Pitfalls of Demographic Forecasts of US Elections

Richard Calvo, *University of California Berkeley*
Vincent Pons, *Harvard University, CEPR, and NBER*
Jesse M. Shapiro, *Harvard University and NBER*[*]

September 2024

Abstract

Many observers have forecast large partisan shifts in the US electorate based on demographic trends. Such forecasts are appealing because demographic trends are often predictable even over long horizons. We backtest demographic forecasts using data on US elections since 1952. We envision a forecaster who fits a model using data from a given election and uses that model, in tandem with a projection of demographic trends, to predict future elections. Even a forecaster with perfect knowledge of future demographic trends would have performed poorly over this period—worse even than one who simply guesses that each election will have a 50-50 partisan split. Enriching the set of demographics available does not change this conclusion. We discuss both mechanical and economic reasons for this finding, and show suggestive evidence that parties adjust their platforms in accordance with changes in the electorate.

JEL Codes: D72, J11, C53, P00
Keywords: election forecasting, party platforms, electoral competition

[*]We thank our dedicated research assistants for their contributions to this project. We thank Isaiah Andrews, Gemma Dipoppa, Ray Fair, and Ed Glaeser for helpful comments. Vincent Pons is the cofounder of a company in Europe, eXplain. The paper includes the researchers' own analyses based in part on data from the American National Election Studies. American National Election Studies and the relevant funding agency/agencies bear no responsibility for use of the data or for interpretations or inferences based upon such uses. This paper uses the Census Bureau Data API but is not endorsed or certified by the Census Bureau.
E-mail: richard_calvo@berkeley.edu, vpons@hbs.edu, jesse_shapiro@fas.harvard.edu.

1 Introduction

Forecasting elections is a popular sport of scholars (e.g., Fair, 2011) and pundits (e.g., Silver, 2012) alike. In addition to its entertainment value, forecasting is valuable because election outcomes matter for public policies (e.g., Brollo and Troiano, 2016; Fiva et al., 2018; Marx et al., 2022), and so predicted election outcomes can influence markets (e.g., Snowberg et al., 2007; Kelly et al., 2016), and political uncertainty can depress them (Julio and Yook, 2012).

One approach consists of forecasting election results based on demographic changes. The appeal of such forecasts comes from the strong correlations between vote choice and demographic characteristics such as race and education observed in cross sections (e.g., Campbell et al., 1960; *Economist*, 2018; Center, 2023) and from our ability to predict long-term demographic trends caused by factors such as aging, migration, or fertility and mortality rates (Petropoulos et al., 2022).[1] However, subsequent events have often defied election forecasts based on demographics. For instance, in the book *The Emerging Republican Majority*, originally published in 1969, Kevin Phillips argued that demography would doom the Democrats.[2] In the 2002 book *The Emerging Democratic Majority*, John Judis and Ruy Teixeira argued precisely the opposite,[3] before wondering *Where Have All the Democrats Gone?* twenty years later (Judis and Teixeira, 2023).

In this paper, we backtest demographic forecasts systematically. In each presidential election year, we use data from nationally representative samples of US voters provided by the American National Election Study (ANES) to fit a binary logit model relating a person's vote to their age, gender, race, income, education, and the type of area in which they live. We use the fitted model to predict individual vote choices in the next election based

[1] For discussions of the methods and accuracy of demographic forecasting, see, e.g., George et al. (2004), Booth (2006), Girosi and King (2008), Hauer (2019), and Baker et al. (2021).

[2] Phillips writes, "Unluckily for the Democrats, their major impetus is centered in stagnant Northern industrial states—and within those states, in old decaying cities, in a Yankee countryside that has fewer people than in 1900, and in the most expensive suburbs. Beyond this, in the South and West, the Democrats dominate only two expanding voting blocs—Latins and Negroes." (2014).

[3] Judis and Teixeira write, "What makes it likely that a Democratic majority will emerge over the next decade? First of all, as a result of the transition to postindustrial society, each of the McGovern constituencies will continue to grow as a percent of the electorate. And barring a sea change in Republican politics these constituencies will continue to vote Democratic. Second of all, as post industrial areas continue to grow, white working-class and professional voters in these areas are likely to converge on a worldview that is more compatible with the Democrats than with Republicans." (2002).

on the demographic attributes of the voters in that election, and we predict future overall vote shares by averaging these individual forecasts. We forecast the overall Republican share of the two-party vote among survey respondents and measure forecasts' accuracy by computing their root mean squared error (RMSE) relative to the truth.

We find that demographic forecasts of US presidential elections are poor. Forecasts of election results up to five elections in advance perform about as well as a naive forecast simply guessing that the result of the next election will be the same as today's result, and worse than predicting that every election will be an even (50-50) contest between Democrats and Republicans.

Our analysis stacks the deck in favor of the forecaster in several ways. First, we assume the forecaster knows the demographics of the survey sample in the next election, akin to perfect foresight of demographic trends. Second, we focus our analysis on the sample of voters, akin to perfect foresight of trends in turnout in different groups. Third, we focus on predicting the election outcome *in the survey sample*, so that the forecaster is not penalized for departures between this and the official result, though we also show that our results are not sensitive to this choice. Our hypothetical forecaster fails at their task despite these many advantages.

We extend our results in several ways, including predicting future election results based on an extended set of demographic covariates, and allowing for a rich set of interactions among demographic characteristics using regression trees. None of these extensions meaningfully improves the performance of demographic forecasts. Because some prior demographic forecasts are based on aggregate data, we also apply our approach to county-level data, predicting voting at the county level using county-level demographics, and forecasting future elections based on trends in these characteristics. If anything, this approach performs even worse than the one based on survey data. Demographic forecasts likewise do not perform well in predicting congressional elections or party identification.

The inaccuracy of *short-term* demographic forecasts is perhaps not surprising: demographic shifts are far too slow to explain the large shifts in vote share observed from one election to the next. But demographic forecasts of *more distant future elections* do not perform better. Rational-choice theory predicts this finding. We show that in a standard Downsian model of electoral competition between two parties that are office-motivated and only weakly committed to a certain ideology, changes in the composition of the electorate man-

ifest as changes in party platforms rather than as changes in their vote shares (Hotelling, 1929; Downs, 1957; Becker, 1958). We test this prediction by combining ANES data on individual vote choices and issue positions with data on party platforms from the Manifesto Project. We find that changes in US parties' ideology and in the stances they take on different issues, from environmental protection to minority rights, have tracked demographic shifts in the electorate, albeit with a lag.

Others have noted that US politics tends to remain competitive despite changes in the electorate, with the *Economist* (2023a) calling this "The great mystery of American politics."[4] We provide what is to our knowledge the first systematic evidence of this pattern, and the first analysis to observe that it is exactly what is predicted by some of the most classic ideas in rational choice theory. In doing this, our paper contributes to a rich literature on the determinants of election results and on election forecasts.

In modern electoral campaigns, voters, political parties, and investors receive a stream of forecasts from polls and from prediction markets (Forsythe et al., 1992; Wolfers and Zitzewitz, 2004). Unfortunately, these predictions tend to remain volatile while campaign news comes out, and tend to be most reliable only shortly before the election (Wlezien and Erikson, 2002; Berg et al., 2008; Erikson and Wlezien, 2012; Jennings et al., 2020). Furthermore, they tell us which candidate is most likely to win but not why.[5]

Much of the change in incumbents' polling numbers during campaigns can be explained by economic fundamentals such as the level of GDP growth (Gelman and King, 1993; Kaplan et al., 2012). Accordingly, social scientists have sought to forecast election results with models calibrated on past elections and using aggregate factors as predictors (e.g., Fair, 1978; Rosenstone, 1983; Lewis-Beck and Rice, 1984; Abramowitz, 1988; Campbell, 1996; Lewis-Beck, 2005; Fair, 2009). Such models would ideally help identify the main forces influencing election outcomes and predict these results with more lead time than polls, possibly even before parties choose their nominees. In practice, these models often include polls or a closely related variable such as incumbent approval ratings, and they tend to require data from the quarters immediately before the election to achieve maximum

[4]The *Economist* writes that, "Even profound changes in what it means to be a Democrat or Republican seem to return the parties to their equilibrium, as though obeying some thermostat." (2023a)

[5]While polls or prediction markets alone may not say much about the forces shaping election results, they can be used as an ingredient to estimate the effect of debates, shocks, or other events (e.g., Snyder Jr and Yousaf, 2020; Le Pennec and Pons, 2023).

accuracy.[6] Furthermore, because estimation of these models treats each election as a single observation, these models can include only a limited number of explanatory variables.

Compared with models based on economic fundamentals, models forecasting elections based on demographics have two important strengths. First, because demographic changes can plausibly be predicted long in advance—much more so than, say, inflation, unemployment, or economic growth—they can be used to forecast election results with considerable anticipation. For instance, in the aforementioned books, Phillips (2014) and Judis and Teixeira (2002) adopt horizons of years and even decades. Second, as our approach demonstrates, it is possible to use one observation per survey respondent in the calibration stage, which allows us to consider a rich set of demographic factors as well as interactions between them.[7]

But demographic forecasts will only be reliable if the relationship between demographic factors and voting behavior is sufficiently stable over time.[8] There are certainly reasons to believe that the correlations between demographics and vote choices observed in a specific election will persist to some extent afterward: previous research shows that people's demographic characteristics strongly influence their social and political identity (Lazarsfeld et al., 1948; Mason, 2016; Mason and Wronski, 2018), which is highly persistent over time (Campbell et al., 1960; Green et al., 2002).[9] On the other hand, major shifts in the partisanship of certain groups do take place. For instance, educated voters have increasingly rallied to the Democratic party since the 1960s (Gethin et al., 2021) while some minority groups have recently started to peel away (*Economist*, 2023b; Times, 2024). On net, the accuracy of demographic forecasts will depend on the speed at which the size of different groups changes and the speed at which parties' response and other forces lead these groups' partisan preferences to change. This race, we find, is won by the second horse.

[6] Recent synthetic models use Bayesian methods to combine data from polls with forecasts based on fundamentals (e.g., Lock and Gelman, 2010; Linzer, 2013; Lewis-Beck and Dassonneville, 2015).

[7] Though see Kim and Zilinsky (2024) on the limits of this approach in predicting individual vote choice.

[8] Beyond elections, there is mounting evidence that demographic trends affect a wide range of outcomes, from health expenditures (De Meijer et al., 2013) to financial inclusion (Sarma and Pais, 2011), the start-up rate (Karahan et al., 2024), and economic growth (Maestas et al., 2023).

[9] Furthermore, the transmission of both demographic characteristics and partisan attachments across generations may contribute to make the correlations between them durable (Jennings and Niemi, 1968; Black et al., 2005; Bengtson et al., 2009; Jennings et al., 2009; Black and Devereux, 2011).

2 A Model of Electoral Competition with Demographics

To help frame the evidence that follows, we develop a model of two-party electoral competition in which voters' behavior depends on their demographic group, and parties may have non-electoral motivations. The model combines elements that are standard in the literature (e.g., Austen-Smith and Banks, 2005, Chapter 7), and the analysis develops their implications for demographic forecasting. In the model, two parties, denoted L and R, compete in an election. Each party $j \in \{L, R\}$ simultaneously chooses (and publicly announces) a platform $x_j \in \mathcal{X} \subseteq \mathbb{R}^K$, for $\mathcal{X}$ a convex, compact policy space with dimension $K \in \mathbb{N}$.

There are G groups of voters, and each group $g \in \{1, ..., G\}$ has $N_g \in \mathbb{N}$ members, so that $N = \sum_{g=1}^{G} N_g$ is the size of the electorate, and $\mathbf{N} = (N_1, ..., N_G) \in \mathbb{N}^G$ is its composition. We may think of a group g as a demographic cell (e.g., college-educated white men in their 40s), but in principle groups may be even finer than that (e.g., individual voters).

A voter's behavior depends on the voter's group. Each voter in a given group g votes for party $j \in \{L, R\}$ with probability $p_g(x_j, x_{-j})$, where $p_g : \mathcal{X}^2 \to [0, 1]$ is a function continuous in its arguments, satisfying $p_g(x_L, x_R) + p_g(x_R, x_L) \leq 1$ for all $x_L, x_R \in \mathcal{X}$.

Each party $j \in \{L, R\}$ is concerned with its electoral prospects, summarized by its expected plurality, expressed as a share of the electorate

$$P_j(x_j, x_{-j}; \mathbf{N}) = \frac{1}{N} \sum_{g=1}^{G} N_g \left[p_g(x_j, x_{-j}) - p_g(x_{-j}, x_j) \right].$$

Each party $j \in \{L, R\}$ is also concerned with its ideological and other commitments, summarized by a party-specific continuous function $b_j : \mathcal{X} \to \left[-\frac{1}{2}, \frac{1}{2}\right]$ of the party's platform that we may think of as reflecting the platform's coherence with those commitments.

The payoff $\pi_j : \mathcal{X}^2 \to \mathbb{R}$ of party j is then given as

$$\pi_j(x_j, x_{-j}; \mathbf{N}) = (1 - \beta_j) P_j(x_j, x_{-j}; \mathbf{N}) + \beta_j b_j(x_j)$$

where $\beta_j \in [0, 1]$ denotes the importance that party $j \in \{L, R\}$ attaches to nonelectoral motives.

We focus on Nash equilibrium in pure strategies. Proposition 4 in the Online Appendix

gives example sufficient conditions for the existence of such an equilibrium.[10]

Standard ideas in the literature imply that when parties are entirely electorally-motivated, the equilibrium value of the expected plurality in the election does not depend on the composition $\mathbf{N}$ of the electorate.

Proposition 1. (Demographic neutrality under electorally-motivated parties.) *Suppose that parties are electorally-motivated in the sense that $\beta_L = \beta_R = 0$. Then for any composition $\mathbf{N} \in \mathbb{N}^G$ and any platforms (x_L^*, x_R^*) that constitute a Nash equilibrium in pure strategies, the expected plurality is zero, $P_R(x_R^*, x_L^*; \mathbf{N}) = 0$.*

Proof. Because $\beta_L = \beta_R = 0$, Lemma 1 in the Appendix implies an immediate contradiction with either $P_R(x_R^*, x_L^*; \mathbf{N}) > 0$ or $P_R(x_R^*, x_L^*; \mathbf{N}) < 0$. □

When parties are instead entirely ideologically motivated, the equilibrium value of the expected plurality $P_R(x_R, x_L; \mathbf{N})$ in the election depends strongly on the composition $\mathbf{N}$ of the electorate, in the sense that knowing how each group g votes in an election with composition $\mathbf{N}'$ is sufficient to forecast the change in the expected plurality if the composition changes to some $\mathbf{N}''$.

Proposition 2. (Demographic determinism under ideologically-motivated parties.) *Suppose that parties are ideologically-motivated in the sense that $\beta_L = \beta_R = 1$. Then for any platforms (x_L^*, x_R^*) such that $x_j^* \in \arg\max_{x \in \mathcal{X}} b_j(x)$ and any $\mathbf{N}', \mathbf{N}'' \in \mathbb{N}^G$, the platforms (x_L^*, x_R^*) constitute a Nash equilibrium in pure strategies, and the expected pluralities $P_R' = P_R(x_R^*, x_L^*; \mathbf{N}')$ and $P_R'' = P_R(x_R^*, x_L^*; \mathbf{N}'')$ obey*

$$P_R' - P_R'' = \sum_{g=1}^{G} \left(\frac{N_g'}{N'} - \frac{N_g''}{N''} \right) \left[p_g(x_R^*, x_L^*) - p_g(x_L^*, x_R^*) \right].$$

Proof. The result follows immediately from the definition of Nash equilibrium and of the expected plurality. □

More generally, when parties have both electoral and ideological motivations, the extent to which the composition $\mathbf{N}$ influences the equilibrium plurality is limited by the strength of nonelectoral motives.

[10] Sufficient conditions for the existence of a Nash equilibrium in pure strategies in the case where $\beta_L = \beta_R = 0$ can be found in, for example, Austen-Smith and Banks (2005, see Theorems 7.9 and 7.10).

Proposition 3. (Non-electoral motivations bound electoral effects of demographics.) *Suppose that* $(x_L^*, x_R^*) \in \mathcal{X}^2$ *and* $(x_L^{**}, x_R^{**}) \in \mathcal{X}^2$ *constitute Nash equilibria in pure strategies under compositions* $\mathbf{N}' \in \mathbb{N}^G$ *and* $\mathbf{N}'' \in \mathbb{N}^G$*, respectively with expected pluralities* $P_R' = P_R(x_R^*, x_L^*; \mathbf{N}')$ *and* $P_R'' = P_R(x_R^{**}, x_L^{**}; \mathbf{N}'')$*. Then if* $\beta_L, \beta_R \in (0,1)$*, the absolute difference in the expected plurality under the two equilibria is bounded by an increasing function of* β_L, β_R*,*

$$\left|P_R' - P_R''\right| \leq \left(\frac{\beta_L}{1-\beta_L} + \frac{\beta_R}{1-\beta_R}\right).$$

Moreover, if $P_R', P_R'' \geq 0$ *or* $P_R', P_R'' \leq 0$ *then the bound is tighter, respectively* $\left|P_R' - P_R''\right| \leq \frac{\beta_L}{1-\beta_L}$ *or* $\left|P_R' - P_R''\right| \leq \frac{\beta_R}{1-\beta_R}$.

Proof. Lemma 1 in the Appendix implies that $(1-\beta_L)\left|P_R'\right| \leq \beta_L\left(b_L(x_L^*) - b_L(x_R^*)\right) \leq \beta_L$ if $P_R' > 0$ and $(1-\beta_R)\left|P_R'\right| \leq \beta_R\left(b_R(x_R^*) - b_R(x_L^*)\right) \leq \beta_R$ if $P_R' < 0$, and likewise for P_R''. The desired result then follows from the fact that $\beta_L, \beta_R \in (0,1)$, the triangle inequality, and the definition of the absolute value. □

Example 1. Suppose that $K = 1$, that $p_g(x_j, x_{-j}) = \frac{1}{2} + \frac{1}{8}\left[(x_{-j} - \tilde{x}_g)^2 - (x_j - \tilde{x}_g)^2\right]$, and that $b_j(x_j) = \frac{1}{2} - \frac{1}{8}(x_j - \tilde{x}_j)^2$ for $\tilde{x}_g, \tilde{x}_j \in \mathcal{X} = [-1,1]$ group- and party-specific bliss points.[11] Then in any interior equilibrium we have that

$$x_j^* = (1-\beta_j)\,\tilde{x}(\mathbf{N}) + \beta_j \tilde{x}_j,$$

where $\tilde{x}(\mathbf{N}) = \frac{1}{N}\sum_{g=1}^G N_g \tilde{x}_g$ is the average voter's bliss point. We also have that

$$P_R(x_R^*, x_L^*; \mathbf{N}) = \frac{1}{2}(x_R^* - x_L^*)\left[\tilde{x}(\mathbf{N}) - \frac{1}{2}(x_L^* + x_R^*)\right].$$

Then in the special case with $\tilde{x}_R = -\tilde{x}_L = 1$ and $\beta_L = \beta_R = \beta \in [0,1]$, we have that

$$P_R(x_R^*, x_L^*; \mathbf{N}) = \beta^2 \tilde{x}(\mathbf{N}).$$

Intuitively, the plurality depends on the average voter's bliss point $\tilde{x}(\mathbf{N})$ to the extent that

[11] As a microfoundation we may imagine that each voter in group g has expressive utility $u_g(x) = -\frac{1}{8}(x_k - \tilde{x}_g)^2$ from voting for a party with platform $x \in \mathcal{X}$ and an idiosyncratic utility from voting for party R distributed uniformly on $\left[-\frac{1}{2}, \frac{1}{2}\right]$, and that each voter votes for party L if and only if the expressive utility for party L exceeds that for party R by more than the idiosyncratic utility.

the parties are willing to sacrifice votes for ideological or other reasons.

