# THE RISK OF NARROW, DISPUTABLE RESULTS IN THE U.S. ELECTORAL COLLEGE: 1836-2020

Michael Geruso
Dean Spears

WORKING PAPER 27993

NBER WORKING PAPER SERIES

# THE RISK OF NARROW, DISPUTABLE RESULTS IN THE U.S. ELECTORAL COLLEGE: 1836-2020

Michael Geruso
Dean Spears

Working Paper 27993
http://www.nber.org/papers/w27993

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2020, Revised August 2023

We thank Samuel Arenberg, Marika Cabral, Maya Eden, Andrew Gelman, Brendan Kline, David Molitor, Paolo G. Piacquadio, Itai Sher, Ishaana Talesara, Caroline Thomas and seminar participants at the University of Texas at Austin, the 2021 Conference on Social Choice Theory and Applications, and the 2021 Welfare Economics and Economic Policy Seminar for helpful comments. We thank Nathan Franz for excellent research assistance. 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/w27993

The Risk of Narrow, Disputable Results in the U.S. Electoral College: 1836-2020
Michael Geruso and Dean Spears
NBER Working Paper No. 27993
October 2020, Revised August 2023
JEL No. H8,J1,J18,K16

## ABSTRACT

Close elections are important for many reasons, including that consequent election disputes can weaken democratic legitimacy and risk political violence. We quantify the probability of close outcomes in US presidential races with novel applications of empirical election models from several sources. We show that razor-thin margins are very likely under the Electoral College (EC). And we establish that the EC causes this closeness: It would not occur under any plausibly comparable popular vote system. The tendency of the EC to generate close elections is true today and throughout US presidential voting history.

Michael Geruso
University of Texas at Austin
Department of Economics
1 University Station C3100
Austin, TX 78712
and NBER
mike.geruso@utexas.edu

Dean Spears
University of Texas at Austin
Department of Economics
1 University Station C3100
Austin, TX 78712
and UTECS
dspears@utexas.edu

The University of Texas Electoral College Study is available at http:/utecs.org

# 1 Introduction

An election outcome is *close*[1] if changing some small fraction of the votes could reverse the outcome. A discipline-spanning literature in law, history, economics, and political science has been concerned with the consequences of close elections, including unclear results, weakened state capacity, and diminished legitimacy of leaders. Because the contribution of our paper is to comprehensively describe the *probability* and institutional *causes* of close elections in US presidential races, we begin by noting that this prior literature has identified a number of important *consequences*.

The prior literature has argued that: (i) Close elections are more vulnerable to manipulation or fraud, whether by foreign powers or domestic political actors (e.g., Posner, 2001; Hasen, 2005; Hirsch, 2020), and that the expectation of a close outcome may, in fact, incentivize fraud and other bad faith actions (Rapoport and Weinberg, 2000; Grofman and Feld, 2005). (ii) Close elections are more likely to be disputed, undercutting the value of elections in decisively transferring political power (Przeworski, 2018; Hasen, 2020). (iii) Election officials (often in partisan jobs) and judges may replace voters as the pivotal actors in a narrowly decided election (Rakove, 2004; Gelman, Katz and Bafumi, 2004), such as when making determinations about voter intent from partially punched paper ballots or discarding "naked" mail ballots.[2] (iv) Voter perceptions of fairness and election integrity are critical for legitimacy (see, e.g., Weatherford, 1992; Banducci and Karp, 2003), and narrow margins are empirically linked to lower voter confidence in election integrity (Birch, 2008; Sances and Stewart III, 2015). Indeed, many surveyed Americans viewed the results of the close, disputed presidential elections of 2000 and 2020 as illegitimate.[3] These considerations are independent of any normative claim about who should win an election, given a set of voter intentions or voter actions; they are about the effects of closeness *per se*.

In fact, several US presidential elections have been uncannily close. The election in 2000 hinged on a final tally of just 537 votes in a single state, Florida. The 1916 presidential election hinged on 3,773 votes in California. The 1884 election was decided by 1,149 votes in New York. The 1876 election was decided by 889 votes in South Carolina. Of course, disputation, perceived illegitimacy, and conflict can occur whether or not an election is close, but the 1876 Hayes-Tilden election provides an illustrative example of the type of legitimacy crisis that can attend close election outcomes and why a close

[1] We use the terms *close* and *narrow* as synonyms throughout, when referring to realized or potential election outcomes.

[2] Officials in each swing state typically discard thousands of mail ballots in US presidential elections due to signatures, enveloping, or postmarks. Such judgements are subjective and frequently contested. For example, in the 2016 presidential election, FL officials rejected 21,973 absentee ballots (U.S. Election Assistance Commission, 2017).

[3] See, e.g., Gallup (2001); Monmouth University Poll (2021); CNN (2021).

election may not produce a clear winner.[4] In that election, there were credible allegations that fraud favoring Hayes was pivotal to his statewide win in South Carolina. Because the presumptive Electoral College vote tally was within a single ballot nationally—185 for Hayes versus 184 for Tilden—a single Hayes elector lost would mean that neither candidate carried an Electoral College majority, which would trigger the US House to conduct its own vote to determine the presidency.[5] The razor-thin margin presented an opportunity to seize power through post-election maneuvers, and so several states sent opposed slates of electors (e.g., one backed by the state's Republican governor and another backed by its Democratic attorney general). Democrats in the US Senate threatened to filibuster the Electoral College vote count and certification process. Ultimately, the winner of the presidency was resolved via an unwritten, informal political compromise among party elites just a few days before the scheduled inauguration.[6] In short, the close 1876 election was attended by potentially pivotal fraud, several attempted subversions of statewide vote results by elected leaders, the threat of violence (including a shot fired at Hayes' residence and federal troops mobilized in the capital to brace for insurrection), and a president who was arguably selected via a political deal by party leaders, rather than by voters.

A system that generates close elections may, of course, have benefits as well. Electoral competition, even at the extreme of a narrowly-decided race, may cause citizens' preferences to be better reflected in leaders' actions (Griffin, 2006). And there is an unsettled question over whether, how, and in what contexts close elections drive higher turnout and participation (e.g., Cox and Munger, 1989; Vogl, 2014; Cancela and Geys, 2016; Gerber et al., 2020).

The importance of these consequences motivates the question that our research answers: Were the 1876 election and other close, disputed election outcomes for US president statistical flukes? Or are close elections—and whatever consequences they carry—especially likely in the US's two-tiered Electoral College system, relative to a popular vote? Although there has been significant study among historians, legal scholars, and political scientists of the potential consequences of close elections in the US and abroad, such work hasn't established the empirical probability of these events. Nor

[4] Facts about the 1876 election in the remainder of this paragraph are sourced from Haworth (1906) and Holt (2009).

[5] If neither candidate had a majority, the election would be decided by a new vote in the House of Representatives, nullifying all votes cast by citizens in the election. In Oregon, where the vote went for Hayes (a Republican), the state leadership was Democratic and officials attempted to send a Democratic elector to the Electoral College in defiance of the vote outcome. In Louisiana and Florida, the outcome was contested, and each state sent two opposed slates of electors. The election result remained disputed in the weeks leading up to the scheduled inauguration in March of 1877.

[6] Under this compromise, the 1876 presidency would go to the northern Republican Hayes in exchange for the end of Reconstruction and ultimately the freedom for southern Democrats to institute Jim Crow.

has any prior work settled whether the Electoral College system is especially statistically prone to close outcomes, relative to feasible alternative systems. For the last few presidential elections, predictions of the risk of a close election have been byproducts of election forecasting models, which we incorporate here, though no prior work has offered the comprehensive characterization that our research contributes—including how these probabilities have evolved over US history, the stability of these findings to alternative modeling choices, or any counterfactual comparisons to a plausible alternative electoral system.

In this paper we address these questions, characterizing the probability of close outcomes under the US Electoral College (EC) and under the plausible alternative—a National Popular Vote (NPV). We begin by developing intuitions about why the EC is particularly likely to be decided by a small number of first-tier (citizen) votes. To assess the empirical probabilities, we make a novel application of presidential election models from several sources, primarily focusing on three: Gelman, Heidemanns and Morris (2020), Geruso, Spears and Talesara (2022), and Silver (2020). We take the election models—these and others—as basic data in our empirical exercise. Using these sources, we establish that the EC is very likely to generate a close outcome.[7] For example, the probability that an EC winner today is decided by 7,500 ballots or fewer (about 0.005% of votes cast) in a single pivotal state is greater than 1-in-40. The probability that the EC winner is decided by 75,000 ballots or fewer in a single state is about 1-in-6. This probability rises to 1-in-4 if we allow for the 75,000 votes to be spread across two or three states that together would be pivotal in reversing the outcome.

