kt} \cdot w_{pt}\right) \cdot \xi_{pt} | \xi_{g(p)t}, \xi_p] \\ \mathbb{E}[w_{pt} \cdot \xi_{pt} | \xi_{g(p)t}, \xi_p] \\ \mathbb{E}[x^2_{d(p)kt} \cdot \xi_{pt} | \xi_{g(p)t}, \xi_p] \\ \mathbb{E}[\left(x^2_{d(p)kt} \cdot w_{pt}\right) \cdot \xi_{pt} | \xi_{g(p)t}, \xi_p] \end{array}\right\} g_A(\theta) \\ \left.\begin{array}{c} \mathbb{C}[w_{pt}, x_{d(p)kt}] - \widehat{\mathbb{C}}[w_{pt}, x_{d(p)kt}] \\ \mathbb{E}[w_{pt}] - \widehat{\mathbb{E}}[w_{pt}] \end{array}\right\} g_M(\theta) \end{pmatrix},$$

where $\widehat{\mathbb{C}}$ and $\widehat{\mathbb{E}}$ are the empirical counterparts of the observed moments in the survey data, for each value of the parameters θ.

$\mathbb{C}[w_{pt}, x_{d(p)kt}]$ captures the covariance between voters' demographic heterogeneity and the ideology of the candidates they vote for and $\mathbb{E}[w_{pt}]$ the average demographic of voters who chose Democratic candidates.

G.2 log-log specification: Homogeneous voters

Assuming that voters are homogeneous within precincts enables the estimation of a simpler model, following Berry (1994). Specifically, one can write:

One can write:

$$\log(\frac{S_{jpt}}{1 - S_{jpt}}) = \alpha_t + \mathbf{w'_{pt}}\boldsymbol{\alpha}^{\mathbf{w}}_{\boldsymbol{t}} + \sum_k \beta_{kt} x_{d(p)kt} + \mathbf{w'_{pt}}\boldsymbol{\beta}^{\mathbf{w}}_{\boldsymbol{kt}} x_{d(p)kt} + \xi_{pt}, \tag{25}$$

where S_{jpt} denotes observed vote shares. Here, α_t is the average proppensity to vote for Democratic candidates at time t, $\boldsymbol{\alpha}^{\mathbf{w}}\boldsymbol{t}$ represents the effect of precinct-level voter demographics $\mathbf{w'pt}$ on Democratic voting,

	Outcome: ln(s/(1-s))	
	(1)	(2)
	2000-2010	2012-2020
CultDem - CultRep	0.014	0.015
	(0.032)	(0.016)
EconDem - EconRep	-0.011	-0.029
	(0.028)	(0.017)
CultDem - CultRep × Av. Edu	-0.032***	-0.027***
	(0.010)	(0.005)
CultDem - CultRep × Shr. White	0.077	-0.042
	(0.076)	(0.049)
CultDem - CultRep × Av. Edu × Shr. White	-0.050*	-0.042**
	(0.030)	(0.019)
CultDem - CultRep × Av. Age	0.000	0.000
	(0.002)	(0.001)
EconDem - EconRep × Av. Edu	0.004	0.027***
	(0.009)	(0.006)
EconDem - EconRep × Shr. White	0.110	-0.048
	(0.072)	(0.068)
EconDem - EconRep × Av. Edu × Shr. White	0.035	0.000
	(0.031)	(0.022)
EconDem - EconRep × Av. Age	-0.001	0.003**
	(0.002)	(0.001)
Precinct-pair x Year FE	X	X
Precinct FE	X	X
Observations	82,847	124,075

Table A.6: Estimation of Voter Preferences with Homogeneous Voters

Notes: This table shows the coefficient from Equation (25): a regression of candidates' log odds ratio on interactions of precinct-level demographics and candidate positions, by period. Each column controls for precinct fixed effects and precinct-pair by election fixed effects. Standard errors clustered two ways, by congressional district by year, and by precinct, are reported in parentheses.

and $xd(p)kt$ is the difference in candidates' positions on dimension k, $\beta_{kt}^{\mathbf{w}}$ captures voters' heterogeneity along dimension w for dimension k.

Table A.6 reports the coefficients from (Equation (25)) estimated by OLS. Results are quite similar to the coefficients estimate by the BLP specification.

G.3 Alternative identification strategy

As a robustness check, I test an alternative set of fixed effects, using congressional district by election fixed effects instead of precinct pair fixed effects. If I assume that candidates choose their positions based solely on aggregate district-level taste shocks, this approach allows me to accurately identify within-district heterogeneity in ideological preferences. Essentially, this alternative strategy compares how precincts,

	Outcome: ln(s/(1-s))	
	(1)	(2)
	2000-2010	2012-2020
Av. Edu	-0.006	0.153***
	(0.011)	(0.010)
Shr. White	-2.500***	-3.088***
	(0.116)	(0.073)
Av. Edu × Shr. White	0.256***	0.387***
	(0.037)	(0.026)
Av. Age	0.008***	-0.005***
	(0.002)	(0.002)
Observations	82,847	124,075

Table A.7: Coefficients from second step (Equation (25))

Notes: This table shows the estimated coefficients from the second step, with homogeneous voters, regressing the fixed effects on voter demographics. Standard errors clustered two ways, by congressional district by year, and by precinct, are reported in parentheses.

facing the same choice set but differing in demographic composition, voted differently. This strategy would leave the average level of preferences unidentified since it is captured by the district by election fixed effects. Formally, if I re-write the unobserved taste shock as:

$$\xi_{pt} \quad = \quad \xi_p + \xi_{d(p)t} + \widetilde{\Delta\xi_{pt}},$$

where $\xi_{d(p)t} = \frac{1}{P}\sum_{p' \in d(p)} \xi_{pt}$ is the taste shock common to the whole congressional district and $\widetilde{\Delta\xi_{pt}}$ is the precinct-specific deviation in taste shock.

