NBER WORKING PAPER SERIES

PITFALLS OF DEMOGRAPHIC FORECASTS OF US ELECTIONS

Richard Calvo
Vincent Pons
Jesse M. Shapiro

Working Paper 33016
http://www.nber.org/papers/w33016

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2024, Revised June 2025

We thank our dedicated research assistants for their contributions to this project. For helpful comments we thank Isaiah Andrews, Oren Danieli, Tatyana Deryugina, Gemma Dipoppa, Ben Enke, Ray Fair, Jeremy Friedman, Andrew Gelman, Ed Glaeser, Kareem Haggag, Yueran Ma, Torsten Persson, Jim Snyder, and Alex Tabarrok, along with seminar audiences at the University of Chicago, MIT, the University of Toronto, the NBER (especially discussant Ro'ee Levy), NYU, the PRIN Workshop on the Economics of Culture, and Paris Sorbonne. 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. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

At least one co-author has disclosed additional relationships of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w33016

Pitfalls of Demographic Forecasts of US Elections
Richard Calvo, Vincent Pons, and Jesse M. Shapiro
NBER Working Paper No. 33016
October 2024, Revised June 2025
JEL No. C53, D72, J11, P0

ABSTRACT

Many influential 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. Slow demographic change, unstable group preferences, and strategic party responses all help to explain why demography has not been destiny in US politics.

Richard Calvo
University of California, Berkeley
richard_calvo@berkeley.edu

Vincent Pons
Harvard University
Harvard Business School
and CEPR
and also NBER
vpons@hbs.edu

Jesse M. Shapiro
Harvard University
Department of Economics
and NBER
jesse_shapiro@fas.harvard.edu

> "For a long time, we had this Obama era of politics, where demographics would be destiny and a more diversifying nation would definitely go towards Democrats. And that was the belief. And we're seeing—granted, it's still early—that that is not what's necessarily going to happen."
>
> — Lisa Lerer, *The Daily*, November 6, 2024.

1 Introduction

Demographics have long played a central role in Americans' understanding of their national electoral politics. Witness, for example, the discussion of the Latino vote following the 2024 election (Gerbaud et al., 2024; Pena, 2024; Sanchez, 2024), the white working-class vote following the 2016 election (Cohn, 2016; Lamont et al., 2017; Morgan and Lee, 2018; Carnes and Lupu, 2021), or the Catholic vote following the 1960 election (Phillips, 1960). Supporting the prominence of these discussions, studies find that small numbers of demographic and economic variables can explain a large fraction of the variation in the levels and changes in party vote shares at the county level in recent decades (e.g., Levernier and Barilla, 2006; Freund and Sidhu, 2017; Smith, 2021; Brown and Mettler, 2024). And while some studies find that demographic characteristics are only modestly predictive of voting at the individual level (e.g., Zingher, 2014; Kim and Zilinsky, 2024), these characteristics remain tightly connected to the social and political identities that many scholars put at the heart of US political fissures (Lazarsfeld et al., 1948; Mason, 2016; Mason and Wronski, 2018).

Demographic trends are also the main basis for long-term forecasts of US politics. Unlike the business cycle or geopolitical events, demographics evolve somewhat predictably even over long horizons: if we know the number of white 45-year-old men in the US today, we can forecast with reasonable accuracy the number of white 49-year-old men four years hence.[1] Combined with an observation about which groups tend to vote for which party, demographic trends may thus offer a window into the political future. Efforts to peer through that window go back at least to Kevin Phillips's 1969 book *The Emerging Repub-*

[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), Baker et al. (2021), and Petropoulos et al. (2022).

lican Majority,[2] returned to prominence with John Judis and Ruy Teixeira's 2002 book *The Emerging Democratic Majority*,[3] and continue in recent reports from the Brookings Institution and the Center for American Progress on *America's Electoral Future* (Griffin et al., 2018, 2020). Such forecasts have long influenced the thinking and tactics of campaigns, politicians, and parties, such as the Nixon campaign following Phillips, the Obama campaign following Judis and Texeira, and the GOP following Obama's reelection in 2012.[4]

Despite their longevity, prominence, and influence, so far as we know there have been no systematic evaluations of demographic forecasts of US elections. This gap is important not only because political forecasting is important to political actors and market participants (e.g., Snowberg et al., 2007; Julio and Yook, 2012; Kelly et al., 2016), but also because demographic forecasts embed a substantive hypothesis about US politics. For demographic forecasts to perform well, demographic patterns in political preferences would need to be stable enough to outlast demographic change. Are they?

In this paper, we backtest demographic forecasts using data on US elections from 1952 through 2020. To do this, we fit a model of voting in one election year and use that model, in tandem with a forecast of demographic trends, to predict elections in future years. We find that even a forecaster armed with perfect knowledge of future demographic trends would have performed poorly over this period—worse even than one who guesses that each election will have a 50-50 partisan split. We show that slow demographic change, unstable group preferences, and strategic party responses all help to explain why demography is not destiny in US politics.

[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" (1969).

[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, p. 66).

[4]Buchanan (2015, p. 366) recounts how Republicans built the coalition envisioned by Philipps after 1968. Lizza (2012) describes how the Obama campaign sought to mobilize the groups identified by Judis and Teixeira as part of *The Emerging Democratic Majority*. Following Obama's reelection in 2012, the Republican party identified the rise of such groups as a threat to its electoral strength and sought to adapt accordingly (Barbour et al., 2013, p. 7; see also Graham quoted in Helderman and Cohen, 2012). See also Davis (1974, chapter 12), Carville (2009, pp. 35, 148), and Greenberg (2015, pp. 166, 170).

We fix ideas with a simple accounting framework. Each demographic group has a propensity to vote for each party that may depend on the parties' attributes (e.g., platforms). A forecaster predicts the share of the electorate that each group will account for in some future election, while assuming that each group's propensity to vote for each party will remain unchanged. Such a forecaster faces three pitfalls. First, demographic trends may deviate from the forecast. Second, groups' propensities to vote for one party or another may change over time, even if the parties' attributes remain unchanged. Third, the parties may change their attributes, perhaps in response to demographic change. All of these pitfalls are worse if demographics evolve slowly, because in that case, the pitfalls add noise atop an already weak signal.

We next execute the forecasting task. 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 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.

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. These choices eliminate the first pitfall entirely.

Despite these many advantages, we find that our forecaster fails at their task. 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. More flexible models tend to capture more demographic variation in individuals' votes, but not to improve forecasts of future elections: forecasts re-

main poor when we incorporate an extended set of demographic covariates, allow for rich interactions among demographic covariates using regression trees, or estimate models separately by region. Demographic forecasts likewise do not perform well in predicting congressional elections or party identification.

We turn next to understanding why the forecaster fails. One overarching reason is that demographic change is slow, such that, in most elections, the forecaster's model predicts very little demographically-driven change in the vote share, even over fairly long horizons. However, even models that imply large amounts of demographically-driven change perform poorly. Models estimated on county-level data tend to imply much stronger relationships between voting and demographics than those estimated on survey microdata, perhaps due to ecological fallacy (Robinson, 1950; Freedman, 1999). When our forecaster bases their forecast on models fit at the county level, they forecast large changes in the vote share across elections, but these changes do not come to pass. The forecaster likewise fails when they extrapolate recent trends in groups' propensity to vote for one party or another, even though such an extrapolation implies large demographically-driven changes in election outcomes.

Another reason that demographic forecasts fail is that the relationship between demographic characteristics and voting varies from election to election. According to our baseline model, the probability that an individual with given characteristics votes Republican is highly, but imperfectly, correlated between adjacent elections, with the correlation decaying gradually as we move towards more distant elections. The fact that demographic allegiances shift over time resonates with systematic evidence of changing party bases (e.g., Zingher, 2014), as well as specific instances such as the change in the party allegiances of the educated (Gethin et al., 2021) or those in the South (Black and Black, 2009; Aistrup, 2014; Kuziemko and Washington, 2018). Interestingly, we find even greater levels of change in different groups' stated positions on policy issues, suggesting that groups of voters change not only how they vote, but also their opinions on policy issues.

A final reason that demographic forecasts fail is that the parties change their platforms over time. Rather than passively letting forecasts materialize, the parties notice, and respond to, salient demographic changes, such as the rise of the suburbs (Lassiter, 2003) or the rising Latino share of the population (Barbour et al., 2013). To complement narrative evidence of parties reacting to demographic forecasts, we study the relationship between

party platforms and demographic trends quantitatively. Our baseline model implies that, on net, demographic change has moved the country to the right over the period we study. Data on party platforms from the Manifesto Project shows that the parties have moved to the right, too. Looking across issues, we find that the parties systematically move in the direction predicted by demographic change. Interestingly, however, we also find that demographic forecasts of voter issue positions do not tend to pan out, suggesting either that the parties are reacting to trends that never manifest, or that voters' self-reported issue positions are not a reliable guide to trends in their attitudes.

The principal contribution of our paper is to show that, over the period we study, election results did not reliably follow demographic trends, and to unpack why. Several limitations are worth noting. First, we focus on the postwar United States, which has been characterized by fairly competitive national electoral politics, particularly in recent decades.[5] Second, and relatedly, we focus on making statements about how a forecaster *would have* performed over the period we study, not how such a forecaster *will* perform in the future.[6]

Our findings contribute to two main literatures. The first is a literature on the role of demography in shaping electoral politics and public policy. Zingher (2014) finds that the demographic composition of party bases has shifted over time in the US; Glaeser et al. (2005), Kuziemko and Washington (2018), Gethin et al. (2021), and Longuet-Marx (2025) all document changes in particular relationships between sociodemographic characteristics and voting. Several papers show that when the composition of the electorate changes, so does public policy (e.g., Miller, 2008; Cascio and Washington, 2014; Fujiwara, 2015). Our findings likewise suggest that party platforms respond to demographic shifts over long horizons. These findings add new long-term evidence about an important question typically examined with shorter-term data: whether shifts in candidates' and parties' positions precede or follow changes in public opinion (e.g., Zaller, 1992; Adams, 2012; Enke, 2020; Di Tella et al., 2023).

The second is a literature on election forecasting. On the horizon of weeks or months, elections can be usefully forecast with polling, survey, or prediction market data (Campbell, 1996; Wlezien and Erikson, 2002; Berg et al., 2008; Erikson and Wlezien, 2012; Jen-

[5] *The Economist* observes that, over this period, "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). See also Gelman (2014).

[6] See Grimmer et al. (2024) for a discussion of the difficulty of evaluating forecasting methods.

nings et al., 2020; Camatarri, 2024). On the horizon of quarters or years, elections can be forecast with economic trends (Fair, 1978; Rosenstone, 1983; Lewis-Beck and Rice, 1984; Abramowitz, 1988; Gelman and King, 1993; Lewis-Beck, 2005; Fair, 2009; Kaplan et al., 2012; Chernov et al., 2025). To our knowledge, demographics are the main modern basis for forecasts that stretch over entire election cycles. We find that such forecasts would not have been a useful guide to future election results in the postwar United States.

The remainder of the paper is organized as follows. Section 2 lays out our accounting framework and uses it to define the pitfalls of demographic forecasts. Section 3 describes the data we use in our analysis. Section 4 describes our methods for forecasting and evaluation. Section 5 describes our findings on forecast accuracy. Section 6 unpacks the reasons why demographic forecasts perform poorly over the period we study. Section 7 concludes.

2 Demographic Forecasting and Its Pitfalls

Here we define demographic forecasting, and its pitfalls, in a simple accounting framework. We imagine a two-party election with parties $j \in \{L, R\}$. Each party j has some attributes $x_j \in \mathcal{X}$, which might include, for example, the parties' positions on different issues. The electorate is divided into G groups indexed by g, which might consist of, for example, demographic categories (e.g., college-educated white men in their 40s). A voter in group g votes for party R with probability $p_g(x) \in [0, 1]$, which may depend on the attributes $x = (x_L, x_R)$ of the two parties, as well as, implicitly, the preferences of the group's members. Each group g comprises a share n_g of voters, so that

$$V = \sum_{g=1}^{G} n_g p_g(x)$$

is the expected share of the total vote going to party R.

Suppose that a forecaster believes that, in some future election, each group g will comprise a fraction $\hat{n}_g$ of voters. If the forecaster expects the behavior of the groups, and of the parties, to remain unchanged, then their forecast of the share of the vote going to party R is given by

$$\hat{V} = \sum_{g=1}^{G} \hat{n}_g p_g(x).$$

We can think of $\hat{V}$ as a *pure demographic forecast* because it envisions that the only change between the current and future elections will be in the composition of the voting population.

We will focus on pure demographic forecasts in our empirical analysis. Pure demographic forecasts correspond to baseline forecasts in Griffin et al. (2018, see Table 2 and Figure 2) and Griffin et al. (2020, see Table 3 and Figure 5).[7] Of course, a forecaster might combine a forecast of demographic trends with a forecast of other trends, for example, trends in groups' support for each party.[8] Although not our focus, we will consider some particular examples of such forecasts in our empirical analysis.

To measure the extent to which the forecaster expects demographics to change the vote share, we can define the forecast shift $\hat{V} - V$. When the forecast shift is small, demographic change alone is forecast to have only a small effect on the vote share. We will use the forecast shift in our empirical analysis to compare the extent of forecast demographic change across different forecasting models.

To evaluate the quality of the forecast $\hat{V}$, we can compare it to the realized outcome in the future election. Suppose that, in this future election, the parties have attributes $x^* \in \mathcal{X}$ and that group g, now comprising share n_g^* of voters, votes for party R with probability $p_g^*(x^*)$. We then have that the expected share of the total vote going to party R is

$$V^* = \sum_{g=1}^{G} n_g^* p_g^*(x^*).$$

The error $\hat{V} - V^*$ measures how far above or below the realized outcome is the demographic forecast. We can decompose the error as

$$\hat{V} - V^* = \underbrace{\sum_{g=1}^{G} \left(\hat{n}_g - n_g^*\right) p_g^*(x^*)}_{\text{demography}} + \underbrace{\sum_{g=1}^{G} \hat{n}_g \left(p_g(x) - p_g^*(x)\right)}_{\text{voters}} + \underbrace{\sum_{g=1}^{G} \hat{n}_g \left(p_g^*(x) - p_g^*(x^*)\right)}_{\text{parties}} \tag{1}$$

where the first term measures the error in the demographic forecast, the second the change in voters' behavior, holding party attributes fixed, and the third the change in party at-

[7] Such forecasts are also in the spirit of Phillips (1969) and Judis and Teixeira (2002), though these books do not make quantitative forecasts.

[8] Griffin et al. (2018) and Griffin et al. (2020), for example, consider various scenarios for how different groups will vote in the future.

tributes, holding voter behavior fixed.

Equation (1) illustrates the three pitfalls of pure demographic forecasts of elections. The first pitfall is that the forecast demographic trends do not come to pass. In our empirical analysis, we will model a forecaster who has perfect foresight of future demographic trends, and thus avoids this pitfall altogether.

The second pitfall is that voter behavior changes. In our empirical analysis, we will study changes over time in how different groups of voters vote, and in the issue positions they express.

The third pitfall is that party attributes change. Canonical models of political competition (Hotelling, 1929; Downs, 1957; Becker, 1958) tell us to expect such changes.[9] In our empirical analysis, we will study how changes in party positions relate to changes in voter demographics.

3 Data for Demographic Forecasting

We conduct our main analysis on US presidential elections from 1952 through 2020. 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 additional analyses.

3.1 Data for Main Forecasting Exercises

Our main analysis uses data from the American National Election Study (ANES) Time Series Cumulative Data File (2022). 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 6119 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

[9]Online Appendix A shows that, in a symmetric model of party competition for votes, we can expect that $V = V^* = 1/2$, and therefore any forecast $\hat{V} \neq 1/2$ will be systematically wrong.

candidate, or not giving a valid response to the question). For comparison, we also obtain official election results for our sample period.[10]

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.[11] 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, homemaker, student, or retired), occupation group (professional and managerial; clerical and sales worker; skilled, semi-skilled, and service worker; laborer; farmer, forestry, and fisherman; homemaker), religion (Roman Catholic, Protestant, Jewish, or other), religious participation (three categories 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 use finer categories for age (replacing 10-year bins with 5-year bins) and race (adding categories for Asian and Native American). 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.

All demographic covariates enter our analysis as category indicators ("one-hot-encodings"). We explore specifications that allow rich interactions among these. Our analysis sample ex-

[10] These are 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 1976-2020.

[11] 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; 2024).

cludes any respondent for whom data are missing on any covariate. We 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 Data for Additional Forecasting Exercises

We conduct a number of additional forecasting exercises. In one additional exercise, we repeat our main analysis using county-level data on voting and demographic covariates through 2020. 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 additional exercise, 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.[12]

In a third additional exercise, 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).[13] Therefore, we do not include this variable as a demographic covariate and treat it instead as an alternative outcome to vote choice.[14]

3.3 Data on Voter and Party Issue Positions

Motivated in part by the framework in Section 2, we study how parties respond to demographic change. To measure parties' positions, we use 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

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

[13] 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.

[14] In additional sensitivity analysis, we forecast the Electoral College outcome, using data on the Electoral College from CQ Press (2025), with minor adjustments based on other public sources.

party's platform in the given election year.[15] A measure of zero means the platform includes equal numbers of both types of sub-sentences.

We are also interested in the positions that parties and voters take on individual issues. We identified 11 issues for which it is possible to measure parties' positions in the Manifesto Project database and on which the ANES has conducted consistent surveying for a large portion of our sample period. These issues are environmental protection, government strength, internationalism, labor, law and order, market regulation, military, minority rights, protectionism, traditional morality, and welfare. Online Appendix Table 4 details how we assign these issues to topics coded in the Manifesto Project and to survey questions in the ANES.

To measure each party's position on each issue in each election year, we take from the Manifesto Project 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. To measure each ANES respondent's position on each issue, we construct an indicator equal to one when the respondent expresses the traditionally right-wing position on the given issue, and zero otherwise. Online Appendix Table 4 additionally describes measures of the emphasis on each issue in party platforms, and the importance assigned to each issue by voters, which we use in sensitivity analysis.