These results characterize the likelihood of entering scenarios in which a few state officials may be capable of reversing a national election result. Even an election decided by the extremely narrow margin of 150 votes in a single state has a non-negligible chance of occurring: a little less than 0.1%. In other words, the probability that a one-one millionth share of turnout is pivotal in deciding the election is about a one-in-one thousand event. These findings indicate that the recent electoral experience of close presidential outcomes in 2000, 2016, and 2020 (decided by <0.001% of national turnout in one state, 0.060% in three states, and 0.028% in three states, respectively) were not statistical flukes. They were ex-ante likely. We further show that this has not been a transitory, happenstantial

[7] Methodologically, our work is most closely related to studies in positive political science, economics, and mathematics that have estimated the probability that a voter is pivotal in determining an election (including Shapley and Shubik, 1954; Banzhaf III, 1964; Chamberlain and Rothschild, 1981; Gelman, King and Boscardin, 1998; Gelman, Katz and Tuerlinckx, 2002; Gelman, Katz and Bafumi, 2004; Gelman, Silver and Edlin, 2012; Merlin and Nagel, 2021), as well as the response of parties and campaign investment to these probabilities (Strömberg, 2008; Wright, 2009).

phenomenon arising from present politics and demographics. It has been true over the entire history of presidential voting in the US: under different state composition and territorial borders (25 states plus the voting Territory of Michigan at the beginning of our study period), different dominant parties (Whigs, Democrats, Republicans), and different groups of enfranchised voters (black men, women, the poor).

We conclude with a discussion of whether the Electoral College *causes* this closeness. A plausible alternative explanation is that closeness is the product of strong positional competition between two parties. Under such a view, close outcomes could be frequent under any feasible electoral system, including a National Popular Vote. Contrary to this view, we show that in order to match the EC's tendency to yield an election is decided by a few thousand votes, voting under an NPV system would have to generate a distribution of voting outcomes that is implausibly tight. Popular voting would not generate such narrowly-decided election outcomes.

## 2 Illustrative model

Before assessing the empirical facts of interest, in this section we present a toy model intended to provide two key insights. The first insight rebuts a common intuition about the EC—that an Electoral College (EC) system is more likely than a National Popular Vote (NPV) to generate close election outcomes merely because the EC's two-tier aggregation mechanism affords more "bites at the apple" for a close outcome. This popular understanding holds that because the EC is composed of many winner-takes-all, state-level contests at the second tier—each of which is an opportunity for a narrow-margin outcome in a potentially pivotal state—a small number of citizen votes will tend to decide the Electoral College (see, e.g., Koza et al., 2013, and Hirsch, 2020).[8] Though we show in later empirical results that it is indeed correct that the EC is more likely to generate narrow outcomes, this section's model shows that this conjecture, assumption, or intuition about *why* the EC generates hairsbreadth elections is incorrect.

The second insight from the model is that the partisan distribution of citizen votes across states is a key factor determining how likely a close election is under an EC system. If party competition tends toward 50-50 alignment in the pivotal voting unit, then the EC will tend to deliver close election

[8]From the National Popular Vote Interstate Compact group (Koza et al., 2013): "The current state-by-state winner-take-all system repeatedly creates artificial crises because every presidential election generates 51 separate opportunities for a dispute because of an outcome-altering statewide margin."

outcomes, relative to what an NPV system would produce.

For the purposes of the stylized model, assume that there are three equal-sized states ($s \in \{1,2,3\}$) with equal EC representation (here, one EC vote each, though this does not matter). Denote the realized Democratic vote share in each state as $v_s \in [0,1]$ and the expected Democratic vote share—its partisan lean—as $\mu_s$. States follow the "unit rule," so that a simple statewide vote plurality gains all electoral votes in the state. Numbering states according to increasing $\mu_s$ and assuming that the probability that state 2 is not pivotal in the EC is small enough to ignore, a candidate wins in an EC system if and only if they win the tipping point state—that is, if and only if their realized vote share in state 2 is greater than 0.5. A candidate wins in an NPV system if and only if their national vote share ($\bar{v} \equiv \frac{1}{3}\sum_s v_s$) is greater than 0.5.

For simplicity, we are not assuming here that vote shares in each state are endogenous to the voting system. The intention of the model is to understand the *statistical mechanics* of the EC system in a way that can accommodate any endogenous process that would produce these assumed vote shares. Rather than layering endogenous voting behavior onto our toy model of a few symmetric states, the question of how voting occurs in an equilibrium in which parties and voters respond to the electoral systems in which they are embedded is deferred to the empirical investigation below (beginning in Section 3 for the EC system and Section 5 for NPV systems).

To examine probability distributions over outcomes, let the vote shares $v_s$ be normally distributed for each state. Assume identical variance in each state ($\sigma_s^2 = \sigma^2$) and that realizations $v_s$ are independent.[9] These assumptions help to focus the model and enable an analytical solution; we do not retain them in the empirical exercise that follows. Because the sum of independent normal random variables is normal, the national vote (expressed as the Democratic share) is normal with an expectation of $\frac{1}{3}\sum_s \mu_s$ and a variance of $\frac{1}{3}\sigma^2$.[10]

Let each state contain $n$ voters. Then, under the EC system, the vote margin that is pivotal in determining the outcome is merely the margin, $m$, in the pivotal state (state 2):

$$m^{EC} = n\,|0.5 - v_2| \text{ where } v_2 \sim \mathcal{N}\left(\mu_2, \sigma^2\right). \tag{EC}$$

[9]The normal parameterization allows small tail probabilities to fall outside of [0,1]. See Appendix Section A for an alternative model formulation without this feature. This alternative also relaxes the assumption that state 2 is pivotal.

[10]The variance of the national vote, expressed in terms of vote shares (not vote counts), is the variance of the across-state mean share, $\text{Var}\left(\frac{1}{3}(v_1 + v_2 + v_3)\right)$.

Under the NPV system, the margin that is pivotal in determining the outcome is the national margin:

$$m^{NPV} = 3n\,|0.5 - \bar{v}| \text{ where } \bar{v} \sim \mathcal{N}\left(\frac{1}{3}\sum_s \mu_s, \frac{1}{3}\sigma^2\right). \tag{NPV}$$

Taking the expectations of these two expressions shows that which voting system is more likely to deliver close election outcomes depends on how far $\mu_2$ is from 0.5 relative to $\sigma^2$ and how far $\frac{1}{3}\sum_s \mu_s$ is from 0.5 relative to $\frac{1}{3}\sigma^2$.[11] Therefore, if strong party competition endogenously generates convergence towards 50-50 alignment in the pivotal voting unit, or if this statistical convergence happens for any other reason, then the EC will be closer in expectation. In other words, adding variance across the non-pivotal states (or adding more such states) would increase the variance and expected margin of the national popular vote without changing the expected EC margin.

Table 1 illustrates this numerically. The example assumes expected vote shares $\mu_s$ as indicated in the table. The rows consider alternative scenarios, described by electoral maps: A, B, C and D. To ground the example, the parameters in map A were chosen to match the Democratic vote share in the 2020 presidential election in Kansas ($\mu_1$) and Colorado ($\mu_3$), with Pennsylvania ($\mu_2$) serving as the swing state. In map B, these were chosen from the 2020 vote shares to match South Dakota ($\mu_1$) and New York ($\mu_3$), with Minnesota ($\mu_2$) as the swing state. All maps contain the same number of voters, split equally across the states, and are constructed to have the same expected Democratic share of the vote nationally. But the location of the Democratic and Republican voters differs: In maps B through D, the pivotal (i.e., "swing" or "tipping point") state is slightly further from 50-50 in expectation.

Table 1 illustrates that which electoral system is expected to deliver closer outcomes depends on the partisan geography of voters. In map A, the expectation of the decisive number of votes in the EC system ($\mathbb{E}\left[m^{EC}\right] = 179$) is smaller than the expectation of the decisive number in the NPV system ($\mathbb{E}\left[m^{NPV}\right] = 374$) because the tipping point state is close to 50-50 in expectation. Thus, in map A, the Electoral College system will produce more narrowly-decided elections. In map B the opposite is true: The EC system is more likely to be decided by a greater vote margin than the NPV system ($\mathbb{E}\left[m^{EC}\right] = 403$ versus $\mathbb{E}\left[m^{NPV}\right] = 374$). Map C—which innovates on B by changing the state expectations of the non-swing states $\mu_1$ and $\mu_3$ to be closer to 50-50—reveals that nothing in the A/B comparison relies on the degree of partisan lean in "safe states."

Finally, map D adds two states and is calibrated to match expected national vote tallies of the

[11] In particular, the expectations $\mathbb{E}\left[m^{EC}\right]$ and $\mathbb{E}\left[m^{NPV}\right]$ can be calculated as the means of folded normals.

other maps. The fact that $\mathbb{E}\left[m^{EC}\right]$ and $\mathbb{E}\left[m^{NPV}\right]$ are identical across B, C, and D makes clear that the result depends on the expected partisan lean in the tipping point state ($\mu_2$) and the $\sigma$ terms, but not on the number of states.

Here, the invariance of $\mathbb{E}\left[m^{EC}\right]$ to the number of states considered is built into the stylized model, which assumes a single pivotal voting unit. In the alternative formulation in Appendix A and in the empirical analysis below, no such constraint is imposed. We now turn to empirical distributions of likely presidential election outcomes, estimated from voting and polling data and fully accounting for the empirical patterns of partisan alignment across states—as well as the significant heterogeneity in state sizes, EC representation per capita, and turnout.