This gives the following moment conditions: $\mathbb{E}[x_{d(p)kt} \cdot \widetilde{\Delta\xi_{pt}}] = 0$, which require that precinct temporary deviations in taste shocks are not correlated with differences in candidate positions.

Note that congressional district by election fixed effects exploit a source of variation that is very different to the variation used with precinct-pair by election fixed effects. With precinct-pair by election fixed effects, I am compare arguably similar precincts that were facing a choice between different candidates. In contrast, with district by election fixed effects, I compare different precincts facing a choice between the same candidates. Table A.9 presents the results with congressional-district-by-election fixed effects, note that the average preferences for ideology gets absorbed by the fixed effects.

	Outcome: ln(s/(1-s))	
	(1)	(2)
	2000-2010	2012-2020
CultDem - CultRep	0.000	0.000
	(.)	(.)
EconDem - EconRep	0.000	0.000
	(.)	(.)
CultDem - CultRep × Av. Edu	-0.008**	-0.067***
	(0.004)	(0.005)
CultDem - CultRep × Shr. White	0.009	0.110***
	(0.062)	(0.036)
CultDem - CultRep × Av. Edu × Shr. White	-0.047***	-0.073***
	(0.012)	(0.012)
CultDem - CultRep × Av. Age	-0.000	0.002***
	(0.001)	(0.001)
EconDem - EconRep × Av. Edu	-0.008*	0.035***
	(0.005)	(0.007)
EconDem - EconRep × Shr. White	0.192***	-0.091
	(0.063)	(0.074)
EconDem - EconRep × Av. Edu × Shr. White	-0.017	-0.011
	(0.015)	(0.015)
EconDem - EconRep × Av. Age	0.001	0.002**
	(0.001)	(0.001)
District x Year FE	X	X
Precinct FE	X	X
Observations	417,475	748,089

Table A.8: Estimation of Voter Preferences with Homogeneous Voters (Equation (25))

Notes: This table shows the coefficient from Equation (25): a regression of candidates' log odds ratio on interactions of precinct-level demographics and candidate positions, by period. Each column controls for precinct fixed effects and district-by-election fixed effects. Standard errors clustered two ways, by congressional district by year, and by precinct, are reported in parentheses.

G.4 Additional candidate observable characteristics

While the identification strategy handles any unobservable voters' taste shocks that would span across the congressional district border, one might wonder whether voter preferences for ideology might capture instead preferences of voters for candidate characteristics that vary would with ideology. This section tests the sensitivity of the demand results to the addition of three candidate observable characteristics: incumbency, gender, and race. I obtain candidate gender using name classification algorithms and candidate race and ethnicity from Bouton et al. (2022), which contain candidate's race and ethnicity for House elections after 2006. All elections were either candidates' gender or race and ethnicity are undetermined are excluded, except elections pre-2006 for which the race variables, defined below, are set to zero. While other dimensions such as age likely play a role as well, they are rarely observed for non-incumbents. For each additional observable characteristics, I recover heterogeneity across demographic variables by estimating the following voter random utility model:

$$\begin{aligned} u_{it} = \quad & \sum_k \beta_{ikt} x_{d(i)kt} + \alpha_{it} + \gamma^1_{it}(Inc_{Dd(i)t} - Inc_{Rd(i)t}) + \gamma^2_{it}(Fem_{Dd(i)t} - Fem_{Rd(i)t}) \\ & + \gamma^3_{it}(NonWhite_{Dd(i)t} - NonWhite_{Rd(i)t}) + \xi_{p(i)t} + \epsilon_{it} \end{aligned} \tag{26}$$

where $Inc_{jd(i)t}$ is a dummy equal to one if candidate j is the incumbent, $(Inc_{Dd(i)t} - Inc_{Rd(i)t})$ is therefore the difference between Democratic and Republican candidates incumbent status, which can take values $\{-1, 0, 1\}$, $Fem_{jd(i)t}$ is a dummy equal to one if candidate j is a women, $NonWhite_{jd(i)t}$) if candidate j is non-white, and all the other parameters as defined in equation (6).

I estimate the specification with homogeneous voters within precincts as in equation 25, both with and without the inclusion of additional controls and their interaction with demographic variables. The coefficients on ideology in specifications (2) and (4), which include the extra controls, are only marginally reduced in magnitude and are not statistically different from those in specifications (1) and (3). There is a significant incumbent advantage, which is lower in the second period than in the first. More-educated and white voters are less responsive to this incumbency advantage. Non-white candidates tend to perform better on average than white candidates, and especially so in precincts that have a larger share of non-white voters.