Proposition 3 establishes that the expected plurality is insensitive to demographic composition when nonelectoral motives are weak. Because Proposition 3 establishes only an upper bound on the sensitivity to demographic composition, it allows the expected plurality to be insensitive to demographic composition even when nonelectoral motives are strong. The following example illustrates just such a situation.

Example 2. Continue the setting of Example 1, but now suppose that $b_j(x_R) = \frac{1}{2} + \frac{1}{2}b_j x_j$ for b_j a constant. Then in any interior equilibrium we have that

$$x_j^* = \tilde{x}(\mathbf{N}) + \frac{\beta_j}{(1-\beta_j)} b_j$$

and

$$P_R(x_R^*, x_L^*; \mathbf{N}) = \frac{1}{4}\left[\left(\frac{\beta_L}{(1-\beta_L)} b_L\right)^2 - \left(\frac{\beta_R}{(1-\beta_R)} b_R\right)^2\right]$$

which does not depend on $\mathbf{N}$. In the symmetric case where $\beta_L = \beta_R$ and $b_R = -b_L$, we have that $P_R(x_R^*, x_L^*; \mathbf{N}) = 0$ regardless of $\mathbf{N}$. Intuitively, if parties have equal and opposite ideological motivations, party platforms' deviations from voter preferences are symmetric, so that elections remain competitive in equilibrium regardless of voter demographics or the strength of ideological motivations.

3 Data on Demographics and Voting

We conduct our main analysis on US presidential elections from 1952 through 2016. We collect data on the voting and demographic characteristics of the electorate. This section describes the sources and definitions of these variables as well as those used in extensions.

3.1 Sources and Definitions for Main Analysis

Our main analysis uses data from the American National Election Study (ANES) Time Series Cumulative Data File (2019). These data have the advantage of covering a nationally representative sample of US voters over a long time period. The sample includes between

811 and 4188 voters, depending on the election year. All of our analyses use the survey weights recommended by the data providers to ensure representativeness.

We measure voting with the respondent's self-reported vote in the most recent presidential election. We include in our main analysis only those respondents who report voting for the Democrat or the Republican candidate (instead of not voting, voting for another candidate, or not giving a valid response to the question).

We define two sets of demographic covariates for our analysis. Here we describe these covariates briefly; Online Appendix Table 1 provides more details.

The *main demographic covariates* are age (in 10-year bins), gender, and race (white, Black, Hispanic, or other), which are primary demographic characteristics; education (less than high school, high school, college or more) and income (in terciles), which account for socioeconomic status; and urbanism, which accounts for differences between rural and urban areas.[12] We selected these variables as they are demographic characteristics known to be strong correlates of voting behavior (e.g., Campbell et al., 1960; Wolfinger and Rosenstone, 1980; Brady et al., 1995; Alesina and La Ferrara, 2005; Scala and Johnson, 2017; Gimpel et al., 2020).

The *extended demographic covariates* include the main demographic covariates as well as the respondent's Census region, labor force participation (in labor force either working or seeking work, homemakers, students, or retired), occupation group (professional and managerial; clerical and sales workers; skilled, semi-skilled, and service workers; laborers; farmers, forestry, and fishermen; homemakers), religion (Roman Catholic, Protestant, Jewish, or other), religious participation (based on frequency of attendance), marital status (never married, married, or previously married), whether the respondent is foreign-born, and whether the respondent's parents are foreign-born. The extended demographic covariates also replace 10-year bins with 5-year bins for age. We selected these variables as they often appear in analyses of voting behavior (e.g., Raymond, 2011; *Economist*, 2018; Zingher, 2020; Bellettini et al., 2023; Kim and Zilinsky, 2024) and were also recorded relatively consistently by the ANES even though some (such as occupation) are unavailable in some years.

[12] We define urbanism based on whether the population density of the respondent's congressional district is low (below 1,000 people per square mile), medium (from 1,000 to 2,000 people per square mile), or high (2,000 or more people per square mile). We obtain data on the population density of congressional districts from Ferrara et al. (2022).

All demographic covariates enter our analysis as category indicators ("one-hot-encodings"). We explore specifications that allow rich interactions among these. We omit respondents who have missing data for one or more covariates, and show the sensitivity of our findings to including these respondents and treating missing values as a distinct covariate category. Online Appendix Table 2 reports the frequency of missing data.

3.2 Sources and Definitions for Extensions

In an extension, we repeat our main analysis using county-level data on voting and demographic covariates. We select county-level demographic covariates to match the main individual-level demographic covariates as closely as possible. Online Appendix C describes the sources and definitions for the variables we use in this extension.

In a separate extension, we repeat our main analysis focusing on voting in congressional elections. We study voting in congressional elections in both presidential election years and midterm election years.[13]

In a final extension, we repeat our main analysis focusing on self-reported party identification rather than voting. Party identification is highly predictive of vote choice and reflects people's ideological orientation and political views (e.g., Berelson et al., 1954; Bartels, 2000; Green et al., 2002; Gerber et al., 2010). Therefore, we do not include this variable as a demographic covariate and treat it instead as an alternative outcome to vote choice. We classify respondents as identifying with either the Republican party or the Democratic party, excluding those who identify with neither. We classify respondents who report being independent but closer to one of the two major parties as identifying with that party; Online Appendix Table 1 provides more details.

4 Methods for Forecasting and Evaluation

4.1 Models and Methods for Forecasting Elections

For concreteness, we describe our methods for the application to survey microdata as in our main analysis; analogous concepts apply in the extension to aggregate data. Let $v_{it} \in \{0, 1\}$

[13]The ANES Time Series file includes respondents' self-reported voting in midterm election years from 1958 through 2002.

denote whether respondent i reports voting Republican in election year t. Let $\mathbf{d}_{it}$ be a vector of demographic indicators.[14]

We specify and estimate models of the form

$$\Pr\left(v_{it}=1|\mathbf{d}_{it}\right)=p\left(\mathbf{d}_{it};\theta_t\right)$$

where $p\left(\cdot;\cdot\right)$ is a function known up to the election-specific parameter θ_t. In our main analysis, we assume that $p\left(\cdot;\cdot\right)$ is logistic and estimate θ_t via maximum likelihood. In an extension, we allow that $p\left(\cdot;\cdot\right)$ is an average of regression trees, and we estimate θ_t via the random forest algorithm.[15]

Suppose we wish to forecast the outcome of the election at some horizon $h>0$, i.e., in some future election year $t+h$, based on voter behavior in election t. Given an estimate $\hat{\theta}_t$ of the parameters θ_t, we can use the model to forecast the probability $p\left(\mathbf{d}_{i,t+h};\hat{\theta}_t\right)$ that a given respondent to the survey in election year $t+h$ votes Republican in that year. Taking a sample average of these probabilities yields a forecast $\hat{V}_{t,t+h}$ of the Republican share of the two-party vote in election $t+h$, formed based on voter behavior in election t.

Because the sample of survey respondents in election $t+h$ is representative of the contemporaneous population, the average $\hat{V}_{t,t+h}$ of their predicted votes accounts for all of the changes in demographic covariates between elections t and $t+h$, for example due to aging, changes in education levels, changes in racial and ethnic composition, etc. Moreover, because we focus on a sample of voters, the forecast automatically accounts for changes in turnout among different groups.

4.2 Measuring Forecast Performance

We evaluate a given forecast $\hat{V}_{t,t+h}$ by comparing it to the realized Republican share of the two-party vote in election $t+h$, which we denote by V^*_{t+h}. For our main analysis, we take V^*_{t+h} to be the Republican share of the two-party vote among survey respondents. Focusing on this measure avoids penalizing the forecast for differences between survey-based and official election results due, for example, to survey misreporting (e.g., Wright,

[14] To connect these to the model in Section 2, let each group $g\in\{0,1\}^{\dim(\mathbf{d})}$ represent one possible combination of these indicators.

[15] We optimize the maximum depth and bag size of the random forest algorithm for each year to minimize mean squared error estimated using 10-fold cross-validation.

1993; Atkeson, 1999). For completeness, we also present results based on official election returns.[16]

We can measure the (in)accuracy of a given forecast $\hat{V}_{t,t+h}$ by its Euclidean distance from the realized result V^*_{t+h}, which is $\sqrt{\left(\hat{V}_{t,t+h} - V^*_{t+h}\right)^2} = \left|\hat{V}_{t,t+h} - V^*_{t+h}\right|$. We can measure the average (in)accuracy of a set of forecasts by the the root mean squared error (RMSE) relative to the realized results, which is

$$\sqrt{\frac{1}{|T_h|}\sum_{t\in T_h}\left(\hat{V}_{t,t+h} - V^*_{t+h}\right)^2}.$$

Here, we average over the set T_h of election years for which we observe the realized result at horizon h.

Since perfect forecasting is infeasible, it is helpful to compare the RMSE of a given set of forecasts to that of a feasible alternative. One feasible alternative is to predict that the realized election result V^*_{t+h} in election $t+h$ will be the same as the realized result V^*_t in election t, i.e., to take $\hat{V}_{t,t+h} = V^*_t$. We refer to this benchmark as the *current* forecast. Another feasible alternative is to predict that every election will be an even contest, i.e., to take $\hat{V}_{t,t+h} = 0.5$. We refer to this benchmark as the *even split* forecast.

4.3 Diagnostics and Quantification of Uncertainty

We compute two additional diagnostics to help interpret model performance. The first diagnostic is a measure of how well the fitted model performs in predicting individual voting behavior in the election year on which the model is estimated. We define a given model's *within-election* error as its RMSE in predicting each individual's vote.[17] When

[16] We obtain official election results from the History, Art & Archives, U.S. House of Representatives (2021) for 1952-1972 and from the MIT Election Data and Science Lab (2017) for 1976-2016. Online Appendix Figure 1 shows the relationship between the official and survey-based measures of the Republican share of the two-party vote for the elections in our main sample.

[17] For a given sample I of respondents, this is

$$\sqrt{\frac{1}{|I|}\sum_{i\in I}\left(v_{it} - p\left(\mathbf{d}_{it};\hat{\theta}_t\right)\right)^2}.$$

To guard against overfitting, we estimate this RMSE via 10-fold cross validation. We divide the average estimated RMSE by its counterpart from a model that predicts each voter's vote with the sample mean vote in the given election.

this error is zero, the model predicts each respondent's vote perfectly (but may or may not successfully forecast future elections).

The second diagnostic is a measure of how much the fitted model tends to predict that the two-party vote share will change between elections. As of election t, at horizon h, a given model predicts a change in the Republican share of the two-party vote of $\left(\hat{V}_{t,t+h} - \hat{V}_{t,t}\right)$, where $\hat{V}_{t,t}$ is the model's prediction for the two-party vote share in election t, typically equal to the realized vote share V_t^*. We define a given model's *shift* at horizon h as the root mean square of these changes.[18] When this shift is zero, the model predicts that the changes in demographics between elections t and $t+h$ will not change the Republican share of the two-party vote.

For all values that depend on the survey sample, we quantify uncertainty by reporting a 95% credible interval based on 500 replicates of a Bayesian bootstrap procedure. For each replicate, we draw Dirichlet-distributed weights for all survey respondents, calculate the product of these weights with the provided sampling weights, and recalculate all survey-dependent statistics using the resulting weights.

5 Results on Forecast Performance

5.1 Main Results

Figure 1 presents our main findings on the performance of demographic forecasts of US presidential elections. Online Appendix Table 3 gives more precise magnitudes for the plotted values.

The plot in Panel (a) of Figure 1 presents the results from our main specification. The plot consists of three sections, one describing the performance of benchmark forecasts, the next describing the performance of demographic forecasts, and the last providing diagnostics for the forecasting models.

The first section of the plot describes the performance of our two benchmark forecasts.

[18] For a given set T_h of elections, this is

$$\sqrt{\frac{1}{|T_h|}\sum_{t\in T_h}\left(\hat{V}_{t,t+h} - \hat{V}_{t,t}\right)^2}.$$

Recall that the current forecast predicts that the two-party vote share in the next election will be the same as in the current election. We normalize the RMSE of the current forecast of the next election to one, and normalize the RMSEs of other forecasts by dividing them by the RMSE of the current forecast of the next election. Recall also that the even split forecast predicts that all elections including the next have a two-party vote share of 0.5. The even split forecast achieves a lower RMSE than the current forecast, indicating a better forecast.

The second section of the plot describes the performance of our main demographic forecasts which are based on a logit model using the main demographic covariate set. The plot reports the RMSE of these forecasts one, two, three, four, and five elections in advance. At a one-election horizon, the demographic forecast performs about as well as the current forecast, and 24.5 percent worse than the even split forecast. At longer horizons, forecast performance is no better. The shaded regions represent 95% credible intervals from the Bayesian bootstrap. These intervals all include performance worse than that of the current forecast, and exclude performance as good as the even split forecast.

The third section of the plot describes the diagnostics. The within-election error is 7.1 percent lower than that of a constant model that predicts each voter's vote with the sample mean vote in the same election, indicating that the logit model has nontrivial ability to predict voting behavior in the election in which it is estimated. The average shift at a one-election horizon is about 12.2 percent of the RMSE of the current model, indicating that the model predicts that demographics cause fairly small changes in the two-party vote share between elections. Both of these quantities are quite statistically precise, the within-election error so much so that its credible interval is almost invisible.

The plot in Panel (b) of Figure 1 presents the results when we define forecast accuracy relative to the official election result, rather than to the realized survey result. The conclusions are qualitatively the same as in our main specification. The even-split forecast performs even better in this case than in our main specification, leading to a larger performance gap between the demographic forecasts and the even split forecast.

The plot in Panel (c) of Figure 1 presents the results when we include in the analysis only open-seat elections, i.e., those without an incumbent on the ballot. Demographic forecasts perform even more poorly in this set of elections than in the main specification.

5.2 Extensions and Interpretation

Figure 2 presents our findings on alternative forecasts based on richer demographics. Panel (a) of Figure 2 repeats the results from our main specification. Panel (b) presents results when we use the richer, extended set of demographic covariates. Including these additional covariates does yield improvements in within-election performance, reducing the within-election error by 5.7 percent relative to the model in Panel (a).[19] Including these additional covariates does not yield any meaningful improvement in forecast accuracy, however, showing that improved within-election performance does not guarantee improved forecasting performance.[20]

Panel (c) of Figure 2 presents results when we keep the richer set of demographic covariates and replace the logit model with an average of regression trees estimated using the random forest algorithm. This richer specification reduces the within-election error by a further 5.9 percent relative to the model in Panel (b), but again, does not yield any meaningful improvement in forecast accuracy.

Figure 3 presents our findings on the performance of alternative forecasts based on aggregate data. Panel (a) of Figure 3 repeats the results from our main specification. Panel (b) presents results when we estimate a logistic regression model on county-level data using county-level analogues of the main demographic covariates. These forecasts perform meaningfully worse than the forecasts in our main specification. A clue as to why is in the average shift, which is much larger than in our main specification. County-level regression models tend to imply larger effects of demographic characteristics than do models estimated on survey data, possibly due to ecological fallacy.[21] Because the forecasted changes are large, and do not align with the realized results, the RMSEs are very large as well.

Figure 4 illustrates this interpretation further. Panel (a) shows the realized election result, current forecast based on one election prior, and demographic forecast based on one election prior using our main specification. The plot also shows the difference in error between the current and demographic forecasts. The demographic forecast deviates little

[19] Online Appendix Figure 2 shows the relative importance of each demographic factor to the within-election performance of the model in the specification of Panel (b).

[20] Online Appendix Figure 3 shows that our main results are also not sensitive to alternative ways of treating respondents for whom we are missing information on one or more demographic covariates.

[21] For example, in 2016, the county-level logistic regression implies that changing all adults from high-school graduates to college graduates would have reduced the Republican share of the two-party vote by 56 percentage points, as against 16 percentage points for the main specification.

from the current forecast, often in the wrong direction. Panel (b) uses the demographic forecast based on the county-level logistic regression. Here, the demographic forecast deviates more from the current forecast, but often in the wrong direction, or in the right direction but by too much.

Panel (c) of Figure 3 presents results when we forecast by assuming the Republican share of the two-party vote will remain constant in each county between elections, but that the population will evolve as in the realized data. This corresponds to a forecaster who has perfect foresight about the population of each US county and believes the two-party vote share will remain stable over time within counties even as voters migrate between counties. At some horizons, this forecast outperforms the current forecast, though it does not outperform the even split forecast.

Figure 5 presents our findings on the performance of demographic forecasts of alternative outcomes: vote shares in congressional elections, and party identification. Panel (a) of Figure 5 repeats the results from our main specification. Panels (b) and (c) present results for forecasting the Republican share of the two-party congressional vote in presidential and midterm election years, respectively. In contrast to the results for presidential elections, for congressional elections, the current forecast tends to outperform the even split forecast, particularly in midterm years. The Democrats controlled the House from 1952-1992, after which it was frequently controlled by Republicans. The current forecasting model matches this pattern much better than does an even split. However, in both presidential and midterm years, the demographic forecasts of congressional elections tend to perform no better, and often worse, than the current forecast.

Panel (d) of Figure 5 presents results for forecasting the Republican share of self-reported party identification. Demographic forecasts again perform no better than the current forecast. Here we replace the even split benchmark with a 40-60 benchmark, reflecting the fact that the two parties are more closely competitive in elections than in self-reported party identification. The 40-60 split performs slightly worse than the current forecast and, at most horizons, better than the demographic forecast.

6 Party Adjustment to Demographic Trends

The model in Section 2 highlights that parties may change their positions in response to changes in the composition of the electorate. If parties' responses are large enough, these responses can mute or even negate the effect of changing demographics. In this section, we investigate the importance of this mechanism in our context.

6.1 Background

A large literature tracks changes in parties' positions over time and asks whether they lead or lag changes in voters' views.[22] In the short run, there is evidence that individual candidates adjust their platforms and discourse to the voters they target (e.g., Acree et al., 2020; Enke, 2020; Di Tella et al., 2023). In the longer run, party elites can drive shifts in voters' views (Zaller, 1992; Iversen, 1994; Baum and Groeling, 2009). However, researchers have also found evidence of parties shifting their platforms in response to the platforms and fortunes of rival parties (Adams and Somer-Topcu, 2009) as well as to shifts in public opinion (Erikson et al., 1989; Adams et al., 2004, 2009; Adams, 2012; Klüver and Sagarzazu, 2016; Benefiel and Williams, 2019), particularly among their supporters (Ezrow et al., 2011; Klüver and Spoon, 2016).[23] Parties may respond to changes in public opinion with a lag, especially if they are uncertain about voters' policy preferences until these are reflected at the ballot box (Budge, 1994).