# 3 Data and Methods

## 3.1 Data: External Election Model Output

To examine the probability that election outcomes in US presidential races will be close at some threshold of interest, we take existing probabilistic models of the US elections from several published sources. Two sources are widely-reported-on forecasting models for the 2020 presidential election by *The Economist* (Gelman, Heidemanns and Morris, 2020) and *FiveThirtyEight* (Silver, 2020), which generated probability distributions over the 2020 election outcome, estimated largely from polling data. We use these models as they were produced in late October 2020 just before the 2020 presidential election.

The third data source is Geruso, Spears and Talesara (2022), which is built from historical voting data, rather than polls, and back-casts distributions of potential outcomes in past presidential elections. Geruso, Spears and Talesara (2022) (GST for convenience) includes models describing presidential elections in various discrete periods up to 2016 and extending back to 1836.[12] For example, the Antebellum and post-Reconstruction election models in GST cover the periods 1836–1852 and 1872–1888, respectively, and are intended to describe potential outcomes in a typical election over those periods of relatively stable partisan geographies for a generic pair of Democratic and Republican (or Whig) candidates competing.

The three data sources each generate many thousands of simulated election outcomes as draws

[12] In earlier periods, electors were most commonly decided by state legislatures, rather than citizen voters.

from their estimated statistical processes. Because the EC is a complex statistical object, parameter estimates (such as the conditional covariance of two states' shock process net of regional- and demographics-based covariance) are not in all cases commensurable across the three models, which have different structures and therefore estimate different parameters. The sets of simulated election draws, in contrast, are directly comparable across models and describe each model's implied distribution over election possibilities. For simplicity, we often refer to the simulation output that characterizes these probability distributions as the *model*.

Each simulation draw represents an observation in our analysis, consisting of a vector of state-level vote shares for each of the two major-party contenders. These state-level outcomes for the presidential vote, combined with data on apportioned electors and turnout by state, completely describes an election possibility in terms of the allocation of a state's electoral votes, the Electoral College outcome, and the popular vote.[13,14] Because the election outcome in 2020 revealed significant polling bias that was reflected in the forecasting models, we subtract a 2 percentage point ex-post bias favoring Democrats from every state-level simulated outcome in the 2020 *Economist* and *FiveThirtyEight* forecasts, but also report results using non-demeaned model output.[15]

Figure 1 describes the data. The *Economist* and *FiveThirtyEight* models each consist of 40,000 simulation draws. GST consists of 100,000 draws per time-period-specific model. Only the most recent period model from GST (1988–2016) is included in Figure 1. Panel (A) plots the probability distributions over the popular vote across the three models. Panel (B) plots the implied probability distributions over the EC outcome. The distributions in both panels are calculated simply as the frequency of discrete events in each set of simulation draws.

The Gelman, Heidemanns and Morris (2020), Geruso, Spears and Talesara (2022), and Silver (2020) models disagree on many particulars—central tendencies and higher-order moments are significantly different. Indeed, the three models are not even attempting to represent the same phenomena. The forecasts—Gelman, Heidemanns and Morris (2020) and Silver (2020)—are predicting a particular event

[13]States are apportioned electoral votes equal to the size of the congressional delegation—US senators plus US representatives. See Appendix B for additional institutional detail.

[14]To facilitate comparability between the three datasets, we treat the 5 Electoral College ballots at stake at the congressional district level in Maine and Nebraska as if these were determined at large.

[15]So a simulated Democratic vote share of 0.56 in a state becomes 0.54. The 2 percentage point vote share adjustment is based on comparing the realization of vote shares nationally and in battleground states to the model predictions. See also Gelman and Morris "Comparing election outcomes to our forecast and to the previous election" (https://statmodeling.stat.columbia.edu/2020/11/06/comparing-election-outcomes-to-our-forecast-and-to-the-previous-election/) for the assessment that the average state prediction in the *Economist* model was off by about 2.5 percentage points. See Figures A1 and A4 for non-demeaned versions of the data and main results.

one week before the outcome was resolved. These were estimated after the candidates were selected via the party primaries and after a global pandemic and a sharp recession occurred, among other facts. Differences between these two forecasting models, although the subject of much attention and debate when they were made, are unimportant for our purposes and are small in comparison to differences between either forecasting model and the non-forecasting model by Geruso, Spears and Talesara (2022). GST is methodologically and conceptually different. It estimates probability distributions that describe likely outcomes for a *generic* Democratic and Republican presidential candidate pair over a specified window (in Figure 1, 1988–2016). The GST back-casting models thus are intended to build-in higher-order uncertainty. This is clearly reflected in the dispersion differences in Figure 1.

Unsurprisingly, for many outcomes of interest, these modeling differences yield very different predictions. Yet, as we show below, our main finding regarding the probability of a close outcome in the Electoral College relative to the probability of a close outcome in a single-tier popular vote is not very sensitive to model choice. The types of fine differences important for forecasters—e.g., does some candidate have a 40%, 50%, or 60% chance of winning?—are unimportant here. That is because our findings are concerned with orders of magnitude: We show that voting in the EC system is about 100 times more likely to result in a race decided by a few hundred or few thousand votes than voting in a single-tier system. If diverse plausible models all entail the result that the EC is especially likely to deliver close outcomes, then there need not be agreement on the one "right" model to make progress on understanding this phenomenon.

### 3.2 Identifying narrow outcomes

We compute the probability of a *close* election by observing how frequently, within the set of simulation draws of each model, discarding a small number of votes from the simulated winner's tally could overturn the election result. In particular, for each simulated election draw $q^j \in Q^j$ for election model $j$, we find the minimum number of discarded votes needed to change the outcome of the election.

Our focus on counts of votes contrasts with Merlin and Nagel (2021), who also examine close outcomes in the Electoral College but focus on statutory thresholds for *state recount* rules, which are defined in terms of a fraction of the state vote total.[16] The view of vulnerability in our paper is not particularly tied to recounts, and, for example, would include the possibility of administrative

[16]The results in Merlin and Nagel (2021) are in the same direction (that the EC is prone to close outcomes) but are far more modest in magnitude.

judgements of election officials to discard batches of mail-in ballots, which don't have the same statutory link to margins-as percents. It is likely that both the margin-as-percent and margin-as-votes, which will be imperfectly correlated across states, matter for the disputability of elections.

For our vote-count measure, consider discarding some number of votes, $n$, from the winning candidate in a single state. The presidential election outcome could be altered by discarding those votes only if it were the case that $n$ was greater than the difference in vote totals received by the state-level winner and loser and only if that state's electoral votes at the second tier were individually pivotal in deciding the election. In our baseline results, we focus on cases in which the electoral outcome hinges on a small number of votes in a *single* state. Historical examples include 537 votes in Florida in 2000 and 889 votes in South Carolina in 1876. In other results, we allow the possibility that an election could hinge on discarding a small number of votes split across two or three carefully chosen states—like 2020, in which 43,000 votes split across AZ, GA, and WI decided the outcome.

To allow comparisons across time periods in which vastly different numbers of votes were cast—1.5 million in 1836 versus more than 150 million in 2020—we report closeness thresholds that are normalized by the national turnout. When we characterize the probability that an election is resolved by some margin $m$ or less, we define $m$ as the number of votes dropped from the winner's tally divided by the total number of votes cast. We report results for various values of $m$ from 0.00001% of turnout to 0.1% of turnout. For context, the 2000, 2016, and 2020 races—the closest in recent history—were decided by 0.0005%, 0.060%, and 0.028% of the national two-party vote, respectively.

# 4 Results

Figure 2 presents our main result: the probability of a disputably narrow outcome, in which a small number of votes would change the winner of the presidency in the EC system. Separately for each data source, we plot probabilities that the election is decided by the indicated vote margin (indexed along the horizontal axis) in a single pivotal state. These first results focus on contemporary politics, using the 2020 forecasting models and the GST model covering the 1988–2016 period.

Figure 2 and the corresponding tabulations in Table 2 show that, in the EC system, the probability an election would be reversed by a small share of votes is large and is not different across the three data sources. Scaling to 2020 turnout, the probability of an election decided by 150,000 votes or fewer (about 0.1 percent of turnout) is at least 18% in all models. At a threshold of 7,500 votes, the corresponding probability is about 3% in all models. Even the small margin of 150 votes has a non-negligible chance of occurring: 0.073% in the GST model (the only model precise enough to estimate that statistic).[17] In other words, the probability that less than a one-one millionth share of turnout is pivotal in deciding the election is nearly a one-in-one thousand event.[18]

These results are consistent with US voting history. For example, since 1836 there have been four elections—1876, 1884, 1916, and 2000—that were decided by less than 0.02% of the national two-party vote in a single pivotal state. These elections were decided by 889, 1,149, 3,773, and 537 votes, respectively. There have been 47 races since 1836—or 46 if leaving out the 1864 election during the Civil War—so these four account for 9% of presidential elections. Even though the models in Table 2 weren't calibrated on any data prior to 1988, applying the results in Panel A to this long history indicates that we should expect such outcomes in 8, 9 or 10 percent of elections, depending on which of the three models one considers. Thus, according to these models, these close historical election outcomes were not flukes.