4 Methods for Forecasting and Evaluation

Here we lay out our methods for forecasting and evaluation. Where possible, we follow notation analogous to that in Section 2.

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 our analysis of aggregate data. Let $v_{it} \in \{0, 1\}$ denote whether respondent i reports voting Republican in election year t. Let $\mathbf{d}_{it}$ be a vector of demographic indicators.[16]

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

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

The first step is to specify and estimate a model of contemporaneous voter behavior in each election t. Given a vector of covariates $\mathbf{d}_{it}$, we can write any such model as

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

where $p\left(\cdot;\cdot\right)$ is a function known up to the election-specific parameter θ_t, which may implicitly depend on aggregate factors such as the parties' attributes x. In our main analysis, we assume that $p\left(\cdot;\cdot\right)$ is logistic and estimate θ_t via maximum likelihood.[17] In additional analyses, we assume that $p\left(\cdot;\cdot\right)$ is an average of regression trees and estimate θ_t via the random forest algorithm.[18]

The next step is to extrapolate the model of voter behavior in election t to forecast the election at some horizon $h>0$, i.e., in some future election year $t+h$. Given an estimate $\hat{\theta}_t$ of the parameter 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.[19] Taking a sample average of these probabilities yields a pure demographic 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.

The forecast $\hat{V}_{t,t+h}$ entirely avoids the first pitfall described in Section 2. 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. Notice, in particular, that this means that the forecast accounts for changes in the share of voters in a group (e.g., women) whose share in the full population does not change much over time.

[17] When all respondents in a given covariate category vote for the same party, we exclude respondents in that category from maximum likelihood estimation, and use their empirical frequency of voting Republican as the prediction function. In sensitivity analysis, we present results where we instead estimate via penalized maximum likelihood using the Jeffreys prior penalty suggested by Firth (1993; see also Mansournia et al., 2018; Kosmidis and Firth, 2021).

[18] 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.

[19] To ensure that this calculation is well-defined, for the purpose of estimating θ_t, we include in $\mathbf{d}_{it}$ only those covariates that are available in all elections $t, t+1, ...$ for which we calculate a demographic forecast.

4.2 Measuring Forecast Performance

We can 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, 1993; Atkeson, 1999). For completeness, we also present results based on official election returns.[20]

To measure forecast performance, recall from Section 2 that the error of a given forecast $\hat{V}_{t,t+h}$ is its difference from the realized result, $\hat{V}_{t,t+h} - V^*_{t+h}$. To summarize the magnitude of the error across multiple elections, we calculate the root mean squared error (RMSE),

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

where the average is taken over the set T_h of election years for which we observe the realized result at horizon h. The RMSE is a standard metric of performance for point forecasts (e.g., Hewamalage et al., 2023).

Because 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 captures individual voting behavior in the election year on which the model is estimated. For a given sample I of respondents in

[20] 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.

election t, the RMSE of the model's fit to individual respondents' votes is

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

We estimate this RMSE using 10-fold cross-validation to guard against overfitting. We define a given model's *within-election* error as the ratio of this estimated RMSE, averaged across elections, to the counterpart for a constant model that fits each respondent's vote with the sample mean vote in the same election.[21] When the within-election error is one, the model in question fits respondents' votes no better than the constant model. When the within-election error is zero, the model in question fits each respondent's vote perfectly. In either case, the model in question 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 due to demographic change. As of election t, at horizon h, a given model predicts a demographic *shift* 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 summarize the magnitude of the demographic shifts predicted by a given model, at some given horizon h, by their root mean square,

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

When the root mean square 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 quantities that depend on the survey sample, we communicate uncertainty by reporting a 95 percent posterior credible interval based on 1000 replicates of a Bayesian bootstrap procedure.[22] A 95 percent posterior credible interval contains the true value of the given quantity with probability 0.95 under a given posterior distribution. In this case, the true value is the one that we would calculate if the ANES had surveyed the entire US

[21] That is, the constant-only model fits v_{it} with $\frac{1}{|I|}\sum_{i\in I} v_{it}$.

[22] 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.

population, and the posterior distribution corresponds to a particular class of uninformative Dirichlet process priors (see, e.g., Rubin, 1981; Gasparini, 1995; Chamberlain and Imbens, 2003; Andrews and Shapiro, 2025). Importantly, and consistent with our aim of backtesting demographic forecasts, when we communicate uncertainty about forecast performance, we are communicating uncertainty about how a given forecast *would have* performed over our sample period, rather than about how that same forecasting method *will* perform in the future.

5 Evidence on the Performance of Demographic Forecasts

In this section we lay out our findings on the performance of demographic forecasts in our main specification and many additional specifications. We defer issues of interpretation to later sections. We focus in the main text on visual summaries of our findings; Online Appendix Table 3 gives more precise magnitudes for the plotted values.

5.1 Main Findings

Figure 1 presents our main findings on the performance of demographic forecasts of US presidential elections. 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. 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. The RMSE of the current forecast is therefore both a useful benchmark of forecast performance and a direct measure of the variability of vote shares between elections. 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 therefore does not depend on the forecast horizon. 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 22.1 percent worse than the even split forecast. At longer horizons, forecast performance is no better. The shaded regions represent 95 percent 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 6.9 percent lower than that of a constant model that fits each voter's vote with the sample mean vote in the same election, indicating that the logit model has modest but nontrivial fit to individual voting behavior in the election in which it is estimated. The average shift at a one-election horizon is about 12.6 percent of the RMSE of the current forecast, 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. Because these elections do not have an incumbent on the ballot, votes may depend more on party preferences than on evaluations of the incumbent's performance, potentially increasing the impact of long-run demographic trends. Demographic forecasts perform even more poorly in this set of elections than in the main specification.

Online Appendix Figure 2 compares forecast performance in the first half of our sample period (from 1952 to 1984) to that in the second half of our sample period (from 1988 to 2020). Elections were easier to forecast in the second half of our sample period, in part because many elections were close to an even split. Even in the second half of our sample period, though, demographic forecasts do not convincingly outperform our benchmarks.

5.2 Findings from Richer Predictive Models

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 5.9 percent relative to the model in Panel (a).[23] Including these additional covariates does not yield a clear improvement in forecast accuracy, however.

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 6.4 percent relative to the model in Panel (b), but again, does not improve forecast accuracy.

To summarize, even a model that performs 18 percent better than a constant-only model in fitting individual voting within an election performs no better than the current or even-split forecast in forecasting future elections. Moreover, enriching the models to improve within-election fit does not tend to improve their forecasting performance.

Online Appendix D presents additional results that further probe in this direction. To allow for regional heterogeneity, Online Appendix Figure 4 shows results where we enrich our main specification by allowing the main demographic covariates to have different coefficients for different Census regions. This change reduces the within-election error relative to the main specification, but does not yield any meaningful improvement in forecast accuracy. To guard against bias and overfitting, Online Appendix Figure 5 re-estimates both our baseline specification and our specification with the extended set of demographic covariates using penalized maximum likelihood with either a Jeffreys (Firth) or L1 (lasso) penalty. Forecast accuracy is not improved relative to our baseline specification. Finally, because incorporating more demographic covariates creates more scope for missing data, Online Appendix Figure 6 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.

[23] Online Appendix Figure 3 shows the relative importance of each demographic factor to the within-election performance of the model in the specification of Panel (b).

5.3 Findings for Alternative Outcomes

Figure 3 presents our findings on the performance of demographic forecasts of alternative outcomes: vote shares in congressional elections, and party identification. Panel (a) of Figure 3 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 3 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 traditional numerical advantage of the Democrats in two-party identification (e.g., Converse et al., 1961, p. 274; see also Burden and Klofstad, 2005). The 40-60 split performs slightly worse than the current forecast and, at many horizons, better than the demographic forecast.

Online Appendix Figure 7 presents results for forecasting the Republican share of the Electoral College. Demographic forecasts are no better, and often worse, than the current forecast, and meaningfully worse than the even split forecast.

6 Interpreting the Failure of Demographic Forecasts

6.1 Small Demographically Driven Shifts in Vote Shares

One clue to the poor performance of demographic forecasts is in the small value of the shift reported in Panel (a) of Figure 1. The estimated shift implies that, between adjacent elections, demographic change alone is forecast to shift the Republican share of the two-party vote by only 12.6 percent of the RMSE of the current forecast. The RMSE of the current forecast, in turn, measures the variability in the vote share between adjacent elections.

Panel (a) of Figure 4 illustrates the implications of the small estimated shift. The plot shows the realized election result, current forecast based on one election prior, and demographic forecast based on one election prior (using our main specification). Because the shift is small, the demographic forecast deviates little from the current forecast. The plot also shows the difference in error between the current and demographic forecasts. Because the demographic forecast is not systematically closer to the realized vote share than the current forecast, the demographic forecast performs no better than the current forecast.

Intuitively, the shift reflects both the pace of demographic change and the strength of the estimated relationship between demographics and voting. Perhaps if we look over a period with larger demographic change, or estimate a predictive model that implies a stronger relationship between demographics and voting, demographic forecasts will perform better.

To see how demographic forecasts perform when demographic change is larger, we can extend the forecasting horizon. Panel (b) shows the realized election result, the result from three elections prior, and the demographic forecast based on three elections prior. At a three-election horizon, the demographic forecast is more different from the past election result than at a one-election horizon, though the two remain similar.[24] If anything, though, the demographic forecast performs worse over the longer horizon (see again Panel (a) of Figure 1).

To see how demographic forecasts perform when the predictive model implies a stronger relationship between demographics and voting, we can estimate the predictive model on aggregate data. We use county-level data to fit a logistic regression of the Republican share of the two-party vote on a vector of population shares analogous to the individual-level demographic covariates that we use in our main specification. Past evidence suggests that such a model will imply large explanatory power for demographics.[25]

As expected, the county-level logistic regression tends to imply a stronger relationship between demographics and voting than does our main specification. For example, in 2020, 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 73 percentage points, as against 22 percentage points for our main specification.[26]

[24] The root mean square shift is 25 percent greater at a three-election horizon than at a one-election horizon.

[25] Again see, for example, Levernier and Barilla (2006), Freund and Sidhu (2017), Smith (2021), and Brown and Mettler (2024).

[26] For changing all adults from high-school graduates to college graduates, across all elections, the average

Such differences are in line with the literature on ecological fallacy (e.g., Robinson, 1950; Freedman, 1999), which notes that sorting of individuals across areas can lead to aggregate relationships that are much stronger than their individual-level counterparts.

The county-level logistic regression implies a shift of 66.4 percent of the RMSE of the current forecast, which is much larger than the 12.6 percent estimated in our main specification.[27] Panels (c) and (d) of Figure 4 illustrate the size of the shift by plotting the analogues of Panels (a) and (b) but using the county-level regression as the predictive model. We see that the county-level logistic regression implies large demographically-driven changes in the two-party vote share even on a one-election horizon, and even larger changes at a three-election horizon.

Because these large changes do not manifest in the realized election results, forecasts based on the county-level logistic regression perform poorly. Panel (b) of Online Appendix Figure 8 reports that forecasts based on the county-level logistic regression perform consistently worse than both the current and even split forecasts, with an RMSE more than 3.2 times that of the current forecast at the longest horizon we consider.[28]

To summarize, demographic forecasts perform poorly even—and perhaps especially—when they imply large demographically driven changes in vote shares. Small implied shifts are therefore only part of the story of their poor performance. To understand the rest, we turn to the second two pitfalls highlighted in Equation (1).

6.2 Changing Voter Behavior

The second pitfall highlighted in Equation (1) is changing voter behavior, defined as a change in groups' propensity to vote Republican, holding party attributes fixed.

To study this pitfall, we construct a measure of the stability of the function $p(\mathbf{d};\theta)$ across elections. Specifically, we calculate the correlation between $p\left(\mathbf{d}_{it};\hat{\theta}_t\right)$ and $p\left(\mathbf{d}_{it};\hat{\theta}_{t+h}\right)$

ratio of the absolute change between the county-level and individual-level models is 3.2. For changing all adults from urban to rural residents the average ratio is 0.6 and for changing all adults from male to female it is 33.8.

[27] Similarly, the county-level logistic regression has a within-election error 33.1 percent below that of a constant-only model, as compared to 6.9 percent for our main specification.

[28] Panel (c) of Online Appendix Figure 8 presents results from a different approach to forecasting with aggregate data. There, we mimic 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. This forecaster performs no better than the current forecast, and worse than the even split forecast.

for different elections t and horizons h, and take the average correlation at each horizon h as our measure of the stability of the function at that horizon. To gauge the importance of the richness of the demographic covariates, we calculate the correlation both for the predictive model using the main demographic covariates and for the predictive model using the extended demographic covariates.

Panel (a) of Figure 5 shows our findings. The correlation for the current election, horizon $h = 0$, is mechanically one.[29] Voter behavior changes meaningfully at a horizon of one election, with correlations of 0.82 (using the main demographic covariates) or 0.77 (using the extended demographic covariates) at that horizon. The correlation declines more gradually as we move to longer forecasting horizons, reaching values of 0.72 and 0.67, respectively, at a five-election horizon. The correlation is consistently lower for the model using the extended demographic covariates than for the model using the baseline demographic covariates.

These findings suggest several reasons why demographic forecasts are poor. One is that the relationship between demographics and voting is partly transitory, changing meaningfully even between adjacent elections. Another is that, even apart from this transitory component, the relationship tends to be less stable at longer forecasting horizons. And a third is that the relationship is less stable when we use richer demographic covariates, suggesting a tradeoff between the richness of the predictive model and the usefulness of its predictions.

A limitation of the evidence in Panel (a) of Figure 5 is that the function $p(\mathbf{d}; \theta)$ implicitly depends on both the behavior of voters and the attributes of parties, which may enter the function indirectly via the parameter θ_t. One way to limit the role of the parties is to model issue positions rather than voting.[30] To do this, we estimate versions of our predictive models but treating as the outcome variable the respondent's position (right or left) on each of a set of policy issues. We again calculate the correlation in the predicted val-

[29] The correlation at horizon $h = 0$ is inflated by overfitting. In practice we find that the effect of overfitting is small, with a correlation at horizon $h = 0$ of 0.98 (main demographic covariates) or 0.97 (extended demographic covariates) when calculated using 10-fold cross-validation.

[30] Changes in party positions may lead voters to change which party they identify with and which candidate they vote for while leaving their own issue positions unaffected (Downs, 1957; Jackson, 1975; Fiorina, 1981; Franklin and Jackson, 1983). In practice, the relationship between voters' issue positions and party identification runs both ways (Campbell et al., 1960; Converse and Markus, 1979; Goren, 2005; Converse, 2006; Broockman and Butler, 2015; Highton and Kam, 2016).

ues over different elections and horizons, and, for each horizon, average these correlations across elections and then across issues.

Panel (b) of Figure 5 shows our findings. We find that, if anything, the relationship between demographic covariates and issue positions is less stable than the relationship between demographic covariates and voting. This suggests that the instability of the relationship between demographic characteristics and voting is not simply a result of changing party attributes, but rather reflects a more fundamental instability in the political orientation of different demographic groups.

A forecaster aware of this instability might seek to address it by using a more sophisticated forecast. We consider two sets of experiments in that direction. In the first set of experiments, the forecaster assumes that the future value of θ_t will be given by the average estimated value of θ_t over the current and K most recent elections, so that $\hat{\theta}_{t+h} = \frac{1}{K}\sum_{k=0}^{K}\hat{\theta}_{t-k}$. These experiments address the transitory nature of the relationship between demographics and voting by averaging over multiple elections. In the second set of experiments, the forecaster assumes that the value of θ_t will evolve linearly in the future, and estimates the rate of change based on the difference $\left(\hat{\theta}_t - \hat{\theta}_{t-k}\right)$ between the current estimated value and its value in the k^{th} preceding election. These experiments address the evolving nature of the relationship between demographics and voting by extrapolating a trend from past elections.

Online Appendix Figure 9 presents the results of these experiments. We find that the first set of experiments improve the forecasts marginally at best, and the second set of experiments make them worse.

To summarize, changing voter behavior appears to be an important pitfall of demographic forecasts. And, it is not one that a forecaster can easily avoid.

6.3 Changing Party Attributes

The third and final pitfall highlighted in Equation (1) is changing party attributes. Classical models of electoral competition imply that, if parties are sufficiently responsive to voter demographics, demographic forecasts will fail systematically (see Online Appendix A).

There is good evidence that the parties respond to changing voter demographics, and even to forecast changes. In the *Emerging Republican Majority*, Phillips (1969) argued

that the rise of the suburbs would benefit the Republicans. Grohsgal and Kruse (2019) describe how Phillips influenced the Nixon campaign, and Lassiter (2003) documents that both parties responded to the rise of the suburbs, with the Democrats in particular ultimately repositioning themselves to appeal to suburban voters.[31]

Similarly, in the *Emerging Democratic Majority*, Judis and Teixeira (2002) argued that the rising Latino share of the population would benefit the Democrats. Following its loss in the 2012 presidential election, the Republican party launched its *Growth and Opportunity Project* (Barbour et al., 2013) with a charge to "grow the Party and improve Republican campaigns." Citing projections that whites would be in the minority by 2050, the project's 2013 report declared that "America is changing demographically" and that "the changes tilt the playing field even more in the Democratic direction," making it essential for the party to connect with minority voters, and Latino voters in particular (Barbour et al., 2013, p. 7).[32]

For a more quantitative view, we turn next to studying the connection between demographic trends, as measured by our predictive model, and party platforms, as coded by the Manifesto Project (Lehmann et al., 2023). To measure the overall demographic shift towards Republicans implied by our predictive model, we calculate the cumulative shift

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

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.

Panel (a) of Figure 6 plots the estimated value of Δ_τ over the sample period. The plot

[31] 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 districts and the fiscally responsible managers of the 'new economy' has revitalized the competitiveness of the center in a postliberal political order" (2003, p. 328).

