Electoral predictors of polling errors

Sina Chen, John Körtner, Peter Selb, Jens Wiederspohn

May 17, 2023

Abstract

To understand when polls are accurate and when they fail, we adopt a Bayesian hierarchical modeling approach that separates poll bias and variance at the election level, and links error components to a broad range of election features including mobilization, candidacies, polarization, and electoral conduct. An empirical study of 9,298 pre-election polls across the 367 U.S. Senate elections, 1990-2022, reveals an overall trend toward smaller but more uniform errors over time, a negative association between poll variance and mobilization and polarization, and a tendency to underestimate more ideologically extreme Republican candidates. Moreover, Republican poll bias has a modestly positive link with the level of state democracy. Contrary to theoretical expectations, we find little evidence that female or minority candidates are overestimated in polls. While large parts of the variance remain unaccounted for, the empirical approach is promising and easily extended to include other potential error sources.

1 Introduction

Failures to predict landmark elections such as the 2016 U.S. presidential election and the Brexit vote have, rightly or wrongly, rattled public confidence in election polls and the survey method at large (Johnson 2018). Beginning with the polling debacle of the Truman–Dewey presidential race in 1948 (Mosteller et al. 1949), expert committees have been convened in the aftermath of such incidents to investigate potential error sources (for an overview of British and U.S. studies, see Prosser and Mellon 2018). The natural focus of such case studies has been on within-election differences in poll accuracy across firms, sponsors, time, geography, sampling frames, survey modes, fieldwork efforts, response rates, and adjustment methods. Yet poll features alone are not enough to account for poll accuracy (DeSart and Holbrook 2003). As Tudor and Wall (2021) demonstrate in their analysis of more than 20,000 polls across 400 national elections worldwide, the bulk of variance in poll accuracy could be observed between (and not within) elections. The crucial question of why the polls failed in a given election but not in others has often remained a matter of speculation. For instance, nonresponse patterns found in investigations of the 2016 U.S. presidential election polls have been interpreted as the merit of a controversial candidate who stoked "anti-media, anti-elite, and even anti-pollster sentiment" (Gelman and Azari 2017, 3). No matter how plausible, a credible test of such a conjecture would necessitate comparisons across electoral contexts with varying candidacies. In their study of the 2018 and 2020 U.S. Senate elections, Chen and Körtner (2022) found no indication that like-minded candidates endorsed by Donald Trump were underestimated to a greater degree than other Republicans in the polls.

The aim of our study is to develop a contextual understanding of polling errors and their triggers. We adopt a Bayesian hierarchical modeling approach following Shirani-Mehr et al. (2018), which allows us to disentangle systematic and random errors at the election level and to extrapolate error to the election day. We then extend the model to include candidate- and election-level features. Whereas previous studies of U.S. Senate elections examined overall discrepancies between polls and election results, and thus confound poll bias and sampling variance (e.g. Crespi 1988; Hopkins 2009; Stout and Kline 2015), our approach allows us to specifically test triggers that are hypothesized to be linked with bias or variance. In the selection of potential predictors, we go beyond extant large-scale comparative work (e.g. Jennings and Wlezien 2018; Sohlberg and Branham 2020; Tudor and Wall 2021), and cover a broad range of electoral features that pundits and scholars have suspected of encouraging polling errors, including mobilization, candidacies, polarization, and electoral conduct.

We apply the approach to 9,298 pre-election polls across 367 U.S. Senate elections in the period 1990–2022 and test a range of popular hypothesis about predictors.[1] To operationalize features, we rely on a wide variety of data sources and recent measurement advances in political science: data on voting-eligible population and turnout data, data on campaign expenditures, Wikipedia entries, predictions from name recognition (Xie 2022) and facial recognition models (Clarifai Inc. 2022), common-space campaign finance scores (Bonica 2014), data on state control and Grumbach's (2021) state democracy index.

The paper proceeds as follows. Section 2 outlines the total survey error (TSE) framework, which guides our theoretical discussion and empirical analysis of polling errors. Section 3 depicts the approach to model error components and provides descriptives of the distribution of estimated errors across federal states and electoral cycles over 1990–2022. Section 4 expounds our approach to include covariates in the statistical model, theories of polling errors offered in the literature and the media, specifies how we measured the constructs involved, and presents empirical results. The final section 5 summarizes and concludes.

2 Total error and its components

Pre-election polls—especially those conducted long before an election—are often said to be snapshots of public opinion rather than forecasts. As election day approaches, however, voting intentions should crystallize into actual voting behavior, and polls should reflect the

[1] By looking at two-party Republican vote shares, we avoid the difficulties associated with measuring polling errors across different electoral rules and party systems (e.g. Arzheimer and Evans 2014).

subsequent election result more and more accurately (e.g. Gelman and King 1993; Kaplan, Park and Gelman 2012). In survey methodological terms, the election result is the population parameter to be estimated with (late) polls, and the overall discrepancy between a poll estimate and the election result is the total error of the estimate. Statistical theory can be used to determine the sampling variability of an estimator (a.k.a. margin of error), but empirical studies demonstrate that the actual error of polls is, on average, about twice as large as that implied by reported margins (e.g. Buchanan 1986; Schnell and Noack 2014; Shirani-Mehr et al. 2018; Selb et al. 2023).

The concept of TSE evolved in the 1940s among applied statisticians who realized that survey inferences based on sampling theory alone ignore important complications in survey practice, and therefore overstate the accuracy of estimates (see Groves and Lyberg 2010, for a historical overview). The TSE approach distinguishes between errors of measurement, where reported voting intentions do not correspond to future voting behavior, and errors of representation, where the pool of respondents does not properly reflect the target population (i.e. future voters), due to coverage, sampling, or nonresponse issues. Both error types may occur randomly, thus increasing the variance surrounding a survey statistic, or they may systematically pull a statistic in one particular direction, thus introducing bias.

Why is the distinction of polling errors along the representation–measurement and the bias–variance dimensions important for our understanding of polling failures? For one thing, it raises our awareness of the possibility that the same contextual features may impinge on different error components, either in mutually reinforcing or offsetting ways (see Selb and Munzert 2013). The worst-case scenario arguably materializes in contexts that foster both uniform bias and low variance (and thus high confidence) in estimates, as happened with the 2016 and 2020 presidential elections and the Brexit vote (see Jackson 2018).

3 Modeling error components

While conceptually valuable, the variance–bias distinction is not normally identified empirically. Even in situations where we know the population parameter (on declaration of the election result, in the case of polls), we still cannot observe the sampling distribution of an estimator with a single survey. Consequently, most studies of polling error only look at total error (or transformations thereof), thereby confounding error components. The proliferation of election polling over the past three decades, which replicates more or less the same sampling process over and over, has created a rare opportunity to observe the sampling distribution and thus to identify the decomposition.

3.1 Statistical model

Shirani-Mehr et al. (2018) propose a Bayesian hierarchical model that is fit to numerous polls per election to disentangle bias and variance in poll estimates at the election level. They model the two-party Republican vote share p_j measured in poll j as a random draw from a normal distribution with mean π_j and variance σ_j^2:

$$p_j \sim \text{Normal}(\pi_j, \sigma_j^2), \tag{1}$$

$$\text{logit}(\pi_j) = \text{logit}(P_{r[j]}) + \alpha_{r[j]} + \beta_{1r[j]} t_j, \tag{2}$$

$$\sigma_j^2 = \pi_j(1 - \pi_j)/n_j + \phi_{r[j]}^2. \tag{3}$$

In equation (2) the logit[2] of the mean π_j is decomposed into the logit of the actual two-party vote for the Republicans, $P_{r[j]}$, where $r[j]$ identifies the election for poll j, an election-specific bias term, $\alpha_{r[j]}$, and an election-specific time trend, $\beta_{r[j]} t_j$, to account for changes in public opinion over the course of the campaign. The variance σ_j^2 (equation (3)) is composed of the analytic sampling variance of a binomial proportion under simple random sampling (SRS), $\pi_j(1 - \pi_j)/n_j$, where n_j is the sample size, and $\phi_{r[j]}^2$ is variance in excess

[2] The logit scale ensures that estimated poll support is bound between 0 and 1.

of SRS variance due to cluster sampling, nonresponse weighting, and measurement error in the survey variable (see Frankel 2010). See section **??** in the *Supplementary materials* for details on prior specifications.