In the US, in the early 1990s, the elites of the Republican and Democratic parties adopted positions on same-sex relationships that followed the views of their electorates, which had diverged in the 1980s (Fernández and Parsa, 2022).[24] Earlier, in the 1970s, both parties had adjusted their platforms to address the demands of "silent majority" middle-class residents of the suburbs (Lassiter, 2003).[25] More recently, the rising share of Latinos

[22] Parties' stances can either be measured directly, based on parties' official platforms and on politicians' speeches, websites, ads, and votes (e.g., Poole and Rosenthal, 1985; Gentzkow et al., 2019; Danieli et al., 2022), or inferred from the partisan leanings of voters holding different views (e.g., Krasa and Polborn, 2014).

[23] For instance, mainstream parties in Europe have adapted their policy agenda in response to the growing success of green and far-right parties (Abou-Chadi, 2016; Abou-Chadi and Krause, 2020).

[24] Similarly, Chen et al. (2008) argue that party elites followed voter opinion on racial politics in the 1960s .

[25] Lassiter writes, "Although the Republican party initially benefited from the grassroots surge of middle-class consciousness, the populist revolt of the center transcended the conservative mobilization of the New Right. The reinvention of the 'New Democrats' as the champions of quality-of-life issues in suburban swing

in the electorate has led both parties to court these voters by highlighting issues and values that may resonate with them (Paz and Jennings, 2022; Carranza, 2024). Below, we ask more systematically whether and how the parties have adjusted their policy positions in response to demographically-driven shifts in public opinion.

6.2 Trends in Overall Party Positions

We turn now to examining parties' responses to changes in the composition of the electorate. Using our main specification, we quantify the change in the composition of the electorate between adjacent elections t and $t+1$ by the change $\left(\hat{V}_{t,t+1} - \hat{V}_{t,t}\right)$ in the Republican share of the two-party vote predicted based on voter behavior in election t. We quantify the change in the composition of the electorate through election τ by the cumulative sum of the predicted election-specific changes, i.e., by

$$\Delta_\tau = \sum_{\{t \in T_1 : t+1 \leq \tau\}} \left(\hat{V}_{t,t+1} - \hat{V}_{t,t}\right)$$

where recall that T_1 is the set of elections for which we observe the result in the subsequent election. We can interpret Δ_τ as the cumulative change in the electoral advantage of Republicans, through election τ, if the model accurately forecasts the change in the election results one election ahead. In the language of the model in Section 2, Δ_τ measures the electoral effect of the change in the composition of the electorate, holding constant the parties' positions.

Panel (a) of Figure 6 plots the estimated value of Δ_τ over the sample period. The plot shows a cumulative electoral advantage to Republicans of about 4.6 percentage points. The statistical uncertainty in this measure is large. Online Appendix Figure 4 shows how the estimated value of Δ_τ changes when we exclude groups of covariates from the forecaster's model. A forecaster ignoring the role of race would have forecast a more rightward shift over our sample period. A forecaster ignoring the role of urbanism would have forecast a more leftward shift.[26] These findings align with the emphases of Judis and Teixeira (2002)

districts and the fiscally responsible managers of the 'new economy' has revitalized the competitiveness of the center in a postliberal political order." (2003)

[26] A forecaster ignoring the role of education would have forecast a more leftward shift as well, consistent with the finding in Gethin et al. (2021) that more educated people held more conservative positions in the earlier part of our sample period.

and Phillips (2014), respectively.

We can also quantify the change in parties' positions. We do this using data from the Manifesto Project (Lehmann et al., 2023), which provides consistent measurement over our sample period. The project makes available a measure of each party's position on a left-right scale given by the difference between the shares of right-wing and left-wing sub-sentences in the given party's platform in the given election year.[27] A measure of zero means the platform includes equal numbers of both types of sub-sentences. In the language of the model in Section 2, this measure tracks the position of each party in a one-dimensional policy space.

Panel (b) of Figure 6 plots the position of each party over the sample period. According to the index, both parties' platforms moved to the right over most of the sample period, with the Democrats moving later, and reversing the trend in the final two election cycles of the sample. The overall trend appears consistent with the parties reacting, albeit with delay, to the demographic shift visible in Panel (a). It is interesting that the Democrats appear to adjust more slowly than the Republicans, consistent with parties moving more nimbly in their ideologically-preferred direction.

6.3 Trends in Party Positions on Specific Issues

If party positions respond to demographic change, an additional implication is that parties will shift differently on different issues depending on the direction and magnitude of demographically-driven changes in voters' positions. To study this possibility, we selected a set of issues on which the ANES has conducted consistent surveying and for which it is possible to measure the corresponding platform positions in the Manifesto Project database. These issues are environmental protection, government authority, internationalism, labor, law and order, market regulation, military, minority rights, protectionism, traditional morality, and welfare. Online Appendix Table 4 details how we assign these concepts to survey questions in the ANES and topics coded in the Manifesto Project.

For each issue, we repeat the forecasting procedure described in Section 4.1 for the sample of two-party voters, where now the dependent variable is not whether the respondent votes Republican, but rather whether the respondent expresses the traditionally right-wing

[27] This measure is called the *ri*ght-*le*ft or "rile" index and is commonly attributed to Laver and Budge (1992).

position on the given issue. This procedure allows us to forecast positions on the issue and to construct an analogue of the cumulative demographic shift Δ_τ through the 2016 election for each issue. We scale this measure by the number of decades in the sample so that it can be interpreted as a per-decade cumulative shift.

For each issue, we also take the difference in the proportion of right-wing sentences on the issue and the proportion of left-wing sentences on the issue in each party's platform, analogous to our measure of overall party positions. Pooling the series for the two parties we estimate a linear regression of the position on a time trend.[28] We use the estimated coefficient on the time trend, scaled in decadal units, as our measure of the overall trend in parties' positions on the issue.

Figure 7 plots the estimated time trend in party positions on each issue (y-axis) against the estimated cumulative demographic shift in voter positions on each issue (x-axis). The two estimated trends have a Spearman rank correlation of 0.64. In the upper right of the plot, we see that the parties have tended to move to the right on matters of government authority (e.g., the desire for a strong government), and that, based on demographics, voters are predicted to have moved in the same direction. In the lower left of the plot, we see that the parties have tended to move to the left on matters of minority rights and law and order and again that, based on demographics, voters are predicted to have moved in the same direction. Online Appendix Figure 5 shows results where we use an alternative scaling of party platforms, and, separately, where we measure the salience of issues to voters and parties, rather than their positions on a left-right scale.

Because the shifts in voter positions depicted on the x-axis of Figure 7 are, by construction, driven by changes in demographics, they are unlikely to be directly caused by the trends in party platforms depicted on the y-axis or by other political factors such as changes in the news media. Of course, it remains possible that trends in party platforms correlate with demographic shifts in voter positions for reasons other than parties' strategic response. For example, it may be that as the demographics of voters change, so do those of party elites, leading to change in party platforms "from the top." Either way, platforms trend as if in response to demographic shifts in voter positions. In tandem with the model in Section 2, this pattern may help to explain why demographic shifts are not useful in

[28] Letting x_{jt} for $j \in \{L, R\}$ denote the parties' positions in election t, we estimate an ordinary least squares regression of x_{jt} on t, and scale the resulting coefficient to be in units of change per decade.

election forecasting.

7 Conclusion

At any point in time, characteristics like age, gender, race, ethnicity, religion, and education are related both to individuals' self-interest and to their group identity, both of which influence their voting behavior. At any point in time, some demographic changes—such as the aging of the population, or the growing education levels of the workforce—are foreseeable. This combination of factors makes demographic forecasting of elections a tempting activity.

We do find that demographic characteristics explain a meaningful share of the variation in individual vote choices in the *current* election. However, demographic forecasts of *future* election results perform poorly even if we assume perfect foresight of future demographic trends. Demographic forecasts do worse than predicting that every presidential election will be an even contest and than guessing that the result of future congressional elections will be the same as today's. They are poor whether we run the analysis at the individual or county level, and irrespective of the set of demographic covariates and the functional form we use to explain and predict vote choices.

In the short run, demographic changes are simply too slow to account for the dramatic shifts in vote shares observed across elections. In the longer run, parties adjust what they say on different issues in step with demographic shifts, consistent with models of electoral competition. Overall, we conclude that demographic forecasts of the sort we test are not useful for planning.

Proofs of Results Stated in Main Text

Lemma 1. *Suppose that for some composition* $\mathbf{N} \in \mathbb{N}^G$ *the platforms* (x_L^*, x_R^*) *constitute a Nash equilibrium in pure strategies. Then if* $P_j\left(x_j^*, x_{-j}^*; \mathbf{N}\right) < 0$ *for some party* $j \in \{L, R\}$ *we must have* $\beta_j\left(b_j\left(x_j^*\right) - b_j\left(x_{-j}^*\right)\right) \geq -\left(1-\beta_j\right) P_j\left(x_{-j}^*, x_j^*; \mathbf{N}\right) \geq 0$.

Proof. The equilibrium payoff for party j is

$$\pi_j\left(x_j^*, x_{-j}^*; \mathbf{N}\right) = (1-\beta_j)\, P_j\left(x_j^*, x_{-j}^*; \mathbf{N}\right) + \beta_j b_j\left(x_j^*\right).$$

A feasible deviation for party j is to select $x_j = x_{-j}^*$ which yields deviation payoff

$$\pi_j\left(x_{-j}^*, x_{-j}^*; \mathbf{N}\right) = (1-\beta_j)\, P_j\left(x_{-j}^*, x_{-j}^*; \mathbf{N}\right) + \beta_j b_j\left(x_{-j}^*\right) = \beta_j b_j\left(x_{-j}^*\right)$$

where we have used the fact that $P_j\left(x, x; \mathbf{N}\right) = 0$ for all $x \in \mathcal{X}$ and $\mathbf{N} \in \mathbb{N}^G$. For the deviation to be weakly unprofitable requires that

$$(1-\beta_j)\, P_j\left(x_j^*, x_{-j}^*; \mathbf{N}\right) + \beta_j b_j\left(x_j^*\right) \geq \beta_j b_j\left(x_{-j}^*\right)$$

from which the conclusion follows. □

References

Abou-Chadi, T. (2016). Niche party success and mainstream party policy shifts–how green and radical right parties differ in their impact. *British Journal of Political Science 46*(2), 417–436.

Abou-Chadi, T. and W. Krause (2020). The causal effect of radical right success on mainstream parties' policy positions: A regression discontinuity approach. *British Journal of Political Science 50*(3), 829–847.

Abramowitz, A. I. (1988). An improved model for predicting presidential election outcomes. *PS: Political Science & Politics 21*(4), 843–847.

Acree, B. D., J. H. Gross, N. A. Smith, Y. Sim, and A. E. Boydstun (2020). Etch-a-sketching: Evaluating the post-primary rhetorical moderation hypothesis. *American Politics Research 48*(1), 99–131.

Adams, J. (2012). Causes and electoral consequences of party policy shifts in multiparty elections: Theoretical results and empirical evidence. *Annual Review of Political Science 15*(1), 401–419.

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 34*(4), 589–610.

Adams, J., A. B. Haupt, and H. Stoll (2009). What moves parties? The role of public opinion and global economic conditions in Western Europe. *Comparative Political Studies 42*(5), 611–639.

Adams, J. and Z. Somer-Topcu (2009). Policy adjustment by parties in response to rival parties' policy shifts: Spatial theory and the dynamics of party competition in twenty-five post-war democracies. *British Journal of Political Science 39*(4), 825–846.

Alesina, A. and E. La Ferrara (2005). Preferences for redistribution in the land of opportunities. *Journal of Public Economics 89*(5), 897–931.

American National Election Studies (2019). ANES Time Series Cumulative Data File (1948-2016). `https://web.archive.org/web/20210910003456/https://electionstudies.org/data-center/anes-time-series-cumulative-data-file`. Accessed in Sep 2021.

Atkeson, L. R. (1999). "Sure, I voted for the winner!" Overreport of the primary vote for the party nominee in the National Election Studies. *Political Behavior 21*, 197–215.

Austen-Smith, D. and J. S. Banks (2005). *Positive Political Theory II: Strategy and Structure*, Volume 2. University of Michigan Press.

Baker, J., D. Swanson, and J. Tayman (2021). The accuracy of Hamilton–Perry population projections for census tracts in the United States. *Population Research and Policy Review 40*, 1341–1354.

Bartels, L. M. (2000). Partisanship and voting behavior, 1952-1996. *American Journal of Political Science 44*(1), 35–50.

Baum, M. A. and T. J. Groeling (2009). *War stories: The causes and consequences of public views of war*. Princeton University Press.

Becker, G. S. (1958). Competition and democracy. *Journal of Law and Economics 1*, 105–109.

Bellettini, G., C. Berti Ceroni, E. Cantoni, C. Monfardini, and J. Schafer (2023). Modern family? The gendered effects of marriage and childbearing on voter turnout. *British Journal of Political Science 53*(3), 1016–1040.

Benefiel, C. and C. J. Williams (2019). Taking official positions: How public policy preferences influence the platforms of parties in the united states. *Electoral Studies 57*, 71–78.

Bengtson, V. L., C. E. Copen, N. M. Putney, and M. Silverstein (2009). A longitudinal study of the intergenerational transmission of religion. *International Sociology 24*(3), 325–345.

Berelson, B. R., P. F. Lazarsfeld, and W. N. McPhee (1954). *Voting: A Study of Opinion Formation in a Presidential Campaign*. The University of Chicago Press.

Berg, J., R. Forsythe, F. Nelson, and T. Rietz (2008). Results from a dozen years of election futures markets research. *Handbook of Experimental Economics Results 1*, 742–751.

Black, S. E. and P. J. Devereux (2011). Recent developments in intergenerational mobility. *Handbook of Labor Economics 4*, 1487–1541.

Black, S. E., P. J. Devereux, and K. G. Salvanes (2005). Why the apple doesn't fall far: Understanding intergenerational transmission of human capital. *American Economic Review 95*(1), 437–449.

Booth, H. (2006). Demographic forecasting: 1980 to 2005 in review. *International Journal of Forecasting 22*(3), 547–581.

Brady, H. E., S. Verba, and K. L. Schlozman (1995). Beyond SES: A resource model of political participation. *American Political Science Review 89*(2), 271–294.

Brollo, F. and U. Troiano (2016). What happens when a woman wins an election? Evidence from close races in Brazil. *Journal of Development Economics 122*, 28–45.

Budge, I. (1994). A new spatial theory of party competition: Uncertainty, ideology and policy equilibria viewed comparatively and temporally. *British Journal of Political Science 24*(4), 443–467.

Campbell, A., P. E. Converse, W. E. Miller, and D. E. Stokes (1960). *The American Voter*. John Wiley.

Campbell, J. E. (1996). Polls and votes: The trial-heat presidential election forecasting model, certainty, and political campaigns. *American Politics Quarterly 24*(4), 408–433.

Carranza, R. (2024). GOP makes its case to latino voters at convention, sets sights past 2024 election. *USA Today* (July 18th). https://www.usatoday.com/story/news/politics/elections/2024/07/18/gop-makes-its-case-to-latino-voters-at-convention-focus-on-economy/74443860007/. Accessed in Aug 2024.

Center, P. R. (2023). Republican gains in 2022 midterms driven mostly by turnout advantage. (July 12th). https://www.pewresearch.org/politics/2023/07/12/republican-gains-in-2022-midterms-driven-mostly-by-turnout-advantage/. Accessed in Aug 2024.

Chen, A. S., R. W. Mickey, and R. P. Van Houweling (2008). Explaining the contemporary alignment of race and party: Evidence from California's 1946 ballot initiative on fair employment. *Studies in American Political Development 22*(2), 204–228.

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

De Meijer, C., B. Wouterse, J. Polder, and M. Koopmanschap (2013). The effect of population aging on health expenditure growth: A critical review. *European Journal of Ageing 10*, 353–361.

Di Tella, R., R. Kotti, C. Le Pennec, and V. Pons (2023). Keep your enemies closer: Strategic platform adjustments during us and french elections. *NBER Working Paper* (w31503).

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

Economist (2018). How to forecast an American's vote. (November 3rd). https://www.economist.com/graphic-detail/2018/11/03/how-to-forecast-an-americans-vote. Accessed in Aug 2024.

Economist (2023a). The great mystery of American politics. (January 5th). https://www.economist.com/united-states/2023/01/05/the-great-mystery-of-american-politics. Accessed in Dec 2023.

Economist (2023b). Why non-white voters are abandoning the Democratic party. (November 17th). https://www.economist.com/culture/2023/11/17/why-non-white-voters-are-abandoning-the-democratic-party. Accessed in Dec 2023.

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

Erikson, R. S. and C. Wlezien (2012). *The timeline of presidential elections: How campaigns do (and do not) matter*. University of Chicago Press.

Erikson, R. S., G. C. Wright, and J. P. McIver (1989). Political parties, public opinion, and state policy in the united states. *American Political Science Review 83*(3), 729–750.

Ezrow, L., C. De Vries, M. Steenbergen, and E. Edwards (2011). Mean voter representation and partisan constituency representation: Do parties respond to the mean voter position or to their supporters? *Party Politics 17*(3), 275–301.

Fair, R. (2011). *Predicting presidential elections and other things*. Stanford University Press.

Fair, R. C. (1978). The effect of economic events on votes for president. *Review of Economics and Statistics*, 159–173.

Fair, R. C. (2009). Presidential and congressional vote-share equations. *American Journal of Political Science 53*(1), 55–72.

Fernández, R. and S. Parsa (2022). Gay politics goes mainstream: Democrats, republicans and same-sex relationships. *Economica 89*, S86–S109.

Ferrara, A., P. Testa, and L. Zhou (2022). New area- and population-based geographic crosswalks for US counties and congressional districts, 1790-2020. *SSRN Electron. J.*.

Fiva, J. H., O. Folke, and R. J. Sørensen (2018). The power of parties: Evidence from close municipal elections in norway. *Scandinavian Journal of Economics 120*(1), 3–30.

Forsythe, R., F. Nelson, G. R. Neumann, and J. Wright (1992). Anatomy of an experimental political stock market. *American Economic Review*, 1142–1161.

Gelman, A. and G. King (1993). Why are American presidential election campaign polls so variable when votes are so predictable? *British Journal of Political Science 23*(4), 409–451.

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.

George, M. V., S. K. Smith, D. A. Swanson, and J. Tayman (2004). Population projections. In J. S. Siege and D. A. Swanson (Eds.), *The Methods and Materials of Demography*, pp. 561–602. San Diego: Elsevier Academic Press.

Gerber, A. S., G. A. Huber, and E. Washington (2010). Party affiliation, partisanship, and political beliefs: A field experiment. *American Political Science Review 104*(4), 720–744.

Gethin, A., C. Martínez-Toledano, and T. Piketty (2021). Brahmin left versus merchant right: Changing political cleavages in 21 western democracies, 1948–2020. *Quarterly Journal of Economics 137*(1), 1–48.

Gimpel, J. G., N. Lovin, B. Moy, and A. Reeves (2020). The urban–rural gulf in American political behavior. *Political Behavior 42*, 1343–1368.