For reference, we also plot in Figure 2 the closeness of the national popular vote as it occurs in the same election simulation data. The figure shows that, for any threshold of interest, the popular vote is less likely to be close than the EC outcome by nearly two orders of magnitude. Despite close agreement on the probability of a close outcome in the Electoral College, the three models are dispersed in their predictions of the popular vote, highlighting the EC's complex mapping from citizen votes to the electoral votes that determine the presidency. The popular vote statistics are, of course, not a projection of what *would* happen under a National Popular Vote system. Instead, they describe likely outcomes for the popular vote as it is generated under the EC system.

We also calculate in Table 2 (and plot in Figure A2) the probability that an $m$ share of the national votes is pivotal in determining the outcome, if that share is split across two or three carefully chosen states. Except at the smallest margins of a few hundred or few thousand votes—where the probabilities are essentially identical to the single-state case—allowing for the pivotal votes to be split across two or

[17] We cannot produce estimates at similarly low thresholds for the Gelman et al. and Silver models in part because fewer simulation draws are provided for those models (40,000 each), but primarily because the precision of their published data are truncated at 4 digits: A reported Republican share of 0.5000 for state $s$ in simulation $q$ does not distinguish between a state race won by two votes or two hundred votes.

[18] The close correspondence in Figure 2 across the several models of the EC we examine may give an impression of more precision in these results than is warranted. As we show in Appendix Figure A1, it is possible to construct election models where the risk at each threshold $m$ is very high, but somewhat lower than in Figure 2. There we repeat the exercise of Figure 2 for non-demeaned versions of the Gelman, Heidemanns and Morris (2020) and Silver (2020) models and for one of the earliest presidential simulation models for which simulation draws were obtainable (Gelman, Silver and Edlin, 2012).

three states significantly raises the odds of a close election. For example, for a margin less than 0.1% (less than 150,000 votes by 2020 turnout), the probability rises from at least 18% when considering a single pivotal state to at least 35% when allowing three states to be jointly pivotal. Table 2 makes clear that the 1960, 1976, 2016 and 2020 races, each decided by a small number of votes split across two or three states, were not statistical flukes but were likely ex ante. Column 4, which displays result for a model trained on data that stops at 2016, predicts in Panel C that observing a margin in 2020 that was as small as the realized 2020 margin was a 1-in-6 event, ex ante.

Is this risk conveyed by these results large? Over a voting lifetime (15 presidential elections in 60 years), the results imply greater than a 50% chance that a voter will experience a presidential election decided by 0.01% of turnout (today about 15,000 votes or fewer) in a single pivotal state. Of course, the ex post probabilities of experiencing close election outcomes is 100% for voters who experienced the 2000 or 2020 elections, but it is important to establish whether such outcomes were flukes or the results we should routinely expect of the EC system. Figure 2 and Table 2 reveal that a high risk of close elections (and their consequences) is built into the US electoral system.

In Figure A3 and Table A1, we repeat the analysis of Figure 2 for earlier time periods, adding in GST's election models for the Antebellum (1836–1852), Post-Reconstruction (1872–1888), and twentieth century and later periods (1916–1932, 1936–1956, 1964–2016). That analysis, which mirrors the patterns in Figure 2, establishes that the tendency of the EC to produce narrow elections is an enduring empirical feature, true for every political geography that the US has to date produced. The ex ante risk of a close, disputable election has been high for as long as citizens have participated in US presidential elections. It has persisted in periods characterized by radically different geographic patterns of partisan politics—including the addition of new states and voting territories, new enfranchised voters (non-whites, the poor, women), and the birth and death of US political parties. Importantly, the fact that the expected closeness of EC outcomes changes only slightly *and non-monotonically* with the count of states, beginning with 25 in 1836 and ending with 50 plus DC today, further refutes the common "bites of the apple" intuition discussed in Section 2.

## 5 Is the EC Uniquely Vulnerable to Close Elections?

Does the EC *cause* its razor-thin margins, or would US presidential politics generate very narrow wins and losses even under a National Popular Vote? The empirical models considered above do not necessarily answer this. That is because they implicitly incorporate the equilibrium party and voter behavior that the EC produces. In this section, we characterize the counterfactual properties of a popular vote election for US president that would be necessary for it to match the EC in terms of close election risk. We then ask whether such properties are plausible.

We begin with a simple statistical model of potential vote outcomes under a counterfactual NPV system for US president. We force the process be conservative—i.e., more likely to produce a close outcome than the true process would—in two ways. First, we assume it follows a normal distribution, rather than the fat-tailed distributions sometimes preferred by election forecasters (e.g., Silver, 2016). Second, we set the expectation at a 50-50 tie. Both assumptions make a closely-decided NPV race more likely. So is it likely that, even under such assumptions, an NPV would produce the same kind of hairsbreadth election outcomes that the EC does?

Expressing vote shares in percentage points, the distribution of votes in the NPV system is $F \sim N(0, \sigma^2)$, where the mean has been normalized to zero by subtracting 50 points. Given this distribution, it is straightforward to calculate the standard deviation $\sigma$ that generates the same probability as does the EC of arriving at a voting outcome within some margin of interest $m$. Formally, we find the $\sigma$ that solves $\int_{m/2\sigma}^{m/2\sigma} \phi(x)dx = Pr.^{EC}(m)$, where $\phi$ is the standard normal density and $Pr.^{EC}(m)$ is the probability of a within-margin-$m$ result in the EC system, as estimated and displayed in Figure 2 and Table 2.[19]

For example, we found above that $Pr.^{EC}(.01\text{pp})$ is about 5% (Table 2, Panel A). That is, the EC generates about a 1-in-20 chance of an outcome decided by 15,000 votes or fewer in a single state (scaling margins by 2020 turnout). The calibrated NPV distribution that would generate a matching probability of such a narrow outcome has a $\sigma$ equal to 0.076 percent of turnout.[20] To emphasize: the calibrated $\sigma$ is equal to a little over 100,000 votes; we do not mean 7.6 percent.[21]

This implied spread of potential election results under the hypothetical NPV system is very small. To put the number in context, Figure 3 plots this popular vote counterfactual ($F \sim N(0, 0.076^2)$) against estimates of the empirical popular vote distributions for US president today (Panel A) and against the typical spread of votes in other popular-vote elections in the US and abroad (Panel B).

[19] In the integral limits, the margin $m$ is expressed by the range $-m/2$ to $+m/2$. A within-0.01% margin is generated by Republican vote shares in the range 49.995 to 50.005 points.

[20] In particular, we match the estimate for a .01 percentage point margin in the GST column in Panel A of Table 2, which is 5.3%. Of the three models, GST generates the most conservative (lowest) estimates of close-election risk in the EC.

[21] Matching other EC statistics, such as the probability of a race decided by less than .0001%, or .001% of turnout, generates similarly small calibration values for $\sigma$. We tabulate these in Appendix Table A2.

[22] The typical variance of vote margins in US Senate races, US gubernatorial races, and single-tier presidential elections in other OECD countries is large relative to a $\sigma$ of 0.076 points.

The figure illustrates that the distribution of vote possibilities in an NPV system would have to be exceedingly narrow—far outside of historical experience with popular vote elections—in order to match the closeness of the EC system. If the popular vote counterfactual shown in Figure 3 were a correct statistical model summarizing campaign and voter behavior under a National Popular Vote for US president, then experiencing a presidential race in which the winner gained more than 50.5% of the two-party popular vote would be an extreme statistical fluke—about seven standard deviations above the mean, occurring with probability $< 10^{-10}$. In contrast to this prediction, popular-vote elections for US senators, US governors, and OECD presidents routinely have (two-party) victory margins larger than 50.5% to 49.5%. This suggests the NPV counterfactual cannot be a correct statistical model and that a National Popular Vote would be much less likely to yield close election outcomes than the present EC system.

Although we cannot rule out the abstract possibility that an NPV system for the US presidency would always generate such tight outcomes, we can conclude that it would be inconsistent with all historical experience of popular voting in the US and elsewhere—and on that basis deeply implausible. National popular voting, under plausible assumptions, simply cannot match the EC's actual, historical tendency toward narrow outcomes. In this sense, the EC *causes* close election outcomes, relative to the feasible alternative system.

## 6 Conclusion

Our work informs a perennial policy debate over the desirability of maintaining the EC system—though without settling any question. Proponents of a National Popular Vote and other reforms might note that these findings about the likelihood of narrow, disputable elections add to a list of undesirable EC features that include inversions ("wrong winners"), the possibility of a contingent election in the US House, and the relegation of most states to spectator status. Proponents of the status quo or more modest reforms might note that, under an NPV, actors in all states, not merely swing states, could have incentives to engage in vote fraud and manipulation. And of course, election disputes can occur for other structural or happenstantial reasons unrelated to vote margins, such as the choice of voting

[22] Appendix C describes the popular vote elections from which we calculate the standard deviations used here.

technologies and procedures, which could be improved without constitutional change (Alvarez et al., 2012).

Nonetheless, we hope that our findings on the statistical tendency of the EC to deliver close election results may serve to clarify thinking on one source of perceived illegitimacy and litigation risk following a presidential race. Writing about the 2000 presidential election, Lempert (2016) suggested that the tractability of quarantining recounts to specific states is an advantage of the Electoral College: "Whoever won, Bush or Gore, it was going to be by a hairsbreadth. Because of the Electoral College, we did not have to recount the whole nation. Instead we could focus on a more manageable task—recounting the state of Florida." This argument is built on an empirically mistaken assumption. The error here is that although a hypothetical national recount might have been difficult, it would not have been needed and would not have occurred. In statistical expectation, only the Electoral College generates hairsbreadth outcomes.