	2000-2010		2012-2020	
	(1)	(2)	(3)	(4)
CultDem - CultRep	0.013	0.002	0.015	0.021
	(0.032)	(0.027)	(0.016)	(0.018)
EconDem - EconRep	-0.011	0.068**	-0.029	-0.045**
	(0.028)	(0.027)	(0.017)	(0.019)
Female Cand. Dem - Rep		0.016		-0.013
		(0.022)		(0.012)
Incumbent Dem - Rep		0.378***		0.212***
		(0.028)		(0.017)
NonWhite Cand. Dem - Rep		0.069		0.047***
		(0.049)		(0.017)
CultDem - CultRep × Av. Edu	-0.032***	-0.030***	-0.027***	-0.015**
	(0.010)	(0.009)	(0.005)	(0.006)
CultDem - CultRep × Shr. White	0.077	-0.033	-0.042	0.007
	(0.076)	(0.094)	(0.049)	(0.067)
CultDem - CultRep × Av. Edu × Shr. White	-0.050*	-0.026	-0.042**	-0.041
	(0.030)	(0.035)	(0.019)	(0.025)
CultDem - CultRep × Av. Age	0.000	-0.000	0.000	-0.001
	(0.002)	(0.002)	(0.001)	(0.001)
EconDem - EconRep × Av. Edu	0.004	-0.006	0.027***	0.022***
	(0.009)	(0.008)	(0.006)	(0.008)
EconDem - EconRep × Shr. White	0.110	0.080	-0.048	-0.002
	(0.072)	(0.097)	(0.068)	(0.084)
EconDem - EconRep × Av. Edu × Shr. White	0.035	-0.020	0.000	-0.024
	(0.031)	(0.038)	(0.022)	(0.027)
EconDem - EconRep × Av. Age	-0.001	0.003	0.003**	0.001
	(0.002)	(0.002)	(0.001)	(0.001)
Female Cand. Dem - Rep × Av. Edu		-0.010		0.009*
		(0.009)		(0.005)
Female Cand. Dem - Rep × Shr. White		-0.048		-0.122*
		(0.085)		(0.065)
Female Cand. Dem - Rep × Av. Edu × Shr. White		0.067*		-0.011
		(0.039)		(0.021)
Female Cand. Dem - Rep × Av. Age		-0.002		0.000
		(0.002)		(0.001)
Incumbent Dem - Rep × Av. Edu		-0.007		-0.003
		(0.007)		(0.005)
Incumbent Dem - Rep × Shr. White		-0.404***		-0.207***
		(0.121)		(0.068)
Incumbent Dem - Rep × Av. Edu × Shr. White		-0.031		0.002
		(0.032)		(0.021)
Incumbent Dem - Rep × Av. Age		0.005***		0.001
		(0.002)		(0.001)
NonWhite Cand. Dem - Rep × Shr. White		-0.362***		-0.205***
		(0.138)		(0.058)
NonWhite Cand. Dem - Rep × Av. Edu × Shr. White		-0.017		0.028
		(0.056)		(0.024)
NonWhite Cand. Dem - Rep × Av. Age		0.003		-0.001
		(0.003)		(0.001)
Precinct-pair x Year FE	X	X	X	X
Precinct FE	X	X	X	X
Observations	82,847	62,814	124,075	88,948

Table A.9: Robustness to the Inclusion of Candidate Observable Characteristics

Notes: This table reproduced table A.6 including additional candidate observable characteristics. Specification (1) and (3) are identical to the specifications in the main paper.

G.5 Voters' turnout decision

While the model estimated in the main paper abstracts from voters' turnout decisions, candidates' positions are likely to influence whether voters choose to vote or not, making it an interesting margin to study. The model presented in Section 4 can be easily generalized to a model with turnout decision, where voter's utility to vote for candidate j is written as:

$$u_{ijt} = \sum_k \beta_{ikt} x_{jd(i)kt} + \alpha_{iDt} \mathbb{1}_{j=D} + \alpha_{iRt} \mathbb{1}_{j=R} + \omega_{it} + \xi_{jp(i)t} + \epsilon_{ijt}, \tag{27}$$

where $x_{jd(i)kt}$ is the position of candidate $j = \{0, D, R\}$ on topic k in congressional district $d(i)$ in election t. The parameter β_{ikt} captures the preferences of voter i on dimension k. Further, α_{iDt} (resp. α_{iRt}) is voter i's utility when voting for a Democratic (resp. Republican) candidate independently of their positions, ω_{it} is voter's utility of turning out, independently of the candidate chosen, $\xi_{jp(i)t}$ is a precinct-level taste shock for candidate j in election t, and ϵ_{ijt} is an individual-level taste shock in favor of candidate j in election t, which I assume follows a type-I extreme value distribution. I normalize the utility of not voting to $u_{i0t} = \epsilon_{i0t}$, which gives the following specification, assuming voters are homogeneous within precincts:

$$\ln(\frac{\widetilde{S_{jpt}}}{1 - S_{0pt}}) = \mathbf{w'_{pt}} \mathbb{1}_{j=D} \boldsymbol{\alpha}^{\mathbf{w}}_{\boldsymbol{Dt}} + \mathbf{w'_{pt}} \mathbb{1}_{j=R} \boldsymbol{\alpha}^{\mathbf{w}}_{\boldsymbol{Rt}} + \sum_k \beta_{kt} x_{jd(p)kt} + \mathbf{w'_{pt}} \boldsymbol{\beta}^{\mathbf{w}}_{\boldsymbol{kt}} x_{jd(p)kt} + \xi_{jpt}, \tag{28}$$

where $\widetilde{S_{jpt}}$ denote observed overall vote shares and $1 - S_{0pt}$ the turnout in precinct p at election t.

I estimate equation (28) using the same identification strategy as in the main paper, with precinct-pair fixed effects. Since there are two candidates, I include precinct-pair by party by election fixed effects, as well as precinct by party fixed effects. The corresponding estimates are reported in Table A.10. The estimates on the effect of ideology are smaller but lead to the same conclusions as the specification without turnout: educated voters prefer more progressive cultural policies but more conservative economic policies, especially in the second period. The gradients are generally larger for white voters. The coefficients on the demographic variables indicate voters' average utility of voting, independent of candidates' party. In both periods, educated, white, and older voters have a higher turnout.