Girosi, F. and G. King (2008). *Demographic Forecasting*. Princeton University Press.

Green, D., B. Palmquist, and E. Schickler (2002). *Partisan Hearts and Minds: Political Parties and the Social Identities of Voters*. Yale University Press.

Hauer, M. E. (2019). Population projections for US counties by age, sex, and race controlled to shared socioeconomic pathway. *Scientific Data 6*(1), 1–15.

History, Art & Archives, US House of Representatives (2021). Election statistics: 1920 to present. `https://history.house.gov/Institution/Election-Statistics`. Accessed in Sep 2022.

Hotelling, H. (1929). Stability in competition. *Economic Journal 39*(153), 41–57.

Iversen, T. (1994). The logics of electoral politics: Spatial, directional, and mobilizational effects. *Comparative Political Studies 27*(2), 155–189.

Jennings, M. K. and R. G. Niemi (1968). The transmission of political values from parent to child. *American Political Science Review 62*(1), 169–184.

Jennings, M. K., L. Stoker, and J. Bowers (2009). Politics across generations: Family transmission reexamined. *Journal of Politics 71*(3), 782–799.

Jennings, W., M. Lewis-Beck, and C. Wlezien (2020). Election forecasting: Too far out? *International Journal of Forecasting 36*(3), 949–962.

Judis, J. B. and R. Teixeira (2002). *The Emerging Democratic Majority*. Simon and Schuster.

Judis, J. B. and R. Teixeira (2023). *Where Have All the Democrats Gone? The Soul of the Party in the Age of Extremes*. Henry Holt and Company.

Julio, B. and Y. Yook (2012). Political uncertainty and corporate investment cycles. *Journal of Finance 67*(1), 45–83.

Kaplan, N., D. K. Park, and A. Gelman (2012). Understanding persuasion and activation in presidential campaigns: The random walk and mean reversion models. *Presidential Studies Quarterly 42*(4), 843–866.

Karahan, F., B. Pugsley, and A. Şahin (2024). Demographic origins of the start-up deficit. *American Economic Review 114*(7), 1986–2023.

Kelly, B., L. Pástor, and P. Veronesi (2016). The price of political uncertainty: Theory and evidence from the option market. *Journal of Finance 71*(5), 2417–2480.

Kim, S.-y. S. and J. Zilinsky (2024). Division does not imply predictability: Demographics continue to reveal little about voting and partisanship. *Political Behavior 46*, 67–87.

Klüver, H. and I. Sagarzazu (2016). Setting the agenda or responding to voters? Political parties, voters and issue attention. *West European Politics 39*(2), 380–398.

Klüver, H. and J.-J. Spoon (2016). Who responds? Voters, parties and issue attention. *British Journal of Political Science 46*(3), 633–654.

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

Lassiter, M. D. (2003). Suburban strategies: The volatile center in postwar american politics. In J. S. Siege and D. A. Swanson (Eds.), *The Democratic Experiment: New Directions in American Political History*, pp. 327–349. Princeton: Princeton University Press.

Laver, M. and I. Budge (1992). *Party Policy and Government Coalitions*. St. Martin's Press.

Lazarsfeld, P. F., B. Berelson, and H. Gaudet (1948). *The People's Choice: How the Voter Makes Up His Mind in a Presidential Campaign*. Columbia University Press.

Le Pennec, C. and V. Pons (2023). How do campaigns shape vote choice? Multicountry evidence from 62 elections and 56 TV debates. *Quarterly Journal of Economics 138*(2), 703–767.

Lehmann, P., S. Franzmann, T. Burst, S. Regel, F. Riethmüller, A. Volkens, B. Weßels, and L. Zehnter (2023). The Manifesto Data Collection. Manifesto Project (mrg/cmp/marpor). `https://doi.org/10.25522/manifesto.mpds.2023a`. Accessed in Nov 2023.

Lewis-Beck, M. S. (2005). Election forecasting: Principles and practice. *British Journal of Politics and International Relations 7*(2), 145–164.

Lewis-Beck, M. S. and R. Dassonneville (2015). Forecasting elections in Europe: Synthetic models. *Research & Politics 2*(1), 2053168014565128.

Lewis-Beck, M. S. and T. W. Rice (1984). Forecasting presidential elections: A comparison of naive models. *Political Behavior 6*, 9–21.

Linzer, D. A. (2013). Dynamic bayesian forecasting of presidential elections in the states. *Journal of the American Statistical Association 108*(501), 124–134.

Lock, K. and A. Gelman (2010). Bayesian combination of state polls and election forecasts. *Political Analysis 18*(3), 337–348.

Maestas, N., K. J. Mullen, and D. Powell (2023). The effect of population aging on economic growth, the labor force, and productivity. *American Economic Journal: Macroeconomics 15*(2), 306–332.

Marx, B., V. Pons, and V. Rollet (2022). Electoral turnovers. *NBER Working Paper*.

Mason, L. (2016). A cross-cutting calm: How social sorting drives affective polarization. *Public Opinion Quarterly 80*(S1), 351–377.

Mason, L. and J. Wronski (2018). One tribe to bind them all: How our social group attachments strengthen partisanship. *Political Psychology 39*, 257–277.

MIT Election Data and Science Lab (2017). US President 1976–2020. `https://doi.org/10.7910/DVN/42MVDX`. Accessed in Sep 2021.

Paz, C. and N. Jennings (2022). A practical guide to winning Latino voters. *Vox* (September 20th). `hhttps://www.vox.com/the-highlight/23331662/latino-voters-how-to-win-democrats-republicans`. Accessed in Aug 2024.

Petropoulos, F., D. Apiletti, V. Assimakopoulos, M. Z. Babai, D. K. Barrow, S. B. Taieb, C. Bergmeir, R. J. Bessa, J. Bijak, J. E. Boylan, et al. (2022). Forecasting: Theory and practice. *International Journal of Forecasting 38*(3), 705–871.

Phillips, K. P. (2014). *The Emerging Republican Majority: Updated Edition*. Princeton University Press.

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

Raymond, C. (2011). The continued salience of religious voting in the united states, germany, and great britain. *Electoral Studies 30*(1), 125–135.

Rosenstone, S. (1983). *Forecasting Presidential Elections*. Yale University Press.

Sarma, M. and J. Pais (2011). Financial inclusion and development. *Journal of International Development 23*(5), 613–628.

Scala, D. J. and K. M. Johnson (2017). Political polarization along the rural-urban continuum? The geography of the presidential vote, 2000–2016. *The ANNALS of the American Academy of Political and Social Science 672*(1), 162–184.

Silver, N. (2012). *The Signal and the Noise: Why So Many Predictions Fail-But Some Don't*. Penguin.

Snowberg, E., J. Wolfers, and E. Zitzewitz (2007). Partisan impacts on the economy: Evidence from prediction markets and close elections. *Quarterly Journal of Economics 122*(2), 807–829.

Snyder Jr, J. M. and H. Yousaf (2020). Making rallies great again: The effects of presidential campaign rallies on voter behavior, 2008-2016. Technical report, National Bureau of Economic Research.

Times, F. (2024). The Latino swing voters who could decide the us election. (August 8th). `https://www.ft.com/content/a0847d8b-6301-4248-b1cd-e9b79b426552`. Accessed in Aug 2024.

Wlezien, C. and R. S. Erikson (2002). The timeline of presidential election campaigns. *Journal of Politics 64*(4), 969–993.

Wolfers, J. and E. Zitzewitz (2004). Prediction markets. *Journal of Economic Perspectives 18*(2), 107–126.

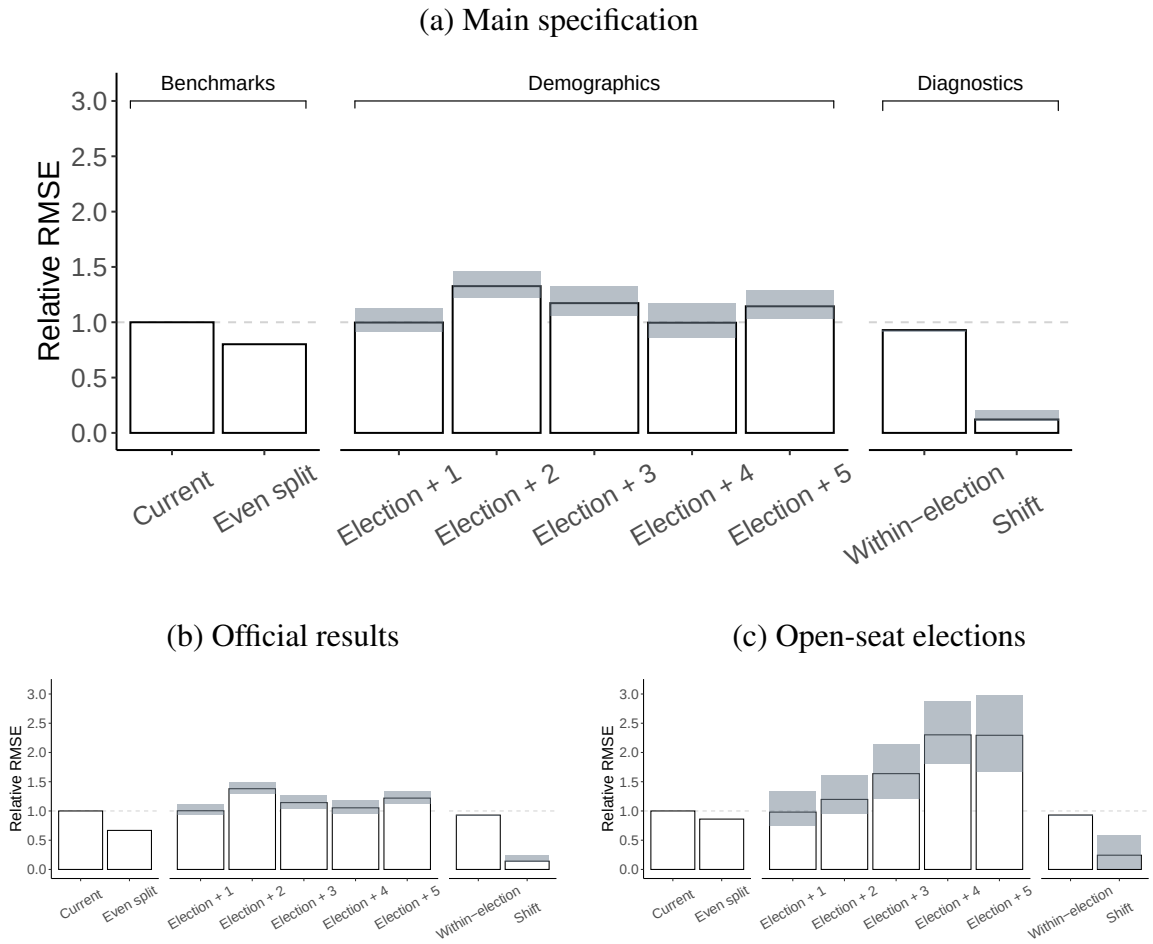
Wolfinger, R. E. and S. J. Rosenstone (1980). *Who Votes?* Yale University Press.

Wright, G. C. (1993). Errors in measuring vote choice in the National Election Studies, 1952-88. *American Journal of Political Science*, 291–316.

Zaller, J. (1992). *The Nature and Origins of Mass Opinion*. Cambridge University.

Zingher, J. N. (2020). On the measurement of social class and its role in shaping white vote choice in the 2016 US presidential election. *Electoral Studies 64*, 102119.

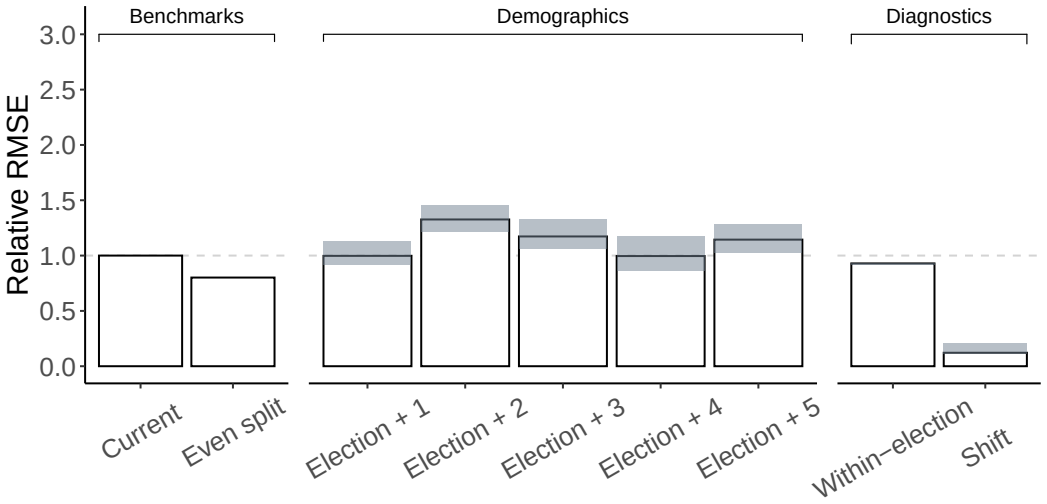
Figure 1: Performance of Demographic Forecasts of US Presidential Elections



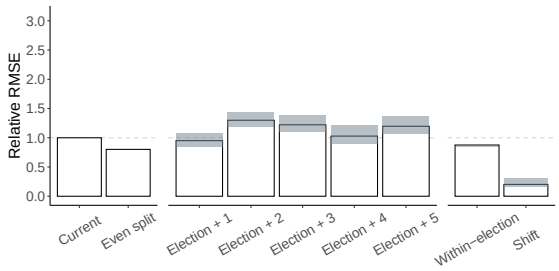
Notes: Each plot displays the relative root mean squared error (RMSE) of election forecasts across varying specifications. The first section of each plot presents the RMSE of the current forecast of the next election, and of the even split forecast, as defined in Section 4.2. The second section of each plot presents the RMSE of demographic forecasts up to five elections in the future, as defined in Section 4.1. The third section of each plot presents the within-election error and average shift at a one-election horizon, as defined in Section 4.3. The within-election error is normalized by dividing by the within-election error of a model that predicts each vote with the sample mean vote. All other values are normalized by dividing by the RMSE of the current forecast of the next election. Shaded regions depict 95 percent credible intervals calculated based on a Bayesian bootstrap. Panel (a) presents results for our main specification. Panel (b) presents results when we benchmark forecast accuracy relative to the official election result. Panel (c) presents results when we restrict attention to open-seat elections.

Figure 2: Performance of Demographic Forecasts of US Presidential Elections: Richer Demographics

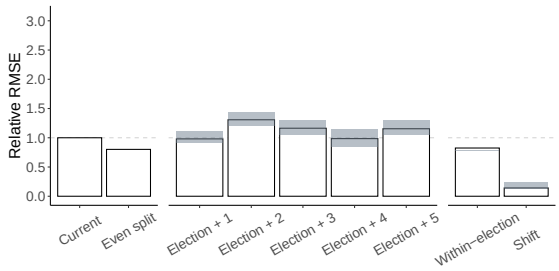
(a) Main specification



(b) Extended demographic covariates



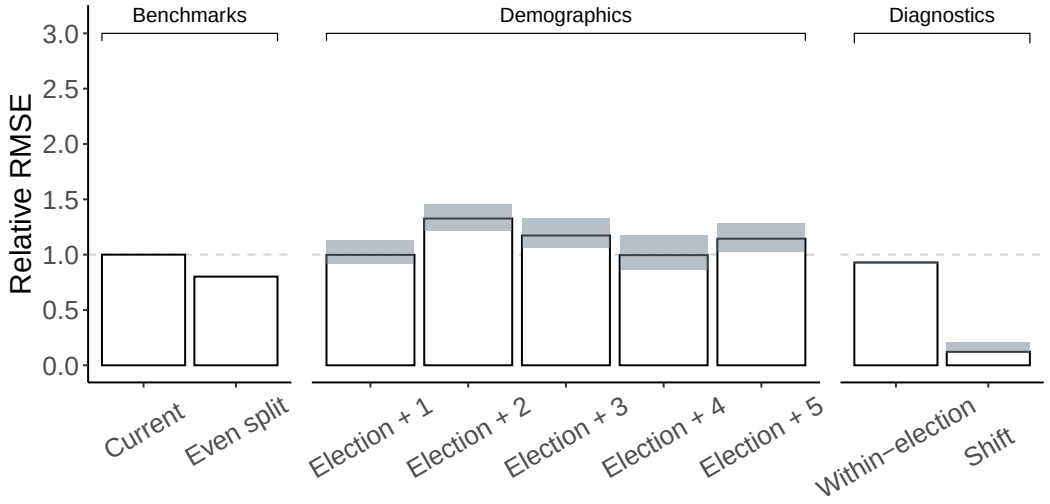
(c) Regression trees



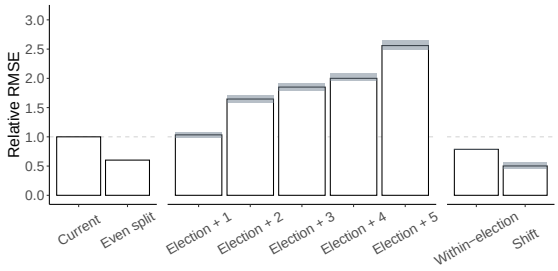
Notes: Each plot displays the relative root mean squared error (RMSE) of election forecasts across varying specifications. The first section of each plot presents the RMSE of the current forecast of the next election, and of the even split forecast, as defined in Section 4.2. The second section of each plot presents the RMSE of demographic forecasts up to five elections in the future, as defined in Section 4.1. The third section of each plot presents the within-election error and average shift at a one-election horizon, as defined in Section 4.3. The within-election error is normalized by dividing by the within-election error of a model that predicts each vote with the sample mean vote. All other values are normalized by dividing by the RMSE of the current forecast of the next election. Shaded regions depict 95 percent credible intervals calculated based on a Bayesian bootstrap. Panel (a) presents results for our main specification. Panel (b) presents results when we use the extended set of demographic covariates. Panel (c) presents results when we use the extended set of demographic covariates and replace the logit model with an average of regression trees estimated using the random forest algorithm. We optimize the maximum depth and bag size of the random forest algorithm for each year to minimize mean squared error estimated using 10-fold cross-validation, and hold these hyperparameters fixed over replicates of the Bayesian bootstrap.