[32] The report stated, "The Republican Party must focus its efforts to earn new supporters and voters in the following demographic communities: Hispanic, Asian and Pacific Islanders, African Americans, Indian Americans, Native Americans, women, and youth. This priority needs to be a continual effort that affects every facet of our Party's activities, including our messaging, strategy, outreach, and budget" (Barbour et al., 2013, p. 12). A 2015 follow-up reported progress in these efforts (Priebus and Day, 2015).

shows a cumulative electoral advantage to Republicans peaking in 1976 at 6.7 percentage points and ending the sample period at about 1.2 percentage points, though with large statistical uncertainty. According to the estimates, Phillips (1969) published as Republicans' advantage waxed, and Judis and Teixeira (2002) as it waned.[33]

Panel (b) of Figure 6 plots the left-right position of each party over the sample period as measured by the Manifesto Project (Lehmann et al., 2023). According to this measure, 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 three election cycles of the sample period. The overall trend appears consistent with the parties reacting, albeit with delay, to the demographic shift visible in Panel (a). The plot also suggests that the parties moved more nimbly in their ideologically preferred directions, with the Republicans moving to the right in line with demographic trends, and ahead of the Democrats, and the Democrats later moving back to the left, again in line with demographics, and without a corresponding move by the Republicans during the sample period.

In addition to adjusting their overall ideological orientation, the parties can react to demographic change by adjusting their positions on particular policy dimensions, such as quality-of-life or fiscal responsibility in the case of suburban voters (Lassiter, 2003), or family values in the case of Latino voters (Carranza, 2024). To measure such adjustments, we conduct an exercise analogous to the one depicted in Figure 6, this time looking across policy issues. For each issue, we estimate a version of our predictive model but treating as the outcome variable the respondent's position (right or left) on the issue. We then compute an analogue of the cumulative demographic shift Δ_τ through the 2020 election for each issue. We scale this measure by the number of decades in the sample so that it can be interpreted as a measure of how much voters' positions would have moved to the right on the given issue, in an average decade, if the model provided an accurate forecast. We compare this measure to a measure of the trend in each party's position on the issue, computed as the coefficient from a linear regression of the party's measured position on a decadal time trend.[34]

[33] Online Appendix Figure 10 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 urbanism would have forecast a more leftward shift over our sample period. A forecaster ignoring the role of race would have forecast a more rightward shift. These findings align with the emphases of Phillips (1969) and Judis and Teixeira (2002), respectively.

[34] Letting x_{jt} denote party j's position on the given issue in election t, we estimate an ordinary least squares

Panel (a) of 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). We show the trend in party positions separately for the two parties, as well as pooled together. We observe an overall positive relationship, with the parties tending to move to the right on issues on which the voters are also predicted to have done so. The pooled measure of the trend in party positions has a Spearman rank correlation of 0.50 with the demographic shift in voters' positions, though with wide statistical uncertainty. We also again encounter evidence that parties move more nimbly in their preferred direction, with the Republicans tending to move more aggressively on the issues in the upper right of the plot, and Democrats even more so on the issues in the lower left. Notice that, because the x-axis variable is determined entirely by demographics, reverse causality does not seem like a plausible explanation of the patterns in the plot.

The simplest interpretation of the evidence in Panel (a) of Figure 7 is that demographic change brought about changes in Americans' policy preferences, and the parties adjusted. The story is not quite that simple, however. Panel (b) of Figure 7 replaces the y-axis variable with the estimated time trend in voters' realized positions on each issue, computed as the coefficient from a linear regression of the share of voters expressing a right-wing view of the issue on a decadal time trend. If the changes in issue positions predicted from the demographic model came to pass, we would observe a positive relationship between the realized (y-axis) and predicted (x-axis) trends in voters' issue positions. If anything, the relationship is slightly negative.

We see two plausible interpretations of this pattern. One interpretation is that the parties overreacted to demographic change, and that demographic forecasts are no better at predicting voters' issue positions than at predicting their votes. Consistent with this interpretation, Online Appendix Figure 11 shows that demographic forecasts of issue positions tend to perform poorly (see also again Panel (b) of Figure 5.)

Another interpretation is that the demographic trends reflect real change in voters' issue positions that do not manifest in voters' stated positions. A possible reason for this pattern could be that voters' stated issue positions reflect their preferences relative to prevailing policies or to the parties' platforms, such that they are not a reliable guide to the absolute policy preferences that drive party platforms. Consistent with this interpretation, Online

regression of x_{jt} on t, and scale the resulting coefficient to be in units of change per decade.

Appendix Table 4 reports that some of the survey questions assessing issue positions explicitly ask for the respondent's opinion relative to the status quo.[35]

Online Appendix D presents alternative versions of Figure 7. Online Appendix Figure 12 uses an alternative measure of party positions and reaches qualitatively similar conclusions. Online Appendix Figure 13 focuses on the importance that parties and voters assign to different issues, rather than the positions they take on these issues. These data are more sparse but are consistent with parties putting more emphasis on issues that are predicted to become more important to voters.

In addition to adjusting their policy platforms, the parties may adjust their voter outreach to target growing groups of the electorate. Online Appendix Figure 14 presents suggestive, albeit statistically imprecise, evidence that the parties increasingly target members of growing groups, consistent with the narrative evidence that we discuss above.

To summarize, the parties appear to adjust their attributes in the direction in which voters are predicted to move based on demographic trends. Whether parties' adjustments are warranted, or instead represent an excessive response to forecast changes in voter opinion, is an open question. Either way, parties' adjustments will tend to further undermine demographic forecasts.

7 Conclusion

Demographic forecasts have long been the main method for forecasting US electoral trends over long horizons and, as such, an important input into party strategy. These forecasts embed an important hypothesis about politics, namely that demographic groups' partisan leanings are more stable than their relative sizes.

We find that demographic forecasts perform poorly even when the forecaster has perfect knowledge of future demographic trends. Demographic forecasts fare worse than simply predicting that every presidential election will be an even contest, and no better than guessing that future election results will mirror current ones. These forecasts are poor irrespective of the set of demographic covariates and the functional form we use to explain and predict vote choices.

[35] Even for questions without an explicit reference to the status quo, respondents may interpret such questions relative to the status quo (Franklin and Wlezien, 1997; Jennings, 2009; Wlezien, 2017, pp. 407-409).

The inaccuracy of short-term demographic forecasts is perhaps not surprising, as demographic change may be swamped by other factors that vary from election to election. We find, though, that demographic forecasts perform poorly even when they imply large demographically driven political changes across elections. Even more importantly, demographic forecasts do not perform better over the long term. Changing voter preferences and changing party platforms counteract the electoral swings that demographic trends would otherwise predict.

We conclude that demographic forecasts of the sort we test are not useful for planning—and caution against assuming that demography is political destiny.

References

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

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.

Aistrup, J. A. (2014). *The Southern Strategy Revisited: Republican Top-Down Advancement in the South*. University Press of Kentucky.

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 (2022). ANES Time Series Cumulative Data File (1948-2020). `https://web.archive.org/web/20240926231149/https://electionstudies.org/data-center/anes-time-series-cumulative-data-file/`. Accessed in Sep 2024.

Andrews, I. and J. M. Shapiro (2025). Communicating scientific uncertainty via approximate posteriors. *NBER Working Paper* (32038).

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.

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.

Barbaro, M., N. Cohn, L. Lerer, and A. W. Herndon (2024, Nov). Trump, again. `https://www.nytimes.com/2024/11/06/podcasts/the-daily/trump-election.html`. Accessed in April 2025.

Barbour, H., S. Bradshaw, A. Fleischer, Z. Fonalledas, and G. McCall (2013, March). Growth & opportunity project. Technical report, Republican National Committee. `https://www.wsj.com/public/resources/documents/RNCreport03182013.pdf`. Accessed in April 2025.

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

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.

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, E. and M. Black (2009). *The Rise of Southern Republicans*. Harvard University Press.

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.

Broockman, D. E. and D. M. Butler (2015). The causal effects of elite position-taking on voter attitudes: Field experiments with elite communication. *American Journal of Political Science 59*(4).

Brown, T. E. and S. Mettler (2024). Sequential polarization: The development of the rural-urban political divide, 1976–2020. *Perspectives on Politics 22*(3), 630–658.

Buchanan, P. J. (2015). *The Greatest Comeback: How Richard Nixon Rose from Defeat to Create the New Majority*. PRH Christian Publishing.

Burden, B. C. and C. A. Klofstad (2005). Affect and cognition in party identification. *Political Psychology 26*(6), 869–886.

Camatarri, S. (2024). Predicting popular-vote shares in US presidential elections: A model-based strategy relying on ANES data. *PS: Political Science & Politics*, 1–5.

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.

Carnes, N. and N. Lupu (2021, March). The white working class and the 2016 election. *Perspectives on Politics 19*(1), 55–72.

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.

Carville, J. (2009). *40 More Years: How the Democrats Will Rule the Next Generation*. Simon and Schuster.

Cascio, E. U. and E. Washington (2014). Valuing the vote: The redistribution of voting rights and state funds following the Voting Rights Act of 1965. *Quarterly Journal of Economics 129*(1), 379–433.

Chamberlain, G. and G. W. Imbens (2003). Nonparametric applications of Bayesian inference. *Journal of Business and Economic Statistics 21*(1), 12–18.

Chernov, M., V. Elenev, and D. Song (2025). The comovement of voter preferences: Insights from US presidential election prediction markets beyond polls. *NBER Working Paper* (33339).

Cohn, N. (2016, November). Why Trump won: Working-class whites. *The New York Times*. https://www.nytimes.com/2016/11/10/upshot/why-trump-won-working-class-whites.html. Accessed in April 2025.

Converse, P. E. (2006). The nature of belief systems in mass publics (1964). *Critical Review 18*(1-3), 1–74.

Converse, P. E., A. Campbell, W. E. Miller, and D. E. Stokes (1961). Stability and change in 1960: A reinstating election. *American Political Science Review 55*(2), 269–280.

Converse, P. E. and G. B. Markus (1979). Plus ça change...: The new CPS election study panel. *American Political Science Review 73*(1), 32–49.

CQ Press (2025). Voting and elections collection. https://library.cqpress.com/elections/#_=_. Accessed in May 2025.

Davis, L. J. (1974). *The Emerging Democratic Majority: Lessons and Legacies From the New Politics*. Stein and Day.

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.

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.

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.

Ferrara, A., P. A. Testa, and L. Zhou (2022). New area- and population-based geographic crosswalks for U.S. counties and congressional districts, 1790-2020. Inter-university Consortium for Political and Social Research (ICPSR). V4. `https://doi.org/10.3886/E150101V4`. Accessed in April 2024.

Ferrara, A., P. A. Testa, and L. Zhou (2024). New area- and population-based geographic crosswalks for U.S. counties and congressional districts, 1790-2020. *Historical Methods: A Journal of Quantitative and Interdisciplinary History 57*(2), 67–79. Accessed in April 2024.

Fiorina, M. (1981). *Retrospective Voting in American National Elections*. Yale University Press.

Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika 80*(1), 27–38.

Franklin, C. H. and J. E. Jackson (1983). The dynamics of party identification. *American Political Science Review 77*(4).

Franklin, M. and C. Wlezien (1997). The responsive public: Issue salience, policy change, and preferences for European unification. *Journal of Theoretical Politics 9*(3), 347–363.

Freedman, D. A. (1999). Ecological inference and the ecological fallacy. *International Encyclopedia of the Social & Behavioral Sciences 6*(4027-4030), 1–7.

Freund, C. and D. Sidhu (2017). Manufacturing and the 2016 election: An analysis of US presidential election data. *Peterson Institute for International Economics*.

Fujiwara, T. (2015). Voting technology, political responsiveness, and infant health: Evidence from Brazil. *Econometrica 83*(2), 423–464.

Gasparini, M. (1995). Exact multivariate Bayesian bootstrap distributions of moments. *Annals of Statistics 23*(3), 762–768.

Gelman, A. (2014). The twentieth-century reversal: How did the Republican states switch to the Democrats and vice versa? *Statistics and Public Policy 1*(1), 1–5.

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.

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.

Gerbaud, G., C. Harrison, and K. Robertson (2024, November). How Latinos voted in the 2024 U.S presidential election. *Americas Society/Council of the Americas*. `https://www.as-coa.org/articles/how-latinos-voted-2024-us-presidential-election`. Accessed in April 2025.

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.

Glaeser, E. L., G. A. M. Ponzetto, and J. M. Shapiro (2005). Strategic extremism: Why Republicans and Democrats divide on religious values. *Quarterly Journal of Economics 120*(4), 1283–1330.

Goren, P. (2005). Party identification and core political values. *American Journal of Political Science 49*(4), 881–896.

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

Greenberg, S. B. (2015). *America Ascendant: A Revolutionary Nation's Path to Addressing Its Deepest Problems and Leading the 21st Century*. Thomas Dunne Books.

Griffin, R., W. H. Frey, and R. Teixeira (2020, October). America's electoral future. *Center for American Progress*. https://www.americanprogress.org/article/americas-electoral-future-3/. Accessed in April 2025.

Griffin, R., R. Teixeira, and W. H. Frey (2018, April). America's electoral future. *Brookings Institution*. https://www.brookings.edu/articles/americas-electoral-future_2018/. Accessed in April 2025.

Grimmer, J., D. Knox, and S. Westwood (2024). Assessing the reliability of probabilistic US presidential election forecasts may take decades. *Stanford University Working Paper*. https://doi.org/10.31219/osf.io/6g5zq. Accessed in Sep 2024.

Grohsgal, D. and K. M. Kruse (2019, August). How the Republican majority emerged. *The Atlantic*. https://www.theatlantic.com/ideas/archive/2019/08/emerging-republican-majority/595504/. Accessed in May 2025.

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.

Helderman, R. S. and J. Cohen (2012, Aug). As Republican convention emphasizes diversity, racial incidents intrude. *The Washington Post*. https://www.washingtonpost.com/politics/2012/08/29/b9023a52-f1ec-11e1-892d-bc92fee603a7_story.html. Accessed in May 2025.

Hewamalage, H., K. Ackermann, and C. Bergmeir (2023). Forecast evaluation for data scientists: Common pitfalls and best practices. *Data Mining and Knowledge Discovery 37*(2), 788–832.

Highton, B. and C. D. Kam (2016). The long-term dynamics of partisanship and issue orientations. *Journal of Politics*.

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.

Jackson, J. E. (1975). Issues, party choices, and presidential votes. *American Journal of Political Science 19*(2), 313–333.

Jennings, W. (2009). The public thermostat, political responsiveness and error correction: Border control and asylum in Britain, 1994-2007. *British Journal of Political Science 39*(4), 847–870.

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.

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.

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.

Kosmidis, I. and D. Firth (2021). Jeffreys-prior penalty, finiteness and shrinkage in binomial-response generalized linear models. *Biometirka 108*, 71–82.

Kuziemko, I. and E. Washington (2018). Why did the Democrats lose the South? Bringing new data to an old debate. *American Economic Review 108*(10), 2830–67.

Lamont, M., B. Y. Park, and E. Ayala-Hurtado (2017). Trump's electoral speeches and his appeal to the American white working class. *The British Journal of Sociology 68*(S1), S153–S180.

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.

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.

Levernier, W. and A. Barilla (2006, Dec). The effect of region, demographics, and economic characteristics on county-level voting patterns in the 2000 presidential election. *Review of Regional Studies 36*, 427–447.

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

Lizza, R. (2012, Oct). The final push. *The New Yorker*. `https://www.newyorker.com/magazine/2012/10/29/the-final-push`. Accessed in May 2025.

Longuet-Marx, N. (2025). Party lines or voter preferences? Explaining political realignment. `https://nicolaslonguetmarx.github.io/PartyLines_NLM.pdf`. Accessed in April 2025.

Mansournia, M. A., A. Geroldinger, S. Greenland, and G. Heinze (2018). Separation in logistic regression: causes, consequences, and control. *American Journal of Epidemiology 187*(4), 864–870.

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.

Miller, G. (2008). Women's suffrage, political responsiveness, and child survival in American history. *Quarterly Journal of Economics 123*(3), 1287–1327.

MIT Election Data and Science Lab (2017). U.S. President 1976-2020. Harvard Dataverse. V8. https://doi.org/10.7910/DVN/42MVDX. Accessed in February 2024.

Morgan, S. L. and J. Lee (2018, April). Trump voters and the white working class. *Sociological Science 5*, 234–245.

Pena, A. (2024, November). We welcomed them: G.O.P embraces new Latino voters in America. *The New York Times*. https://www.nytimes.com/2024/11/11/world/americas/arizona-latino-vote.html. Accessed in April 2025.

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, C. (1960, November). About the election: A second look. *The New York Times*, 230. https://timesmachine.nytimes.com/timesmachine/1960/11/20/99968139.html. Accessed in April 2025.

Phillips, K. P. (1969). *The Emerging Republican Majority* (First ed.). New Rochelle, N.Y.: Arlington House.

Priebus, R. and S. Day (2015, Mar). Growth & opportunity project: Post-election progress report. Technical report, Republican National Committee. https://prod-cdn-static.gop.com/media/documents/GROW-OPP_1426636924.pdf. Accessed in May 2025.

Raymond, C. (2011). The continued salience of religious voting in the United states, Germany, and Great Britain. *Electoral Studies 30*(1), 125–135.

Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review 15*(3), 351–357.

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

Rubin, D. B. (1981). The Bayesian bootstrap. *Annals of Statistics 9*(1), 130–134.

Sanchez, G. R. (2024, November). A deep dive into the 2024 Latino male electorate. *Brookings Institution*. `https://www.brookings.edu/articles/a-deep-dive-into-the-2024-latino-male-electorate/`. Accessed in April 2025.

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.

Smith, G. (2021). Presidential voting and county composition; a research note on the 2020 election. *APSA Preprints*.

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.

Wlezien, C. (2017). The thermostatic model: The public, policy and politics. In J. Fisher, E. Fieldhouse, M. Franklin, R. Gibson, M. Cantijoch, and C. Wlezien (Eds.), *The Routledge Handbook of Elections, Voting Behavior and Public Opinion* (First ed.), Volume 1, pp. 404–415. Routledge.