	Outcome: $\ln(\widetilde{S_j/(1-S_0)})$	
	(1)	(2)
	2000-2010	2012-2020
Cultural Position	-0.012	0.019
	(0.025)	(0.019)
Economic Position	0.023	-0.024
	(0.025)	(0.021)
Dem=1 × Av. Edu	-0.024	-0.024
	(0.021)	(0.020)
Dem=1 × Shr. White	-0.461**	-1.104***
	(0.205)	(0.231)
Dem=1 × Av. Edu × Shr. White	0.048	-0.142**
	(0.069)	(0.067)
Dem=1 × Av. Age	-0.003	-0.006
	(0.006)	(0.004)
Av. Edu	0.013	0.025*
	(0.016)	(0.014)
Shr. White	0.640***	1.797***
	(0.180)	(0.275)
Av. Edu × Shr. White	0.028	0.123***
	(0.051)	(0.044)
Av. Age	0.028***	0.010***
	(0.005)	(0.003)
Cultural Position × Av. Edu	-0.026***	-0.026***
	(0.009)	(0.009)
Cultural Position × Shr. White	0.006	0.044
	(0.081)	(0.058)
Cultural Position × Av. Edu × Shr. White	0.006	-0.042
	(0.035)	(0.030)
Cultural Position × Av. Age	-0.001	-0.006***
	(0.003)	(0.002)
Economic Position × Av. Edu	0.010	0.026**
	(0.011)	(0.010)
Economic Position × Shr. White	0.012	-0.211**
	(0.090)	(0.087)
Economic Position × Av. Edu × Shr. White	0.009	0.001
	(0.038)	(0.039)
Economic Position × Av. Age	-0.001	0.006**
	(0.003)	(0.002)
Precinct-pair x Party x Year FE	X	X
Precinct x Party FE	X	X
Observations	216,078	301,630

Table A.10: Robustness to the Estimation of Turnout Decision.

Notes: This table reproduced table presents the estimated coefficients from equation 28.

H Alternative supply specification

While Section 5 has estimated a supply model where candidates maximize their vote shares while complying with party discipline, this section explores the robustness of this section to having instead candidates maximizing their probability of winning the election.

I write each candidate's objective function as Π_{jt}, which depends on the candidate's share of votes, the distance between their chosen position and the party leadership.

$$\Pi_{jt}(\mathbf{x_{jt}}) = \underbrace{P_j(\mathbf{x_{jt}}, \mathbf{x_{-jt}})}_{\text{probability of winning}} - \underbrace{\lambda_{jt}||\mathbf{x_{jt}} - \mathbf{N_{jt}}||^2}_{\text{distance from national party platform}} + \eta_{jt} \tag{29}$$

where the parameter λ_{jt} captures candidates' cost of deviating from the party line.

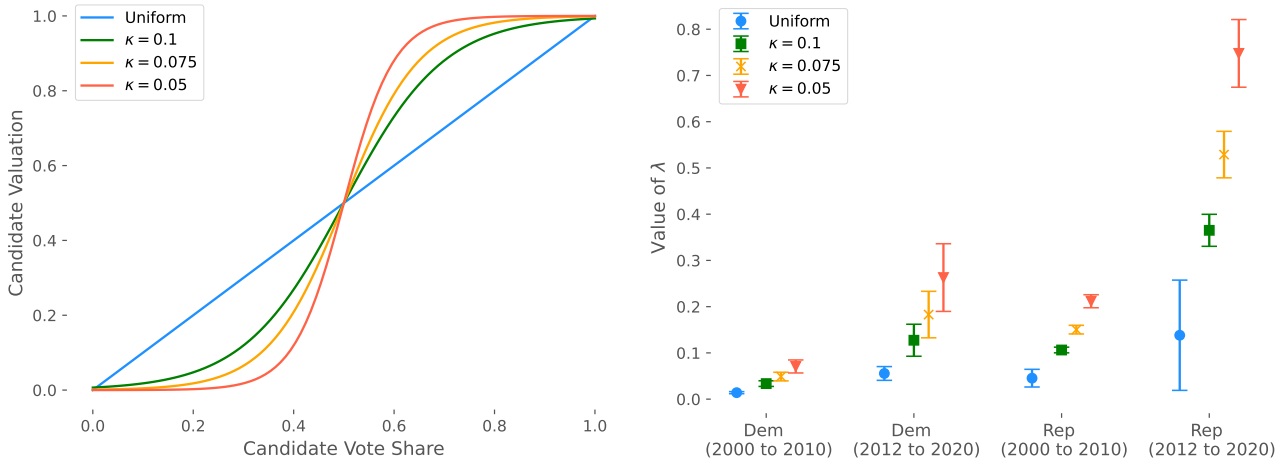
There is an aggregate shock ζ_j that creates some uncertainty around candidates' probability of winning.

If, instead of a uniform distribution as used in the main paper, ζ follows a Logistic distribution, which allows the probability of winning to be a non-linear function of the vote share, the probability of winning can be re-written as:

$$\begin{aligned} P_j(\mathbf{x_{jt}}, \mathbf{x_{-jt}}) &= \mathbb{P}(s_j(\mathbf{x_{jt}}, \mathbf{x_{-jt}}) + \zeta_j \geq 0.5) \\ &= 1 - \mathbf{P}(\zeta_j \leq 0.5 - s_j(\mathbf{x_{jt}}, \mathbf{x_{-jt}})) \\ &= \frac{1}{1 + \exp\big(-\frac{s_j - 0.5}{\kappa}\big)} \end{aligned}$$

where κ is the scale of the Logistic distribution.