Figure 3: Performance of Demographic Forecasts of US Presidential Elections: Aggregate Data

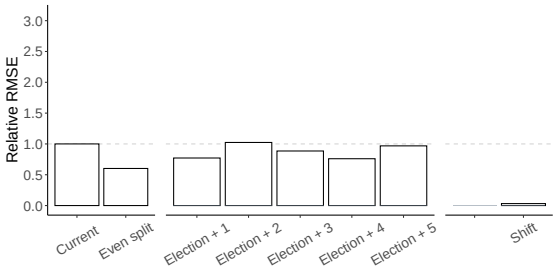
(a) Main specification



(b) County-level logistic regression



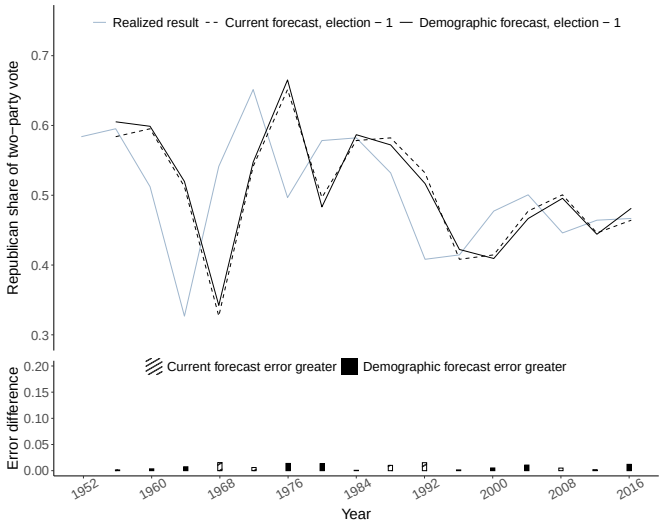
(c) County-level population change



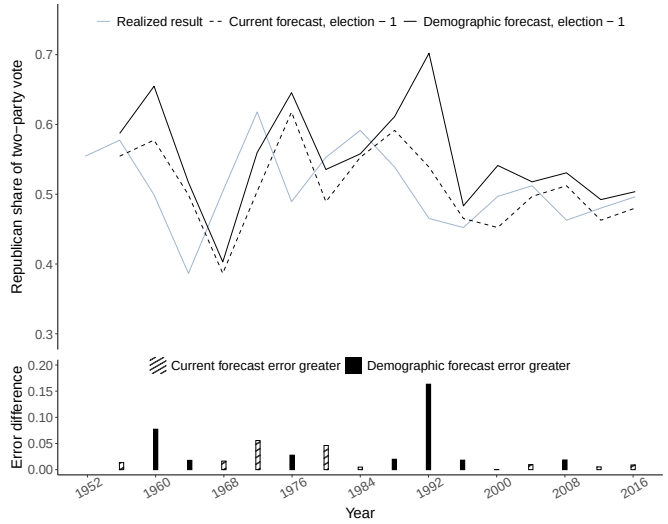
Notes: Each plot displays the relative root mean squared error (RMSE) of election forecasts across varying specifications. The first section of each plot presents the RMSE of the current forecast of the next election, and of the even split forecast, as defined in Section 4.2. The second section of each plot presents the RMSE of demographic forecasts up to five elections in the future, as defined in Section 4.1. The third section of each plot presents the within-election error and average shift at a one-election horizon, as defined in Section 4.3. The within-election error is normalized by dividing by the within-election error of a model that predicts each vote with the sample mean vote. All other values are normalized by dividing by the RMSE of the current forecast of the next election. Shaded regions depict 95 percent credible intervals calculated based on a Bayesian bootstrap. Panel (a) presents results for our main specification. Panel (b) presents results when we use a county-level logistic regression on the main demographic covariates as the basis for forecasting. Panel (c) presents results when we forecast elections by assuming each county's vote remains the same between elections, but allowing each county's population to evolve as in the actual data. For the specification in Panel (c) the within-year error is not well-defined.

Figure 4: Forecasting with Individual-Level vs. Aggregate Data

(a) Individual-level data (main specification)



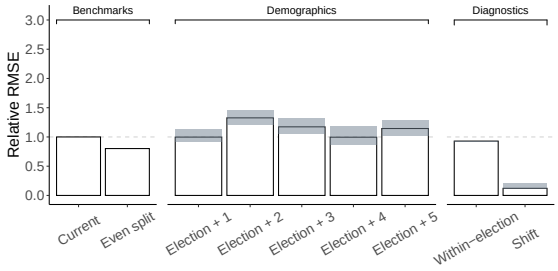
(b) Aggregate data (county-level logistic regression)



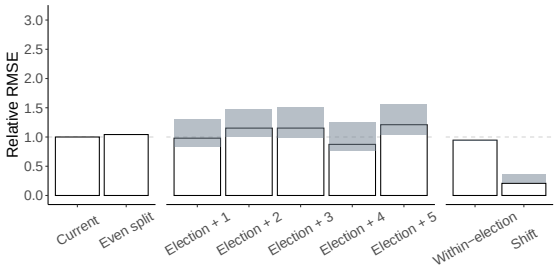
Notes: The top portion of each plot is a time series that shows the election result (gray series), a current forecast equal to the result in the previous election (dashed black series), and a demographic forecast based on data in the previous election (solid black series). The bottom portion of each plot is a bar plot that shows the difference in (absolute) error between the current and demographic forecasts, hatched when the error is greater for the current forecast, and solid when the error is greater for the demographic forecast. Panel (a) uses the main specification from Panel (a) of Figure 1. Panel (b) uses the specification based on a county-level logistic regression in Panel (b) of Figure 3.

Figure 5: Performance of Demographic Forecasts of Congressional Elections and Party Identification

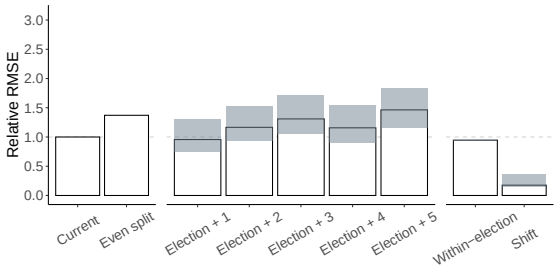
(a) Main specification



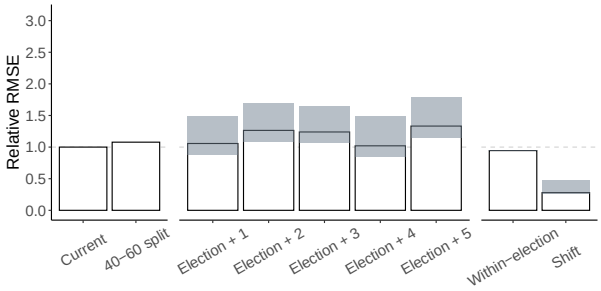
(b) Congressional elections in presidential years



(c) Congressional elections in midterm years



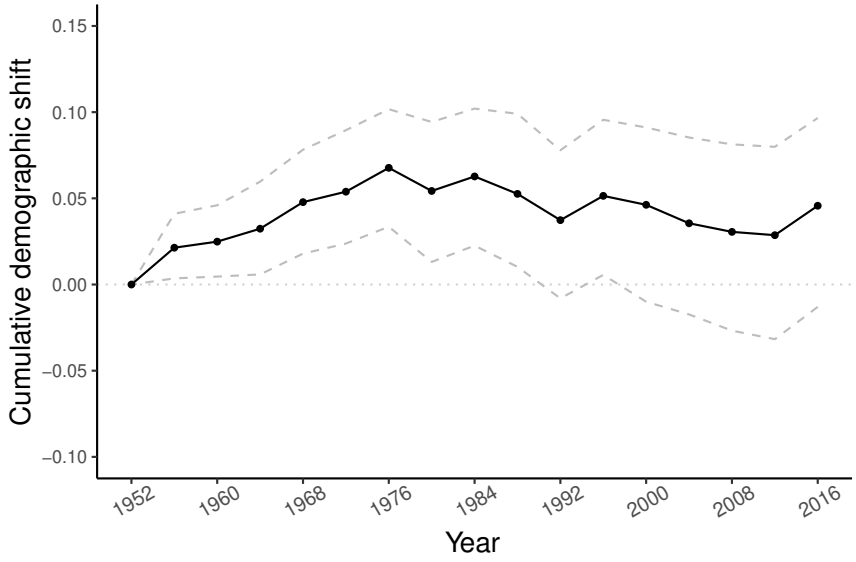
(d) Party identification



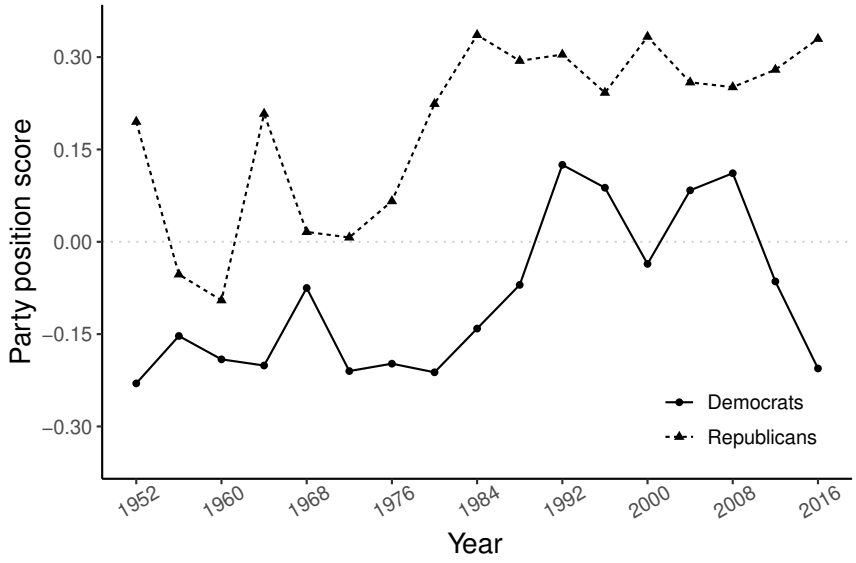
Notes: Each plot displays the relative root mean squared error (RMSE) of election forecasts across varying specifications. The first section of each plot presents the RMSE of the current forecast of the next election, and of the even split forecast, as defined in Section 4.2. The second section of each plot presents the RMSE of demographic forecasts up to five elections in the future, as defined in Section 4.1. The third section of each plot presents the within-election error and average shift at a one-election horizon, as defined in Section 4.3. The within-election error is normalized by dividing by the within-election error of a model that predicts each vote with the sample mean vote. All other values are normalized by dividing by the RMSE of the current forecast of the next election. Shaded regions depict 95 percent credible intervals calculated based on a Bayesian bootstrap. Panel (a) presents results for our main specification. Panel (b) and (c) present results for forecasting the Republican share of the two-party vote in congressional elections in presidential and midterm election years, respectively. Panel (d) presents results for forecasting the Republican share of party identification, where we replace the even split benchmark with one in which party identification is always 40-60 in favor of Democrats.

Figure 6: Shifts in Demographics and Party Positions

(a) Predicted cumulative shift in the electorate

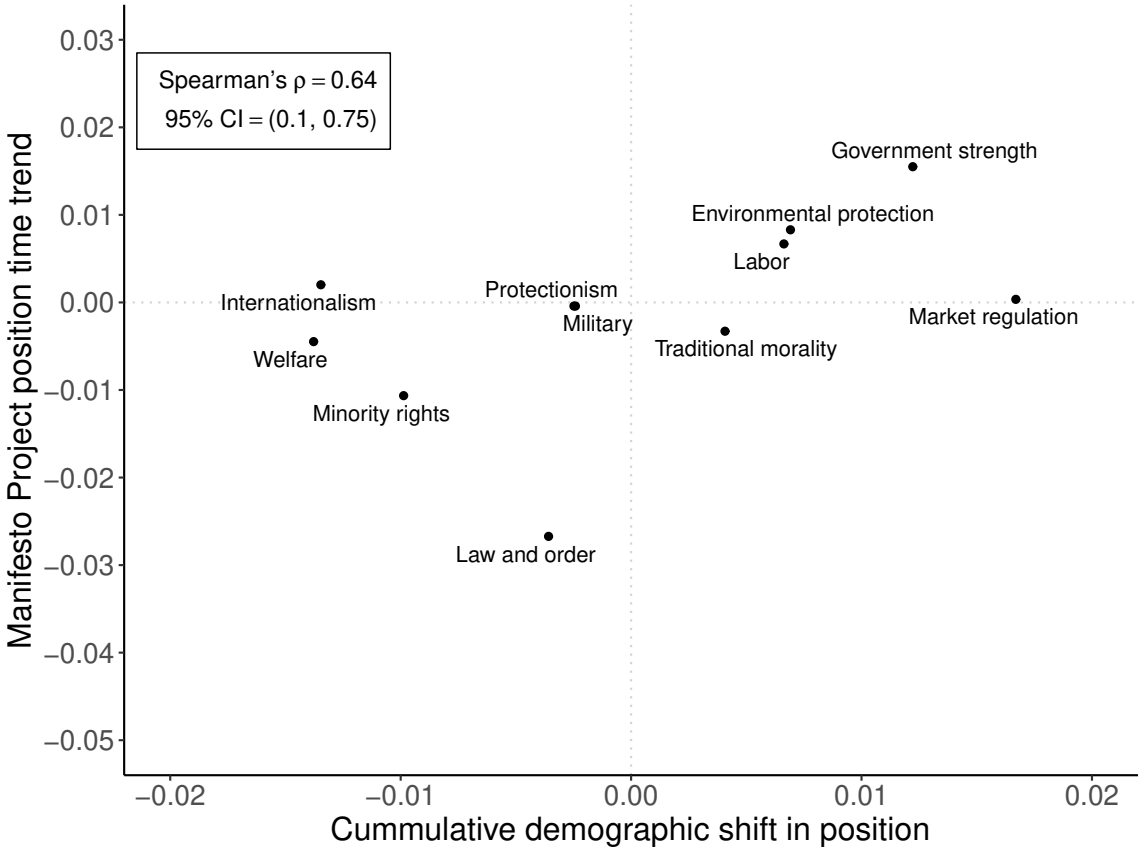


(b) Measured shift in party positions



Note: Panel (a) shows the cumulative predicted change Δ_τ, according to our main specification, in the Republican share of the two-party vote due to demographic change, as defined in Section 6.2. The dashed series depicts pointwise 95% credible intervals calculated based on a Bayesian bootstrap. Panel (b) shows the trend in the position of each party's national platform, measured by the difference between the shares of right-wing and left-wing sub-sentences in the party's platform, as defined in Section 6.2.

Figure 7: Shifts in Demographics and Party Positions on Issues



Note: The plot is a scatterplot. Each point represents an issue for which we can estimate voter positions in the survey data and party positions in the Manifesto Project data, as described in Panel A of Online Appendix Table 4. The y-axis variable is the estimated per-decade linear time trend in the difference between the shares of right-wing vs. left-wing sentences on the issue in party platforms, as defined in Section 6.3. The x-axis variable is the estimated per-decade change in voters' probability of supporting the right-wing position on the issue, as defined in Section 6.3. In the top left we display the Spearman rank correlation between the y-axis variable and the x-axis variable as well as a corresponding 95 percent credible interval calculated based on a Bayesian bootstrap.

Online Appendix for
Pitfalls of Demographic Forecasts of US Elections

Richard Calvo, *University of California Berkeley*
Vincent Pons, *Harvard University, CEPR, and NBER*
Jesse M. Shapiro, *Harvard University and NBER*[1]

A Additional Theoretical Results

Proposition 4. (Sufficient Conditions for Existence of an Equilibrium in Pure Strategies) *Suppose that for each group $g \in \{1, .., , G\}$, the function $p_g(\cdot, \cdot)$ is strictly concave in its first argument and strictly convex in its second, and that the functions $b_L(\cdot)$,$b_R(\cdot)$ are strictly concave. Then there exists a Nash equilibrium and any Nash equilibrium is in pure strategies.*

Proof. Because the sum of strictly concave functions is strictly concave, for any $j \in \{L, R\}$ the payoff function $\pi_j(\cdot)$ is strictly concave. As a result, each party j's best response to any strategy by party $-j$ is a singleton, implying that any equilibrium is in pure strategies.

Because the functions $p_g(\cdot, \cdot)$, $b_L(\cdot)$, $b_R(\cdot)$ are continuous, the payoff function $\pi_j(\cdot)$ is continuous for $j \in \{L, R\}$. This implies that at any $x_{-j} \in \mathcal{X}$, the best response

$$x_j^*(x_{-j}) = \arg\max_{x \in \mathcal{X}} \pi_j(x, x_{-j}; \mathbf{N})$$

is continuous in x_{-j}.

Because the space $\mathcal{X}^2$ is convex and compact, and the best response mapping $(x_L^*, x_R^*) : \mathcal{X}^2 \to \mathcal{X}^2$ is continuous, existence of an equilibrium follows from Brouwer's fixed point theorem. □

B Survey Data Description

[1]E-mail: richard_calvo@berkeley.edu, vpons@hbs.edu, jesse_shapiro@fas.harvard.edu.

Online Appendix Table 1: Survey Question Wording and Coding

Panel A: Outcome variables

Outcome	Coverage	Relevant ANES Question	Coding Notes
Vote for President	1952-2016	VCF0704: 1952-1964: (IF R VOTED:) Who did you vote for President? 1968-1976: (IF R VOTED:) Who did you vote for in the election for President? 1980-LATER: (IF R VOTED:) How about the election for President? Did you vote for a candidate for President? (IF YES:) Who did you vote for?	Coded as "Republican, "Democrat", or excluded.
Vote for Congressman	1952-2004, 2008, 2012, 2016	VCF0707: 1952-1970: COUNTY OF REGISTRATION NOT DETERMINED [NO BALLOT CARD]: How about the vote for Congressman. Did you vote for a candidate for Congress? (IF YES:) Who did you vote for? Which party was that? 1972: COUNTY OF REGISTRATION NOT DETERMINED [NO BALLOT CARD]: How about the election for Congressman-- that is, for the House of Representatives in Washington? Which party's candidate did you vote for for Congressman? 1974,1976: COUNTY OF REGISTRATION NOT DETERMINED: [NO BALLOT CARD]: How about the election for Congressman-- that is, for the House of Representatives in Washington? Did you vote for a candidate for Congress? Whom did you vote for? Which party was that? 1978: ALL CASES [BALLOT CARD]: Here is a list of candidates for the major races in this district. How about the election for the House of Representatives in Washington? Did you vote for a candidate for the U.S. House of Representatives? (IF YES:) Who did you vote for? I. 1980-LATER - REGISTERED IN IW COUNTY: 1980-1982,1984 PERSONAL,1986-1996,1998 PERSONAL,2000 PERSONAL [BALLOT CARD]: Here is a list of candidates for the major races in this district. How about the election for the House of Representatives in Washington? Did you vote for a candidate for the U.S. House of Representatives? (IF YES:) Who did you vote for? 1984 TELEPHONE [NO BALLOT CARD]: I am going to read a list of candidates for the major races in your district. In the election for the House of Representatives, the ballot listed: [Names and party affiliations of all House candidates on the Ballot Card]. Did you vote for a candidate for the U.S. House of Representatives? (IF YES:) Who did you vote for?	Coded as "Republican, "Democrat", or excluded.