The fitted model allows us to estimate several election-level quantities of interest: *Election-day bias* b_{0r} is obtained by setting the temporal distance to election day t_j to 0,

$$\mathrm{logit}(\pi_{0r}) = \mathrm{logit}(P_r) + \alpha_r, \tag{4}$$

and then subtracting the Republican two-party vote P_r,

$$b_{0r} = \pi_{0r} - P_r.$$

The *expected election-day bias*, $\mathrm{E}(b_{0r})$, is defined by replacing α_r in equation (4) with the expectation of the election-specific bias μ_α. Positive (negative) values of b_{0r} indicate that polls would, on average, overestimate (underestimate) the Republican candidate on election day. An advantage of estimating bias by extrapolating to election day is that we can utilize polls conducted well ahead of the election for parameter estimation. In contrast, most previous studies are limited to final polls to avoid mistaking swings in public opinion for polling errors, thus discarding valuable data (e.g. Panagopoulos 2021). Finally, *excess standard deviation*, ϕ_r, is used to measure random fluctuation (as opposed to bias), with *expected excess standard deviation* $\mathrm{E}(\phi_r)$ being defined as $\sqrt{\mu_\phi}$.

3.2 Data and overall patterns

To illustrate how the statistical model operates, figure 1 plots observed total error distributions (A), and estimated election-day biases and standard deviations (B) in 367 senate elections by electoral cycle covering 1990–2022.

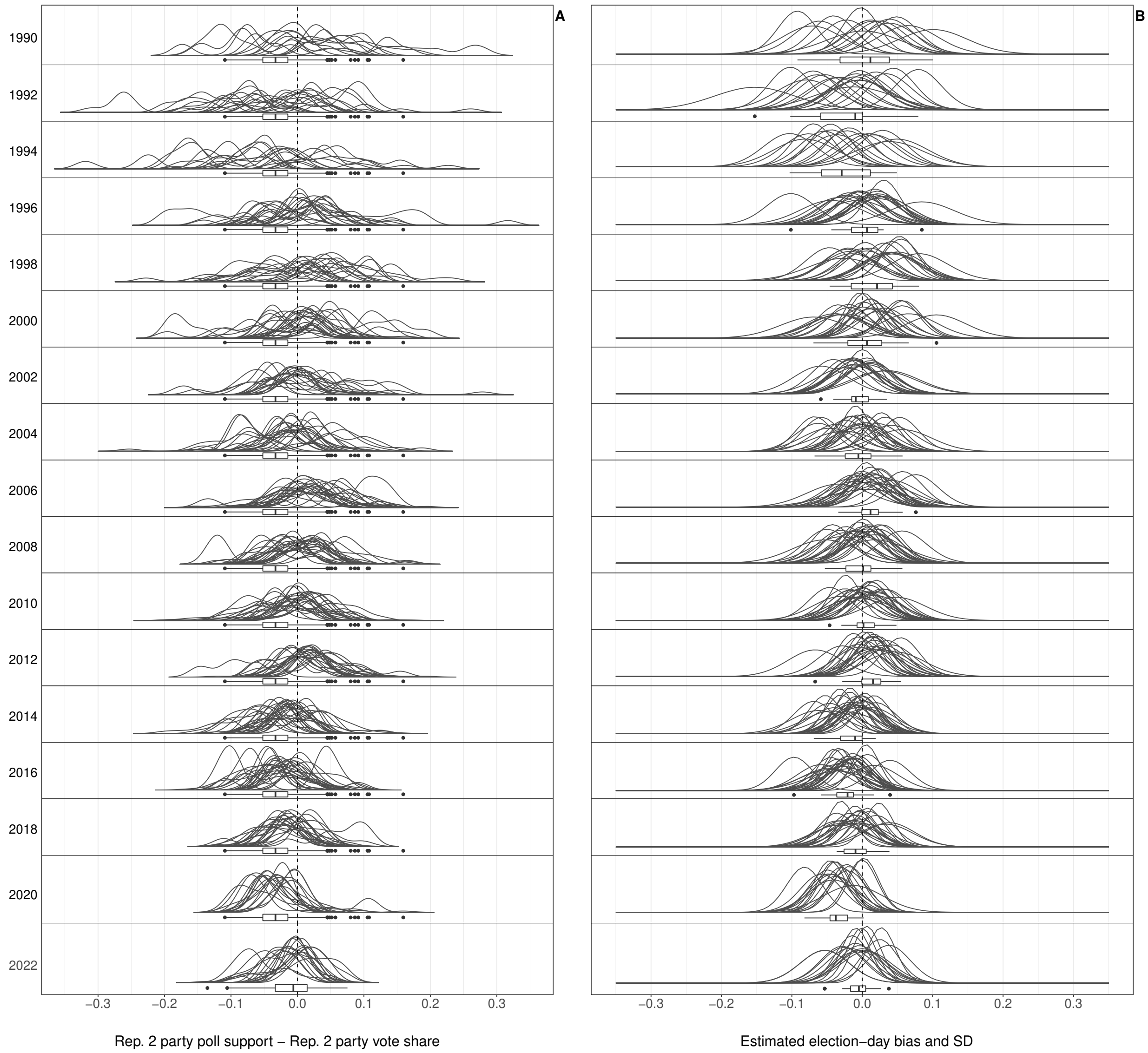

Figure 1: Observed TSE distributions over election cycles, 1990–2022 (A). Each density curve represents one election. Boxplots show the median and scatter of TSE for each election year. Estimated election-day bias and standard deviation (B). Each (normal) distribution with mean and scale based on these estimated quantities represents one election. Boxplots show the median and scatter of estimated election-day bias for each election cycle.

Election-day standard deviations are obtained by inserting π_{0r} into equation (3) and averaging all polls of the respective election. We analyze a total of 9,298 polls averaging 25 polls per election, with a minimum of five and a maximum of 128. Pre-election polling data from 1990 to 2020 was kindly provided by FiveThirtyEight upon request. For 2022, pre-election polling data from FiveThirtyEight is openly accessibly on their website.[3]

Evidently, estimated election-day biases are less extreme than observed average total errors. This occurs for two reasons: First, the election-specific bias parameters, α_r, are given distributions (for details, see section **??** in the *Supplementary materials*), which effectively shrink their values toward their mean. Second, by setting the time trend, $\beta_{1r[j]}t_j$, to zero, estimated election-day biases account for the fact that polls often converge to the election result as the election day approaches, whereas the TSE distributions indiscriminately include polls conducted long before the election (see section **??** in the *Supplementary materials*).

Focusing on the estimated distributions, the over-time pattern is striking: in the 1990s, estimated election-day biases scattered widely across states, indicating a whole bunch of marked polling failures. Regarding both the magnitude and direction of biases, there has been a trend toward greater uniformity since the early 2000s. At the same time, poll variances tended to decrease over time. This trend culminated during the 2014 to 2020 election cycles, which were characterized by mostly moderate but consistent poll biases against Republican candidates. The bias was found to be independent of the sampling frames and survey modes used (Clinton et al. 2020).