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

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.

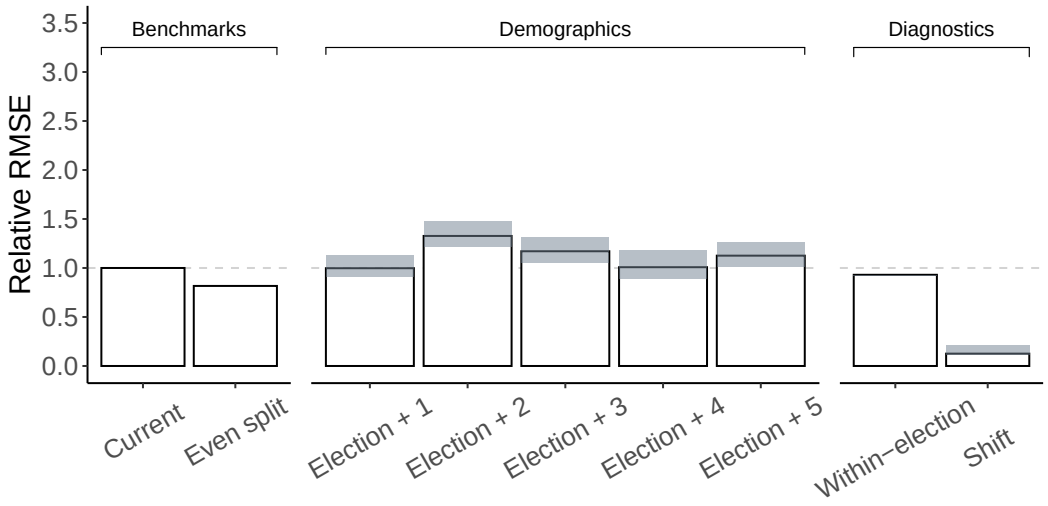
Zaller, J. (1992). *The Nature and Origins of Mass Opinion*. Cambridge University.

Zingher, J. N. (2014). An analysis of the changing social bases of America's political parties: 1952–2008. *Electoral Studies 35*, 272–282.

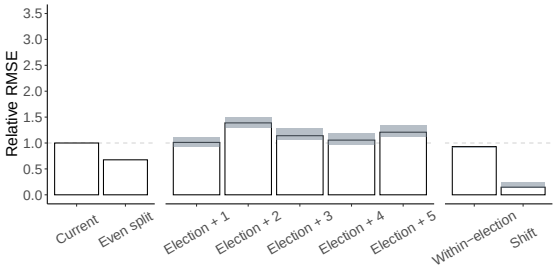
Zingher, J. N. (2020). On the measurement of social class and its role in shaping white vote choice in the 2016 US presidential election. *Electoral Studies 64*, 102119.

Figure 1: Performance of Demographic Forecasts of US Presidential Elections

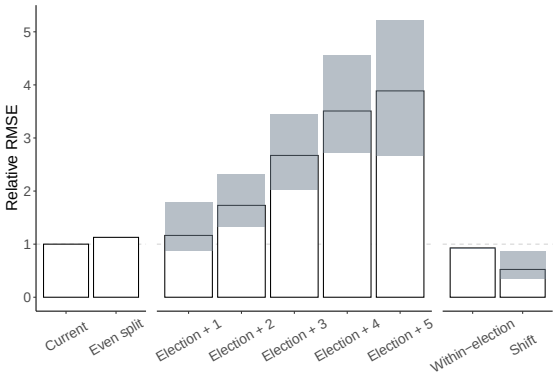
(a) Main specification



(b) Official results



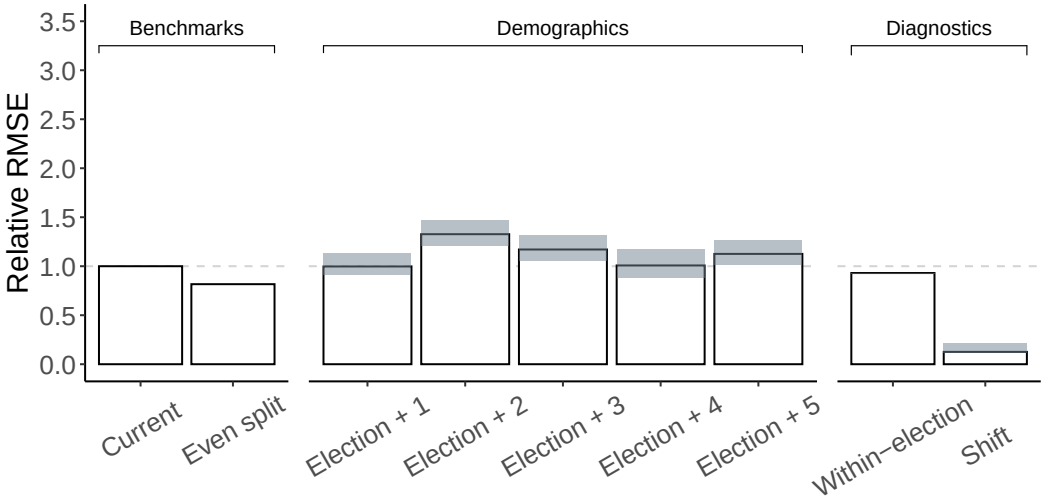
(c) Open-seat elections



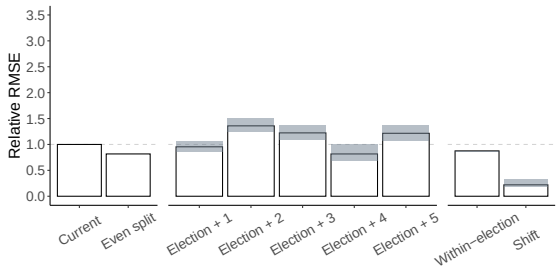
Notes: Each plot displays the relative root mean squared error (RMSE) of forecasts across varying specifications. The first section of each plot presents the RMSE of the current forecast of the next election, and of the even split forecast, as defined in Section 4.2. The second section of each plot presents 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 root mean square 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

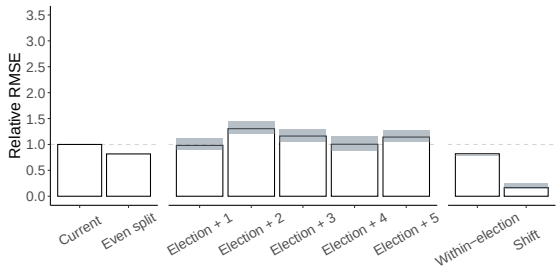
(a) Main specification



(b) Extended demographic covariates



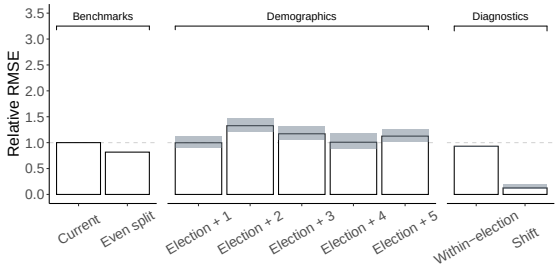
(c) Regression trees



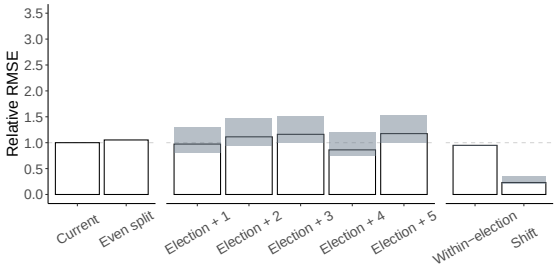
Notes: Each plot displays the relative root mean squared error (RMSE) of forecasts across varying specifications. The first section of each plot presents the RMSE of the current forecast of the next election, and of the even split forecast, as defined in Section 4.2. The second section of each plot presents 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 root mean square 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 Congressional Elections and Party Identification

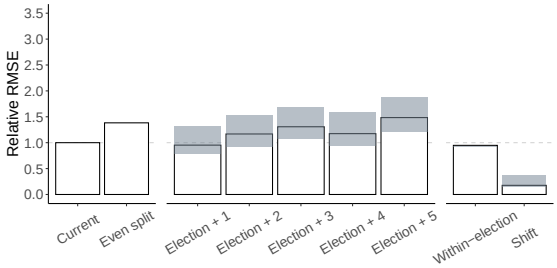
(a) Main specification



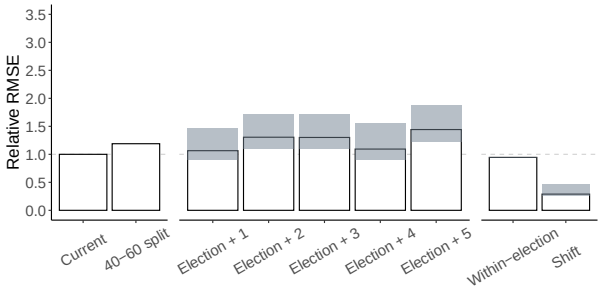
(b) Congressional elections in presidential years



(c) Congressional elections in midterm years

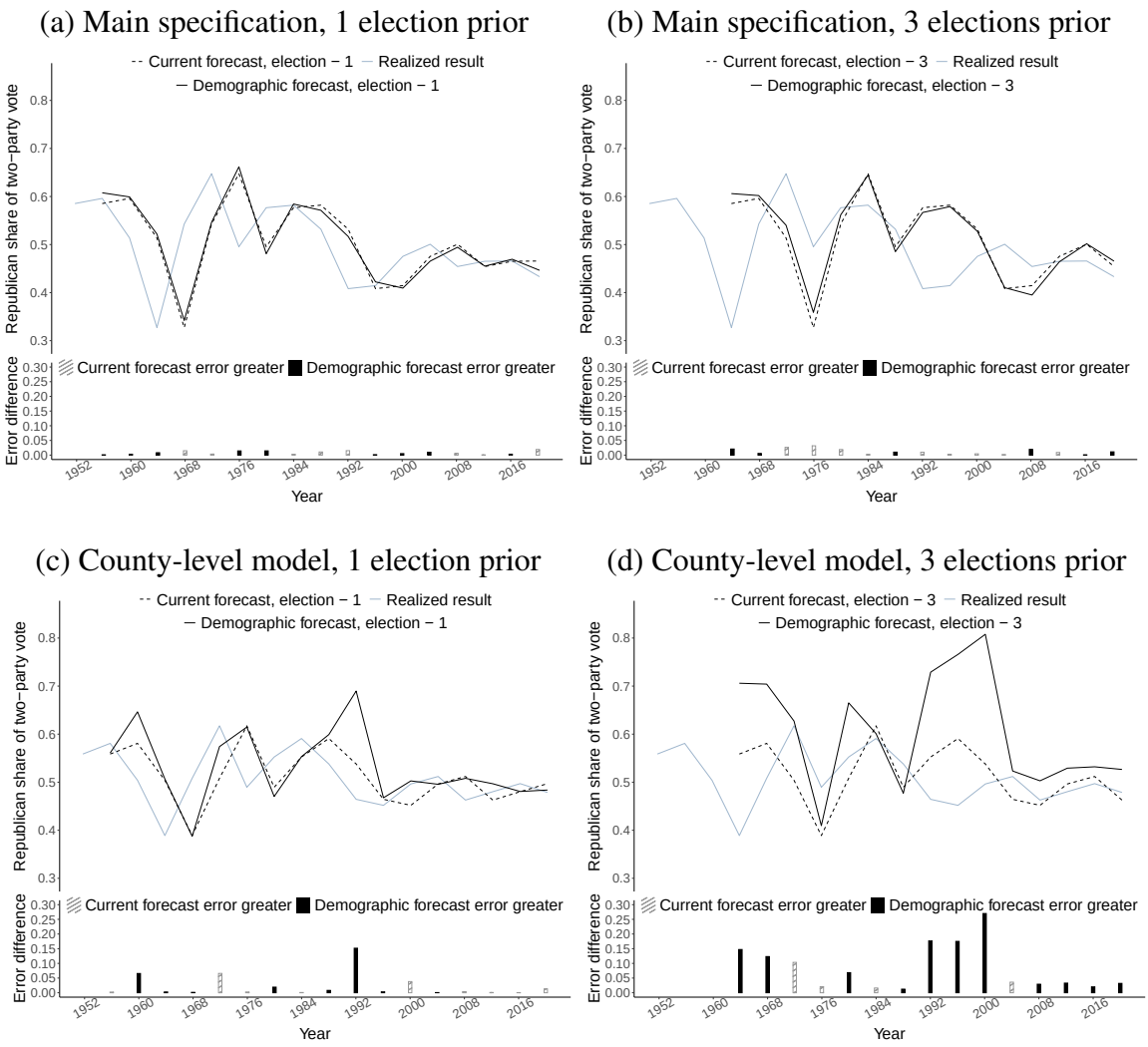


(d) Party identification



Notes: Each plot displays the relative root mean squared error (RMSE) of 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 root mean square 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.

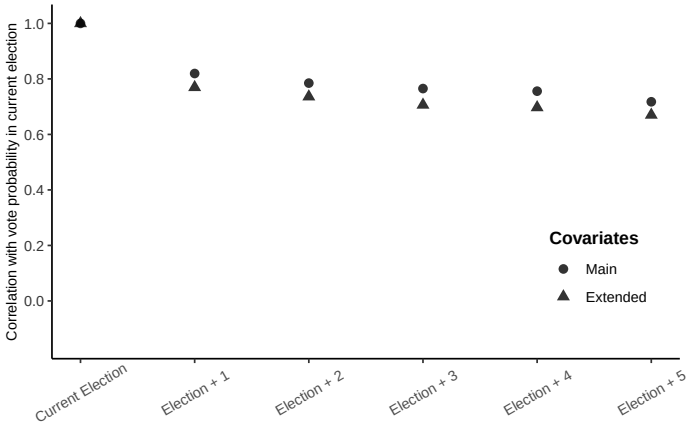
Figure 4: Time-Series Properties of Demographic Forecasts



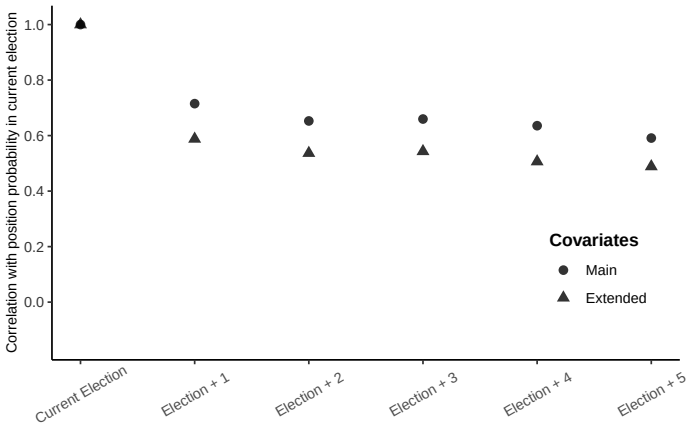
Notes: The top portion of each plot is a time series that shows the election result (gray series), a current forecast equal to the result in a past election (dashed black series), and a demographic forecast based on data from the same past election (solid black series). The bottom portion of each plot 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. Panels (a) and (b) use the main specification from Panel (a) of Figure 1 and forecasts from the one and three elections prior, respectively. Panels (c) and (d) use the county-level logistic specification from Panel (b) of Online Appendix Figure 8 and forecasts from the one and three elections prior, respectively.

Figure 5: Stability of Voting and Issue Position Probabilities Across Elections

(a) Voting probabilities



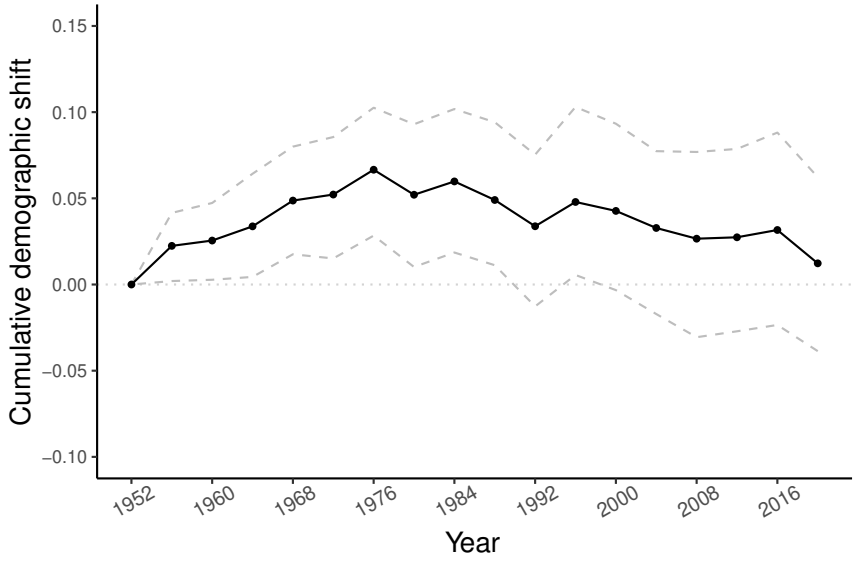
(b) Issue position probabilities



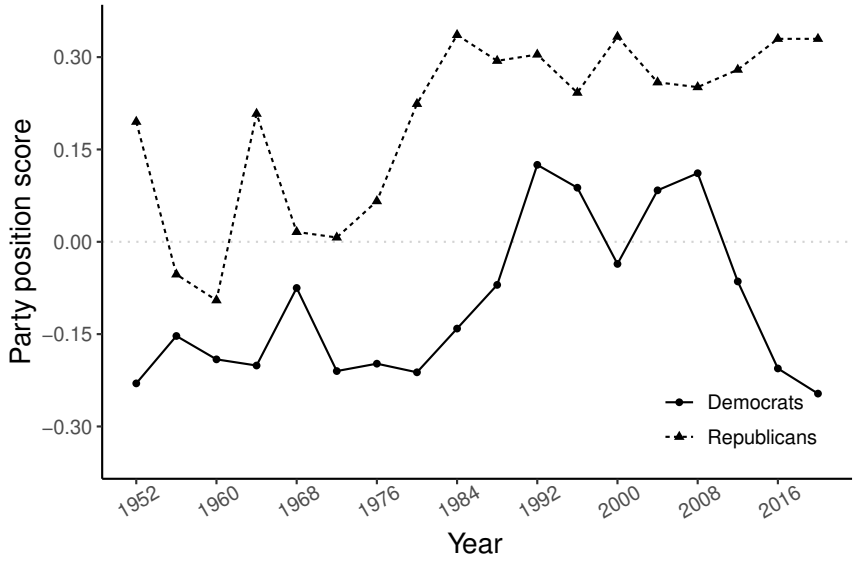
Notes: Panel (a) displays the average (across elections) of the correlation (across survey respondents) in the estimated probability (of voting Republican) at different horizons. Panel (b) displays the average (across elections and then across issues) of the correlation (across survey respondents) in the estimated probability (of expressing a right-wing position on the issue) at different horizons. The correlation of the estimated probability in the current election with itself is mechanically one. The correlation for "Election + 1" is the average (across elections or elections and issues) of the correlation (across survey respondents) between a respondent's estimated probability (of voting Republican or expressing a right-wing position on the issue) according to the predictive model estimated for the current election year, and the same respondent's probability according to the predictive model estimated for the next election year. The predictive model is a binary logit model (Section 4.1) with independent variables given by either the main demographic covariates or the extended demographic covariates (Section 3.1) and dependent variable given by either the vote (Section 3.1) or the issue position (Section 3.3).