## References


**Alvarez, R. Michael, Stephen Ansolabehere, Thad E. Hall, Jonathan N. Katz, Ronald L. Rivest, and Charles Stewart.** 2012. "Voting: What has changed, what hasn't, & what needs improvement."

**Banducci, Susan A, and Jeffrey A Karp.** 2003. "How elections change the way citizens view the political system: Campaigns, media effects and electoral outcomes in comparative perspective." *British Journal of Political Science*, 33(3): 443–467.

**Banzhaf III, John F.** 1964. "Weighted voting doesn't work: A mathematical analysis." *Rutgers L. Rev.*, 19: 317.

**Birch, Sarah.** 2008. "Electoral institutions and popular confidence in electoral processes: A cross-national analysis." *Electoral Studies*, 27(2): 305–320.

**Cancela, João, and Benny Geys.** 2016. "Explaining voter turnout: A meta-analysis of national and subnational elections." *Electoral Studies*, 42: 264–275.

**Chamberlain, Gary, and Michael Rothschild.** 1981. "A note on the probability of casting a decisive vote." *Journal of Economic Theory*, 25(1): 152–162.

**CNN.** 2021. "CNN Presidential Election Poll." *SSRS Research*.

**Cox, Gary W, and Michael C Munger.** 1989. "Closeness, expenditures, and turnout in the 1982 US House elections." *American Political Science Review*, 83(1): 217–231.

**Gallup.** 2001. "Seven out of 10 Americans Accept Bush as Legitimate President."

**Gelman, Andrew, Gary King, and W John Boscardin.** 1998. "Estimating the probability of events that have never occurred: when is your vote decisive?" *Journal of the American Statistical Association*, 93(441): 1–9.

**Gelman, Andrew, Jonathan N Katz, and Francis Tuerlinckx.** 2002. "The mathematics and statistics of voting power." *Statistical Science*, 420–435.

**Gelman, Andrew, Jonathan N Katz, and Joseph Bafumi.** 2004. "Standard voting power indexes do not work: an empirical analysis." *British Journal of Political Science*, 34(4): 657–674.

**Gelman, Andrew, Merlin Heidemanns, and G. Elliott Morris.** 2020. "State and national presidential election forecasting model." https://github.com/TheEconomist/us-potus-model.

**Gelman, Andrew, Nate Silver, and Aaron Edlin.** 2012. "What is the probability your vote will make a difference?" *Economic Inquiry*, 50(2): 321–326.

**Gerber, Alan, Mitchell Hoffman, John Morgan, and Collin Raymond.** 2020. "One in a million: Field experiments on perceived closeness of the election and voter turnout." *American Economic Journal: Applied Economics*, 12(3): 287–325.

**Geruso, Michael, Dean Spears, and Ishaana Talesara.** 2022. "Inversions in US Presidential Elections: 1836?2016." *American Economic Journal: Applied Economics*, 14(1): 327–57.

**Griffin, John D.** 2006. "Electoral Competition and Democratic Responsiveness: A Defense of the Marginality Hypothesis." *The Journal of Politics*, 68(4): 911–921.

**Grofman, Bernard, and Scott L Feld.** 2005. "Thinking about the political impacts of the Electoral College." *Public Choice*, 123(1-2): 1–18.

**Hasen, Richard L.** 2005. "Beyond the margin of litigation: reforming US election administration to avoid electoral meltdown." *Wash. & Lee L. Rev.*, 62: 937.

**Hasen, Richard L.** 2020. *Election Meltdown.* Yale University Press.

**Haworth, Paul Leland.** 1906. *The Hayes-Tilden Disputed Presidential Election of 1876.* Burrows Brothers.

**Hirsch, Alan.** 2020. *A Short History of Presidential Election Crises:(and how to Prevent the Next One).* City Lights Books.

**Holt, Michael.** 2009. *By One Vote: The Disputed Presidential Election of 1876.* Taylor & Francis.

**Koza, John R, Barry Fadem, Mark Grueskin, Michael S Mandell, Robert Richie, Joseph F Zimmerman, John B Anderson, Birch Bayh, John Buchanan, and Tom Campbell.** 2013. *Every Vote Equal: A State-Based Plan for Electing the President by National Popular Vote.* . 4th ed., National Popular Vote Press.

**Leip, Dave.** 2019. "David Leip's Atlas of U.S. Elections."

**Lempert, Richard.** 2016. "Two cheers for the Electoral College: Reasons not to abolish it." *Brookings*.

**Merlin, Vincent, and Jack Nagel.** 2021. "The U.S. Electoral College and the Probability of Disputable Outcomes under Direct and Indirect Elections."

**Monmouth University Poll.** 2021. "National: More Americans Happy About Trump Loss than Biden Win."

**Posner, Richard A.** 2001. *Breaking the deadlock: The 2000 election, the constitution, and the courts.* Princeton University Press.

**Przeworski, Adam.** 2018. *Why bother with elections?* John Wiley & Sons.

**Rakove, Jack N.** 2004. "Presidential selection: Electoral fallacies." *Political Science Quarterly*, 119(1): 21–37.

**Rapoport, David C, and Leonard Weinberg.** 2000. "Elections and violence." *Terrorism and Political Violence*, 12(3-4): 15–50.

**Sances, Michael W, and Charles Stewart III.** 2015. "Partisanship and confidence in the vote count: Evidence from US national elections since 2000." *Electoral Studies*, 40: 176–188.

**Shapley, Lloyd S, and Martin Shubik.** 1954. "A method for evaluating the distribution of power in a committee system." *American political science review*, 48(3): 787–792.

**Silver, Nate.** 2016. "A User's Guide To FiveThirtyEight's 2016 General Election Forecast." *FiveThirtyEight.com*. https://fivethirtyeight.com/features/a-users-guide-to-fivethirtyeights-2016-general-election-forecast/. Accessed: 2019-6-13.

**Silver, Nate.** 2020. "FiveThirtyEight 2020 General Election Forecast." https://projects.fivethirtyeight.com/2020-election-forecast/.

**Strömberg, David.** 2008. "How the Electoral College influences campaigns and policy: the probability of being Florida." *American Economic Review*, 98(3): 769–807.

**U.S. Election Assistance Commission.** 2017. "The Election Administration and Voting Survey: 2016 Comprehensive Report." Report to the 115th Congress.

**Vogl, Tom S.** 2014. "Race and the politics of close elections." *Journal of Public Economics*, 109: 101–113.

**Weatherford, M Stephen.** 1992. "Measuring political legitimacy." *American Political Science Review*, 86(1): 149–166.

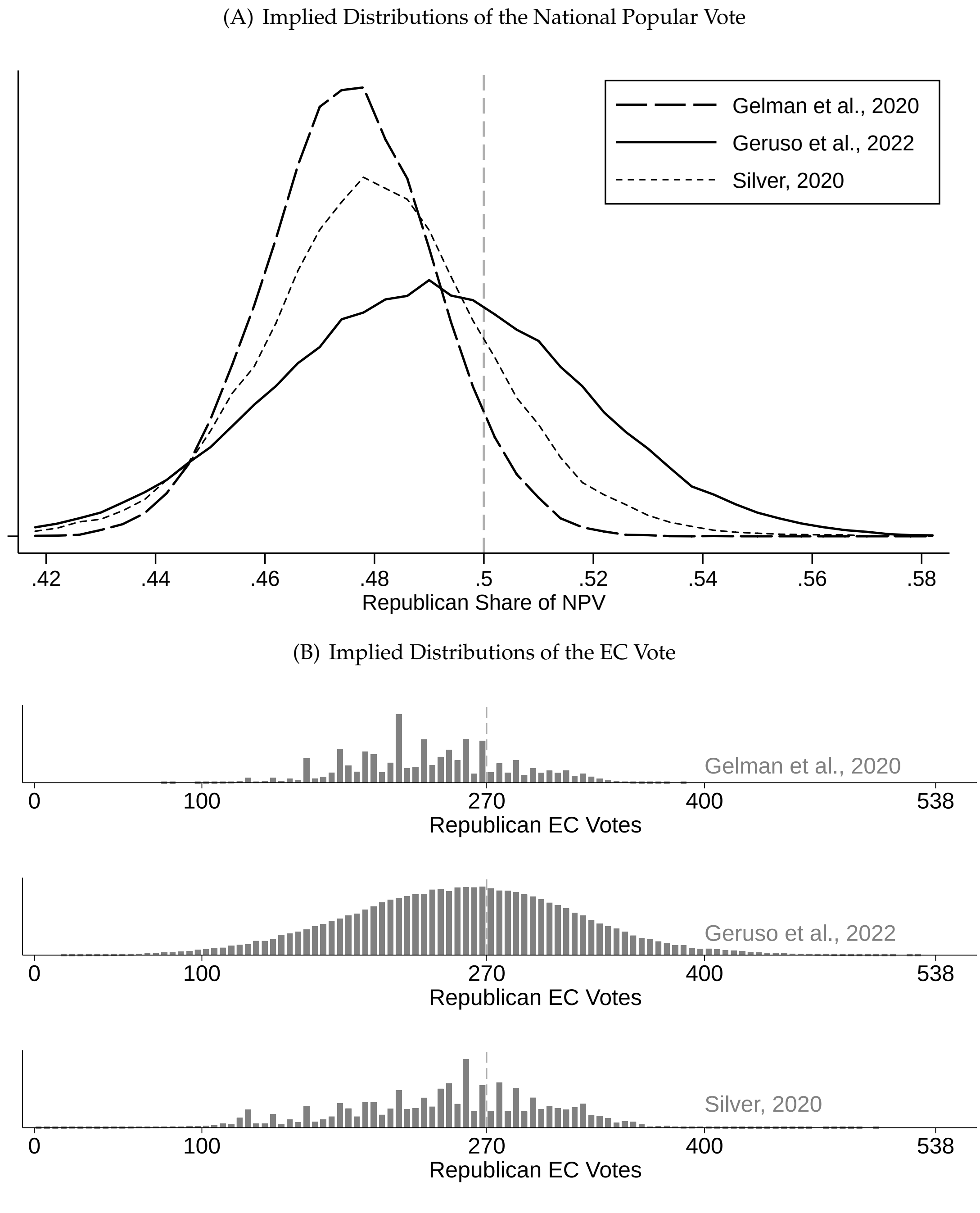
**Wright, John R.** 2009. "Pivotal states in the Electoral College, 1880 to 2004." *Public Choice*, 139(1-2): 21–37.