In 2022, however, the pattern vanished, possibly due to the emergence of a number of Republican-leaning pollsters that overestimated their preferred candidates, thus compensating for the previous pro-Democrat bias (e.g. Cohn 2022). Tendencies toward smaller but more uniform poll biases and decreased poll variances might be indicative of various phenomena, including improvements in and standardization of polling methods, pollsters adjusting their results according to the results of other polls ("herding"), or a convergence of contextual factors and methodological problems across statewide contests (Abramowitz and Webster 2016). The following section discusses such contextual factors and their potential linkages to polling error components.

[3] `https://github.com/fivethirtyeight/data/tree/master/polls`.

4 Features of the electoral contest and their linkages to error components

In this section, we extend the statistical model described in section 3 to allow the inclusion of covariates. We then review the theory and previous empirical work that attribute polling errors to features of the electoral contest. We describe how we measured the relevant constructs for the 367 U.S. senate elections under scrutiny, and estimate their association with relevant error components.[4] To encourage the reader to make sense of the data on their own, we use visual tools to present the results as informatively as possible.

4.1 Including covariates

We extend the model described in section 3 to include election-level covariates as predictors of bias α_r and excess variance ϕ_r^2, which are defined in section 3.1. Hence, we re-specify α_r as

$$\alpha_r \sim \text{Normal}(\mu_\alpha + \beta_2 \text{feature}_r, \sigma_\alpha^2),$$

with feature_r being a feature varying between elections. If it additionally varies between parties, the mean of α_r is defined as $\mu_\alpha + \beta_2 \text{featureDEM}_r + \beta_3 \text{featureREP}_r + \beta_4 \text{featureDEM}_r$ featureREP_r, with featureDEM_r being the measurement for the Democrat candidate and featureREP_r for the Republican candidate.[5] To estimate election day poll bias while taking into account contextual features, this re-specification of α_r is used to approximate the election-day mean in equation (4) from which then the Republican two-party vote P_r is subtracted.

[4]Note that we include covariate measurements in the above model specifications one by one in order to avoid computational difficulties associated with sparse data.

[5]Note that $\beta_4 \text{featureDEM}_r \text{featureREP}_r$ drops in cases, like incumbency, where we cannot observe a feature for both candidates at the same time.

The excess variance, ϕ_r^2, is redefined as

$$\phi_r^2 \sim \text{Normal}(\mu_\phi + \gamma \text{feature}_r, \sigma_\phi^2).$$

To estimate the excess standard deviation, ϕ_r, while accounting for contextual features, we take the square root of this expression. For estimating the expected excess standard deviation, the square root of $\mu_\phi + \gamma\text{feature}_r$ is taken, respectively.

4.2 Linking contextual features and error components: theory and evidence

In this section, we summarize contextual theories of polling errors offered in the literature and the media, report previous evidence, specify how we measured the constructs involved, and present our own empirical results.

4.2.0.1 Electoral mobilization. Pollsters often blame unusual turnout for polling failures (see Asher 2016) but classical theories of electoral mobilization offer conflicting implications. Campbell's (1960) surge-and-decline model suggests that high-turnout elections should be more difficult to predict due to the inclusion of peripheral voters with low turnout propensities and unstable preferences. In contrast, Lazarsfeld, Berelson and Gaudet's (1968) notion of election campaigns as activators of voters' latent preferences implies that increased mobilization efforts enhance voter information and reduce measurement variance in voting intentions. Tingsten's (1963) law of dispersion maintains that the higher the turnout, the more evenly distributed electoral participation will be across social groups, which could help pollsters sort out likely voters and adjust for nonresponse (Sohlberg and Branham 2020). All arguments have in common that they do not imply any immediate directional effect on poll bias favoring one political party or another. Rather, Campbell (1960) suggests that electoral mobilization is positively related with measurement variance in poll estimates, while

Lazarsfeld, Berelson and Gaudet (1968) and Tingsten (1963) imply that higher turnout limits measurement variance.

The empirical evidence so far is inconclusive. In his seminal study of media-sponsored pre-election polls for offices at different federal levels in the U.S., Crespi (1988) finds the absolute polling error for the winning candidate to be negatively correlated with turnout, suggesting accuracy gains with higher turnout. Another regression analysis of the absolute total error in elections across 44 countries yields no significant (linear) relationship (Sohlberg and Branham 2020). Looking at elections across countries Daoust (2021) finds a (modestly) negative relationship between absolute poll error and turnout. Indirect evidence comes from Selb and Munzert (2013), who find that vote overreporting is positively associated with actual turnout. They conjecture that higher turnout intensifies social desirability pressure, making it harder to filter out nonvoters and ultimately leading to increased polling error.

In our analysis, we focus on (excess) standard deviation. We use two measures of electoral mobilization: actual turnout and campaign intensity. Actual turnout is measured using data about the size of the voting-eligible population from the U.S. Elections Project (`https://www.electproject.org`) and state-level turnout figures are drawn from the MIT Election Data + Science Lab (`https://electionlab.mit.edu`). To measure campaign intensity, we rely on logged total per capita (per state inhabitant) expenditures of the Republican and Democrat candidates in each senate race, adjusted for inflation. To that end, we scraped the total 1990–2022 campaign expenditures of each U.S. Senate candidate from the Federal Election Commission website (`https://www.fec.gov`). Unlike actual turnout, campaign intensity can be measured ahead of an election and could therefore be used to predict polling error in advance.

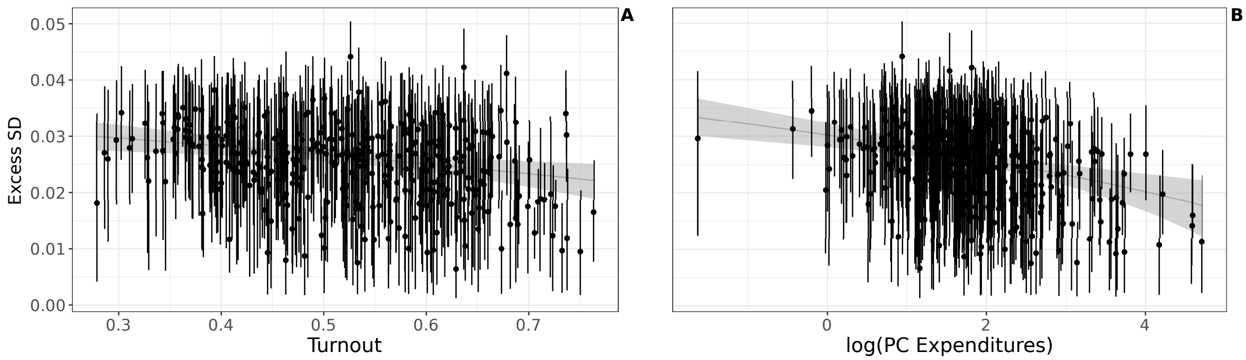

Figure 2: Estimated excess standard deviation vs. turnout (A) and logged per capita expenditures (B). Each point represents one election, with vertical lines showing the 95% credible intervals. Gray lines and shaded areas show the estimated expected excess standard deviation and 95% credible interval across all elections.

Figure 2 suggests negative, though noisy, associations between both indicators of electoral mobilization (turnout, log campaign expenditures per capita) and poll variance as measured by excess standard deviation. These findings tentatively support claims that increased electoral mobilization improves voter information and thus reduces measurement variance in voting intentions.