Figure 6: Shifts in Demographics and Party Positions

(a) Predicted cumulative shift in the electorate



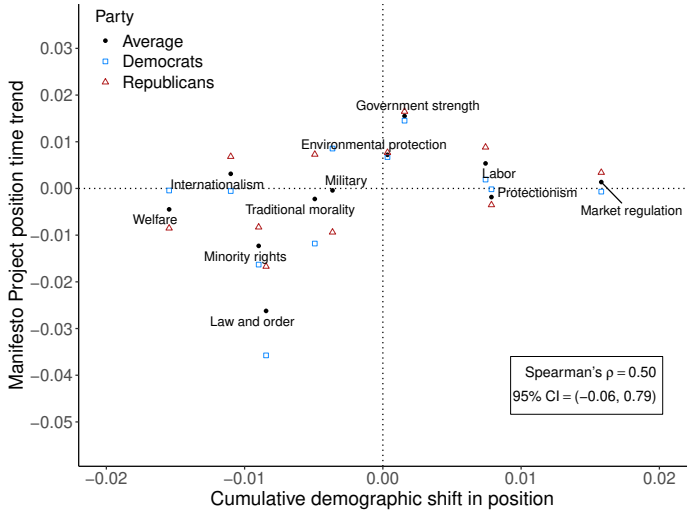
(b) Measured shift in party positions



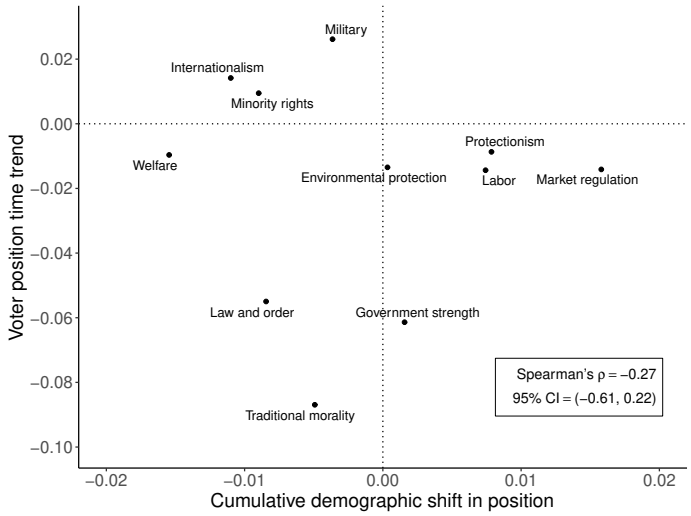
Notes: Panel (a) shows the cumulative predicted change Δ_τ, according to our main specification, in the Republican share of the two-party vote due to demographic change, as defined in Section 6.3. The dashed series depicts pointwise 95 percent credible intervals calculated based on a Bayesian bootstrap. 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 3.3.

Figure 7: Shifts in Demographics and Party Positions on Issues

(a) Shifts in parties' issue positions



(b) Shifts in voters' issue positions



Notes: The plots are scatterplots. In Panel (a), 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 Section 3.3. 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 3.3. The triangle is for the Republican party, the square is for the Democratic party, and the dot is for both parties pooled together. The x-axis variable is the estimated per-decade cumulative demographic shift in voters' probability of supporting the right-wing position on the issue, as defined in Section 6.3. In the bottom right we display the Spearman rank correlation between the pooled y-axis variable and the x-axis variable as well as a corresponding 95 percent credible interval calculated based on a Bayesian bootstrap. Panel (b) replaces the y-axis variable with the estimated per-decade linear time trend in voters' stated positions on the given issue, as defined in Section 6.3.

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 Demographic Change in a Model of Electoral Competition

Here we show that, in a symmetric model of party competition for votes, we can expect election outcomes to be competitive regardless of demographic change, and, therefore, pure demographic forecasts to be systematically wrong. To see this, suppose, in the model of Section 2, that the two parties each simultaneously choose platforms $x_L, x_R \in \mathcal{X}$ and then receive a payoff equal to their share of the vote. Suppose, further, that voter behavior treats the parties symmetrically in the sense that $p_g(x) = \frac{1}{2}$ for all $g \in \{1, ..., G\}$ whenever $x_L = x_R$. It is immediate that in any pure-strategy equilibrium, the two parties must expect an equal share of the vote, because if not, whichever party is at a disadvantage has a profitable deviation to the platform of the other party.[2] It follows that $V = V^* = \frac{1}{2}$ if all elections constitute pure-strategy equilibria of the game.[3]

To see the implications for forecasting, notice that if we impose the conditions for a pure-strategy equilibrium, the error becomes

$$\hat{V} - V^* = \sum_{g=1}^{G} \hat{n}_g \left(p_g(x) - \frac{1}{2} \right). \tag{2}$$

Equation (2) states that the essential basis of a pure demographic forecast—heterogeneity in different groups' propensity to vote for party R—is a source of error once we consider the implications of electoral competition. Intuitively, any such heterogeneity creates the (mistaken) impression that changes in group shares will influence the future vote.

[1]E-mail: richard_calvo@berkeley.edu, vpons@hbs.edu, jesse_shapiro@fas.harvard.edu.
[2]Sufficient conditions for the existence of a Nash equilibrium in pure strategies in a similar model can be found in, for example, Austen-Smith and Banks (2005, see Theorems 7.9 and 7.10).
[3]Notice that this reasoning does not imply that the parties adopt identical platforms in any pure-strategy equilibrium, only that they receive equal shares of the total vote. It may be, for example, that one party receives relatively more of the vote of one group, and another party, more of the vote of another group.

B Survey Data Description

Online Appendix Table 1: Survey Question Wording and Coding

Panel A: Outcome variables

Outcome	Coverage	Relevant ANES Question	Coding Notes
Vote for President	1952-2020	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, 2020	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, 2020	... 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	1952-2020	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 A: Outcome variables

Outcome	Coverage	Relevant ANES Question	Coding Notes
Contacted by Any Political Party	1956-2020	VCF9030a: Did PARTY contact R about election campaign?	Coded as "Yes" or "No".
Contacted by Democratic Party	1956-2020	VCF9030b: Did Democratic Party contact R about election campaign?	Coded as "Yes" or "No".
Contacted by Republican Party	1956-2020	VCF9030c: Did Republican Party contact R about election campaign?	Coded as "Yes" or "No".

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)	1952-2020	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	1952-2020	VCF0104: Respondent gender	Coded as female or male/other ("Other" introduced in 2016).
Race	1952-2020	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	1952-2020	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	1952-2020	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.)	1952-2020	... 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	1956-2020	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; 2024) 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	1952-2020	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	1952-2020	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)	1952-2020	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	1952-2020	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	1952-2020	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	1952-2020	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	1952-2020	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	885
1956	1,762	1,266	1,178	1,156
1960	1,181	898	880	829
1964	1,571	1,111	1,059	1,021
1968	1,557	911	889	853
1972	2,705	1,587	1,532	1,503
1976	2,248	1,322	1,222	1,164
1980	1,614	877	784	736
1984	2,257	1,376	1,238	1,203
1988	2,040	1,195	1,078	1,047
1992	2,485	1,357	1,225	1,197
1996	1,714	1,034	929	920
2000	1,807	1,120	916	906
2004	1,212	811	721	710
2008	2,322	1,539	1,367	1,359
2012	5,914	4,188	3,847	3,831
2016	4,270	2,609	2,420	2,410
2020	8,280	6,119	5,700	5,667

Notes: 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 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), with minor adjustments based on other public sources.

Age

Our data on age come from the US Decennial Census accessed via Social Explorer (US Census Bureau, 2010) for 1950-2010 and via the US Census Bureau Population Division (US Census Bureau, 2023) for 2020. 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 2020.

Gender

Our data on gender come from the US Decennial Census accessed via Social Explorer (US Census Bureau, 2010) and via the US Census Bureau Population Division (US Census Bureau, 2023) for 2020. 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 2020.

Urbanism

Our data on urbanism come from the US Decennial Census accessed via Social Explorer (US Census Bureau, 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, 2010) and via the US Census Bureau Population Division (US Census Bureau, 2023) for 2020. 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 2020.

Education

Our data on education in 1950 come from the US Decennial Census accessed via Social Explorer (US Census Bureau, 2010). Our data on education in 1970, 1980, 1990, and 2000 come from the US Decennial Census accessed via the US Department of Agriculture Economic Research Service (US Census Bureau, 2000). Our data on education in 2010 and 2020 come from American Community Survey 5-Year Estimates for 2008–2012 and 2018-2022 accessed via the US Department of Agriculture Economic Research Service (US Census Bureau, 2024). 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.[4] We observe these variables in 1950, 1970, 1980, 1990, 2000, and 2010.

Income

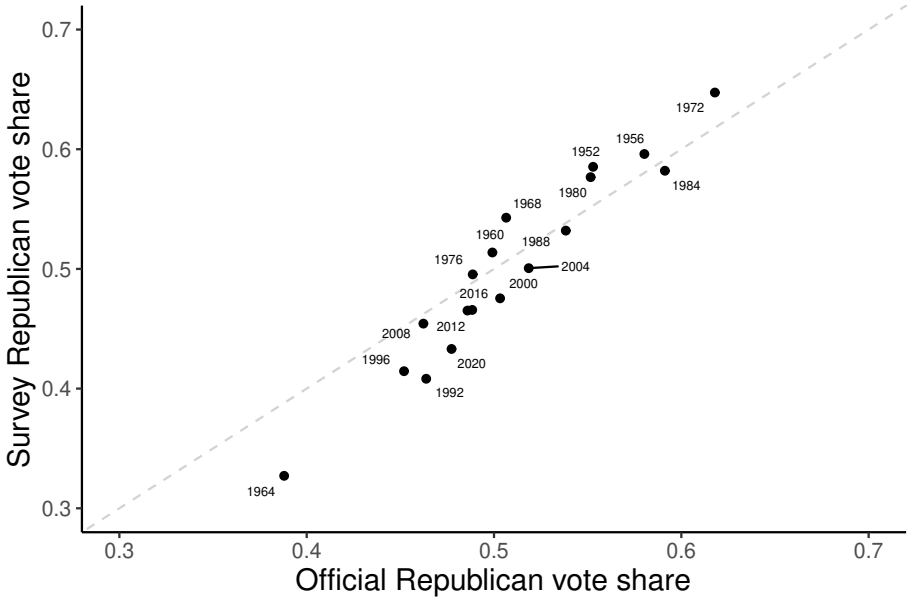
Our data on income come from two sources: for 1952, 1962, and 1972, the US Census *County and City Data Books* (Inter-university Consortium for Political and Social Research, 2008; Haines and Inter-university Consortium for Political and Social Research, 2010) provide median family income, and for each year from 1969 through 2016, the US Bureau of Economic Analysis (BEA) Regional Economic Accounts (US Bureau of Economic Analysis, 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 harmonize variables across years, we regress log per capita personal income in 1972 on log median family income in 1972 (the year in which the two

[4]For the 1950 data, we encode those with "No school years completed," "At least some elementary," and "1-3 years high school" as having less than a high school education, those with "4 years high school" as having only a high school education, those with "1-3 years college" as having some college, and those with "4 years college" as having a college degree, and we exclude those with "Unknown years of school." For other years we directly observe the fraction of residents aged 25 or older in each category.

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 analysis.

D Additional Empirical Results

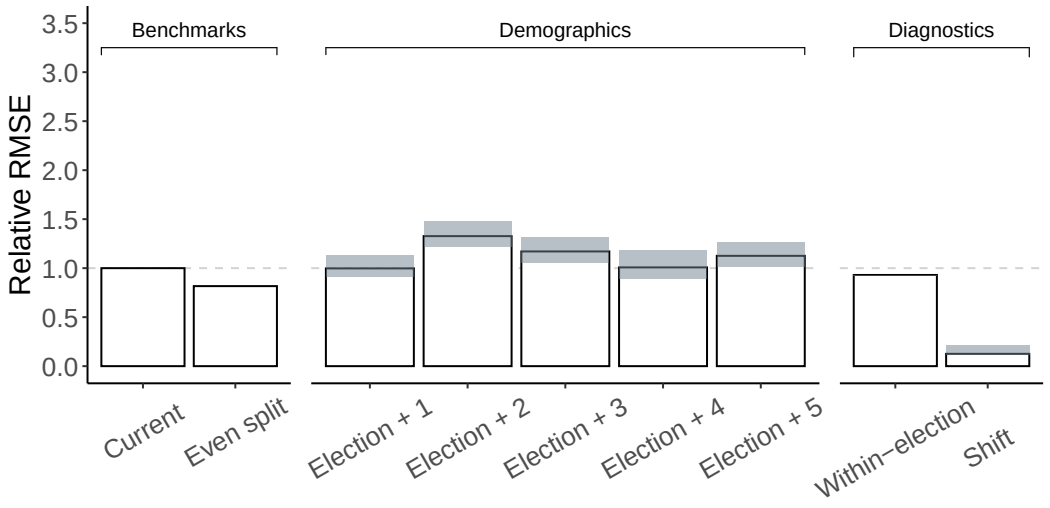
Online Appendix Figure 1: Republican Share of Two-Party Vote, Survey vs. Official Results



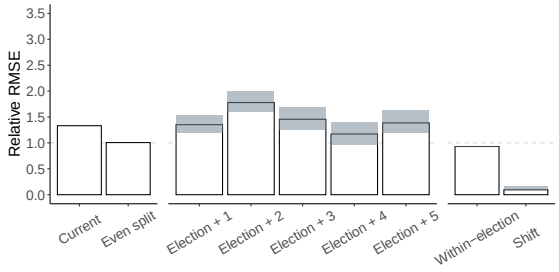
Notes: The plot is a scatterplot. The unit of analysis is the presidential election. The y-axis depicts the Republican share of the two-party vote among survey respondents. The x-axis depicts the Republican share of the two-party vote from official election results. The dashed line is a 45-degree line.

Online Appendix Figure 2: Performance of Demographic Forecasts of US Presidential Elections: Earlier vs. Later Elections

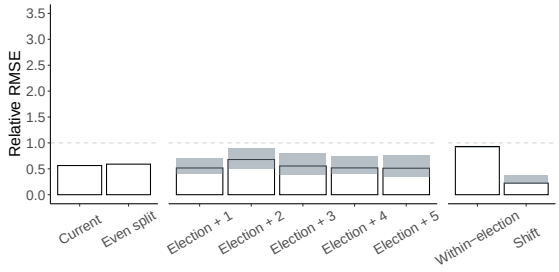
(a) Main specification



(b) First half of sample period

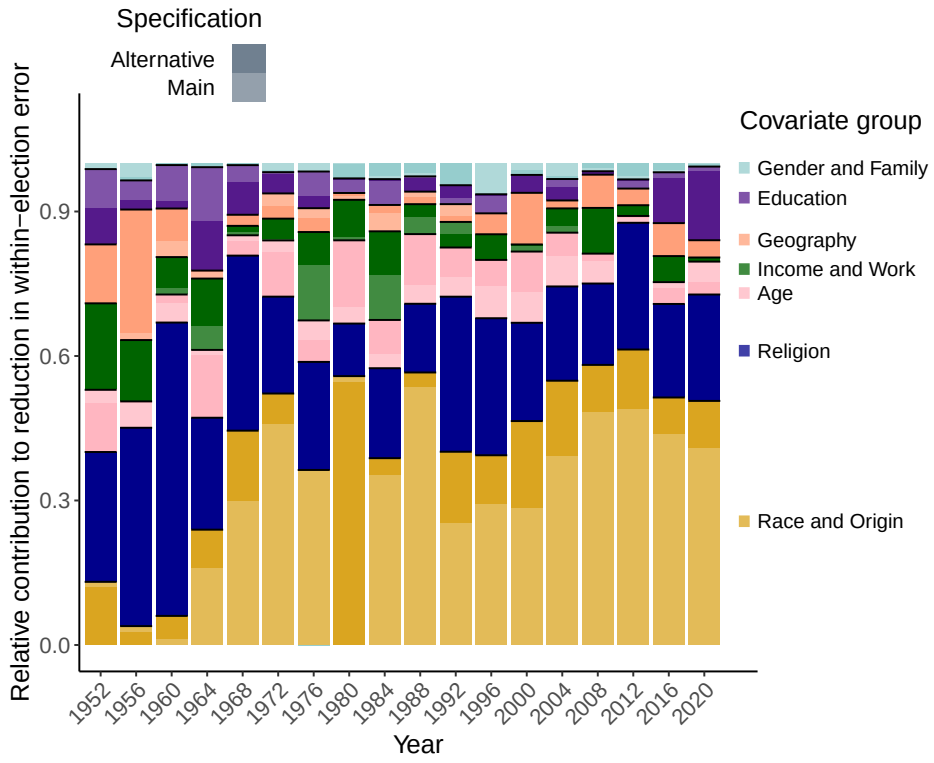


(c) Second half of sample period



Notes: Each plot displays the relative root mean squared error (RMSE) of 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 root mean square 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 restrict attention to elections from 1952 to 1984. Panel (c) presents results when we restrict attention to elections from 1988 to 2020. All normalizing constants are calculated using the full sample.