**Table 1:** Disputed Election Risk in the EC and NPV: An Illustrative Model

| | | | | | | | | expected number of votes needed to reverse outcome if held under: | |
|---|---|---|---|---|---|---|---|---|---|
| | state | | state expectations | | | | | EC system | NPV System |
| map | sizes | $\mu_0$ | $\mu_1$ | $\mu_2$ | $\mu_3$ | $\mu_4$ | NPV, $\mathbb{E}[\bar{v}]$ | $\mathbb{E}\left[m^{EC}\right]$ | $\mathbb{E}\left[m^{NPV}\right]$ |
| A | 10,000 | | 45% | 51% | 57% | | 51% | <u>179</u> | 374 |
| B | 10,000 | | 37% | 54% | 62% | | 51% | 403 | <u>374</u> |
| C | 10,000 | | 42% | 54% | 57% | | 51% | 403 | <u>374</u> |
| D | 6,000 | 37% | 42% | 54% | 60% | 62% | 51% | 403 | <u>374</u> |

**Note:** Table shows how the number of votes that are pivotal in expectation in determining the EC winner depends on the geographic distribution of support. The table considers alternative "maps" of where Democratic and Republican voters reside. All maps hold the Democratic share of the national popular vote, $\mathbb{E}[\bar{v}]$, fixed at 51%. All maps contain 30,000 voters and equally sized states. In A, B, and C, the standard deviation of vote outcomes is 2% ($\sigma_1, \sigma_2, \sigma_3 = 200$ votes). In map D, which adds two states, the $\sigma_s$ terms are calibrated so that the expected margins nationally ($\mathbb{E}\left[m^{NPV}\right]$) and in state 2 ($\mathbb{E}\left[m^{EC}\right]$) match maps B and C. Appendix Section A presents a more complex model that relaxes various assumptions imposed here but produces the same substantive results.

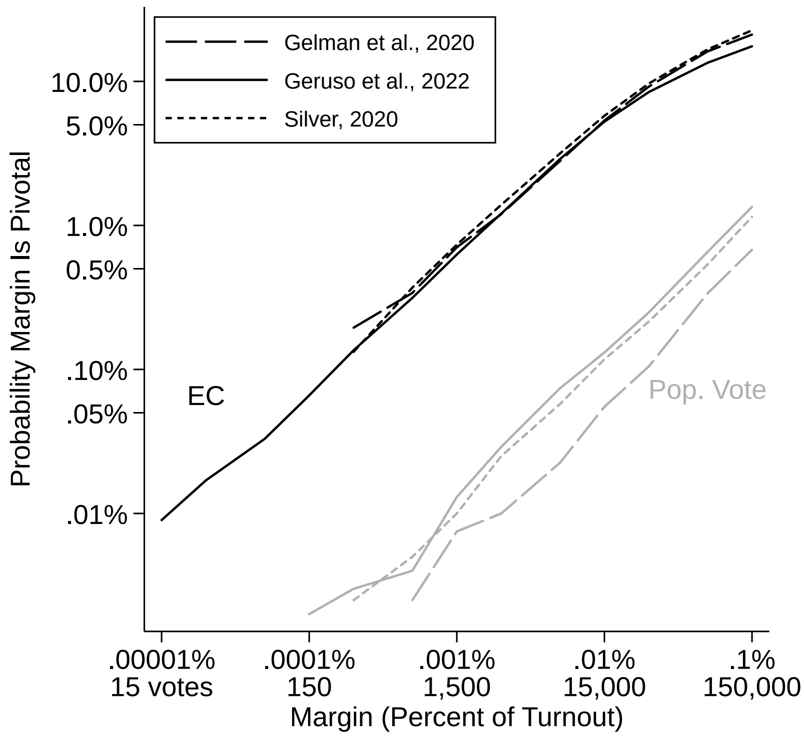
**Figure 1:** Data and Descriptive Statistics: Distributions of Likely Election Outcomes from Several Sources




**Note:** Figure shows the probability distributions over electoral outcomes implied by three datasources, constructed from the 40,000 or 100,000 simulation draws in each datasource. Panel (A) plots the distribution of popular vote outcomes. Panel (B) plots the distribution of electoral vote outcomes.

**Figure 2:** Main Result: Close Election Risk in the Electoral College

(A) Probabilities in Three Models (Single Pivotal State)




**Note:** Figure plots the probability that a national election outcome would be reversible by changing a small number of votes in a single state. The horizontal axis indicates the closeness margin being evaluated, labelled in the top line as a percent of overall turnout, and labelled in the bottom line as vote counts, assuming turnout of 150 million voters (roughly 2020 turnout). The vertical axis indicates the probability that the indicated margin is pivotal. Underlying data are derived from the simulation models of Gelman, Heidemanns and Morris (2020), Geruso, Spears and Talesara (2022) (modern period, 1988–2016), and Silver (2020). For the same set of simulation models, the lighter gray lines indicate the probabilities that the corresponding popular vote generated by each model is within the indicated margin.

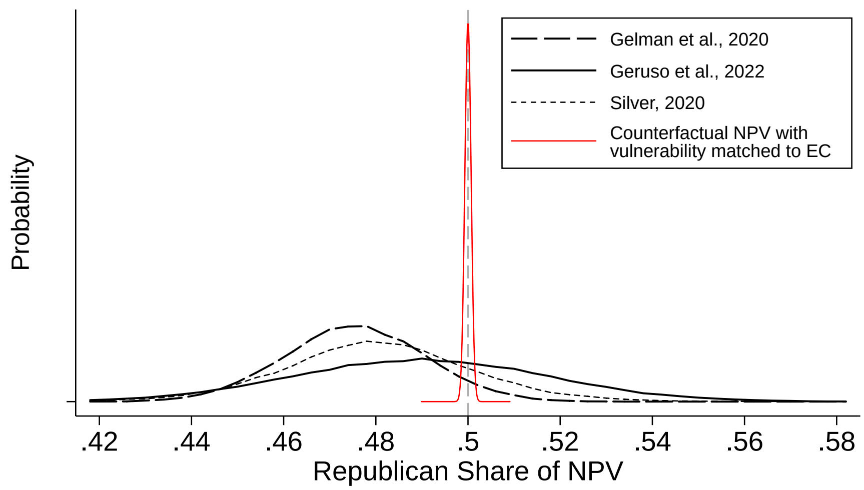
**Table 2:** Main Result: Close Election Risk in the Electoral College