4.2.0.2 Frontrunners and incumbents. Gelman et al. (2016) hold that supporters of trailing candidates are less likely to participate in polls than supporters of the leading candidate, thus exaggerating the expected margin of victory of the presumed winner. Although the authors do not commit to a particular mechanism, there are obvious parallels with bandwagon dynamics (see Barnfield 2020, for a recent overview), meaning that candidates performing better in polls attract additional support, either through voter mobilization or conversion, merely due to their poll performance. If trailing candidates lose support due to their poor poll performance, this is sometimes referred to as a "spiral-of-silence" or "Titanic" effect. The supposed psychological mechanisms include gratification from winning, conformity pressure, and decision heuristics such that people rely on majority judgments. To be sure, bandwagon dynamics are thought to affect voter turnout and candidate choice at the election, so it does not seem far-fetched to assume that similar processes drive response behavior in polls. If bandwagon dynamics affect polls but not elections, polls should overestimate frontrunners. If they affect elections but not polls, polls should underestimate frontrunners. But if bandwagon dynamics affected both in equal measure, then there should be no frontrunner bias.

Similar arguments have been made regarding incumbents running for re-election. Not only do they enjoy an electoral advantage, see Mattei (1998), they are also overrated in the polls. Of the many mechanisms allegedly responsible for the incumbency advantage in elections (for an overview, see Mayhew 2008), only name recall and recognition seem to explain the overestimation of incumbents in polls, especially early on when challengers are not yet known (Kam and Zechmeister 2013).

Gelman et al. (2016) use a combination of traditional cross-sectional surveys and a huge high-frequency panel survey fielded during the 2012 U.S. presidential election campaign. They find that daily sample composition varied more than voting intentions in response to campaign events. They conclude that volatility in polls during the campaign were more likely due to differential sample composition than swing voters. Kennedy et al. (2018), on the other hand, do not find higher nonresponse rates (as an indication of spiral-of-silence or Titanic effects) in staunchly pro-Trump areas during the 2016 presidential race. Analyzing polls on 180 gubernatorial and senate elections over the period 1989–2006, Hopkins (2009) finds that frontrunners are regularly overestimated which, according to the above logic, might indicate bandwagon dynamics in polls but not votes. Mattei (1998) shows that U.S. House incumbents were overestimated in the 1996 ANES pre-election survey. Further, by comparing winners in open House districts with incumbent winners he finds that the overestimation of incumbents in the ANES postelection surveys in 1982–1996 is not a hidden frontrunner effect.

To distinguish between bandwagon dynamics in polls and voter participation, we measure frontrunners both at the beginning and at the end of the election campaign. We classify early frontrunners as those candidates who consistently lead the first three polls in the run-up to an election with more than five percentage points. At the end of the campaign, we measure the margin in the final poll between the Republican and Democrat candidates. Information about incumbency was scraped from Wikipedia.

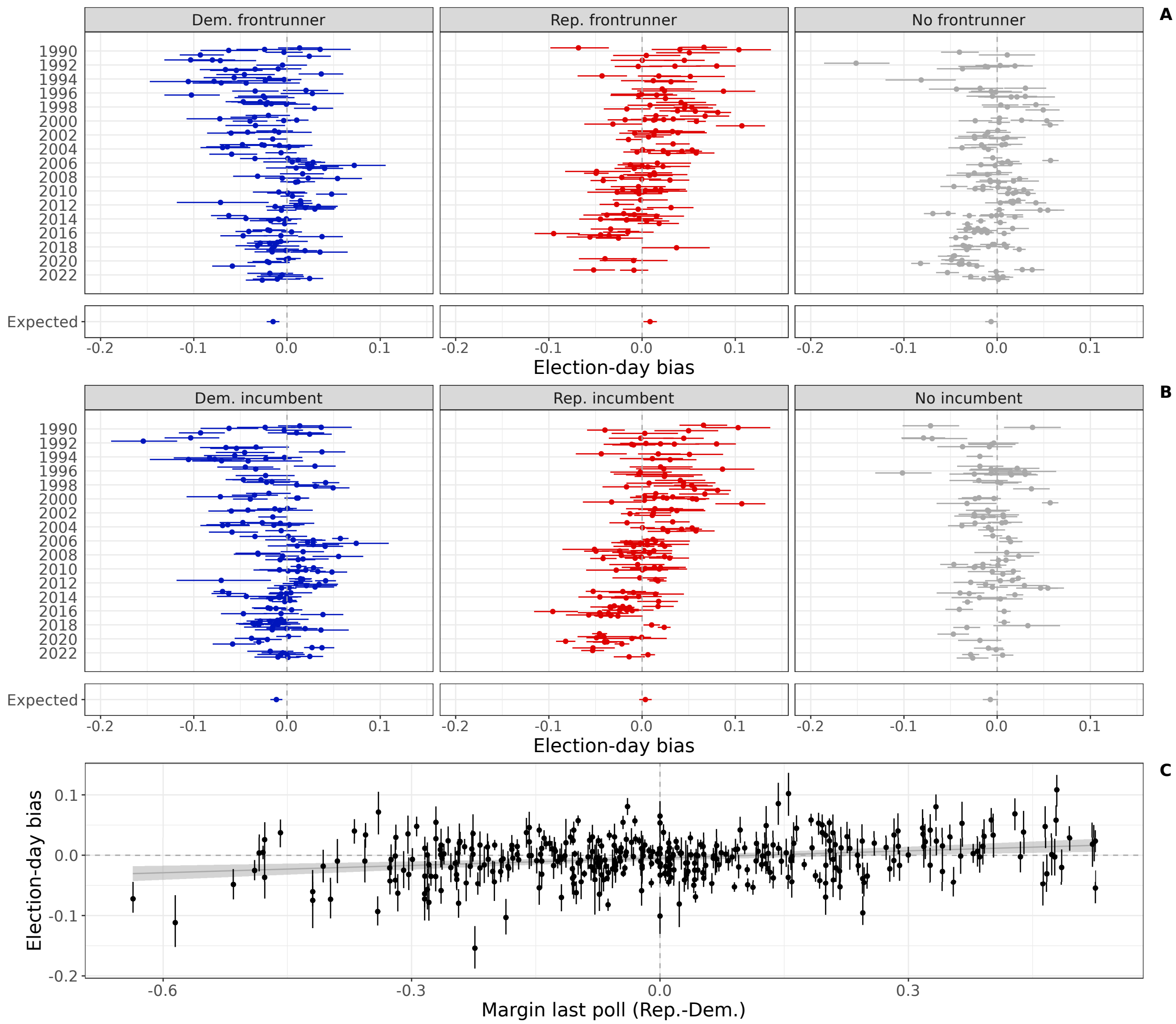

Figure 3: Estimated election-day bias by frontrunner status (A) and incumbency (B). Each point represents one election, with horizontal lines showing the 95% credible intervals. The *Expected* row shows the average estimated expected election-day bias across all elections. Estimated election-day bias vs. margin of last poll (C) with vertical lines showing the 95% credible intervals. The gray line and shaded area show the average estimated expected election-day bias and 95% credible interval across all elections.

The results shown in figure 3 indicate that up to 2004 there was an overestimation of frontrunners (pane A) as well as incumbents (pane B) supporting bandwagon effects and incumbency advantage claims. However, this association seems to vanish from the mid-2000s on. There is even a tendency to underestimate Republican frontrunners and incumbents, but this might be due to the overall pattern we observe for U.S. Senate pre-election polls. Further, the patterns in panes A and B are very similar, which can be explained by the high degree of overlap between frontrunners and incumbents (Johnston and Lachance 2022). Pane C shows that the more the Republican candidate leads in the last poll before election day, the more his or her vote share will be overestimated, in line with bandwagon dynamics.