Online Appendix Table 1: Survey Question Wording and Coding (continued)

Panel A: Outcome variables

Outcome	Coverage	Relevant ANES Question	Coding Notes
Vote for Congressman (Cont.)	1952-2004, 2008, 2012, 2016	... 1998 TELEPHONE [BALLOT CARD]: Please take out the (color) sheet of paper that was folded inside your booklet. There you see a list of candidates for the major race(s) in this district. How about the election for the HOUSE OF REPRESENTATIVES in Washington? Did you vote for a candidate for the U.S. House of Representatives? (IF YES:) Who did you vote for? 2000 TELEPHONE,2002 [NO BALLOT CARD]: How about the election for the House of Representatives in Washington. Did you vote for a candidate for the U.S. House of Representatives? Did you vote for (the Democrat, [NAME], or) (the Republican, [NAME]) (IF IND/3RD PARTY CANDIDATE: or the [PARTY] candidate, [NAME])? II. 1980-LATER - REGISTERED OUTSIDE IW COUNTY: 1980-1996,1998 PERSONAL,2000,2002 [NO BALLOT CARD]: How about the election for the House of Representatives in Washington? Did you vote for a candidate for the U.S. House of Representatives? (IF YES:) Who did you vote for? (2000: Which party was that?) 1998 TELEPHONE [NO BALLOT CARD]: [TELL RESPONDENT, IF NECESSARY, 'We won't need to use the ballot card in your booklet since you are in a different city/town/county).'] How about the election for the HOUSE OF REPRESENTATIVES in Washington? Did you vote for a candidate for the U.S. House of Representatives? (IF YES:) Who did you vote for?	
Party Identification	1952-2016	VCF0301: Generally speaking, do you usually think of yourself as a Republican, a Democrat, an Independent, or what? (IF REPUBLICAN OR DEMOCRAT) Would you call yourself a strong (REP/DEM) or a not very strong (REP/DEM)? (IF INDEPENDENT, OTHER [1966 AND LATER: OR NO PREFERENCE; 2008: OR DK) Do you think of yourself as closer to the Republican or Democratic party?	Coded as Republican ("Independent - Republican", "Weak Republican", or "Strong Republican"), Democrat ("Independent - Democrat", "Weak Democrat", or "Strong Democrat"), or excluded.

Online Appendix Table 1: Survey Question Wording and Coding (continued)

Panel B: Main demographic covariates

Characteristic	Coverage	Relevant ANES Question	Coding Notes
Age (10-year bins)	1952-2016	VCF0102: 1964-1976: What is your date of birth? 1978-1982: What is the month and year of your birth? 1984-LATER: What is the month, day and year of your birth?	We use the age bins as coded by ANES: 17-24, 25-34, 35-44, 45-54, 55-64, 65-74, and 74+.
Gender	1952-2016	VCF0104: Respondent gender	Coded as female or male/other ("Other" in 2016 only).
Race	1952-2016	VCF0105b: 1948,1952,1956-1970: Interviewer observation of Race. 1972-1976: Interviewer observation of Race. In addition to being American, what do you consider your main ethnic group or nationality group? 1978: Interviewer observation of Race. Interviewer observation: R of Hispanic origin. In addition to being American, is there another nationality or ethnic group that you feel you belong to? (IF YES:) What group is that? 1980,1982,1984,1986: Interviewer observation of Race. Interviewer observation: R of Hispanic origin. In addition to being American, what do you consider your main ethnic group or nationality group? 1988-1998: Interviewer observation of Race. In addition to being American, what do you consider your main ethnic group or nationality group? [IF HISPANIC ETHNIC GROUP NOT MENTIONED] Are you of Spanish or Hispanic origin or descent? 2000-2008: What racial or ethnic group or groups best describes you? [MULTIPLE MENTIONS CODED BY IWR] In addition to being American, what do you consider your main ethnic group or nationality group? [IF HISPANIC ETHNIC GROUP NOT MENTIONED] Are you of Spanish or Hispanic origin or descent? 2012,2016: Are you Spanish, Hispanic, or Latino? FTF ONLY: I am going to read you a list of five race categories. Please choose one or more races that you consider yourself to be: [MULTIPLE MENTIONS]: White / Black or African-American / American Indian or Alaska Native / Asian / Native Hawaiian or other Pacific Islander / Other. Answers have been summarized into 4 categories.	Coded as "White non-Hispanic", "Black non-Hispanic", "Hispanic", or "Other or multiple races, non-Hispanic".
Education	1952-2016	VCF0110: 1952-1972: How many grades of school did you finish? 1974,1976: What is highest grade of school or year of college you have completed? Did you get a high school diploma or pass a high school equivalency test? Do you have a college degree? (IF YES:) What degree is that? 1978-1984: What is highest grade of school or year of college you have completed? Did you get a high school diploma or pass a high school equivalency test? Do you have a college degree? (IF YES:) What is the highest degree that you have earned? 1986-2008: What is highest grade of school or year of college you have completed? Did you get a high school diploma or pass a high school equivalency test? What is the highest degree that you have earned? 2012: What is the highest level of school you have completed or the highest degree you have received? Answers have been summarized into 4 categories.	Coded as "Grade school or less (0-8 grades)", "High school (12 grades or fewer, incl. non-college training if applicable)", "Some college (13 grades or more but no degree; 1948 ONLY: college, no identification of degree status)", or "College or advanced degree (no cases 1948)".

Online Appendix Table 1: Survey Question Wording and Coding (continued)

Panel B: Main demographic covariates

Characteristic	Coverage	Relevant ANES Question	Coding Notes
Income Group	1952-2016	VCF0114: 1952,1956-1960: About what do you think your total income will be this year for yourself and your immediate family? 1962: Would you tell me how much income you and your family will be making during this calendar year, 1962. I mean, before taxes. 1964,1968: About what do you think your total income will be this year for yourself and your immediate family. Just give me the number/ letter) of the right income category. 1966,1970: Many people don't know their exact (1966/1970) income yet; but would you tell me as best you can what you expect your (1966/1970) income to be-- before taxes? You may just tell me the letter of the group on this card into which your family income will probably fall. 1972-1990, 1992 LONG-FORM,1994-2008 EXC. 2000 TELEPHONE: Please look at this card/page (2000 FTF: the booklet) and tell me the letter of the income group that includes the income of all members of your family living here in [previous year] before taxes. This figure should include salaries, wages, pensions, dividends, interest, and all other income. (IF UNCERTAIN:) What would be your best guess? 1992 SHORT FORM: Can you give us an estimate of your total family income in 1991 before taxes? This figure should include salaries, wages, pensions, dividends, interest and all other income for every member of your family living in your house in 1991. First could you tell me if that was above or below $24,999? (IF UNCERTAIN: what would be your best guess?) (IF ABOVE/BELOW $24,999:) I will read you some income categories, could you please stop me when I reach the category that corresponds to your family situation? 2000 TELEPHONE: I am going to read you a list of income categories. Please tell me which category best describes the.total income of all members of your family living in your house in 1999 before taxes. This figure should include salaries, wages, pensions, dividends, interest, and all other income. Please stop me when I get to your family's income. 2012: Information about income is very important to understand how people are doing financially these days. Your answers are confidential. Would you please give your best guess? The next question is about [the total income of all the members of your family living here / your total income] in 2011, before taxes. This figure should include income from all sources, including salaries, wages, pensions, Social Security, dividends, interest, and all other income. What was [the total income in 2011 of all your family members living here / your total income in 2011]?	Coded as low ("0 to 16 percentile" or "17 to 33 percentile"), medium ("34 to 67 percentile"), or high ("68 to 95 percentile" or "96 to 100 percentile").

Online Appendix Table 1: Survey Question Wording and Coding (continued)

Panel B: Main demographic covariates

Characteristic	Coverage	Relevant ANES Question	Coding Notes
Income Group (Cont.)	1952-2016	... September 10, 2019 35 (IF DK/RF:) Was it $40,000 or more, or less than that? (IF LESS THAN 40,000:) Was it $20,000 or more, or less than that? (IF LESS THAN 40,000 AND LESS THAN 20,000:) Please mark the answer that includes the income of all members of your family living here in 2011 before taxes. (IF LESS THAN 40,000 BUT MORE THAN 20,000:) Please mark the answer that includes the income of all members of your family living here in 2011 before taxes. (IF MORE THAN 40,000:) Was it $70,000 or more, or less than that? (IF MORE THAN 40,000 BUT LESS THAN 70,000:) Please mark the answer that includes the income of all members of your family living here in 2011 before taxes. (IF MORE THAN 40,000 AND MORE THAN 70,000:) Was it $100,000 or more, or less than that? (IF MORE THAN 40,000, MORE THAN 70,000, BUT LESS THAN 100,000:) Please mark the answer that includes the income of all members of your family living here in 2011 before taxes. (IF MORE THAN 40,000, MORE THAN 70,000 AND MORE THAN 100,000:) Please mark the answer that includes the income of all members of your family living here in 2011 before taxes.	
Urbanism	1956-2016	VCF0900: Congressional district of interview.	We use this ANES question and data on the population and area of each congressional district from Ferrara et al. (2022) to obtain the population density of each respondent's congressional district. We create indicators for low (less than 1,000 people per square mile), medium (between 1,000 and 2,000 people per square mile), and high (2,000 or more people per square mile) density.

Online Appendix Table 1: Survey Question Wording and Coding (continued)

Panel C: Extended demographic covariates

Characteristic	Coverage	Relevant ANES Question	Coding Notes
Census Region	1952-2016	VCF0112: Region - U.S. Census	Coded as "Northeast" (CT, ME, MA, NH, NJ, NY, PA, RI, VT), "North Central" (IL, IN, IA, KS, MI, MN, MO, NE, ND, OH, SD, WI), "South" (AL, AR, DE, D.C., FL, GA, KY, LA, MD, MS, NC, OK, SC,TN, TX, VA, WV)", or "West" (AK, AZ, CA, CO, HI, ID, MT, NV, NM, OR, UT, WA, WY).
Labor Force Participation	1952-2016	VCF0118: 1968-1970: Are you presently employed, or are you unemployed, or retired, (a housewife), (a student), or what? 1972-1978: (1972: We'd like to know if you are looking for work, working now) (1974-1978: We'd like to know if you are working now, or are you unemployed,) retired, (a housewife) a (student), or what? (IF HOMEMAKER OR STUDENT) Are you doing any work for pay at the present time? (IF R IS HOMEMAKER OR STUDENT AND R IS WORKING FOR PAY:) About how many hours do you work on your job in the average week? 1980 AND LATER EXC. 2002 FRESH CROSS: We'd like to know if you are working now, temporarily laid off, or are you unemployed, retired, permanently disabled, (a homemaker), (a student), or what? (STUDENT OR HOMEMAKER:) Are you doing any work for pay at the present time? (RETIRED 1980,1982, 1988 AND LATER:) Are you doing any work for pay at the present time? (DISABLED 1982,1988 AND LATER:) Are you doing any work for pay at the present time? (STUDENT OR HOMEMAKER:) About how many hours do you work on your job in the average week? (RETIRED AND ANSWERED WORKING FOR PAY 1980:) In an average week do you work 20 or more hours on that job? (RETIRED OR DISABLED, AND R ANSWERED WORKING FOR PAY 1982,1988 AND LATER:) About how many hours do you work on your job in the average week? (RETIRED OR DISABLED AND R VOLUNTEERED WORKING FOR PAY 1984,1986:) About how many hours do you work on your job in the average week? 2002 FRESH CROSS: We'd like to know if you are working now, or are you unemployed, retired, a homemaker, (a student), or what? (MULTIPLE RESPONSES) Answers have been summarized into 5 categories.	Coded as labor force participants ("Employed" or "Not employed: laid off, unemployed, on strike, permanently disabled, other" in ANES), "Retired", "Homemaker", or "Student".

Online Appendix Table 1: Survey Question Wording and Coding (continued)

Panel C: Extended demographic covariates

Characteristic	Coverage	Relevant ANES Question	Coding Notes
Occupation Group	1952-2000, 2004	VCF0115: 1952-1964: What is your occupation. I mean, what kind of work do you do? (IF NOT CLEAR OR OBVIOUS [1958,1960,1964 only]:) What exactly do you do on your job? (IF NOT ASCERTAINED:) What kind of business is that? (IF R IS UNEMPLOYED:) What kind of work do you usually do? (IF R IS RETIRED:) What kind of work did you do before you retired? 1968-1970: (IF EMPLOYED OR ON STRIKE:) What kind of work do you do? [What exactly do you do on your job?] (IF UNEMPLOYED OR RETIRED:) What kind of work did you do when you were employed? [What exactly did you do on your job?] 1972-1982: (IF R IS WORKING NOW OR IS TEMPORARILY LAID OFF:) What is your main occupation [What sort of work do you do? Tell me a little more about what you do.] (IF R IS UNEMPLOYED:) What kind of work did you do on your last regular job [What was your occupation?] (IF R IS RETIRED OR DISABLED:) What kind of work did you do when you worked [What was your main occupation?] 1984 AND LATER: (IF R IS WORKING NOW OR IS TEMPORARILY LAID OFF:) What is your main occupation [What sort of work do you do?] What are your most important activities or duties? (IF R IS RETIRED/UNEMPLOYED /DISABLED:) What kind of work did you do on your last regular job [What was your occupation?] What were your most important activities or duties? Answers have been summarized into 6 categories.	Coded as “Professional or clerical”, “Clerical and sales workers”, “Skilled, semi-skilled and service workers”, “Laborers, except farm”, “Farmers, farm managers, farm laborers and foremen; forestry and fishermen”, or “Homemakers”.
Age (5-year bins)	1952-2016	VCF0101: 1964-1976: What is your date of birth? 1978-1982: What is the month and year of your birth? 1984-LATER: What is the month, day and year of your birth?	Coded in 5-year bins: 17-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49,50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85-89, 90-94, 95+.

Online Appendix Table 1: Survey Question Wording and Coding (continued)

Panel C: Extended demographic covariates

Characteristic	Coverage	Relevant ANES Question	Coding Notes
Religion	1952-2016	VCF0128: 1952-1964: Is your Church (1962: religious) preference Protestant, Catholic or Jewish? 1966-1968: Are you Protestant, Catholic or Jewish? 1970-1988,2002: Is your religious preference Protestant, Catholic, Jewish, or something else? 1990 AND LATER, exc. 2002: (IF R ATTENDS RELIGIOUS SERVICES:) Do you mostly attend a place of worship that is Protestant, Roman Catholic, Jewish or what? (IF R DOESN'T ATTEND RELIGIOUS SERVICES:) Regardless of whether you now attend any religious services do you ever think of yourself as part of a particular church or denomination? (IF YES:) Do you consider yourself Protestant, Roman Catholic, Jewish or what?	Coded as "Protestant", "Catholic [Roman Catholic]", "Jewish", or "Other or none".
Religiosity	1952-2016	VCF0130: 1970-1988: (IF ANY RELIGIOUS PREFERENCE) Would you say you/do you go to (church/synagogue) every week, almost every week, once or twice a month, a few times a year, or never? 1990 AND LATER: Lots of things come up that keep people from attending religious services even if they want to. Thinking about your life these days, do you ever attend religious services, apart from occasional weddings, baptisms or funerals? (IF YES:) Do you go to religious services every week, almost every week, once or twice a month, a few times a year, or never? VCF0131: 1952-1968: Would you say you go to church regularly, often, seldom or never?	Coded as low, medium, or high. For responses pre-1970: low is defined as "seldom" or "never" attend church or "no religious preference"; medium is defined as "often" attend church; high is defined as "regularly" attend church. For responses in 1970 or later: low is defined as "never" attend church, attend church "a few times a year", or "no religious preference"; medium is defined as attending church "once or twice a month"; high is defined as attending church "every week" or "almost every week".

Online Appendix Table 1: Survey Question Wording and Coding (continued)

Panel C: Extended demographic covariates

Characteristic	Coverage	Relevant ANES Question	Coding Notes
Marital Status	1952-2016	VCF0147: 1952: Are you married? 1986: Are you married now and living with your husband/wife-- or are you widowed, divorced, separated, or have you never married? Are you now living with someone as a couple, though not married? 1956-2004: Are you married now and living with your husband/ wife (2002: spouse)-- or are you widowed, divorced, separated, or have you never married? 2008: Are you married now and living with your husband/wife-- or are you widowed, divorced, separated, or have you never married? / Are you married, divorced, separated, widowed, or have you never been married? 2012,2016: Are you now married, widowed, divorced, separated or never married? Are you currently living with a partner, or not?	Coded as "Married", never married ("Never Married" or "Partners; not married" in ANES), or previously married ("Divorced", "Separated", or "Widowed" in ANES).
Foreign Born Parents	1952-2016	VCF0143: Were both your parents born in this country?	Coded as "Yes" or "No".
Respondent Foreign Born	1952-1994	VCF0142: Where were you born? (IF UNITED STATES:) Which state?	Coded as "Yes" or "No".

Notes: Table displays the coverage, ANES survey question wording, and coding for all outcome variables (Panel A), main demographic covariates (Panel B), and extended demographic covariates (Panel C).

Online Appendix Table 2: Sample Sizes in Survey Data

Year	All respondents	Respondents reporting a presidential vote for a major party		
			Demographic covariates non-missing	
		All	Main	Extended
1952	1,899	1,235	1,143	882
1956	1,762	1,266	1,178	1,144
1960	1,181	898	880	828
1964	1,571	1,111	1,059	1,016
1968	1,557	911	889	833
1972	2,705	1,587	1,532	1,463
1976	2,248	1,322	1,222	1,135
1980	1,614	877	784	720
1984	2,257	1,376	1,238	1,182
1988	2,040	1,195	1,078	1,027
1992	2,485	1,357	1,225	1,161
1996	1,714	1,034	929	901
2000	1,807	1,120	916	885
2004	1,212	811	721	696
2008	2,322	1,539	1,367	1,359
2012	5,914	4,188	3,844	3,828
2016	4,270	2,609	2,420	2,410

Note: For each election year the table reports, respectively, the number of respondents in the ANES ("All respondents"), the number of respondents reporting a vote for a major party presidential candidate ("Respondents reporting a presidential vote for a major party: All"); of these, the number of respondents with non-missing values of all main demographic covariates ("Demographic covariates non-missing: main") and all extended demographic covariates ("Demographic covariates non-missing: extended"), as defined in Section 3.1.

C County-Level Data Description

We collect data on aggregate county-level data on voting and six demographic characteristics—age, gender, urbanism, race, education, and income—chosen to align as closely as possible with the variables used in our main specification. When data are not available for each election year, we use the data for the most recent available year. In this section, we provide information on our data sources and variable definitions.

Voting

Our data on county-level US presidential election results come from Leip (2016).

Age

Our data on age come from the US Decennial Census accessed via Social Explorer (US Census Bureau 1950-2010). For each county, we measure the fraction of residents in each of the following age bins: 0–4, 5–9, 10–14, 15–24, 25–34, 35–44, 45–54, 55–64, 65–74, and 75 and older. We observe these variables every ten years starting in 1950 and ending in 2010.

Gender

Our data on gender come from the US Decennial Census accessed via Social Explorer (US Census Bureau 1950-2010). For each county, we measure the fraction of residents who are female and include this variable in our model. We observe this variable every ten years starting in 1950 and ending in 2010.

Urbanism

Our data on urbanism come from the US Decennial Census accessed via Social Explorer (US Census Bureau 1950-2010). For each county, we observe population density (measured in people per square mile), which we use to construct urbanism categories using the same thresholds as in the main specification. We observe this variable every ten years starting in 1950 and ending in 2010.

Race

Our data on race come from the US Decennial Census accessed via Social Explorer (US Census Bureau 1950-2010). For each county, we measure the fraction of residents who are white and include this variable in our model. We observe this variable every ten years starting in 1950 and ending in 2010.