| Margin (Winner's Tally - Loser's Tally) | | Probability | | |
|---|---|---|---|---|
| Margin in Points (Shares*100) | Margin in Votes, Assuming 150M Two-Party Turnout | Gelman et al. (2020) | Geruso et al. (2022) | Silver (2020) |
| (1) | (2) | (3) | (4) | (5) |
| **Panel A: Race Decided within Margin in a Single Pivotal State** | | | | |
| 0.00001 | 15 | - | 0.01% | - |
| 0.00002 | 30 | - | 0.02% | - |
| 0.00005 | 75 | - | 0.03% | - |
| 0.0001 | 150 | - | 0.07% | - |
| 0.0002 | 300 | 0.20% | 0.14% | 0.13% |
| 0.0005 | 750 | 0.34% | 0.31% | 0.37% |
| 0.001 | 1,500 | 0.71% | 0.63% | 0.74% |
| 0.002 | 3,000 | 1% | 1% | 1% |
| 0.005 | 7,500 | 3% | 3% | 3% |
| 0.01 | 15,000 | 5% | 5% | 6% |
| 0.02 | 30,000 | 9% | 8% | 10% |
| 0.05 | 75,000 | 16% | 13% | 17% |
| 0.1 | 150,000 | 21% | 18% | 23% |
| **Panel B: Race Decided within Margin Split Across Two States or Fewer** | | | | |
| 0.00001 | 15 | - | 0.01% | - |
| 0.00002 | 30 | - | 0.02% | - |
| 0.00005 | 75 | - | 0.03% | - |
| 0.0001 | 150 | - | 0.07% | - |
| 0.0002 | 300 | 0.20% | 0.14% | 0.13% |
| 0.0005 | 750 | 0.34% | 0.32% | 0.38% |
| 0.001 | 1,500 | 0.71% | 0.64% | 0.74% |
| 0.002 | 3,000 | 1% | 1% | 1% |
| 0.005 | 7,500 | 3% | 3% | 3% |
| 0.01 | 15,000 | 6% | 6% | 6% |
| 0.02 | 30,000 | 11% | 11% | 12% |
| 0.05 | 75,000 | 22% | 21% | 23% |
| 0.1 | 150,000 | 32% | 29% | 33% |
| **Panel C: Race Decided within Margin Split Across Three States or Fewer** | | | | |
| 0.00001 | 15 | - | 0.01% | - |
| 0.00002 | 30 | - | 0.02% | - |
| 0.00005 | 75 | - | 0.03% | - |
| 0.0001 | 150 | - | 0.07% | - |
| 0.0002 | 300 | 0.20% | 0.14% | 0.13% |
| 0.0005 | 750 | 0.34% | 0.32% | 0.38% |
| 0.001 | 1,500 | 0.71% | 0.64% | 0.74% |
| 0.002 | 3,000 | 1% | 1% | 1% |
| 0.005 | 7,500 | 3% | 3% | 3% |
| 0.01 | 15,000 | 6% | 6% | 7% |
| 0.02 | 30,000 | 11% | 12% | 12% |
| 0.05 | 75,000 | 24% | 24% | 25% |
| 0.1 | 150,000 | 38% | 35% | 39% |

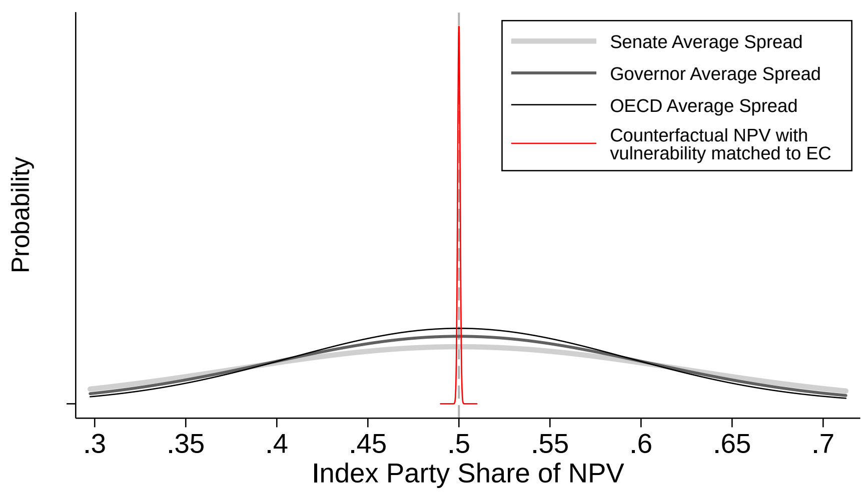
**Note:** Table lists the probabilities that a national election outcome would be reversible by changing a small number of votes. Tabulations correspond to the results in Figures 2 and A2. See corresponding figure notes for additional detail. We cannot produce estimates at very low *closeness* thresholds for the Gelman, Heidemanns and Morris (2020) and Silver (2020) models because the precision of their published data is truncated at 4 digits.

**Figure 3:** A Counterfactual National Popular Vote Distribution that Would Rival EC in *Closeness* Is Implausibly Tight

(A) NPV Counterfactual Compared to the Actual Popular Vote Distribution Generated by the EC System




(B) NPV Counterfactual Compared to the Spreads in Actual Popular Vote Outcomes in Popular Vote Systems




**Note:** Figure compares several empirical distributions of popular vote outcomes with the distribution generated by a hypothetical NPV for US president that would deliver close outcomes with similar expected frequency to that of the present EC system. The comparison popular vote distributions in Panel A are repeated from Panel A of Figure 1. The comparison popular vote distributions in Panel B are typical of vote margin spreads in US Governor and US Senate races, as well as Presidential races in other OECD countries, as described in Appendix C. These popular-vote races are frequently decided by margins of several points and have large standard deviations of vote share outcomes around their means. The height of the counterfactual NPV spike is not drawn to scale with other distributions; the spike height is lower relative to truth in order to preserve some visual detail in the other distributions plotted.

# ONLINE APPENDIX for "The Risk of Narrow, Disputable Results in the U.S. Electoral College: 1836-2020"

by Geruso and Spears

## A Computational robustness check of illustrative model without simplifying assumptions

Section 2 presented an illustrative model of the fact that, in principle, either a two-stage or a one-stage election system could be *closer*, on average and in the sense of this paper, depending on the partisan polarization of states, modeled here as the state-specific expectations for the two-party vote share. In order to present simple analytical expressions, Section 2 made two simplifying assumptions: that the state with the middle mean was always the tipping point state in the second-stage and that vote shares within states were normally distributed, which permitted the possibility that vote shares could be outside of $[0, 1]$.

Here we present Monte Carlo simulations of a version of the model without those assumptions and reach the same conclusion. We let the log-odds of each state vote share be normally distributed around a mean $\mu_s^\ell$ such that $\mu_s^\ell = \ln\left(\frac{\mu_s}{1-\mu_s}\right)$, where $\mu_s$, for each state and map, is just as in Section 2. Each of these normal distributions has a standard deviation of 0.202 logit points, which would correspond to the 5 percentage points used in the analytical model for even odds.

We compute 1 million Monte Carlo simulations, which include cases where the order of the states changes. Table A3 presents the results, which are quantitatively similar and substantially identical to those in Table 1 for the simpler model.

## B Institutional Detail

The two-tiered EC system is established in Article II, Section 1 of the Constitution, and affected by various constitutional amendments and state laws. In current practice, the first tier involves citizens voting for electors pledged to a candidate. Today, in all states except Maine and Nebraska, the statewide popular vote winner is awarded all of the state's EC representation, though there is no constitutional requirement to award electors in this way (or to involve citizens in presidential elections at all), and the state-specific arrangements meant to bind electors to vote for their pledged candidate are imperfect. At the second tier, electoral votes are tallied.

States are apportioned electoral votes equal to the size of the congressional delegation (US senators plus US representatives). The number of electors are thus linked to the Apportionment Acts of Congress, which determine the number of US House seats. In earlier periods that we study, fewer states and different apportionment law generated different sizes of the US Electoral College. For example, in 1836 when Whigs competed with Democrats across 25 states plus the Territory of Michigan, there were 294 EC ballots, and a winning majority was 148. Today there are 538 electors: 435 corresponding to US House members, 100 corresponding to senators, and 3 for Washington DC. The candidate with a simple majority of 270 votes in the electoral college wins the presidency. If neither candidate secures 270 Electoral College votes, the US House of Representatives choses a candidate to become president, under special rules.

# C Descriptive statistics of single-tier elections in the US and abroad

We use publicly available data to describe the probability of close elections in US Senate races, US gubernatorial races, and single-tier presidential elections in the OECD countries that have them. The uncertainty in a single-tier election can be well summarized in two parameters: the mean and variance of possible vote outcomes. For our purposes, which is to describe the spread of vote shares, we focus only on the variance: We artificially replace the mean vote shares with 0.5 when examining popular-vote elections, in order to bias our computations in favor of finding that these elections are likely to be close. Thus we ask whether the empirical dispersion of likely outcomes in any observed popular-vote election is narrow enough to produce outcomes as close as the EC in expectation, even when we assume these popular vote elections end in a tie in expectation.

Table A4 reports summary statistics that describe two important single-tier electoral contests in the US: governorships and US Senate seats. For these, we use the Leip Election Atlas (Leip, 2019) and include all available data on races since 1960.[23] A useful summary statistic for understanding the likelihood of a close outcome is the standard deviation of the historical vote shares. The table lists, for governor and Senate races in each state (combining Senate seats in a state), the standard deviation of the Republican share of the two-party vote, expressed in percentage points. We do not include in the calculation any races in which a third party won. From Table A4, the standard deviation of voting outcomes for a US governor or US Senate race is typically greater than 10 percentage points.

To more precisely compare the probabilities of close outcomes in US popular vote elections against the EC, we convert the empirical standard deviations reported in Table A4 to probabilities that each state race is decided by various margins (0.1 point, 0.01 point, or 0.001 point) by making two conservative assumptions. First, we demean the vote share outcomes for every elected seat and recenter these at 50-50, so that the expectation of these (mean-shifted) elections is a tie. Ignoring the empirical partisan lean for each seat increases the probability density around a close outcome and allows us to sidestep any objection that the US presidency, to which we will compare, is more likely to center national party politics around a 50-50 Republican-Democrat split than is true of state races.

Second, we fit the observed Republican vote shares to a normal distribution of vote possibilities, rather than the fat-tailed distributions preferred by election forecasters (e.g., Silver, 2016). These assumptions are conservative in the sense that we claim that the EC generates closer outcomes in expectation, and these assumptions overstate the likely closeness of the popular-vote systems they are meant to describe.