4.2.0.3 Minority and female candidates. The "Bradley", "Wilder", "Whitman", or "Hillary" effect refers to the overestimation of minority or female candidates in pre-election polls. This bias arises because respondents may hesitate to admit their unwillingness to vote for these candidates due to perceived social acceptance of certain views, called socially desirable reporting (SDR). SDR is a well-established concept in survey research (e.g. Tourangeau and Yan 2007). Brown-Iannuzzi, Najle and Gervais (2019) provide evidence for SDR in self-reported voting preferences from a survey about political candidates, showing that respondents are less willing to vote for minorities when measured indirectly. Analyzing polls for elections contested by minority candidates and randomly sampled elections without minority candidates, Hopkins (2009) does not find support for an overestimation of female candidates, and only finds evidence of overestimated support for black candidates before 1996, suggesting that the effect might have been present in the past.

Stout and Kline (2008) formulate misreporting within a framework of preference falsification: if the individual utility from voicing conforming preferences out-weights the utility gained by expressing true beliefs, individuals misstate their voting intentions. In Stout and Kline (2008, 2011, 2015), they study polls for U.S. senate and gubernatorial elections with female and black candidates and a matched sample without. They find an underestimation of female candidates compared to their white male counterparts. For black candidates, they find an overestimation in line with the general theory. However, the coefficient is only significant conditional on the salience of ethnicity and the electoral strength of the candidate.

To code ethnicity, we label the candidates based on their names, pictures, and on background information from the internet where we looked for cues about descent (Chandra 2006). Based on this information, we apply a restrictive definition of minority status, since for this group of candidates we would expect the strongest association with election-day bias: we label candidates with predominantly black, Asian, Indian, Hispanic (excluding Cuban), and middle eastern cues. To validate our results, we relied on a name recognition model developed in Xie (2022) to predict ethnicity based on the first and last name of the candidates. In addition, we use a multi-model demographics workflow that detects faces based on facial recognition and estimates demographic characteristics of those faces (Clarifai Inc. 2022).

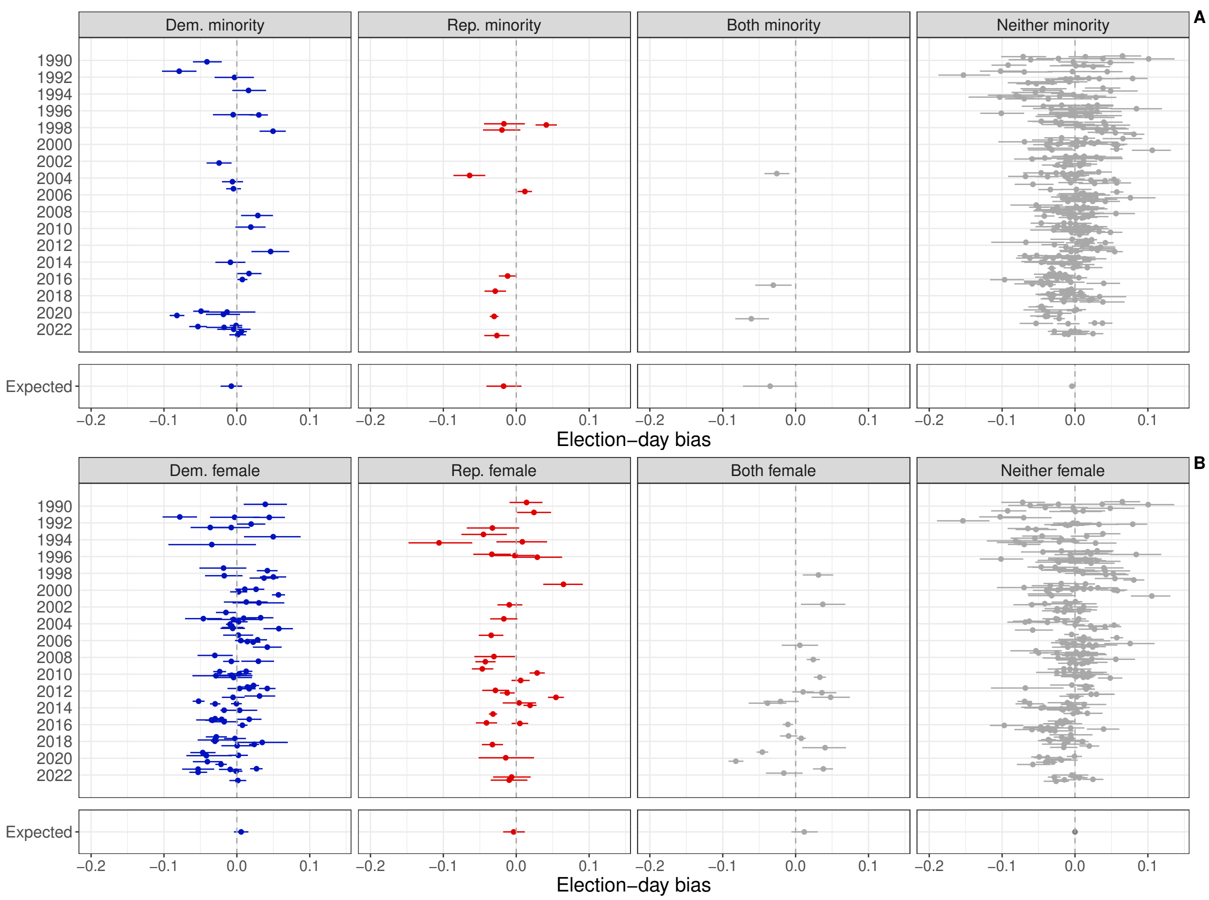

Figure 4: Estimated election-day bias by minority status (A) and gender (B). Each point represents one election, with horizontal lines showing the 95% credible intervals. The *Expected* row shows the average estimated expected election-day bias across all elections.

The results using our main labels are reported in figure 4.[6] To code a binary definition of gender, we rely on hand-coded labels (we also generated labels based on facial recognition, but the discrepancies could clearly be attributed to error).[7] We separately model the *Both minority* and *Both female* categories to ensure that our results are not biased by these elections.

In contrast with common social psychological reasoning, we do not find support for an overestimation of minority or female candidates. A possible interpretation of this finding is that the importance of identity differs across contexts (Stout and Kline 2015; Hopkins 2009). In line with our general strategy, however, we do not test for conditionality beyond the main predictors to avoid overfitting.

4.2.0.4 Candidate extremity and political polarization. Political ideology among electoral candidates can lead to partisan bias through several channels. First, supporters of extremist candidates might misreport their preference or refuse to participate due to SDR. Brownback and Novotny (2018) conducted list experiments to study SDR during the 2016 presidential campaign and found evidence that explicit polling overstated agreement with Clinton relative to Trump. Coppock (2017), also using list experiments during the same campaign, found no evidence of "shy" Trump supporters. Enns, Lagodny and Schuldt (2017) found evidence for the existence of "hidden" Trump supporters based on a forced-choice measure in a representative survey during the 2016 presidential campaign. Chen and Körtner (2022) compare U.S. Senate pre-election polls of Trump-endorsed candidates to other Republican candidates and find no evidence of an underestimation of extremist candidates.