Education

Our data on education in 1950 come from the US Decennial Census accessed via Social Explorer (US Census Bureau 1950-2010). For 1970, 1980, 1990, and 2000, our data come from the US Decennial Census accessed via the US Department of Agriculture Economic Research Service (US Census Bureau 1970-2000). Our data on education in 2010 come from the American Community Survey 5-Year Estimates for 2008–2012 accessed via the US Department of Agriculture Economic Research Service (US Census Bureau 2008-2012). For each county, we measure the fraction of residents aged 25 or older who have less than a high school education, who have only a high school education, who have some college education, and who have a college degree. For the 1950 US Decennial Census data, we construct these categories using the following raw variables:

- Less than high school: "No school years completed," "At Least Some Elementary," and "1-3 years high school"
- Only high school: "4 years high school"
- Some college: "1-3 years college"
- College degree: "4 years college"

When calculating the fraction of residents aged 25 or older in each of these categories, we exclude from the denominator any resident with "Unknown years of school." For subsequent years, the data source comes with these four variables already defined, so we do not need to construct them ourselves. Between the two datasets, we observe these variables in 1950, 1970, 1980, 1990, 2000, and 2010.

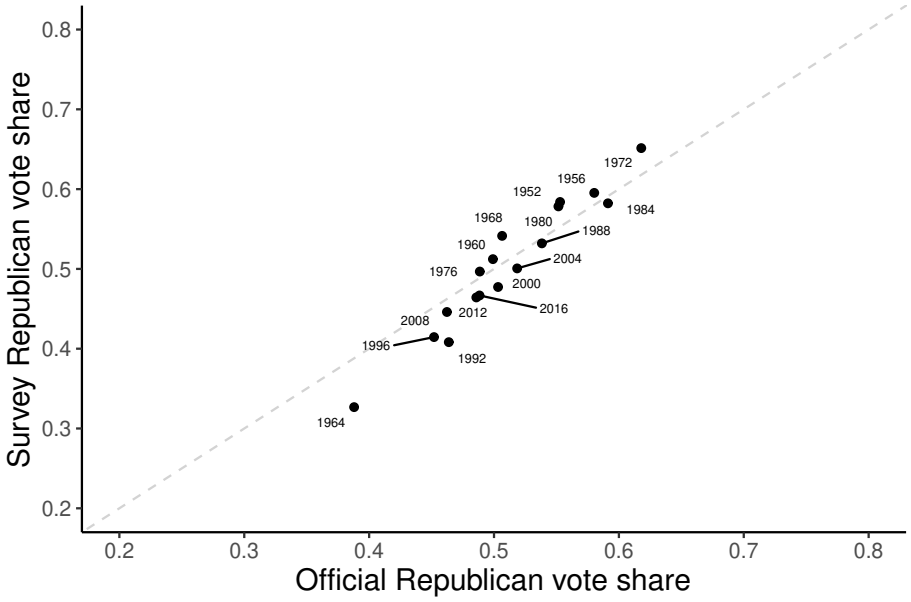
Income

Our data on income come from two sources: for 1952, 1962, and 1972, the US Census County and City Data Books (US Census Bureau 1952-1972) provide median family in-

come, and for each year from 1969 through 2016, the US Bureau of Economic Analysis (BEA) Regional Economic Accounts (US Bureau of Economic Analysis 1969-2022) provide per capita personal income. We convert both variables from nominal into real terms using the January value of the Consumer Price Index (US Bureau of Labor Statistics 2024). To allow for consistency in our predictors across years, we regress log per capita personal income in 1972 on log median family income in 1972 (the year in which the two datasets overlap) and use the estimated linear model to impute log per capita personal income for 1952 and 1962 from log median family income. We then use the imputed values of real log per capita personal income for 1952 and 1962, along with the true values of real log per capita personal income for 1969 onwards, in our model.

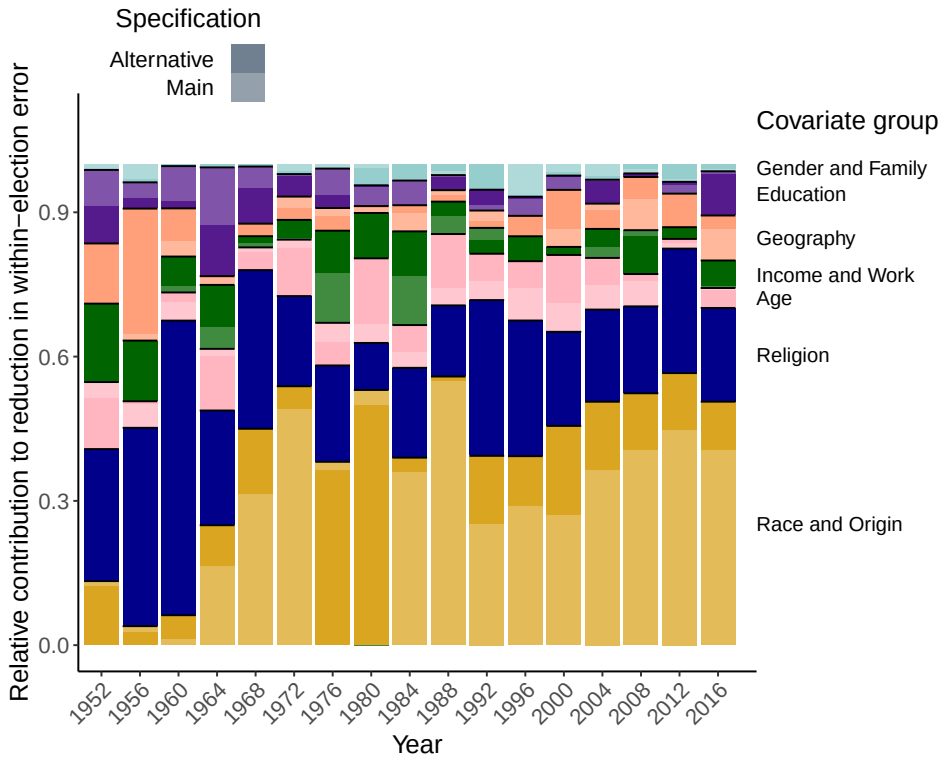
D Additional Empirical Results

Online Appendix Figure 1: Republican Share of Two-Party Vote, Survey vs. Official Results



Note: The plot is a scatterplot. The unit of analysis is the presidential election. The y-axis depicts the Republican share of the two-party vote among survey respondents. The x-axis depicts the Republican share of the two-party vote from official election results. The dashed line is a 45-degree line.

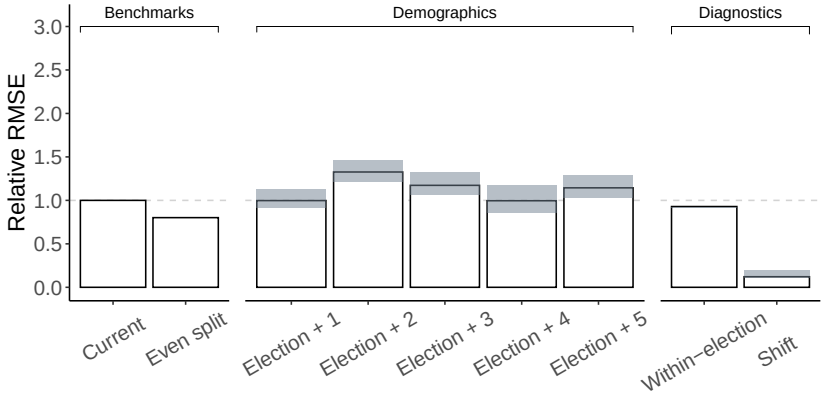
Online Appendix Figure 2: Contributions of Covariate Groups to Reduction in Within-election Error



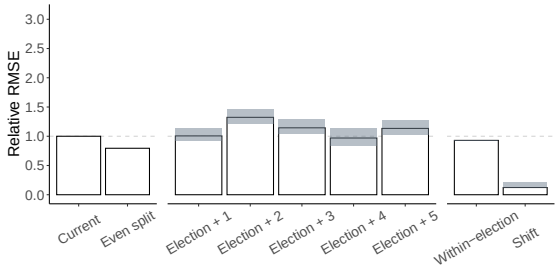
Note: The plot shows the contribution of each group of variables to the reduction in within-election error, as defined in Section 4.3. To calculate the contribution of a given group of variables, we re-estimate the binary logit model without the given group of variables and calculate the increase in within-election error, expressed relative to the sum of contributions across all groups of variables that we consider. We calculate the contribution of each group of variables under the main specification in Panel (a) of Figure 1, which uses the main set of demographic covariates, and under the alternative specification in Panel (b) of Figure 2, which uses the extended set of demographic covariates. For each group of variables, the lighter shaded portion of the bar corresponds to the main specification, and the darker shaded portion of the bar corresponds to the alternative specification. The lower shaded portion of the bar denotes the smaller of the two contributions, and the upper shaded portion of the bar denotes the difference between the smaller and greater contributions. The groups of variables are, "Age", "Education," "Gender and Family" (which includes gender and marital status), "Income and Work" (which includes income, labor force participation, and occupation), "Race and Origin" (which includes race, own foreign-born status, parents' foreign-born status), "Geography" (which includes urbanism and Census region), and "Religion" (which includes religion and religious participation).

Online Appendix Figure 3: Performance of Demographic Forecasts of US Presidential Elections, Alternative Treatment of Missing Covariates

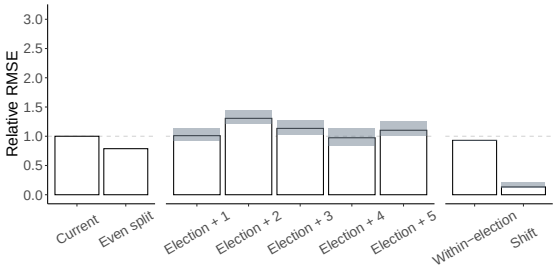
(a) Main specification



(b) Include respondents missing extended covariates

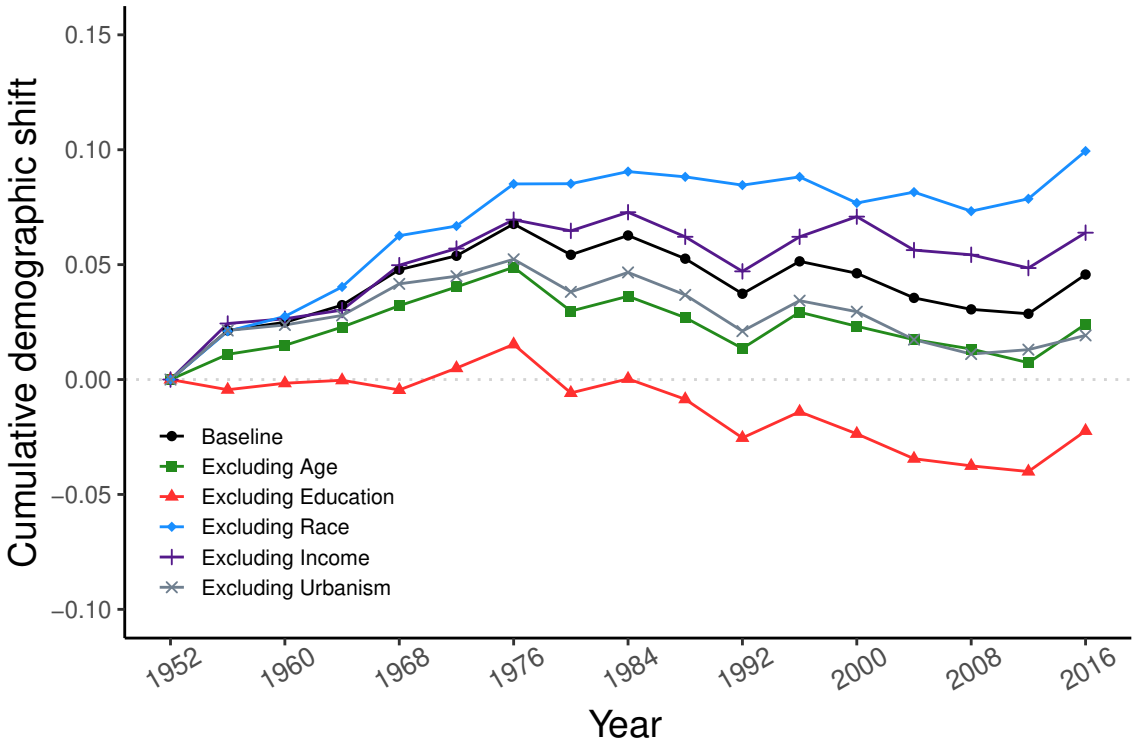


(c) Include respondents missing main covariates



Notes: Each plot displays the relative root mean squared error (RMSE) of election forecasts across varying specifications. The first section of each plot presents the RMSE of the current forecast of the next election, and of the even split forecast, as defined in Section 4.2. The second section of each plot presents the RMSE of demographic forecasts up to five elections in the future, as defined in Section 4.1. The third section of each plot presents the within-election error and average shift at a one-election horizon, as defined in Section 4.3. The within-election error is normalized by dividing by the within-election error of a model that predicts each vote with the sample mean vote. All other values are normalized by dividing by the RMSE of the current forecast of the next election. Shaded regions depict 95 percent credible intervals calculated based on a Bayesian bootstrap. Panel (a) presents results for our main specification in which we include respondents who voted for a candidate in a major party and for whom we have information on all demographic covariates in the extended set. Panel (b) presents results when we include respondents for whom we have information on all demographic covariates in the main set, even if we are missing information for some demographic covariates in the extended set. Panel (c) presents results when we further include respondents for whom we are missing information for some demographic covariates in the main set. To do this, for each demographic covariate in the main set, we add to the predictive model an indicator for whether the covariate value is missing, and impute all other indicators to zero when the covariate value is missing.

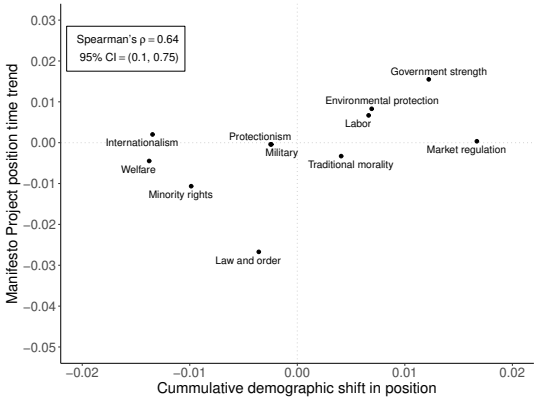
Online Appendix Figure 4: Predicted Cumulative Shift in the Electorate, Excluding Covariates



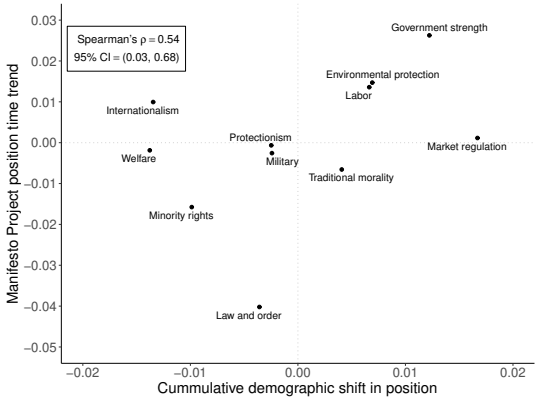
Note: The baseline series corresponds to the predicted cumulative shift in the electorate, as depicted in Panel (a) of Figure 6 and defined in Section 6.2. Each other series corresponds to an experiment in which we exclude the given covariate from the set used to estimate the predictive model, and recompute the predicted cumulative shift in the electorate.

Online Appendix Figure 5: Shifts in Demographics and Party Positions on Issues, Additional Analysis

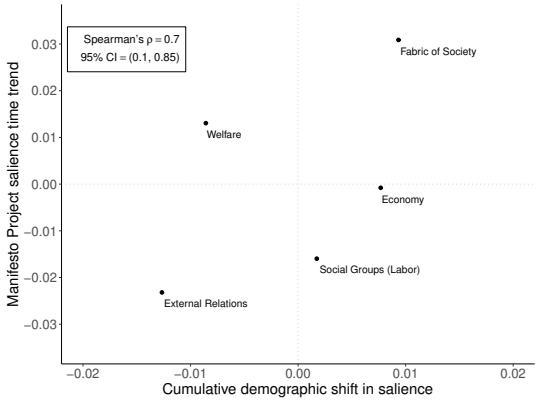
(a) Issue Positions, Baseline



(b) Issue Positions, Alternative Scaling



(c) Issue Salience



Note: Panel (a) repeats the scatterplot from Figure 7. Each point represents an issue for which we can estimate voter positions in the survey data and party positions in the Manifesto Project data, as described in Panel A of Online Appendix Table 4. The y-axis variable is the estimated per-decade linear time trend in the difference between the shares of right-wing vs. left-wing sentences on the issue in party platforms, as defined in Section 6.3. The x-axis variable is the estimated per-decade change in voters' probability of supporting the right-wing position on the issue, as defined in Section 6.3. Panel (b) replaces the y-axis variable with one that restricts attention to the portions of each party's platform that are coded as ideological by the Manifesto Project. Panel (c) replaces both the y-axis and x-axis variables with counterparts based on issue salience. Each point represents an issue for which we can estimate importance to voters in the survey data and emphasis by parties in the Manifesto Project data, as described in Panel B of Online Appendix Table 4. The y-axis variable is the estimated per-decade linear time trend in the share of sub-sentences that refer to the given issue in party platforms. The x-axis variable is the estimated per-decade change in voters' probability of listing the given issue as the one most important to them. The upper left of each plot reports the Spearman rank correlation between the y-axis variable and the x-axis variable as well as a corresponding 95 percent credible interval calculated based on a Bayesian bootstrap.