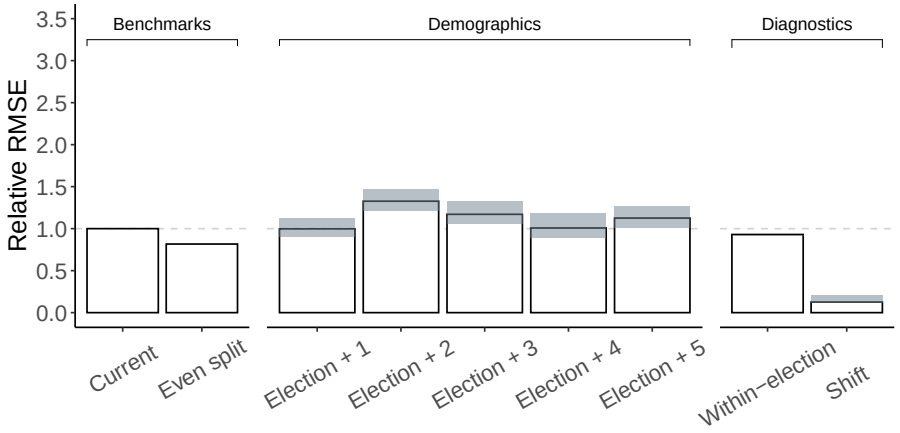
Online Appendix Figure 3: Contributions of Covariate Groups to Reduction in Within-election Error



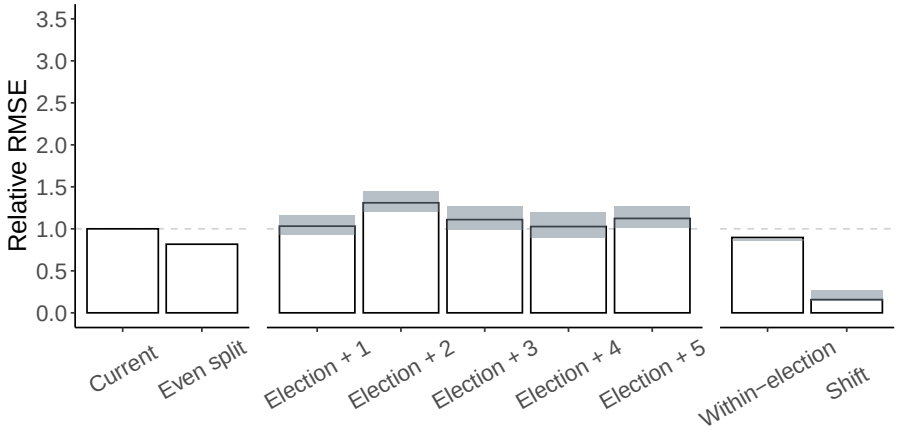
Notes: The plot shows the contribution of each group of variables to the reduction in within-election error, as defined in Section 4.3. To calculate the contribution of a given group of variables, we re-estimate the binary logit model without the given group of variables and calculate the increase in 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 4: Performance of Demographic Forecasts of US Presidential Elections, Separate Model by Region

(a) Main specification



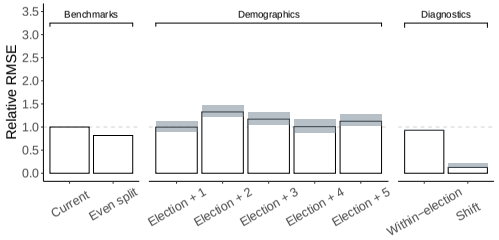
(b) Separate model by region



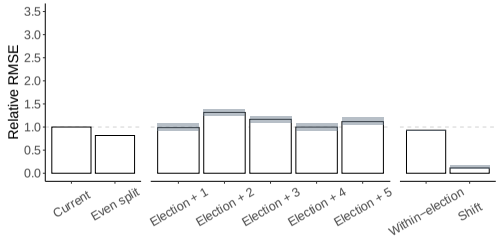
Notes: Each plot displays the relative root mean squared error (RMSE) of forecasts across varying specifications. The first section of each plot presents the RMSE of the current forecast of the next election, and of the even split forecast, as defined in Section 4.2. The second section of each plot presents 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 root mean square 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 modify our main specification by allowing the parameters of the predictive model to differ by Census region.

Online Appendix Figure 5: Performance of Demographic Forecasts of US Presidential Elections, Penalized Estimation

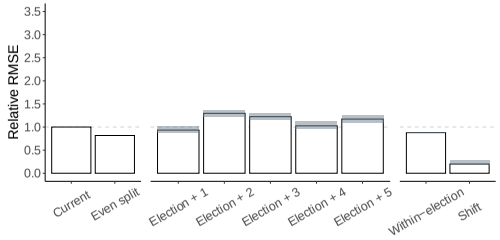
(a) Main specification



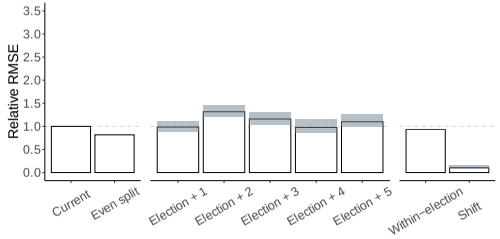
(b) Main demographic covariates, Jeffreys penalty



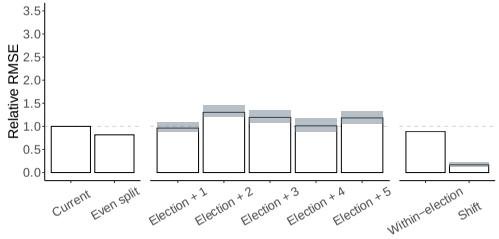
(c) Extended demographic covariates, Jeffreys penalty



(d) Main demographic covariates, L1 penalty



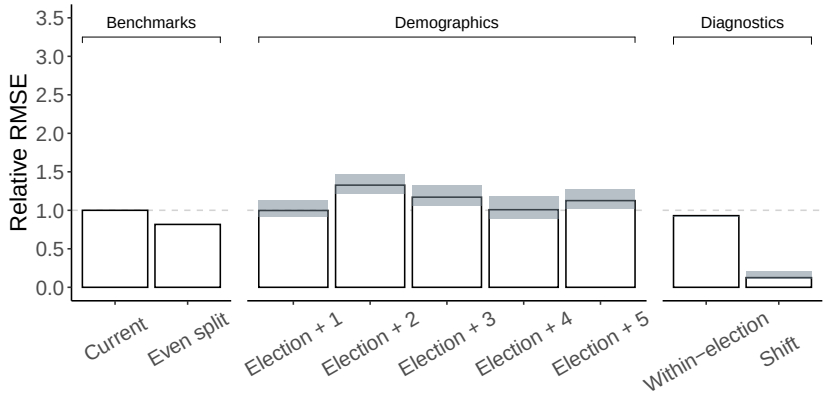
(e) Extended demographic covariates, L1 penalty



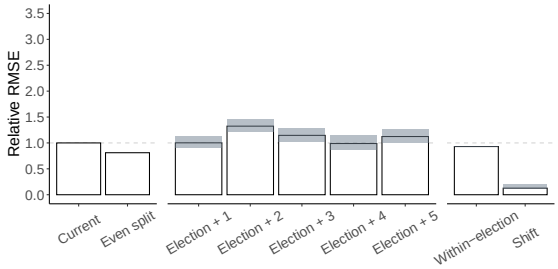
Notes: Each plot displays the relative root mean squared error (RMSE) of forecasts across varying specifications. The first section of each plot presents the RMSE of the current forecast of the next election, and of the even split forecast, as defined in Section 4.2. The second section of each plot presents 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 root mean square 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. Panels (b) and (d) present results when we estimate our predictive model using penalized maximum likelihood and the main demographic covariates. Panel (c) and (e) present results when we estimate our predictive model using penalized maximum likelihood and the extended demographic covariates. In panels (b) and (c), the penalty is a Jeffreys prior penalty. In panels (d) and (e), the penalty is an L1-penalty with penalty parameter chosen via 10-fold cross-validation.

Online Appendix Figure 6: Performance of Demographic Forecasts of US Presidential Elections, Alternative Treatment of Missing Covariates

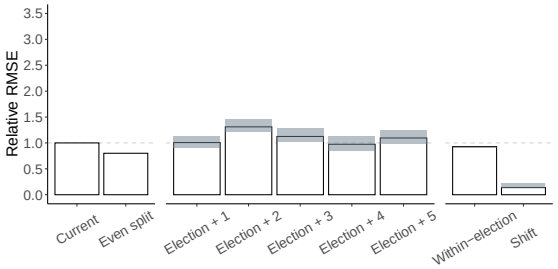
(a) Main specification



(b) Include respondents missing extended covariates



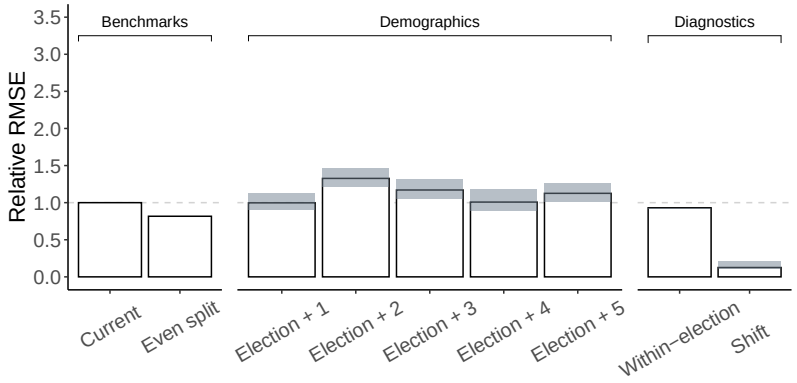
(c) Include respondents missing main covariates



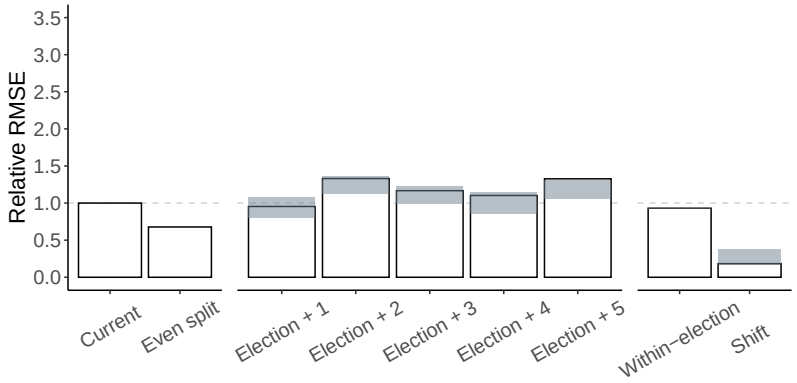
Notes: Each plot displays the relative root mean squared error (RMSE) of forecasts across varying specifications. The first section of each plot presents the RMSE of the current forecast of the next election, and of the even split forecast, as defined in Section 4.2. The second section of each plot presents 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 root mean square 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 7: Performance of Demographic Forecasts of the US Electoral College

(a) Main specification



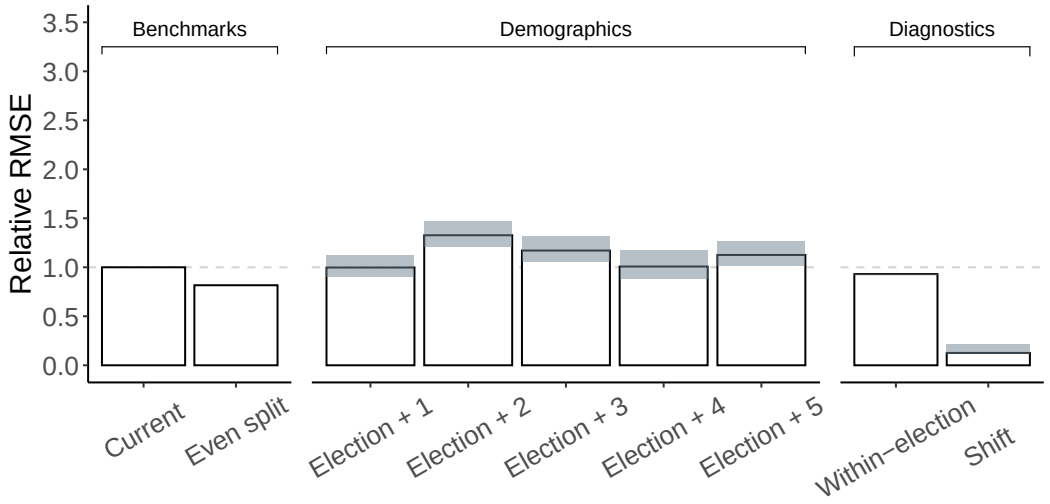
(b) Electoral College



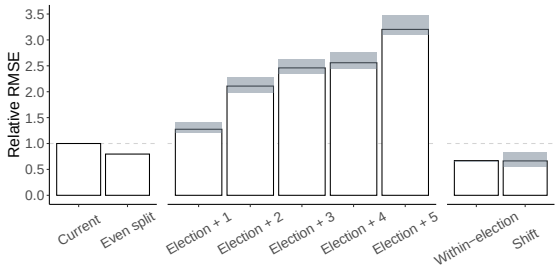
Notes: Each plot displays the relative root mean squared error (RMSE) of forecasts across varying specifications. The first section of each plot presents the RMSE of the current forecast of the next election, and of the even split forecast, as defined in Section 4.2. The second section of each plot presents 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 root mean square 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 where we forecast the Republican share of the Electoral College. To do this, we assume that each voter votes independently according to the predictive model defined in Section 4.1, and that each survey respondent represents a number of voters in their state inverse to the respondent's survey weight. We approximate the probability of a Republican popular victory in each state via Monte Carlo simulation, assign all of a state's electors to the winner of the state's popular vote, and use the implied expected Republican share of the Electoral College as the forecast.

Online Appendix Figure 8: Performance of Demographic Forecasts of US Presidential Elections: Aggregate Data