Second, a partisan bias can arise from differential trust in the survey sponsor. Right-wing populists frequently attack the media and universities, who are among the main sponsors

[6] The results of the other approaches can be found in **??** of the *Supplementary materials*

[7] The transgender candidate Misty Snow is labeled as female (which is her chosen gender). Snow ran for the Democrats in Utah in 2016 and was overestimated with an approximate average election-day Republican poll bias of -0.035 (credible interval: -0.056, -0.013).

of polls. If the distrust and resulting unwillingness to participate in polls is associated with being right-wing, bias would be the result. Merkle and Edelman (2009) show that, when conducting exit polls in New Jersey and New York in 1997, offering folders and pens with VNS logos — an exit polling consortium formed by media companies from both sides of the political — produced a significant bias in favor of the Democrat candidate. A similar result is found by Bischoping and Schuman (1992) in the context of the 1990 Nicaraguan presidential election. Presser, Blair and Triplett (1992) and Tourangeau, Presser and Sun (2014) find evidence that varying sponsorship significantly affects reported opinions in survey experiments in the context of the campaign for mayor of Marion Barry in Washington in 1990 and in the run-up to the 2012 U.S. general elections, respectively. A third channel through which ideology may act on polling errors is polarization. Polarization between the major parties increases their distinguishability, which might make the election easier to predict because of reduced voter transitions. However, polarization also increases the expected party differential (Downs 1957). If the alternatives are seen to imply important differential consequences, the stimulation to vote will be relatively high, leading to an increased turnout of peripheral voters (Campbell 1960), which in turn can cause an adjustment error.

To measure the ideology of candidates, we rely on common-space campaign finance scores (CFscores) by Bonica (2014) based on campaign donation data, available from 1990 to 2018.[8] This score locates U.S. Senate candidates on a left–right scale. Using campaign donations as a basis for measurement is attractive because the data is available for senators as well as nonelected candidates, whereas standard measurements based on roll-call estimates are only computable for senators. To analyze whether ideologically extreme candidates are more underestimated, we run separate models for the CFscores of Democrat and Republican candidates. The results in pane A of figure 5, with CFscores ranging from liberal (negative values) to conservative (positive values), do not support this claim. We would expect that the more liberal a Democrat candidate is, the more the Republican vote share would be

[8]Results for 2020 and 2022 are based on candidates who had already run in previous elections.

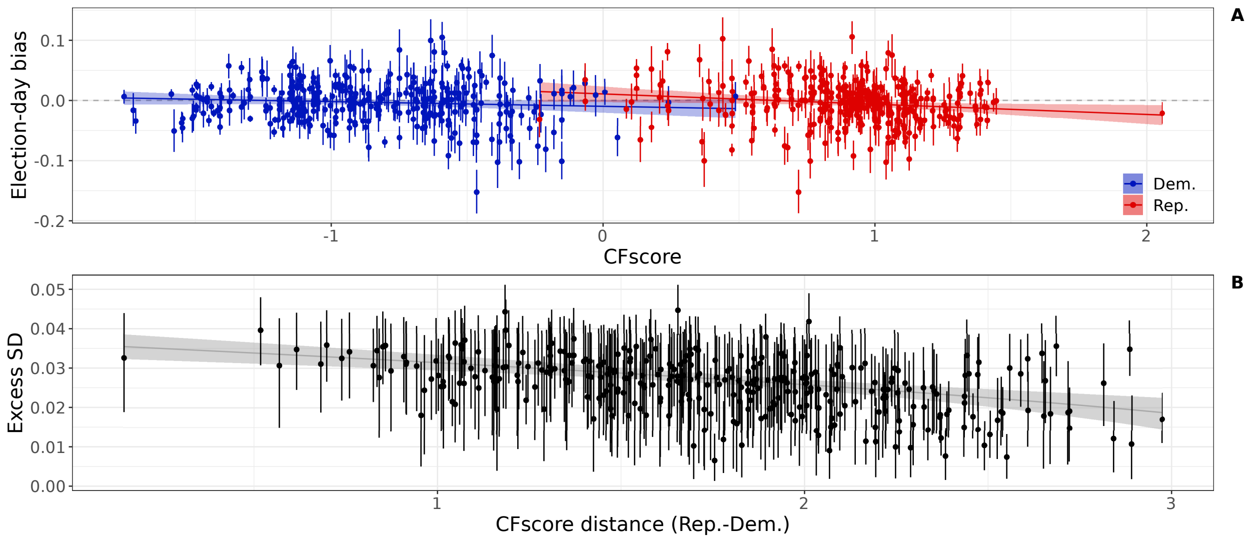

Figure 5: Estimated election-day bias vs. CFscore (A) and estimated excess standard deviation vs. CFscore distance (B). Each point represents one election, with vertical lines showing the 95% credible intervals. The blue and red (A) and gray (B) lines and shaded areas show the (average) estimated expected election-day bias and 95% credible interval across all elections.

overestimated, meaning blue points should be located above the horizontal line at 0. However, there is a slight tendency for conservative Republican candidates to be underestimated in this case.

The hypothesized link between polarization and polling error is one of variance, rather than bias. To approximate election-level polarization, we compute the ideological distance between the two major candidates in an election as the difference in their CFscores. In pane B of figure 5 we do find some support for decreasing excess standard deviation with higher levels of polarization operationalized by CFscore distance in line with the "distinguishability hypothesis". Note that over the last three decades candidates from both parties moved more to the extremes, as measured by CFscores.

4.2.0.5 Electoral conduct. An alternative explanation for deviations between pre-election polls and the observed vote share is that polls give an accurate picture of public opinion, but the actual election results do not. There are various mechanisms by which this "biased" vote can occur. In this study we focus on voter suppression, where parties continually use their discretion to determine voting requirements that disproportionately obstruct specific social groups (Helmke, Kroeger and Paine 2022). Tudor and Wall (2021) note that when individuals who intend to vote find themselves unable to do so, election results might be biased in relation to the true voting intentions reflected in the polls. For example, increasingly strict voter identification laws disproportionately affect minorities and the poor, who traditionally support the Democrats (e.g. Fraga 2018). Using smartphone data Chen et al. (2022) show, that voters in majority black neighborhoods wait substantively longer to cast their ballot compared to voters in majority white neighborhoods.

To measure the potential for voter suppression, we use an index developed in Grumbach (2021) to operationalize state democracy. The index provides yearly measures of the level of democracy for each state in the U.S. based on 51 items associated with gerrymandering,[9] the cost of voting, integrity, and observable democratic outcomes. In addition, we use an indicator of Republican state control, as prior evidence suggests that voter suppression is predominantly pursued by Republicans (e.g. Wang 2012). State control is defined as a majority in all three of the state Senate, state House, and the governorship (trifecta) (Helmke, Kroeger and Paine 2022). Data on state control is available for the whole period under analysis. For 1990, we collect the information on state control from the National Conference of State Legislatures (`https://www.ncsl.org`) and the National Governors Association (`https://www.nga.org`). For 1992 to 2022, the information is available from Ballotpedia (`https://ballotpedia.org`).

A weak, positive association between election-day bias and the State Democracy Index can be found in pane A of figure 6, meaning the lower the level of democracy in a state, the more the Republican candidate is underestimated. This is in line with the "biased" vote hypothesis. If we look at pane B in figure 6, the hypothesis that Republican candidates are underestimated in Republican-controlled states is supported in the years since 2014, but not before.

[9] While Gerrymandering is not an issue in U.S. Senate elections due to fixed state borders, it reflects the overall state of democracy.