Figure A.23 reports different measures of the estimate of party discipline λ for different values of the scale of the Logistic distribution (κ). The findings are quantitatively similar for each value of κ and the ratio between λ for a given value of κ is almost identical. Overall, lower values of κ lead to larger estimates of party discipline since, for most candidates, the slope of candidates' valuation becomes steeper.



(a) Candidate valuation for different κ

(b) Estimated λ

Figure A.23: Estimated λ for different values of κ.

Notes: The Figure shows estimated levels of party discipline when relaxing the linearity assumption of the candidate valuation of their vote share. I re-estimate λ where the voting aggregate shock follows a Logistic distribution instead of a uniform distribution. The Figure reports different estimates of λ for different scale parameters of the Logistic distribution κ with location 0.5. Panel (a) shows different shape of their valuation for different values of κ. Panel (b) shows the corresponding estimates of party discipline.

I Details on Realignment Counterfactuals

This section details the strategy to measure the contributions of each demand-side and supply-side factors to the change in equilibrium outcomes: candidate positioning and voting behavior. Each counterfactual scenario is characterized by a set of parameters that give a unique equilibrium defined by $(\mathbf{x}^*, \mathbf{s}^*) = f\big(\mathbf{m_t}, F_t(\mathbf{w}; p), \boldsymbol{\beta_t}, \mathbf{N_t}, \boldsymbol{\lambda_t} \big| \boldsymbol{\alpha_t}, \boldsymbol{\xi_t}, \boldsymbol{\eta_t}\big)$ where $\mathbf{x}^*, \mathbf{s}^*$ are vectors of candidate positions and vote probabilities, $\mathbf{m_t} = (m_{p,d})_{\forall p,d}$ is a redistricting scenario, that is a map of precincts to districts, $F_t(\mathbf{w}; p)$ is the distribution of demographics in each precinct p, $\boldsymbol{\beta_t}$ captures voters' ideological preferences, $\mathbf{N_t}$ denotes national party leaders' positions, $\boldsymbol{\lambda_t}$ measures the strength of party discipline, $\boldsymbol{\alpha_t}$ accounts for voters' partisan preferences, $\boldsymbol{\xi_t}$ is a vector of precinct-level taste shocks, and $\boldsymbol{\eta_t}$ is a vector of candidate-level ideological shocks.

I consider five factors separately: (a) Changes in the demographic distribution ($F_t(\mathbf{w})$) and the district map ($\mathbf{m_t}$), (b) changes in voter preferences ($\boldsymbol{\beta_t}$), (c) changes in the party leadership positions ($\mathbf{N_t}$), with (c1) changes of cultural positions and (c2) changes of economic positions, and (d) changes in the strength of party discipline ($\boldsymbol{\lambda_t}$).

Each scenario is defined by a combination of factors, evaluated either at their initial or final values. With five factors, this results in $2^5 = 32$ counterfactual scenarios. To assess the contribution of each factor, I compare its effect on the equilibrium outcomes, holding all other factors at various permutations of their initial and final values. The results are aggregated across scenarios using a Shapley value decomposition (Shapley et al., 1953), following the approach of Guriev et al. (2023).

For each factor $z \in \{\boldsymbol{\beta}, \mathbf{m}, \mathbf{N_{cult}}, \mathbf{N_{econ}}, \boldsymbol{\lambda}\}$, its contribution to the change in voter i's probability to vote for the Democratic candidate $s_{ij}(\cdot)$ is given by:

$$\phi_i(z) = \frac{1}{n} \sum_{C} \binom{n}{|C|}^{-1} \big(s_{ij}(\cdot, z_{final}) - s_{ij}(\cdot, z_{initial})\big)$$

for each "coalition" C not including z. A coalition C here denotes the factors that are evaluated at their final rather than initial value. $n = 5$. I then average the contributions across individuals, separately for more-educated and less-educated individuals. I apply the same strategy to decompose the change in candidate positions.

J Details on Counterfactual Support for Environmental Policies

Appendix Figure A.19 shows, for each political topic,[21] the relative weighting of the cultural versus economic dimension, defined as the ratio of coefficients from a linear regression of candidate topic-specific positions on their cultural and economic ideal points:

$$x_{j,env} = \gamma_0 + \gamma_{cult} x_{j,cult} + \gamma_{econ} x_{j,econ} + \zeta_j. \tag{30}$$

The relative weighting parameter is defined as $\rho = \frac{\gamma_{cult}}{\gamma_{cult}+\gamma_{econ}}$. Environmental issues stand out in the figure as the topic that is closest to both the economic and cultural dimensions, with a cultural weighting around 0.5. Additionally, there appears to be a significant difference between the two parties, with a much larger cultural weighting for Democratic candidates. In other words, Democratic candidates who are progressive on environmental issues tend to be progressive on cultural issues rather than economic ones. The reverse is true in the Republican Party, Republican candidates who are conservative on environmental issues tend to be conservative on economic issues rather than cultural ones.

These differing weightings between parties suggest that parties may have some flexibility in how they position themselves on a topic, for a given level of progressiveness. To capture this idea of leadership weighting on one topic, I represent the environmental leadership position as a vector in an (economic, cultural) policy space with the environmental position as the norm and the weighting parameter ρ as the angle. The strength of the cultural and economic dimensions in the leadership position is then obtained by getting the polar coordinates of the vector:

$$N_l^{cult} = |N_l^{environment}| \cdot \sin \rho_l \tag{31}$$

$$N_l^{econ} = |N_l^{environment}| \cdot \cos \rho_l, \tag{32}$$

with $l \in \{D, R\}$.