Online Appendix Table 3: Performance of Demographic Forecasts

Exercise	RMSE		CI Bound		Exercise	RMSE		CI Bound	
	Relative	(Absolute)	Lower	Upper		Relative	(Absolute)	Lower	Upper
Main Specification (Panel (a) of Figure 1)					*Official results (Panel (b) of Figure 1)*				
Current	1.000	(0.099)	-	-	Current	1.000	(0.084)	-	-
Even split	0.801	(0.079)	-	-	Even split	0.667	(0.056)	-	-
Election + 1	0.997	(0.098)	0.915	1.132	Election + 1	1.003	(0.085)	0.929	1.115
Election + 2	1.326	(0.131)	1.218	1.460	Election + 2	1.380	(0.117)	1.298	1.486
Election + 3	1.173	(0.116)	1.062	1.326	Election + 3	1.143	(0.097)	1.047	1.273
Election + 4	0.996	(0.098)	0.860	1.176	Election + 4	1.054	(0.089)	0.955	1.189
Election + 5	1.144	(0.113)	1.029	1.287	Election + 5	1.220	(0.103)	1.121	1.341
Within-election	0.929	(0.459)	0.915	0.925	Within-election	0.929	(0.459)	0.915	0.925
Shift	0.122	(0.012)	0.111	0.203	Shift	0.142	(0.012)	0.130	0.237
Open-seat elections (Panel (c) of Figure 1)					*Extended demographic covariates (Panel (b) of Figure 2)*				
Current	1.000	(0.062)	-	-	Current	1.000	(0.099)	-	-
Even split	0.861	(0.053)	-	-	Even split	0.801	(0.079)	-	-
Election + 1	0.980	(0.060)	0.758	1.340	Election + 1	0.950	(0.094)	0.859	1.082
Election + 2	1.198	(0.074)	0.954	1.618	Election + 2	1.300	(0.128)	1.199	1.443
Election + 3	1.638	(0.101)	1.214	2.143	Election + 3	1.221	(0.121)	1.105	1.388
Election + 4	2.302	(0.142)	1.808	2.881	Election + 4	1.028	(0.102)	0.897	1.218
Election + 5	2.294	(0.141)	1.669	2.986	Election + 5	1.196	(0.118)	1.069	1.361
Within-election	0.930	(0.462)	0.914	0.928	Within-election	0.876	(0.432)	0.848	0.863
Shift	0.244	(0.015)	0.016	0.585	Shift	0.203	(0.020)	0.172	0.314
Regression trees (Panel (c) of Figure 2)					*County-level logistic regression (Panel (b) of Figure 3)*				
Current	1.000	(0.099)	-	-	Current	1.000	(0.094)	-	-
Even split	0.801	(0.079)	-	-	Even split	0.602	(0.056)	-	-
Election + 1	0.981	(0.097)	0.906	1.105	Election + 1	1.034	(0.097)	0.992	1.084
Election + 2	1.307	(0.129)	1.215	1.440	Election + 2	1.646	(0.154)	1.592	1.712
Election + 3	1.163	(0.115)	1.057	1.307	Election + 3	1.851	(0.173)	1.795	1.921
Election + 4	0.985	(0.097)	0.852	1.144	Election + 4	2.000	(0.187)	1.948	2.085
Election + 5	1.154	(0.114)	1.054	1.306	Election + 5	2.560	(0.240)	2.493	2.655
Within-election	0.825	(0.407)	0.779	0.790	Within-election	0.786	(0.103)	0.771	0.783
Shift	0.142	(0.014)	0.122	0.233	Shift	0.502	(0.047)	0.449	0.556
County-level population change (Panel (c) of Figure 3)									
Current	1.000	(0.094)	-	-					
Even split	0.602	(0.056)	-	-					
Election + 1	0.772	(0.072)	-	-					
Election + 2	1.024	(0.096)	-	-					
Election + 3	0.885	(0.083)	-	-					
Election + 4	0.760	(0.071)	-	-					
Election + 5	0.969	(0.091)	-	-					
Within-election	-	-	-	-					
Shift	0.032	(0.003)	-	-					

Online Appendix Table 3: Performance of Demographic Forecasts (cont.)

Exercise	RMSE		CI Bound		Exercise	RMSE		CI Bound	
	Relative	(Absolute)	Lower	Upper		Relative	(Absolute)	Lower	Upper
Main Specification (Panel (a) of Figure 1)					*Congressional elections in presidential years (Panel (b) of Figure 5)*				
Current	1.000	(0.099)	-	-	Current	1.000	(0.053)	-	-
Even split	0.801	(0.079)	-	-	Even split	1.042	(0.055)	-	-
Election + 1	0.997	(0.098)	0.915	1.132	Election + 1	0.980	(0.052)	0.841	1.309
Election + 2	1.326	(0.131)	1.218	1.460	Election + 2	1.152	(0.061)	1.003	1.478
Election + 3	1.173	(0.116)	1.062	1.326	Election + 3	1.152	(0.061)	0.997	1.514
Election + 4	0.996	(0.098)	0.860	1.176	Election + 4	0.874	(0.046)	0.757	1.247
Election + 5	1.144	(0.113)	1.029	1.287	Election + 5	1.210	(0.064)	1.033	1.555
Within-election	0.929	(0.459)	0.915	0.925	Within-election	0.946	(0.468)	0.932	0.941
Shift	0.122	(0.012)	0.111	0.203	Shift	0.208	(0.011)	0.208	0.359
Congressional elections in midterm years (Panel (c) of Figure 5)					*Party Identification (Panel (d) of Figure 5)*				
Current	1.000	(0.062)	-	-	Current	1.000	(0.036)	-	-
Even split	1.371	(0.085)	-	-	40-60 split	1.078	(0.039)	-	-
Election + 1	0.956	(0.060)	0.751	1.315	Election + 1	1.057	(0.038)	0.879	1.482
Election + 2	1.165	(0.073)	0.934	1.528	Election + 2	1.264	(0.045)	1.086	1.697
Election + 3	1.309	(0.082)	1.064	1.719	Election + 3	1.239	(0.044)	1.060	1.639
Election + 4	1.157	(0.072)	0.900	1.548	Election + 4	1.020	(0.036)	0.849	1.490
Election + 5	1.463	(0.091)	1.159	1.843	Election + 5	1.333	(0.048)	1.155	1.797
Within-election	0.946	(0.468)	0.927	0.941	Within-election	0.945	(0.465)	0.934	0.942
Shift	0.176	(0.011)	0.160	0.369	Shift	0.280	(0.010)	0.265	0.475
Include respondents missing extended covariates (Panel (b) of Appendix Figure 3)					*Include all voters (Panel (c) of Appendix Figure 3)*				
Current	1.000	(0.097)	-	-	Current	1.000	(0.097)	-	-
Even split	0.794	(0.077)	-	-	Even split	0.787	(0.076)	-	-
Election + 1	1.006	(0.098)	0.920	1.134	Election + 1	1.010	(0.098)	0.925	1.140
Election + 2	1.323	(0.129)	1.213	1.460	Election + 2	1.306	(0.127)	1.216	1.451
Election + 3	1.144	(0.112)	1.038	1.293	Election + 3	1.134	(0.110)	1.025	1.280
Election + 4	0.970	(0.095)	0.837	1.141	Election + 4	0.974	(0.095)	0.834	1.127
Election + 5	1.134	(0.111)	1.027	1.279	Election + 5	1.103	(0.107)	1.004	1.254
Within-election	0.929	(0.459)	0.916	0.925	Within-election	0.930	(0.459)	0.918	0.928
Shift	0.123	(0.012)	0.113	0.205	Shift	0.134	(0.013)	0.113	0.206

Note: The table reports the statistics plotted in each given panel and figure, with absolute (rather than relative) values in parentheses.

E Issue Positions and Salience Data Description

Online Appendix Table 4: Issue Positions and Salience in Survey and Platform Data

Panel A: Issue Positions

Issue	Years	Source	Question	Coding
Environmental protection	1984-2002, 2016	ANES	VCF9047 - Should federal spending on improving and protecting the environment (2000,2002: environmental protection; 2008,2012,2016: protecting the environment) be increased, decreased, or stay the same?	Coded as right-wing ("Same", "Decreased", "Cut out entirely") or left-wing ("Increased").
		MP	per501 - General policies in favor of protecting the environment, fighting climate change, and other "green" policies. For instance: general preservation of natural resources; preservation of countryside, forests, etc.; protection of national parks; animal rights.	We code this as left-wing.
Government strength	1964-2000	ANES	VCF0829 - Some people are afraid the government in Washington is getting too powerful for the good of the country and the individual person. Others feel that the government in Washington is not getting too strong (1964,1966,1970: has not gotten too strong for the good of the country). 1964-1972: Have you been interested enough in this to favor one side over the other? 1976-1992: Do you have an opinion on this or not? ALL YEARS: (IF YES:) What is your feeling? Do you think the government is too powerful or do you think the government is not getting too strong?	Coded as right-wing ("Opinion: the government has not gotten too strong") or left-wing ("Opinion: the government is getting too powerful").
		MP	per305 - References to the manifesto party's competence to govern and/or other party's lack of such competence. Also includes favorable mentions of the desirability of a strong and/or stable government in general.	Manifesto Project codes this as right-wing in the rile index.

Online Appendix Table 4: Issue Positions and Salience in Survey and Platform Data (continued)

Issue	Years	Source	Question	Coding
Internationalism	1956-2016	ANES	VCF0823 - 1956-1960: (Same introduction as in VCF0805 [CARD WITH RESPONSES SHOWN]). 1968,1980: Now I'd like to read some of the things people tell us when we interview them (1968: and ask you; 1980: As I read, please tell me) whether you agree or disagree with them. 1972: I'd like you to tell me whether you agree or disagree with each of these next six statements. 1976: I am going to read you two statements about US foreign policy and I would like you to tell me whether you agree or disagree with each statement 1984-1988,1992: I am going to read a statement about US foreign policy, and I would like you to tell me whether you agree or disagree. 1990,1994-LATER: Do you agree or disagree with this statement. ALL YEARS: "This country would be better off if we just stayed home and did not concern ourselves with problems in other parts of the world."	Coded as right-wing ("Agree (1956-1960: incl. 'agree strongly' and 'agree but not strongly')") or left-wing ("Disagree (1956-1960: incl. 'disagree strongly' and 'disagree but not strongly')").
		MP	per107 - Need for international co-operation, including co-operation with specific countries other than those coded in 101. May also include references to the: need for aid to developing countries; need for world planning of resources; support for global governance; need for international courts; support for UN or other international organizations.	Manifesto Project codes this as left-wing in the rile index.
		MP	per109 - Negative references to international co-operation. Favorable mentions of national independence and sovereignty with regard to the manifesto country's foreign policy, isolation and/or unilateralism as opposed to internationalism.	We code this as right-wing.

Online Appendix Table 4: Issue Positions and Salience in Survey and Platform Data (continued)

Issue	Years	Source	Question	Coding
Labor	1964-2016	ANES	VCF0210 - Labor unions -- feeling thermometer	Coded as right-wing (thermometer is less than or equal to 50) or left-wing (thermometer is greater than 50).
		MP	per701 - Favorable references to all labour groups, the working class, and un- employed workers in general. Support for trade unions and calls for the good treatment of all employees, including: more jobs; good working conditions; fair wages; pension provisions etc.	Manifesto Project codes this as left-wing.
		MP	per702 - Negative references to labour groups and trade unions. May focus specifically on the danger of unions 'abusing power'.	We code this as right-wing.
Law and order	1984-2016	ANES	VCF0888 - Should federal spending on dealing with crime be increased, decreased or kept about the same?	Coded as right-wing ("Increased") or left-wing ("Same" or "Decreased").
		MP	per605 - Favorable mentions of strict law enforcement, and tougher actions against domestic crime. Only refers to the enforcement of the status quo of the manifesto country's law code. May include: increasing support and resources for the police; tougher attitudes in courts; importance of internal security.	Manifesto Project codes this as right-wing in the rile index.

Online Appendix Table 4: Issue Positions and Salience in Survey and Platform Data (continued)

Issue	Years	Source	Question	Coding
Market regulation	1964-2016	ANES	VCF0209 - Big business -- feeling thermometer	Coded as right-wing (thermometer is greater than 50) or left-wing (thermometer is less than or equal to 50).
		MP	per401 - Favorable mentions of the free market and free market capitalism as an economic model. May include favorable references to: laissez-faire economy; superiority of individual enterprise over state and control systems; private property rights; personal enterprise and initiative; need for unhampered individual enterprises.	Manifesto Project codes this as right-wing in the rile index.
		MP	per403 - Support for policies designed to create a fair and open economic market. May include: calls for increased consumer protection; increasing economic competition by preventing monopolies and other actions disrupting the functioning of the market; defense of small businesses against disruptive powers of big businesses; social market economy.	Manifesto Project codes this as left-wing in the rile index.
Military	1964-2012	ANES	VCF0213 - Military -- feeling thermometer	Coded as right-wing (thermometer is greater than 50) or left-wing (thermometer is less than or equal to 50).
		MP	per104 - The importance of external security and defense. May include statements concerning: the need to maintain or increase military expenditure; the need to secure adequate manpower in the military; the need to modernize armed forces and improve military strength; the need for rearmament and self-defense; the need to keep military treaty obligations.	Manifesto Project codes this as right-wing in the rile index.
		MP	per105 - Negative references to the military or use of military power to solve conflicts. References to the 'evils of war'. May include references to: decreasing military expenditures; disarmament; reduced or abolished conscription.	Manifesto Project codes this as left-wing in the rile index.

Online Appendix Table 4: Issue Positions and Salience in Survey and Platform Data (continued)

Issue	Years	Source	Question	Coding
Minority rights	1970-2016	ANES	VCF0830 - 1970-1984,1986 FORM B, 1988 FORM B: Some people feel that the government in Washington should make every possible effort to improve the social and economic position of blacks (1970: Negroes) and other minority groups (1980: even if it means giving them preferential treatment). Others feel that the government should not make any special effort to help minorities because they should help themselves (1970: but they should be expected to help themselves). 1986 FORM A, 1988 FORM A, 1990 AND LATER: Some people feel that the government in Washington should make every (prior to 1996 only: possible) effort to improve the social and economic position of blacks. (1996-LATER: Suppose these people are at one end of a scale, at point 1). Others feel that the government should not make any special effort to help blacks because they should help themselves. (1996-LATER: Suppose these people are at the other end, at point 7. And, of course, some other people have opinions somewhere in between, at points 2,3,4,5 or 6). ALL YEARS: Where would you place yourself on this scale, or haven't you thought much about it? (7-POINT SCALE SHOWN TO R)	Coded as right-wing (7 point scale is greater than 4) or left-wing (7 point scale is less than or equal to 4).
		MP	per503 - Concept of social justice and the need for fair treatment of all people. This may include: special protection for underprivileged social groups; removal of class barriers; need for fair distribution of resources; the end of discrimination (e.g. racial or sexual discrimination).	We code this as left-wing.
Protectionism	1988-2016	ANES	VCF9231- Some people have suggested placing new limits on foreign imports in order to protect American jobs. Others say that such limits would raise consumer prices and hurt American exports. Do you favor or oppose placing new limits on imports, or haven't you thought much about this?	Coded as right-wing ("Oppose new limits") or left-wing ("Favor new limits").
		MP	per406 - Favorable mentions of extending or maintaining the protection of internal markets (by the manifesto or other countries). Measures may include: tariffs; quota restrictions; export subsidies.	Manifesto Project codes this as left-wing in the rile index.
		MP	per407 - Support for the concept of free trade and open markets. Call for abolishing all means of market protection (in the manifesto or any other country).	Manifesto Project codes this as right-wing in the rile index.

Online Appendix Table 4: Issue Positions and Salience in Survey and Platform Data (continued)

Issue	Years	Source	Question	Coding
Traditional morality	1986-2016	ANES	VCF0853 - ALL YEARS: 'This country would have many fewer problems if there were more emphasis on traditional family ties.' (2000,2004: do you agree strongly, agree somewhat, neither agree nor disagree, disagree somewhat, or disagree strongly with this statement?)	Coded as right-wing ("Agree Strongly" or "Agree somewhat") or left-wing ("Neither agree nor disagree", "Disagree somewhat", or "Disagree strongly").
		MP	per603 - Favorable mentions of traditional and/or religious moral values. May include: prohibition, censorship and suppression of immorality and unseemly behavior; maintenance and stability of the traditional family as a value; support for the role of religious institutions in state and society.	Manifesto Project codes this as right-wing in the rile index.
		MP	per604 - Opposition to traditional and/or religious moral values. May include: support for divorce, abortion etc.; general support for modern family composition; calls for the separation of church and state.	We code this as left-wing.
Welfare	1976-2012	ANES	VCF0220 - People on welfare -- feeling thermometer	Coded as right-wing (thermometer is less than or equal to 50) or left-wing (thermometer is greater than 50).
		MP	per504 - Favorable mentions of need to introduce, maintain or expand any public social service or social security scheme. This includes, for example, government funding of: health care; child care; elder care and pensions; and social housing.	Manifesto Project codes this as left-wing in the rile index.
		MP	per505 - Limiting state expenditures on social services or social security. Favorable mentions of the social subsidiary principle (i.e. private care before state care)	Manifesto Project codes this as right-wing in the rile index.

Online Appendix Table 4: Issue Positions and Salience in Survey and Platform Data (continued)

Panel B: Issue Salience

ANES Most Important Problem	Manifesto Project Rile Domain
Economics; Business; Consumer Issues (includes foreign investment, tariffs/protection of U.S. industries, international trade deficit/balance of payments, immigration, interstate commerce/transportation; does not include unemployment, defense spending, foreign or government spending on domestic social welfare)	Economy (Free Market Economy, Economic Incentives, Market Regulation, Economic Planning, Protectionism, Controlled Economy, Nationalization, Economic Orthodoxy)
Foreign Affairs and National Defense (includes: foreign aid, defense spending, the space program; does not include: international trade deficit)	External Relations (Anti-imperialism, Military, Peace, Internationalism)
Labor Issues (not unemployment)	Social Groups (Labor Groups)
Social Welfare (includes: population, child care, aid to education, the elderly, health care, housing, poverty, unemployment, 'welfare' etc.)	Welfare and Quality of Life (Welfare Expansion/Limitation, Education Expansion)
Public Order (includes: crime, drugs, civil liberties and non racial civil rights, women's rights, abortion rights, gun control, family/social/religious/moral 'decay,' church and state, etc.)	Fabric of Society (National Way of Life, Traditional Morality, Law and Order, Civic Mindedness)

Notes: Tables show mapping between position and salience questions in ANES and content coding in Manifesto Project (MP). Panel A shows mapping between ANES position questions and ANES and Manifesto Project content codes as well as the years covered for each position. Panel B shows the mapping between the answers to the ANES "Most Important Problem" question and the corresponding Manifesto Project rile index domain used to calculate salience. The wording for the ANES "Most Important Problem" question is: "1960: What would you personally feel are the most important problems the government should try to take care of when the new President and Congress take office in January? 1964: As you well know, there are many serious problems in this country and in other parts of the world. The question is, what should be done about them and who should do it. We want to ask you about problems you think the government in Washington should do something about and any problems it should stay out of. First, what would you personally feel are the most important problems the government should try to take care of when the new President and Congress take office in January? 1966: What do you personally feel are the most important problems which the government in Washington should try to take care of? 1968,1980,1982: As you well know, the government faces many serious problems in this country and in other parts of the world. What do you personally feel are the most important problems which the government in Washington should try to take care of? 1970: As you well know, there are many serious problems in this country and in other parts of the world. We'd like to start out by talking with you about some of them. What do you personally feel are the most important problems which the government in Washington should try to take care of? 1972-1978,1984 AND LATER: What do you think are the most important problems facing this country? (IF MORE THAN ONE PROBLEM:) Of all you've told me (1996-LATER: Of those you've mentioned), what would you say is the single most important problem the country faces?"

Appendix References

Leip, D. (2016). Dave Leip U.S. Presidential General County Election Results.

US Bureau of Economic Analysis (1969–2022). Regional economic accounts. Accessed June 2024.

US Bureau of Labor Statistics (2024). Consumer Price Index for all urban consumers: All items in US city average [cpiaucsl]. retrieved from FRED, Federal Reserve Bank of St. Louis. Accessed June 2024.

US Census Bureau (1950–2010). US decennial Census data. Prepared by Social Explorer. Accessed June 2024.

US Census Bureau (1952–1972). *County and City Data Book*. Prepared by the Interuniversity Consortium for Political and Social Research. Accessed June 2024.

US Census Bureau (1970–2000). US decennial Census data. Prepared by the United States Department of Agriculture, Economic Research Service. Accessed June 2024.

US Census Bureau (2008–2012). American Community Survey data. Prepared by the United States Department of Agriculture, Economic Research Service. Accessed June 2024.