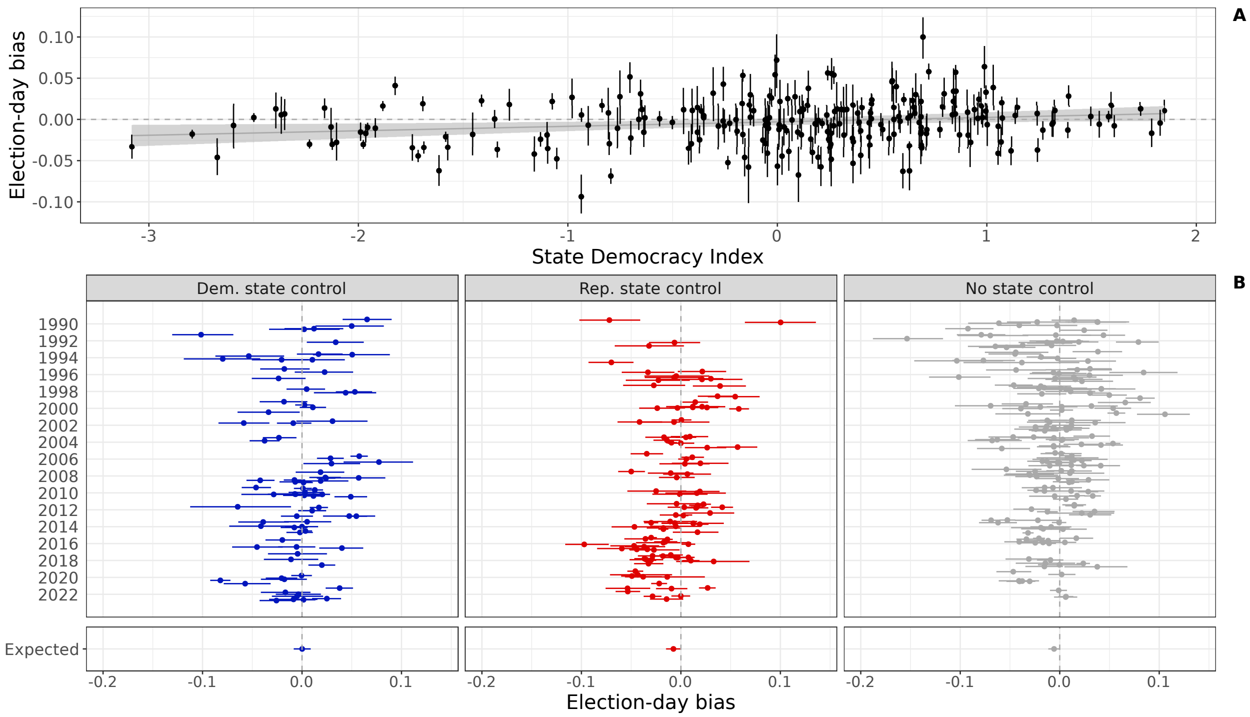

Figure 6: Estimated election-day bias vs. State Democracy Score (A). Each point represents one election, with vertical lines showing the 95% credible interval. The gray line and shaded area show the average estimated expected election-day bias and 95% credible interval across all elections. Estimated election-day bias by state control (B). Horizontal lines show the 95% credible intervals. The *Expected* row shows the average estimated expected election-day bias across all elections.

5 Summary and discussion

Our empirical analysis of more than 9,000 pre-election polls and contextual features revealed increasingly uniform patterns of polling errors across 367 U.S. Senate elections during 1990–2022. Beyond this, we find some discernable patterns of interest. First, estimated excess standard deviation tended to be negatively associated with both turnout and campaign expenditures. This finding lends tentative support to arguments that increased mobilization efforts activate voter preferences and thus reduce measurement variance in polls. Second, frontrunners of both parties were marginally overestimated which suggests some bandwagon dynamics in opinion polls (see Gelman et al. 2016), but not in elections. Third, we found little evidence that female or minority candidates were generally overestimated. Female Republican candidates were, if anything, underestimated. If "Bradley" effects were ever common, it must have been before our period of study (see Hopkins 2009). Fourth, we observed a slight tendency for more ideologically extreme Republican candidates to be underestimated in polls. Both social desirability pressure and anti-pollster sentiment among their supporters have been cited as possible mechanisms behind this phenomenon (e.g. Gelman and Azari 2017; Kennedy et al. 2018). Fifth, poll variance as measured by the estimated excess standard deviation tended to be negatively associated with political polarization, in line with the hypothesis that the clarity of electoral alternatives reduces volatility in voting intentions. Finally, there has been an increasing tendency for polls to underestimate Republican candidates in states that are controlled by the Republican party and/or that score low on the democracy index, which tentatively supports the suspicion that voter suppression may bias elections rather than polls. However, all these empirical results are relatively weak and should be treated with caution.

Several caveats are due here. Both theoretically and empirically, our focus was on marginal associations between single features of the electoral contest and components of polling error. Despite the considerable number of elections we studied, some of the subgroups we looked at were already small. For instance, there were only nine elections that saw Republican minority candidates competing against a white Democrat (see figure 4). Therefore, we could do little to control for potential confounding or effect modification without running the risk of overfitting our model. Our ability to test highly conditional theories, like those formulated in Stout and Kline (2015), was therefore limited. Where possible, we broke down the analyses and visually inspected associations across electoral cycles and states (see section **??** in the *Supplementary materials*). Another limitation is that we did not look at individual-level poll data, thus theoretical implications regarding sample composition and response behavior could not be directly observed (e.g. Gelman et al. 2016; Kaplan, Park and Gelman 2012). Given these limitations, our study could only scratch the surface of election-level sources of polling error and identify ways to approach the problem empirically. Still, our analyses left most of the variability in error components unaccounted for. An inherent difficulty in determining the election-level sources of polling errors is that pollsters constantly adapt their methods in response to failures. In that sense, we are aiming at a moving target. It remains to be seen how far future theoretical and empirical work will take us. Some observers in academia and industry are skeptical and suspect that polling errors are essentially unpredictable (e.g. Campbell 2022). In any case, this paper's approach to model bias and variance in election polls as a function of covariates provides a flexible basis for progress.

References

Abramowitz, Alan I and Steven Webster. 2016. "The rise of negative partisanship and the nationalization of US elections in the 21st century." *Electoral Studies* 41:12–22.

Arzheimer, Kai and Jocelyn Evans. 2014. "A new multinomial accuracy measure for polling bias." *Political Analysis* 22(1):31–44.

Asher, Herb. 2016. *Polling and the public: What every citizen should know*. Cq Press.

Barnfield, Matthew. 2020. "Think twice before jumping on the bandwagon: Clarifying concepts in research on the bandwagon effect." *Political Studies Review* 18(4):553–574.

Bischoping, Katherine and Howard Schuman. 1992. "Pens and polls in Nicaragua: An analysis of the 1990 preelection surveys." *American Journal of Political Science* pp. 331–350.

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

Brown-Iannuzzi, Jazmin L, Maxine B Najle and Will M Gervais. 2019. "The illusion of political tolerance: Social desirability and self-reported voting preferences." *Social Psychological and Personality Science* 10(3):364–373.

Brownback, Andy and Aaron Novotny. 2018. "Social desirability bias and polling errors in the 2016 presidential election." *Journal of Behavioral and Experimental Economics* 74:38–56.

Buchanan, William. 1986. "Election predictions: An empirical assessment." *Public Opinion Quarterly* 50(2):222–227.

Campbell, Angus. 1960. "Surge and decline: A study of electoral change." *Public Opinion Quarterly* 24(3):397–418.

Campbell, W Joseph. 2022. "Misfires and Surprises: Polling Embarrassments in Recent US Presidential Elections." *American Behavioral Scientist* .

Chandra, Kanchan. 2006. "What is ethnic identity and does it matter?" *Annual Review of Political Science* 9:397–424.

Chen, M Keith, Kareem Haggag, Devin G Pope and Ryne Rohla. 2022. "Racial disparities in voting wait times: evidence from smartphone data." *Review of Economics and Statistics* 104(6):1341–1350.

Chen, Sina and John Körtner. 2022. "Did "Trumpists" perform worse in US Senate election polls?" *Significance* 19(3):28–31.

Clarifai Inc. 2022. "Demographics Analysis.". Updated: May 11, 2022.
URL: *https://clarifai.com*

Clinton, Josh, Jennifer Agiesta, Megan Brennan, Camille Burge, Marjorie Connelly, Ariel Edwards-Levy, Bernard Fraga, Emily Guskin, D Sunshine Hillygus, Chris Jackson et al. 2020. "Pre-Election Polling: An Evaluation of the 2020 General Election Poll." *Report of a Task Force* pp. 1–105.