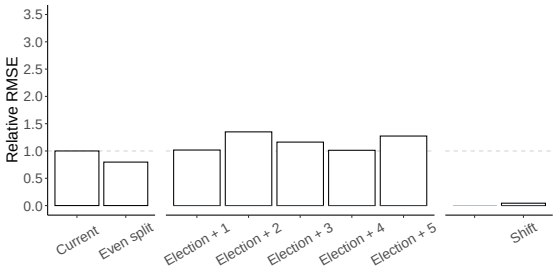
(a) Main specification



(b) County-level logistic regression

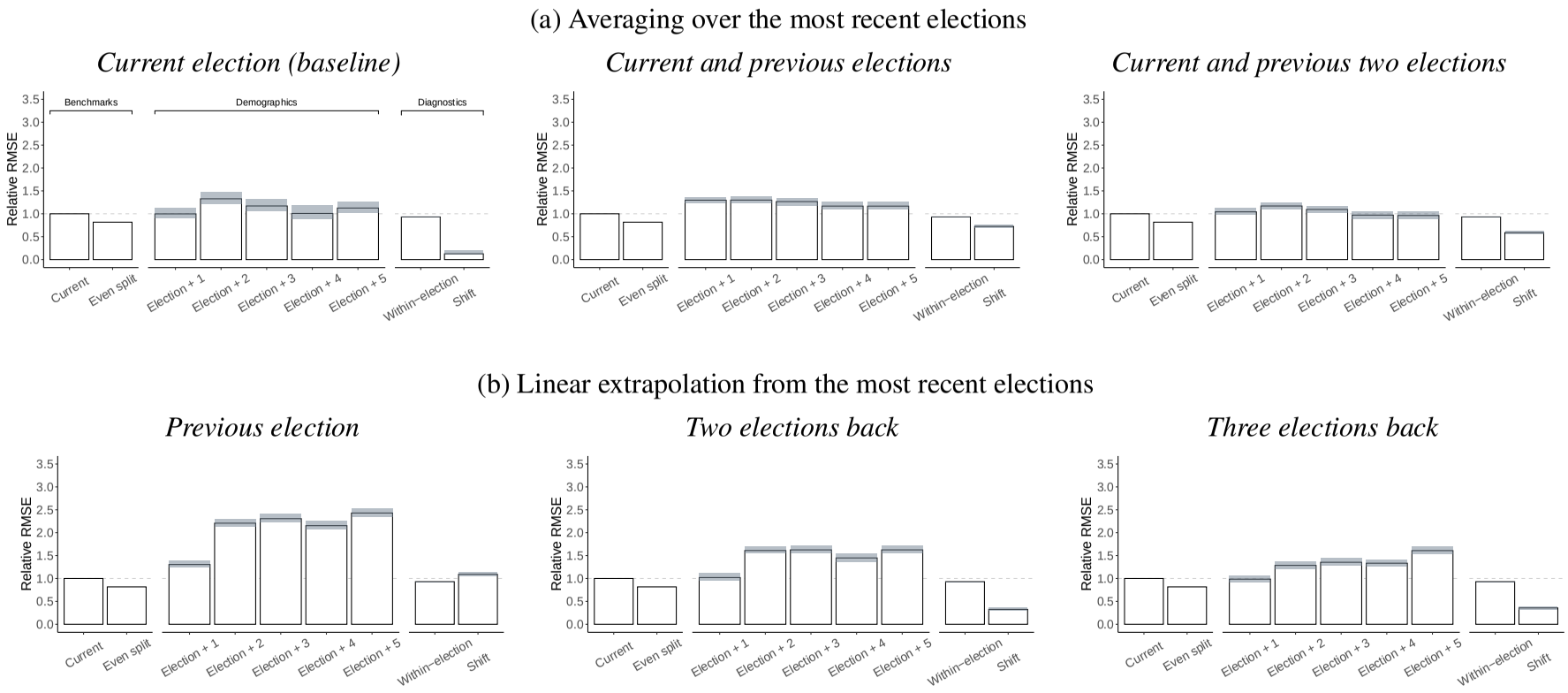


(c) County-level population change



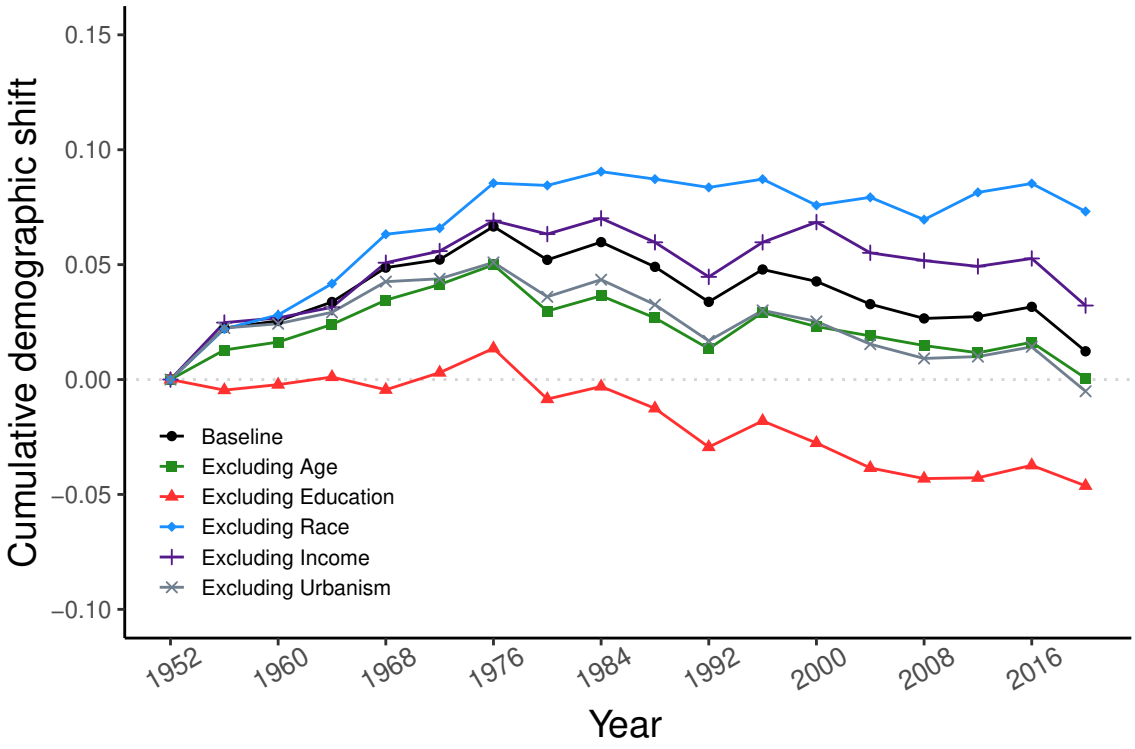
Notes: Each plot displays the relative root mean squared error (RMSE) of forecasts across varying specifications. The first section of each plot presents the RMSE of the current forecast of the next election, and of the even split forecast, as defined in Section 4.2. The second section of each plot presents 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 root mean square 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 share remains the same between elections, but allowing each county's number of voters to evolve as in the actual data. For the specification in Panel (c) the within-election error is not well-defined.

Online Appendix Figure 9: Performance of Demographic Forecasts of US Presidential Elections, Averaging or Extrapolating from Recent Elections



Notes: Each plot displays the relative root mean squared error (RMSE) of 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 root mean square 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 where we assume that future coefficients on demographic covariates are equal to the average over a set of the most recent elections. When there are insufficiently many previous elections, we average over those in the sample. Panel (b) presents results where we assume that coefficients on demographic covariates will evolve linearly going forward, and fit the rate of change by comparing the current election to the coefficient in the previous election, two elections back, or three elections back. When there are insufficiently many previous elections, we do not extrapolate but instead assume that future coefficients will equal current coefficients.

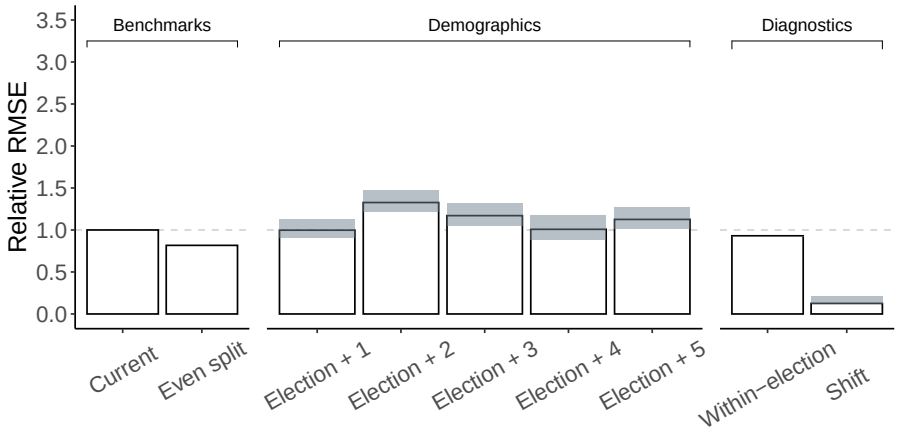
Online Appendix Figure 10: Predicted Cumulative Shift in the Electorate, Excluding Covariates



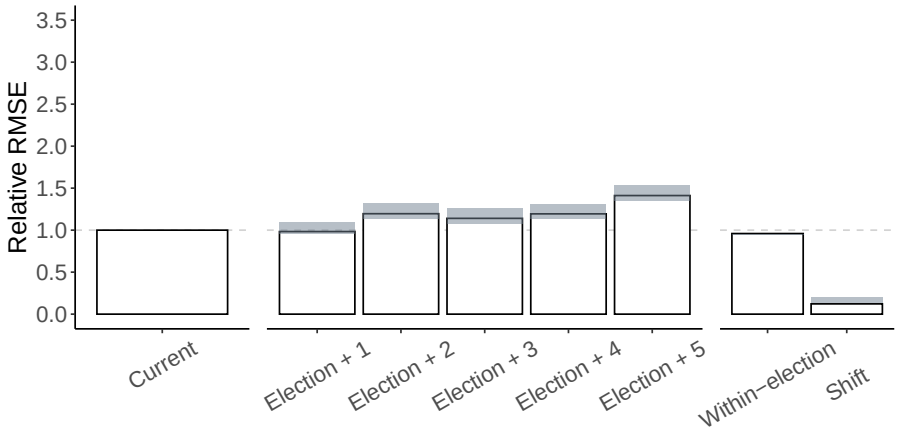
Notes: The baseline series corresponds to the predicted cumulative shift in the electorate, as depicted in Panel (a) of Figure 6 and defined in Section 6.3. Each other series corresponds to an experiment in which we exclude the given covariate from the set used to estimate the predictive model, and recompute the predicted cumulative shift in the electorate.

Online Appendix Figure 11: Performance of Demographic Forecasts of Issue Positions

(a) Main specification



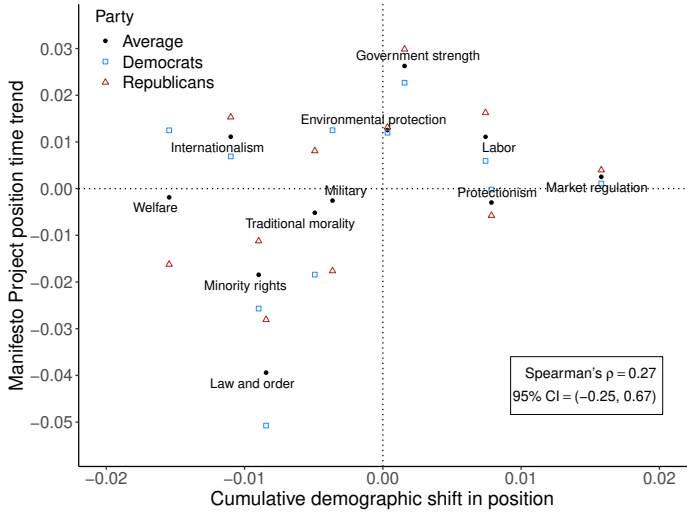
(b) Forecast of issue positions



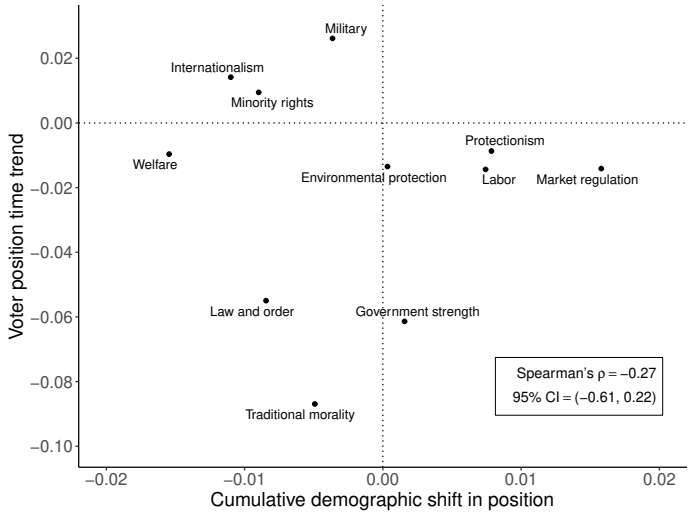
Notes: Each plot displays the relative root mean squared error (RMSE) of forecasts across varying specifications. The first section of each plot presents the RMSE of the current forecast of the next election, and of the even split forecast, as defined in Section 4.2. The second section of each plot presents 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 root mean square 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 forecast voters' stated positions on each of a set of issues described in Section 3.3, and average forecast performance over the issues. We omit the even-split forecast from panel (b) as it is not a useful benchmark for forecasts of issue positions.

Online Appendix Figure 12: Shifts in Demographics and Party Positions on Issues, Additional Analysis

(a) Shifts in parties' issue positions, alternative scaling



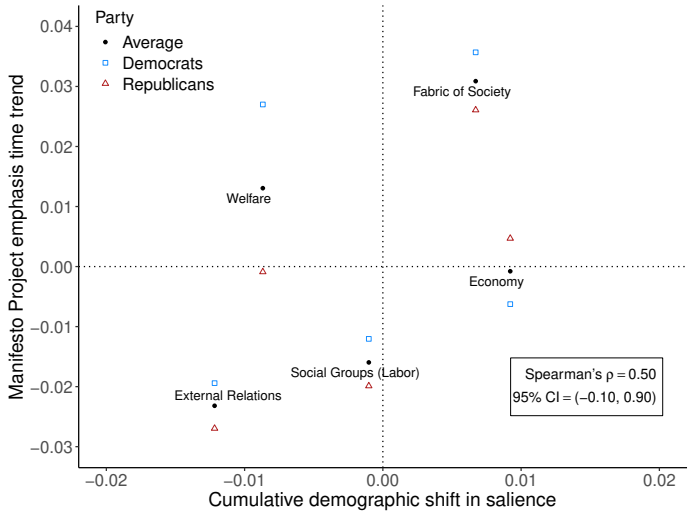
(b) Shifts in voters' issue positions



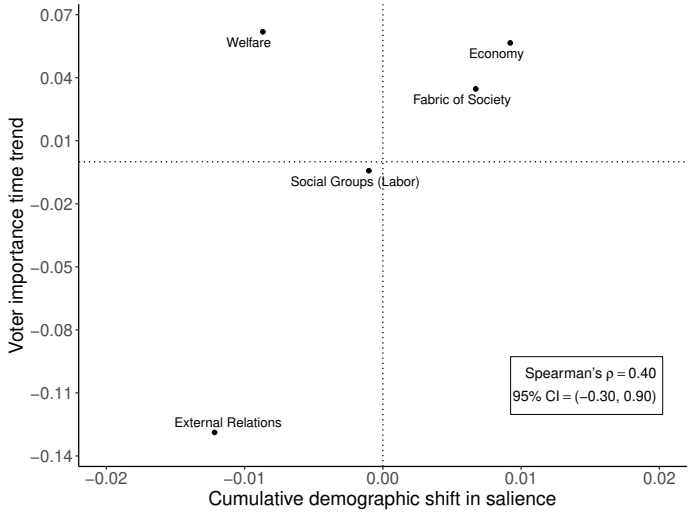
Notes: Panel (a) repeats the scatterplot from Figure 7 Panel (a) but replaces the y-axis variable with one that restricts attention to the portions of each party's platform that are coded as ideological by the Manifesto Project. Panel (b) repeats the scatterplot from Figure 7 Panel (b).

Online Appendix Figure 13: Shifts in Demographics and Party Emphasis on Issues

(a) Shifts in parties' issue emphasis



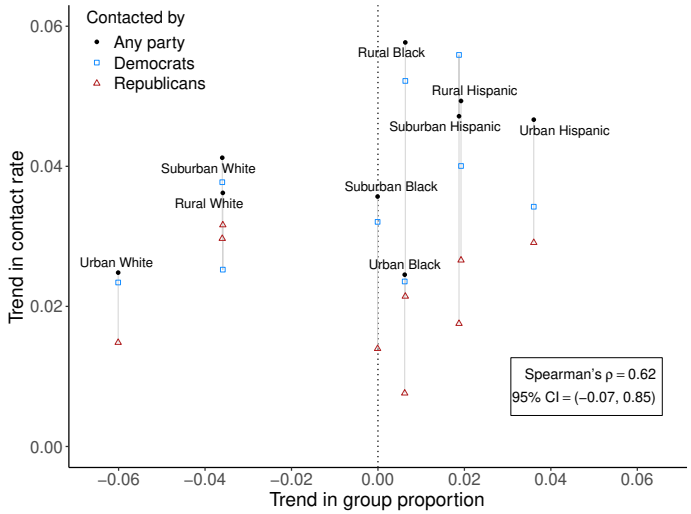
(b) Shifts in voters' issue importance



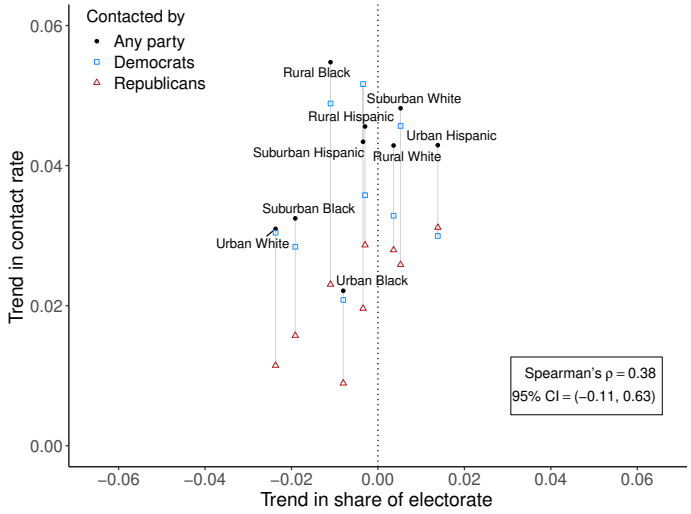
Notes: Panel (a) repeats the scatterplot from Figure 7 Panel (a) but replaces (i) the y-axis variable with the estimated per-decade linear time trend in the share of sub-sentences that refer to the given issue in party platforms and (ii) the x-axis variable with the estimated per-decade cumulative demographic shift in voters' probability of listing the given issue as the one most important to them, as defined in Online Appendix E. Panel (b) replaces the y-axis variable in Panel (a) with the estimated per-decade linear time trend in voters' probability of listing the given issue as the one most important to them.

Online Appendix Figure 14: Trends in Group Sizes and Contact Rates

(a) Trends in contact rates



(b) Trends in contact rates, controlling for initial group size



Notes: The plots are scatterplots. In each panel, each point represents a group of voters defined by categories of urbanism and race, as defined in Section 3.1. The y-axis variable is the estimated per-decade linear time trend in the self-reported contact rate of each group, estimated via a linear regression, with the self-reported contact rate from the survey question described in Online Appendix Table 1. The triangle is for the Republican party, the square is for the Democratic party, and the dot is for contact by either party. The x-axis variable is the estimated per-decade linear trend in the share of the electorate in each group, estimated via a linear regression. In the bottom right we display the Spearman rank correlation between the "any party" y-axis variable and the x-axis variable as well as a corresponding 95 percent credible interval calculated based on a Bayesian bootstrap. Panel (b) residualizes the y-axis and x-axis variables with respect to the share of the electorate in each group at the beginning of the sample period, adding a constant to preserve the sample mean.