There are multiple interpretations to ρ_l. The first is that there are various policy tools available to achieve the same goal (e.g., reducing carbon emissions), and politicians can choose a mix of economic or cultural policies to offer. The second interpretation is a framing one: to motivate a policy, politicians can appeal either to values or to material conditions. Since voters have heterogeneous preferences regarding values and material aspects, different framing strategies lead to different coalitions of support (Enke, 2020; Besley, 2023; Besley and Persson, 2023; Chong and Druckman, 2007).

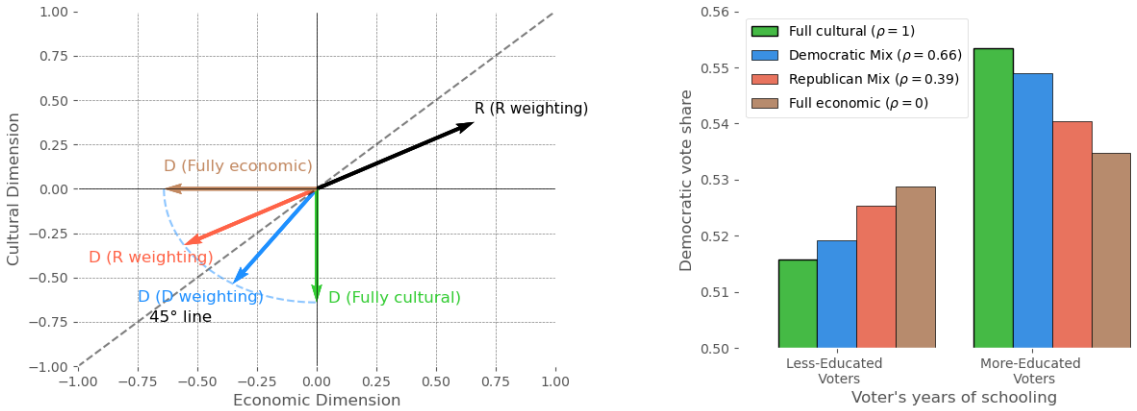
Re-projecting the environmental positions in an (economic,cultural) policy space allows for the evaluation of voters' support for different ρ_l, using the demand estimates from section 4. For example, it is possible to evaluate a counterfactual scenario where Democratic environmental positions remain as left-

[21] Since each topic is originally part of either the economic or cultural dimension, I re-estimate separately an ideal point on each topic and on each dimension, excluding that topic.

wing as they currently are but adopt the same economic weighting as those of the Republican leadership, assessing how this shift would influence voter support.

Previous sections have shown that voters who support progressive cultural policies do not necessarily support progressive economic policies, and vice versa. As a result, for an equally left-wing environmental policy, the type of voters who will back different policies depends on their cultural versus economic weighting. For each counterfactual leadership parameter ρ, I derive new optimal candidate positions using the supply estimates from Section 5, and projected vote shares among different voter groups, assuming the environment is the only issue on which candidates are campaigning.[22]

Panel (a) in Figure 11 presents the various counterfactual scenarios. The first scenario (D weighting) uses the current Democratic leadership weighting and positions. The second scenario applies the current Republican weighting ($\rho = 0.39$) to the Democratic leadership. The remaining scenarios explore leadership positions that are either fully cultural or fully economic .I derive equilibrium candidate positions for both Democratic and Republican candidates across various scenarios, with the Republican leadership position remaining fixed throughout. Although this assumption is an oversimplification, modeling the endogenous selection of leadership positions is beyond the scope of this paper. Panel (b) corresponds to Figure 1 in section 1.



(a) Leadership Counterfactual Environmental Positions

(b) Counterfactual Vote Shares

Figure A.24: Economically Oriented Progressive Environmental Positions Attract More Support from Low-Education Voters

Notes: For each cultural-economic mix, I compute counterfactual levels of support for a Democratic candidate who would campaign only on the environment, using their current position but varying the framing towards economic or cultural. I use the current weighting of Democrats and Republicans as well as a full cultural and a full economic framing. For example, for a Democratic candidate campaigning only on the environment with a leadership weighting of $\rho = 0.66$ (current Democratic mix) would get 51.5% support from less-educated voters and 55% support from more-educated voters.

[22]Studying the environment in isolation is not meant to suggest that it will become the dominant issue but rather to understand the type of voter coalitions that would support different environmental policies. Measuring the relative importance that voters assign to the environment versus other issues is left for future research.



(a) Democratic Sample



(b) Republican Sample

Figure A.25: Example of Website Pages on the Environment

Notes: Sample Democratic and Republican websites corresponding close to the party centroid in the embedding space. I use the 24 embedding dimensions selected by the XG Boost algorithm as predictive of candidates' environmental positions.



Rishi Kumar Favors Climate Action via the Green New Deal

The Green New Deal can address climate change, create jobs and a vibrant economy

Climate change is the single biggest national security threat for the United States and the single biggest threat to worldwide industrialized civilization. We are already seeing the impact climate change is having here in California. Wildfires devastate our communities, leaving families in hardships and ruin.

I think about the future and the state we will leave this world in. It worries me. As our world continues to change, we need to create a system that prepares today's generation and future generations to come. We are tasking these young individuals to solve imperative climate challenges that the world has never witnessed before.

We must act now to combat the effects of climate change in order to give the next generation a chance to succeed. That is why I am a strong supporter of the Green New Deal.

We have a chance to grow our economy and create thousands of new jobs by harnessing the power of green and alternative energy. In our home district, green tech innovators are at the vanguard of new technologies that will help us reduce our reliance on fossil fuels and lead the world in clean energy production.