Cohn, Nate. 2022. "A Key Polling Warning Sign Has Dimmed Since September." *The New York Times* . November 6, 2022.

Coppock, Alexander. 2017. "Did shy Trump supporters bias the 2016 polls? Evidence from a nationally-representative list experiment." *Statistics, Politics and Policy* 8(1):29–40.

Crespi, Irving. 1988. *Pre-election polling: Sources of accuracy and error.* Russell Sage Foundation.

Daoust, Jean-François. 2021. "Blame it on turnout? Citizens' participation and polls' accuracy." *The British Journal of Politics and International Relations* 23(4):736–747.

DeSart, Jay and Thomas Holbrook. 2003. "Campaigns, polls, and the states: Assessing the accuracy of statewide presidential trial-heat polls." *Political Research Quarterly* 56(4):431–439.

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

Enns, Peter K, Julius Lagodny and Jonathon P Schuldt. 2017. "Understanding the 2016 US presidential polls: The importance of hidden Trump supporters." *Statistics, Politics and Policy* 8(1):41–63.

Fraga, Bernard L. 2018. *The turnout gap: Race, ethnicity, and political inequality in a diversifying America*. Cambridge University Press.

Frankel, Martin. 2010. Sampling theory. In *Handbook of Survey Research*, ed. James D Wright and Peter V Marsden. Emerald Group Bingley, UK pp. 83–138.

Gelman, Andrew and Gary 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.

Gelman, Andrew and Julia Azari. 2017. "19 things we learned from the 2016 election." *Statistics and Public Policy* 4(1):1–10.

Gelman, Andrew, Sharad Goel, Douglas Rivers, David Rothschild et al. 2016. "The mythical swing voter." *Quarterly Journal of Political Science* 11(1):103–130.

Groves, Robert M and Lars Lyberg. 2010. "Total survey error: Past, present, and future." *Public Opinion Quarterly* 74(5):849–879.

Grumbach, Jacob M. 2021. "Laboratories of democratic backsliding." *American Political Science Review* pp. 1–18.

Helmke, Gretchen, Mary Kroeger and Jack Paine. 2022. "Democracy by deterrence: Norms, constitutions, and electoral tilting." *American Journal of Political Science* 66(2):434–450.

Hopkins, Daniel J. 2009. "No more wilder effect, never a Whitman effect: When and why polls mislead about black and female candidates." *The Journal of Politics* 71(3):769–781.

Jackson, Natalie. 2018. The Rise of Poll Aggregation and Election Forecasting. In *Oxford Handbook of Polling and Polling Methods*, ed. Lonna Rae Atkeson and R. Michael Alvarez. Oxford: Oxford University Press.

Jennings, Will and Christopher Wlezien. 2018. "Election polling errors across time and space." *Nature Human Behaviour* 2(4):276.

Johnson, Timothy P. 2018. "Presidential Address: Legitimacy, Wicked Problems, and Public Opinion Research." *Public Opinion Quarterly* 82(3):614–621.

Johnston, Richard and Sarah Lachance. 2022. "The predictable campaign: Theory and evidence." *Electoral Studies* 75:102432.

Kam, Cindy D and Elizabeth J Zechmeister. 2013. "Name recognition and candidate support." *American Journal of Political Science* 57(4):971–986.

Kaplan, Noah, David K Park and Andrew Gelman. 2012. "Polls and elections understanding persuasion and activation in presidential campaigns: The random walk and mean reversion models." *Presidential Studies Quarterly* 42(4):843–866.

Kennedy, Courtney, Mark Blumenthal, Scott Clement, Joshua D Clinton, Claire Durand, Charles Franklin, Kyley McGeeney, Lee Miringoff, Kristen Olson, Douglas Rivers et al. 2018. "An evaluation of the 2016 election polls in the United States." *Public Opinion Quarterly* 82(1):1–33.

Lazarsfeld, Paul F, Bernard Berelson and Hazel Gaudet. 1968. The people's choice. In *The people's choice*. Columbia University Press.

Mattei, Franco. 1998. "Winning at the Polls and in the Polls: the Incumbency Advantage in Surveys of US House Voters." *Electoral Studies* 17(4):443–461.

Mayhew, David R. 2008. "Incumbency advantage in US presidential elections: The historical record." *Political Science Quarterly* 123(2):201–228.

Merkle, Daniel M and Murray Edelman. 2009. "An experiment on improving response rates and its unintended impact on survey error." *Survey Practice* 2(3):2964.

Mosteller, Frederick, Herbert Hyman, Phillip J. McCarthy, Eli J. Marks and David B. Truman. 1949. *The pre-election polls of 1948. Report to the Committee on Analysis of Pre-election Polls and Forecasts.* Social Science Research Council.

Panagopoulos, Costas. 2021. "Polls and elections accuracy and bias in the 2020 US general election polls." *Presidential Studies Quarterly* 51(1):214–227.

Presser, Stanley, Johnny Blair and Timothy Triplett. 1992. "Survey sponsorship, response rates, and response effects." *Social Science Quarterly* .

Prosser, Christopher and Jonathan Mellon. 2018. "The twilight of the polls? A review of trends in polling accuracy and the causes of polling misses." *Government and Opposition* 53(4):757–790.

Schnell, Rainer and Marcel Noack. 2014. "The accuracy of pre-election polling of German general elections." *methods, data, analyses* 8(1):20.

Selb, Peter and Simon Munzert. 2013. "Voter overrepresentation, vote misreporting, and turnout bias in postelection surveys." *Electoral Studies* 32(1):186–196.

Selb, Peter, Sina Chen, John Körtner and Philipp Bosch. 2023. "Bias and Variance in Multiparty Election Polls." *Public Opinion Quarterly* .

Shirani-Mehr, Houshmand, David Rothschild, Sharad Goel and Andrew Gelman. 2018. "Disentangling bias and variance in election polls." *Journal of the American Statistical Association* 113(522):607–614.

Sohlberg, Jacob and J Alexander Branham. 2020. "Just a difficult election to poll? How context affects polling accuracy." *Survey Methods: Insights from the Field (SMIF)* .

Stout, Christopher T and Reuben Kline. 2008. "Ashamed not to vote for an African-American; ashamed to vote for a woman: An analysis of the Bradley effect from 1982-2006.".

Stout, Christopher T and Reuben Kline. 2011. "I'm not voting for her: Polling discrepancies and female candidates." *Political Behavior* 33(3):479–503.

Stout, Christopher T and Reuben Kline. 2015. "Racial Salience, Viability, and the Wilder Effect: Evaluating Polling Accuracy for Black Candidates." *Public Opinion Quarterly* 79(4):994–1014.

Tingsten, Herbert. 1963. *Political behavior: Studies in election statistics*. Number 7 Bedminster Press.

Tourangeau, Roger, Stanley Presser and Hanyu Sun. 2014. "The impact of partisan sponsorship on political surveys." *Public Opinion Quarterly* 78(2):510–522.

Tourangeau, Roger and Ting Yan. 2007. "Sensitive questions in surveys." *Psychological Bulletin* 133(5):859.

Tudor, Jack and Matthew Wall. 2021. "A moving target? An analysis of the impact of electoral context on polling error variation in both British and international general elections." *Journal of Elections, Public Opinion and Parties* pp. 1–23.

Wang, Tova. 2012. *The politics of voter suppression: defending and expanding Americans' right to vote*. Cornell University Press.

Xie, Fangzhou. 2022. "rethnicity: An R package for predicting ethnicity from names." *SoftwareX* 17.