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	(0.095)	-	-	Current	1.000	(0.082)	-	-
Even split	0.817	(0.078)	-	-	Even split	0.675	(0.055)	-	-
Election + 1	0.997	(0.095)	0.908	1.128	Election + 1	1.012	(0.082)	0.934	1.118
Election + 2	1.327	(0.126)	1.216	1.472	Election + 2	1.388	(0.113)	1.307	1.503
Election + 3	1.171	(0.112)	1.055	1.319	Election + 3	1.141	(0.093)	1.057	1.282
Election + 4	1.008	(0.096)	0.887	1.179	Election + 4	1.056	(0.086)	0.965	1.194
Election + 5	1.126	(0.107)	1.017	1.267	Election + 5	1.209	(0.099)	1.113	1.338
Within-election	0.931	(0.460)	0.918	0.927	Within-election	0.931	(0.460)	0.918	0.927
Shift	0.126	(0.012)	0.115	0.210	Shift	0.147	(0.012)	0.135	0.245
Open-seat elections (Panel (c) of Figure 1)					*Extended demographic covariates (Panel (b) of Figure 2)*				
Current	1.000	(0.040)	-	-	Current	1.000	(0.095)	-	-
Even split	1.128	(0.045)	-	-	Even split	0.817	(0.078)	-	-
Election + 1	1.164	(0.047)	0.871	1.785	Election + 1	0.955	(0.091)	0.864	1.075
Election + 2	1.730	(0.069)	1.325	2.323	Election + 2	1.359	(0.130)	1.235	1.499
Election + 3	2.671	(0.107)	2.027	3.456	Election + 3	1.223	(0.117)	1.095	1.376
Election + 4	3.508	(0.141)	2.713	4.558	Election + 4	0.817	(0.078)	0.689	1.001
Election + 5	3.887	(0.156)	2.671	5.229	Election + 5	1.215	(0.116)	1.085	1.375
Within-election	0.929	(0.462)	0.911	0.928	Within-election	0.876	(0.432)	0.850	0.863
Shift	0.523	(0.021)	0.349	0.872	Shift	0.220	(0.021)	0.189	0.325
Regression trees (Panel (c) of Figure 2)					*Congressional elections in presidential years (Panel (b) of Figure 3)*				
Current	1.000	(0.095)	-	-	Current	1.000	(0.052)	-	-
Even split	0.817	(0.078)	-	-	Even split	1.053	(0.055)	-	-
Election + 1	0.981	(0.093)	0.891	1.108	Election + 1	0.973	(0.051)	0.805	1.302
Election + 2	1.304	(0.124)	1.199	1.440	Election + 2	1.113	(0.058)	0.937	1.460
Election + 3	1.162	(0.111)	1.058	1.296	Election + 3	1.162	(0.060)	0.997	1.503
Election + 4	1.004	(0.096)	0.881	1.152	Election + 4	0.861	(0.045)	0.747	1.203
Election + 5	1.143	(0.109)	1.040	1.270	Election + 5	1.175	(0.061)	1.002	1.527
Within-election	0.820	(0.405)	0.774	0.785	Within-election	0.949	(0.470)	0.938	0.947
Shift	0.168	(0.016)	0.136	0.252	Shift	0.231	(0.012)	0.212	0.366
Congressional elections in midterm years (Panel (c) of Figure 3)					*Party identification (Panel (d) of Figure 3)*				
Current	1.000	(0.062)	-	-	Current	1.000	(0.035)	-	-
Even split	1.383	(0.086)	-	-	40-60 split	1.189	(0.041)	-	-
Election + 1	0.953	(0.059)	0.780	1.304	Election + 1	1.065	(0.037)	0.899	1.461
Election + 2	1.168	(0.073)	0.924	1.530	Election + 2	1.305	(0.045)	1.097	1.711
Election + 3	1.307	(0.081)	1.071	1.683	Election + 3	1.301	(0.045)	1.105	1.713
Election + 4	1.174	(0.073)	0.936	1.581	Election + 4	1.094	(0.038)	0.901	1.545
Election + 5	1.484	(0.092)	1.214	1.878	Election + 5	1.442	(0.050)	1.229	1.886
Within-election	0.949	(0.470)	0.926	0.940	Within-election	0.945	(0.466)	0.935	0.943
Shift	0.177	(0.011)	0.145	0.369	Shift	0.290	(0.010)	0.261	0.463

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)					*First sample half (Panel (b) of Online Appendix Figure 2)*				
Current	1.000	(0.095)	-	-	Current	1.332	(0.169)	-	-
Even split	0.817	(0.078)	-	-	Even split	1.006	(0.128)	-	-
Election + 1	0.997	(0.095)	0.908	1.128	Election + 1	1.352	(0.172)	1.193	1.530
Election + 2	1.327	(0.126)	1.216	1.472	Election + 2	1.779	(0.226)	1.593	1.995
Election + 3	1.171	(0.112)	1.055	1.319	Election + 3	1.456	(0.185)	1.254	1.693
Election + 4	1.008	(0.096)	0.887	1.179	Election + 4	1.172	(0.149)	0.973	1.403
Election + 5	1.126	(0.107)	1.017	1.267	Election + 5	1.386	(0.176)	1.189	1.625
Within-election	0.931	(0.460)	0.918	0.927	Within-election	0.931	(0.460)	0.921	0.936
Shift	0.126	(0.012)	0.115	0.210	Shift	0.095	(0.012)	0.087	0.158
Second sample half (Panel (c) of Online Appendix Figure 2)					*Region-specific model (Panel (b) of Online Appendix Figure 4)*				
Current	0.563	(0.030)	-	-	Current	1.000	(0.095)	-	-
Even split	0.590	(0.032)	-	-	Even split	0.817	(0.078)	-	-
Election + 1	0.516	(0.028)	0.403	0.702	Election + 1	1.032	(0.098)	0.931	1.168
Election + 2	0.679	(0.036)	0.508	0.902	Election + 2	1.310	(0.125)	1.202	1.444
Election + 3	0.554	(0.030)	0.390	0.796	Election + 3	1.110	(0.106)	0.992	1.272
Election + 4	0.518	(0.028)	0.404	0.737	Election + 4	1.027	(0.098)	0.895	1.197
Election + 5	0.511	(0.027)	0.343	0.748	Election + 5	1.124	(0.107)	1.009	1.269
Within-election	0.931	(0.460)	0.911	0.924	Within-election	0.897	(0.443)	0.862	0.874
Shift	0.224	(0.012)	0.205	0.373	Shift	0.157	(0.015)	0.147	0.273
Main specification with Jeffreys prior penalty (Panel (b) of Online Appendix Figure 5)					*Extended specification with Jeffreys prior penalty (Panel (c) of Online Appendix Figure 5)*				
Current	1.000	(0.095)	-	-	Current	1.000	(0.095)	-	-
Even split	0.817	(0.078)	-	-	Even split	0.817	(0.078)	-	-
Election + 1	0.987	(0.094)	0.921	1.075	Election + 1	0.934	(0.089)	0.869	1.019
Election + 2	1.316	(0.125)	1.252	1.391	Election + 2	1.297	(0.124)	1.228	1.370
Election + 3	1.167	(0.111)	1.098	1.247	Election + 3	1.224	(0.117)	1.148	1.299
Election + 4	0.999	(0.095)	0.924	1.090	Election + 4	1.026	(0.098)	0.951	1.119
Election + 5	1.117	(0.106)	1.048	1.204	Election + 5	1.172	(0.112)	1.101	1.261
Within-election	0.931	(0.460)	0.927	0.936	Within-election	0.878	(0.434)	0.872	0.885
Shift	0.115	(0.011)	0.105	0.178	Shift	0.199	(0.019)	0.178	0.283
Main specification with L1 penalty (Panel (d) of Online Appendix Figure 5)					*Extended specification with L1 penalty (Panel (e) of Online Appendix Figure 5)*				
Current	1.000	(0.095)	-	-	Current	1.000	(0.095)	-	-
Even split	0.817	(0.078)	-	-	Even split	0.817	(0.078)	-	-
Election + 1	0.985	(0.094)	0.899	1.117	Election + 1	0.963	(0.092)	0.874	1.090
Election + 2	1.316	(0.125)	1.201	1.457	Election + 2	1.304	(0.124)	1.196	1.451
Election + 3	1.159	(0.111)	1.029	1.298	Election + 3	1.195	(0.114)	1.077	1.342
Election + 4	0.976	(0.093)	0.852	1.145	Election + 4	1.012	(0.096)	0.886	1.174
Election + 5	1.099	(0.105)	0.995	1.255	Election + 5	1.183	(0.113)	1.067	1.332
Within-election	0.933	(0.461)	0.925	0.937	Within-election	0.888	(0.438)	0.874	0.892
Shift	0.105	(0.010)	0.084	0.157	Shift	0.168	(0.016)	0.136	0.231

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)					*Include respondents missing extended covariates (Panel (b) of Online Appendix Figure 6)*				
Current	1.000	(0.095)	-	-	Current	1.000	(0.095)	-	-
Even split	0.817	(0.078)	-	-	Even split	0.811	(0.077)	-	-
Election + 1	0.997	(0.095)	0.908	1.128	Election + 1	1.002	(0.095)	0.911	1.127
Election + 2	1.327	(0.126)	1.216	1.472	Election + 2	1.324	(0.125)	1.217	1.463
Election + 3	1.171	(0.112)	1.055	1.319	Election + 3	1.146	(0.108)	1.033	1.291
Election + 4	1.008	(0.096)	0.887	1.179	Election + 4	0.989	(0.093)	0.868	1.155
Election + 5	1.126	(0.107)	1.017	1.267	Election + 5	1.124	(0.106)	1.016	1.269
Within-election	0.931	(0.460)	0.918	0.927	Within-election	0.930	(0.460)	0.918	0.927
Shift	0.126	(0.012)	0.115	0.210	Shift	0.127	(0.012)	0.116	0.212
Include all voters (Panel (c) of Online Appendix Figure 6)					*Electoral College (Panel (b) of Online Appendix Figure 7)*				
Current	1.000	(0.095)	-	-	Current	1.000	(0.344)	-	-
Even split	0.800	(0.076)	-	-	Even split	0.678	(0.233)	-	-
Election + 1	1.007	(0.095)	0.915	1.131	Election + 1	0.954	(0.328)	0.803	1.077
Election + 2	1.311	(0.124)	1.203	1.453	Election + 2	1.331	(0.458)	1.119	1.355
Election + 3	1.125	(0.106)	1.022	1.274	Election + 3	1.167	(0.401)	0.985	1.225
Election + 4	0.974	(0.092)	0.855	1.135	Election + 4	1.102	(0.379)	0.861	1.149
Election + 5	1.096	(0.104)	0.989	1.247	Election + 5	1.327	(0.456)	1.059	1.313
Within-election	0.928	(0.459)	0.915	0.925	Within-election	0.931	(0.460)	0.918	0.927
Shift	0.138	(0.013)	0.127	0.222	Shift	0.183	(0.063)	0.166	0.375
County-level logistic regression (Panel (b) of Online Appendix Figure 8)					*County-level population change (Panel (c) of Online Appendix Figure 8)*				
Current	1.000	(0.069)	-	-	Current	1.000	(0.069)	-	-
Even split	0.796	(0.055)	-	-	Even split	0.796	(0.055)	-	-
Election + 1	1.275	(0.088)	1.209	1.411	Election + 1	1.017	(0.070)	-	-
Election + 2	2.109	(0.146)	1.983	2.287	Election + 2	1.350	(0.093)	-	-
Election + 3	2.460	(0.170)	2.347	2.625	Election + 3	1.163	(0.081)	-	-
Election + 4	2.559	(0.177)	2.446	2.765	Election + 4	1.013	(0.070)	-	-
Election + 5	3.203	(0.222)	3.097	3.480	Election + 5	1.273	(0.088)	-	-
Within-election	0.669	(0.082)	0.648	0.665	Within-election	-	-	-	-
Shift	0.664	(0.046)	0.563	0.838	Shift	0.043	(0.003)	-	-
Averaging over current and previous election (Panel (a) of Online Appendix Figure 9)					*Averaging over current and previous two elections (Panel (a) of Online Appendix Figure 9)*				
Current	1.000	(0.095)	-	-	Current	1.000	(0.095)	-	-
Even split	0.817	(0.078)	-	-	Even split	0.817	(0.078)	-	-
Election + 1	1.297	(0.124)	1.238	1.370	Election + 1	1.042	(0.099)	0.983	1.123
Election + 2	1.299	(0.124)	1.229	1.372	Election + 2	1.170	(0.112)	1.104	1.247
Election + 3	1.266	(0.121)	1.184	1.346	Election + 3	1.095	(0.104)	1.023	1.176
Election + 4	1.168	(0.111)	1.096	1.252	Election + 4	0.971	(0.093)	0.892	1.063
Election + 5	1.167	(0.111)	1.094	1.258	Election + 5	0.963	(0.092)	0.888	1.056
Within-election	0.931	(0.460)	0.926	0.935	Within-election	0.931	(0.460)	0.926	0.935
Shift	0.724	(0.069)	0.693	0.766	Shift	0.588	(0.056)	0.556	0.630

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

Exercise	RMSE		CI Bound	
	Relative	(Absolute)	Lower	Upper
Main specification (Panel (a) of Figure 1)				
Current	1.000	(0.095)	-	-
Even split	0.817	(0.078)	-	-
Election + 1	0.997	(0.095)	0.908	1.128
Election + 2	1.327	(0.126)	1.216	1.472
Election + 3	1.171	(0.112)	1.055	1.319
Election + 4	1.008	(0.096)	0.887	1.179
Election + 5	1.126	(0.107)	1.017	1.267
Within-election	0.931	(0.460)	0.918	0.927
Shift	0.126	(0.012)	0.115	0.210
Extrapolating from two elections back (Panel (b) of Online Appendix Figure 9)				
Current	1.000	(0.095)	-	-
Even split	0.817	(0.078)	-	-
Election + 1	1.020	(0.097)	0.954	1.110
Election + 2	1.611	(0.154)	1.544	1.691
Election + 3	1.625	(0.155)	1.542	1.715
Election + 4	1.449	(0.138)	1.369	1.544
Election + 5	1.625	(0.155)	1.552	1.719
Within-election	0.931	(0.460)	0.926	0.935
Shift	0.325	(0.031)	0.294	0.367

Exercise	RMSE		CI Bound	
	Relative	(Absolute)	Lower	Upper
Extrapolating from previous election (Panel (b) of Online Appendix Figure 9)				
Current	1.000	(0.095)	-	-
Even split	0.817	(0.078)	-	-
Election + 1	1.305	(0.124)	1.240	1.391
Election + 2	2.208	(0.210)	2.134	2.291
Election + 3	2.306	(0.220)	2.223	2.403
Election + 4	2.154	(0.205)	2.067	2.250
Election + 5	2.431	(0.232)	2.341	2.524
Within-election	0.931	(0.460)	0.926	0.935
Shift	1.091	(0.104)	1.049	1.133
Extrapolating from three elections back (Panel (b) of Online Appendix Figure 9)				
Current	1.000	(0.095)	-	-
Even split	0.817	(0.078)	-	-
Election + 1	0.985	(0.094)	0.920	1.073
Election + 2	1.286	(0.123)	1.216	1.372
Election + 3	1.354	(0.129)	1.279	1.444
Election + 4	1.334	(0.127)	1.265	1.416
Election + 5	1.609	(0.153)	1.535	1.703
Within-election	0.931	(0.460)	0.926	0.935
Shift	0.357	(0.034)	0.325	0.399

Notes: 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, 2016-2020	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	1956-2020	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	1964-2020	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	1984-2020	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	1964-2020	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	1970-2020	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	1988-2020	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	1986-2020	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 ANES "Most Important Problem" question was included in surveys from 1960 through 2000. 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

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

Ferrara, A., P. A. Testa, and L. Zhou (2022). New area- and population-based geographic crosswalks for U.S. counties and congressional districts, 1790-2020. Inter-university Consortium for Political and Social Research (ICPSR). V4. https://doi.org/10.3886/E150101V4. Accessed in April 2024.

Ferrara, A., P. A. Testa, and L. Zhou (2024). New area- and population-based geographic crosswalks for U.S. counties and congressional districts, 1790-2020. *Historical Methods: A Journal of Quantitative and Interdisciplinary History 57*(2), 67–79. Accessed in April 2024.

Haines, M. R. and Inter-university Consortium for Political and Social Research (2010). Historical, demographic, economic, and social data: The United States, 1790-2002. Inter-university Consortium for Political and Social Research. V3. https://doi.org/10.3886/ICPSR02896.v3. Accessed in April 2024.

Inter-university Consortium for Political and Social Research (2008). County characteristics, 2000-2007 [United States]. Inter-university Consortium for Political and Social Research. V2. https://doi.org/10.3886/ICPSR20660.v2. Accessed in April 2024.

Leip, D. (2016). Dave Leip U.S. Presidential general county election results. Harvard Dataverse. V10. https://doi.org/10.7910/DVN/SUCQ52. Accessed October 2024.

US Bureau of Economic Analysis (1969–2022). Regional economic accounts: CAINC1 - annual personal income by county, 1969–2022. https://apps.bea.gov/regional/downloadzip.htm?_gl=1*5sfuwt*_ga*MjEzODIzMTQ1NS4xNzE3ODc4MjY0*_ga_J4698JNNFT*MTcxOTI0MjIxNS40LjEuMTcxOTI0MjQxOC41My4wLjA. Accessed in June 2024.

US Bureau of Labor Statistics (2024). Consumer Price Index for all urban consumers: All

items in US city average [cpiaucsl]. FRED, Federal Reserve Bank of St. Louis. https://fred.stlouisfed.org/series/CPIAUCSL. Accessed in June 2024.

US Census Bureau (1950–2010). US Decennial Census data. Social Explorer. https://www.socialexplorer.com/tables/C1950CompDS, https://www.socialexplorer.com/tables/C1960CountyDS, https://www.socialexplorer.com/tables/C1970, https://www.socialexplorer.com/tables/C1980, https://www.socialexplorer.com/tables/C1990, https://www.socialexplorer.com/tables/C2000, https://www.socialexplorer.com/tables/C2010. Accessed June 2024.

US Census Bureau (1970-2000). US Decennial Census data. Economic Research Service, United States Department of Agriculture. https://www.ers.usda.gov/data-products/county-level-data-sets/county-level-data-sets-download-data/. Accessed June 2024.

US Census Bureau (2023). Annual county resident population estimates by age, sex, race, and Hispanic origin: April 1, 2020 to July 1, 2023 [CC-EST2023-ALLDATA]. Prepared by the United States Census Bureau, Population Division. Accessed October 2024.

US Census Bureau (2024). American Community Survey data, 2008-2012 and 2018-2022. Economic Research Service, United States Department of Agriculture. https://www.ers.usda.gov/data-products/county-level-data-sets/county-level-data-sets-download-data/. Accessed June 2024.