The United States Congress must play a critical role in supporting these industries and maintain our competitive edge in the global economy, ensuring that windmill manufacturers, solar panel makers, and biofuel producers have the support they need to grow. We have to wean the United States from fossil fuels and curb planet-warming greenhouse gas emissions across our planet and in the workplace. Here is how the Green New Deal enables that:

Figure A.26: Example of Democratic Website Page on the Environment with high economic component

Notes: Sample Economic Democratic website page. The example is obtained by selecting websites with strong embedding predictions in both the top 24 environmental and economic dimensions.



Figure A.27: Example of Republican Website Page on the Environment with high cultural component

Notes: Sample Economic Republican website page. The example is obtained by selecting websites with strong embedding predictions in both the top 24 environmental and cultural dimensions.

K Alternative candidate ideology model

I estimate an ideal point for each candidate for each political topic. I adapted the framework of Vafa et al. (2020) who set up an unsupervised topic model (*Text-Based Ideal Point*). The original model consists in estimating one single ideal point for each candidate across topics. I estimate the ideal point x_{jkt} of each candidate j, in period t on topic k. These ideal points are jointly estimated with the following additional latent variables:

- θ_{dkt}, the per-document d topic k intensity at period t;
- β_{ktv} , the neutral topics k in period t for term v ;
- η_{ktv} , the ideology associated to each term v with respect to topics k in period t.

I place Gamma priors on θ and β and normal priors on η as well as the ideal points x.

All these latent variables interact together to draw the observed count of each term v in document d, authored by $j = a_d$ in period t that I assume follows a Poisson distribution.

$$y_{dv} \sim \text{Pois}\big(\sum_k \theta_{dk}\ \beta_{kv}\ \exp(x_{k,a_d}\eta_{kv})\big)$$

Let's take a politician with an ideal point of $x_{jk} = 0$, meaning that candidate j is completely "neutral" on topic k. To talk about topic k, candidate j will only use words depending on the extent to which they belong to topic k (β_{kv}), independently of their polarization (η_{kv}). However, if candidate j is located on the very right of the political spectrum and has a very positive $x_{j,k}$, they will be more likely to use words which are polarized in the same direction. For instance, on the reproductive rights topic, words "pro-choice" will have a very negative η_{kv} whereas words like "pro-life" will have a very positive η_{kv}, leading right-wing candidates (with positive x_j) to be more likely to use words like "pro-life" and less likely to use words like "pro-choice" to talk about reproductive rights questions.

The four latent variables are estimated by variational inference, fitting an approximate posterior distribution.

I estimate a model with $K = 30$ topics. Out of the 30 topics, 12 have relevant political content. The 18 others are either not directly political (e.g., about contributions, contact information or campaign events) or are not interpretable. I show on Figure A.28 the words associated with the 12 selected topics and the labels I assigned to them. The list of words for the 30 topics is provided in the Appendix.

The central column shows neutral words for each topic (ideal point of 0). The left (resp. right) column shows words with the highest probability for each topic for an ideal point of -1 (resp 1). Note that the signs have been adjusted in order that the average score for Democrats across time is always lower than the average score of Republicans.

Figure A.29 shows the topic proportions for each of the principal topics. The blue dots indicates the average across parties and the bar shows the proportion for the Democratic candidates in particular.



Figure A.28: Neutral and polarized words

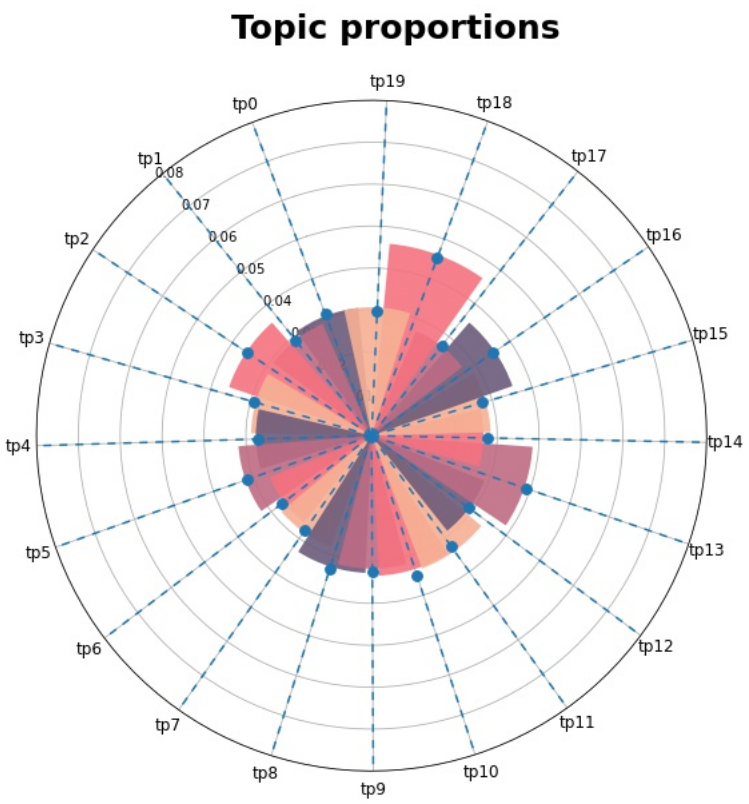


Figure A.29: Topic proportions

Democratic candidates tend to talk more than other candidates about veterans, health care, industry and donations but less about the economy, finance, war and security.

Figures A.30 to A.33 show the distribution of ideal points on different topics. Interestingly, ideal points are not all on a straight line, meaning that candidates differentiate themselves from one topic to another. In other words, the position of the candidate on the economy doesn't pinpoint the position of the candidate on healthcare. The party lines remain, however, clearly defined for most political topics. In particular, reproductive rights, health care, economic topics or education have a very distinct ideal points distribution for the Democratic Party than for the Republican party. That being said, there is no single topic where the most conservative Democrat is less conservative than the most liberal Republican. In other words, on all topics, the distribution of both parties overlap.

Figure A.34 shows the correlation between the ideal points obtained from TBIP and from model developed in the main paper, using both candidate survey and candidate website to recover their ideology. The correlation is 0.45 for the cultural topic and 0.49 for the economic topic.

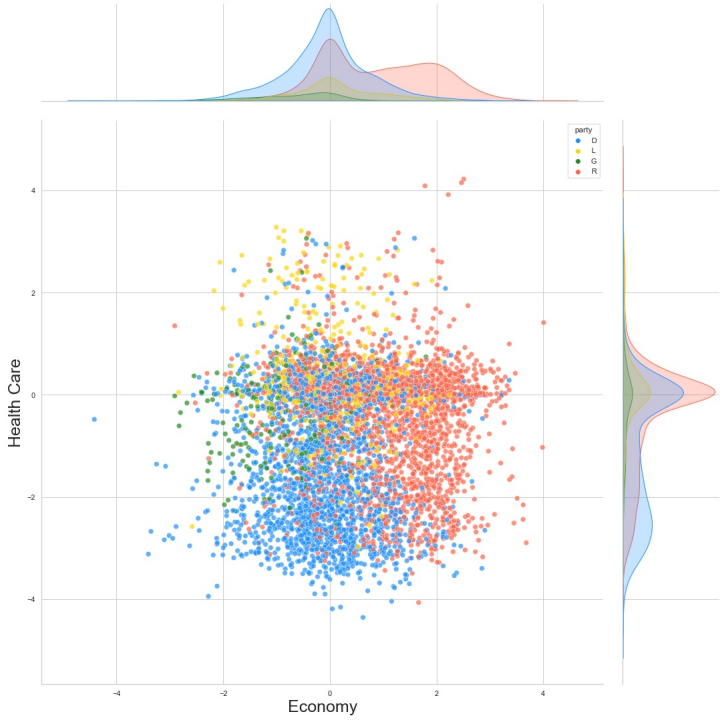


Figure A.30: Ideal point distributions (economy and healthcare)



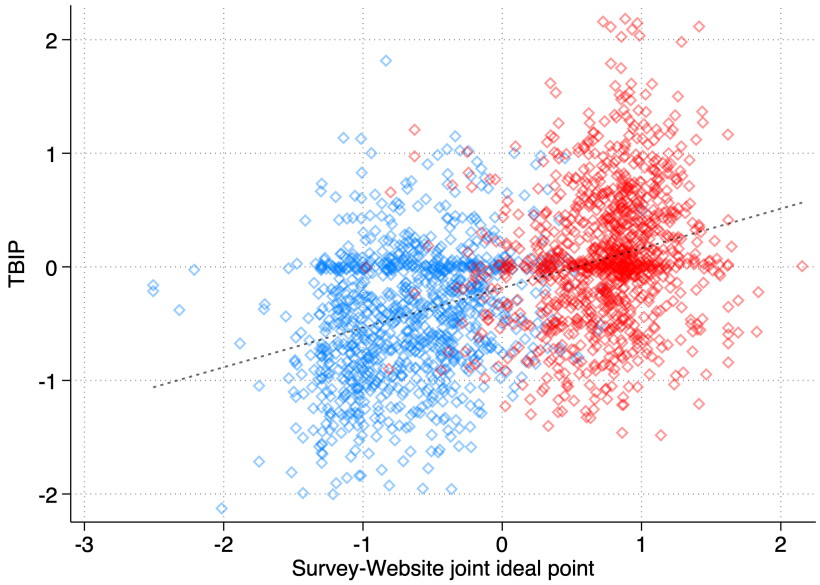
Figure A.31: Ideal point distributions (education and abortion)



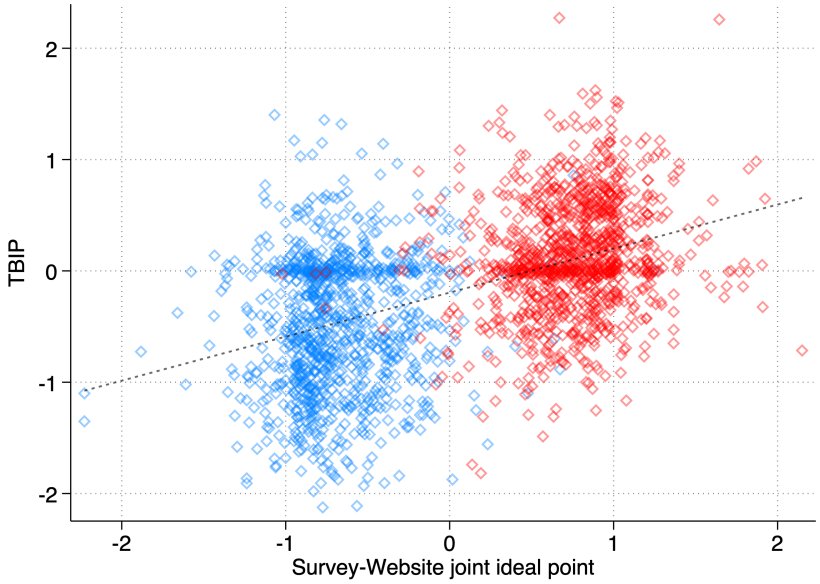
Figure A.32: Ideal point distributions (war and veterans)



Figure A.33: Ideal point distributions (industry and energy)



(a) Cultural topic ($\rho = 0.47$)



(b) Economic topic ($\rho = 0.49$)

Figure A.34: Correlation between survey-website joint ideal point and TBIP