Under these assumptions, Table A4 lists the implied probabilities of close outcomes in popular vote state races. The implied probabilities are low. For both the US Senate and US governorships, the upper bound on the probability that a race is decided by 0.1% of the vote or less averages about 0.4% across the states.

In Table A5, we consider the typical closeness of popular-vote presidential races in other OECD countries. Because OECD countries do not all exhibit stable two party systems, rather than reporting the standard deviation of a fixed party's vote share, we report the standard deviation of the winning party's vote share (among the top two parties).[24] Using the winning party's vote share in these calculations, rather than a fixed party, adds another layer of conservatism to our claim that popular votes are less likely to be close.[25]

In Table A5, we report these statistics for OECD nations that elect their chief of state ("president") via popular vote. We include the outcomes of popular, winner-take-all elections that have taken place

[23] Accessed January 7, 2021.

[24] In some cases, a country uses a two-party runoff, which we use as the two-party election.

[25] Calculating the standard deviation in terms of the winner's vote share, rather than over a fixed party (that could win or lose) tends to understate the variance of the potential outcome. Like the other assumptions, this tends to generate an overestimate of the probability of a close race in a popular vote system.

in a current OECD member country from 1988 to 2000. The table also lists the implied probabilities of close outcomes at various thresholds in these popular vote races, using the same parametric assumptions as in Table A4 for US races (50-50 re-centering and fitting a normal distribution with the empirical standard deviation).

Table A5 shows that the presidential elections of OECD countries in the past 30 years have shown spreads of vote outcomes fairly similar to the spreads of outcomes in US Senate and governor elections. Upper bounds on the probability of a race decided by less than 0.1%, 0.01%, and 0.001% of the vote, respectively, are 0.6%, 0.06%, and 0.01%. Comparing to the Electoral College system, the EC is expected to deliver close results at much higher rates: >34%, >5%, and >0.6%, respectively, for the same thresholds (Table 2, Panel C).

**Table A1:** Probabilities of Close Outcomes Over History (Tabulations from Figure A3)

| | 1964-2016 | | 1936-1956 | | 1916-1932 | | 1872-1888 | | 1836-1852 | |
|---|---|---|---|---|---|---|---|---|---|---|
| Margin in Points Shares*100) | Margin in Votes, Assuming 150M Two-Party Turnout | Probability | Margin in Votes, Assuming 60M Two-Party Turnout | Probability | Margin in Votes, Assuming 40M Two-Party Turnout | Probability | Margin in Votes, Assuming 10M Two-Party Turnout | Probability | Margin in Votes, Assuming 3M Two-Party Turnout | Probability |
| 0.00001 | 15 | 0.00% | 6 | 0.00% | 4 | 0.00% | 1 | 0.00% | 0 | 0.00% |
| 0.00002 | 30 | 0.01% | 12 | 0.00% | 8 | 0.00% | 2 | 0.01% | 1 | 0.01% |
| 0.00005 | 75 | 0.02% | 30 | 0.01% | 20 | 0.01% | 5 | 0.03% | 2 | 0.02% |
| 0.0001 | 150 | 0.04% | 60 | 0.02% | 40 | 0.03% | 10 | 0.06% | 3 | 0.04% |
| 0.0002 | 300 | 0.07% | 120 | 0.04% | 80 | 0.04% | 20 | 0.12% | 6 | 0.08% |
| 0.0005 | 750 | 0.17% | 300 | 0.11% | 200 | 0.10% | 50 | 0.27% | 15 | 0.23% |
| 0.001 | 1,500 | 0.34% | 600 | 0.21% | 400 | 0.18% | 100 | 0.53% | 30 | 0.44% |
| 0.002 | 3,000 | 1% | 1,200 | 0% | 800 | 0% | 200 | 1% | 60 | 1% |
| 0.005 | 7,500 | 1% | 3,000 | 1% | 2,000 | 1% | 500 | 2% | 150 | 2% |
| 0.01 | 15,000 | 3% | 6,000 | 2% | 4,000 | 2% | 1,000 | 4% | 300 | 4% |
| 0.02 | 30,000 | 4% | 12,000 | 2% | 8,000 | 2% | 2,000 | 7% | 600 | 6% |
| 0.05 | 75,000 | 6% | 30,000 | 4% | 20,000 | 4% | 5,000 | 11% | 1,500 | 10% |
| 0.1 | 150,000 | 8% | 60,000 | 5% | 40,000 | 5% | 10,000 | 15% | 3,000 | 14% |

**Note:** Table lists probabilities that election outcomes could be reversible by a small number of votes over various periods in US history. Tabulations correspond to the results in Figure A3. See corresponding figure notes for additional detail.

**Table A2:** Counterfactual National Popular Vote Probability Distributions Calibrated to Match the Expected *Closeness* of the EC System

| Margin in Points (Shares*100) | Margin in Votes, Assuming 150M Two-Party Turnout | Probability of Outcome within Margin | Implied Standard Deviation of Normal NPV with Matching Close-Election Probability (in Points) |
|---|---|---|---|
| 0.00001 | 15 | 0.01% | 0.044 |
| 0.00002 | 30 | 0.02% | 0.047 |
| 0.00005 | 75 | 0.03% | 0.060 |
| 0.0001 | 150 | 0.07% | 0.060 |
| 0.0002 | 300 | 0.14% | 0.059 |
| 0.0005 | 750 | 0.31% | 0.064 |
| 0.001 | 1,500 | 0.63% | 0.064 |
| 0.002 | 3,000 | 1.2% | 0.066 |
| 0.005 | 7,500 | 2.9% | 0.069 |
| 0.01 | 15,000 | 5.3% | 0.076 |
| 0.02 | 30,000 | 8.4% | 0.094 |
| 0.05 | 75,000 | 13.5% | 0.147 |
| 0.1 | 150,000 | 17.5% | 0.226 |

**Note:** Table lists calibration values for a National Popular Vote probability distribution that would match the expected close-election risk of the EC system. In particular, we calculate the standard deviation $\sigma$ that generates the same probability as does the EC of arriving at some close outcome within margin $m$. Formally, we find the $\sigma$ that solves $\int_{m/2\sigma}^{m/2\sigma} \phi(x)dx = Pr.^{EC}(m)$, where $\phi$ is the standard normal density and $Pr.^{EC}(m)$ is the probability of a within-margin-$m$ result in the EC system. Results are calibrated to the GST model from Table 2, Panel A.

**Table A3:** Disputed Election Risk in the EC and NPV: Monte Carlo Robustness Checks

| map | states' sizes | state mean outcomes $\bar{x}_1$ | $\bar{x}_2$ | $\bar{x}_3$ | NPV, $\mathbb{E}\left[\bar{x}\right]$ | mean number of votes needed to reverse outcome if EC system, $\mathbb{E}\left[m^{EC}\right]$ | mean number of votes needed to reverse outcome if NPV system, $\mathbb{E}\left[m^{NPV}\right]$ |
|---|---|---|---|---|---|---|---|
| A | 1,000 | 35.1% | 53.0% | 64.9% | 51.0% | 46.9 | 69.5 |
| B | 1,000 | 20.2% | 62.9% | 69.8% | 51.0% | 129.0 | 61.5 |

## **Table A4:** EC *Closeness* Compared to US Statewide Single-Tier Elections

| | Republican Share Statistics | | Probability of a Race Decided by Less Than: | | | | Republican Share Statistics | | Probability of a Race Decided by Less Than: | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Empirical Mean in points (ignored) | Empirical Standard Deviation in points | 0.1 points | 0.01 points | 0.001 points | | Empirical Mean in points (ignored) | Empirical Standard Deviation in points | 0.1 points | 0.01 points | 0.001 points |
| **US Senate** | | | | | | **US Governor** | | | | | |
| **Senate Average** | **49.5** | **12.3** | **0.38%** | **0.04%** | **0.00%** | **Governor Average** | **49.6** | **10.4** | **0.42%** | **0.04%** | **0.00%** |
| Alabama | 49.3 | 20.9 | 0.19% | 0.02% | 0.00% | Alabama | 40.6 | 20.7 | 0.19% | 0.02% | 0.00% |
| Alaska | 59.6 | 17.4 | 0.23% | 0.02% | 0.00% | Alaska | 49.6 | 9.0 | 0.44% | 0.04% | 0.00% |
| Arizona | 57.3 | 13.5 | 0.30% | 0.03% | 0.00% | Arizona | 51.6 | 7.4 | 0.54% | 0.05% | 0.01% |
| Arkansas | 45.5 | 25.7 | 0.16% | 0.02% | 0.00% | Arkansas | 44.4 | 12.5 | 0.32% | 0.03% | 0.00% |
| California | 41.3 | 14.9 | 0.27% | 0.03% | 0.00% | California | 49.8 | 8.1 | 0.49% | 0.05% | 0.00% |
| Colorado | 52.1 | 6.4 | 0.62% | 0.06% | 0.01% | Colorado | 44.7 | 11.4 | 0.35% | 0.04% | 0.00% |
| Connecticut | 41.8 | 8.1 | 0.49% | 0.05% | 0.00% | Connecticut | 49.8 | 7.6 | 0.52% | 0.05% | 0.01% |
| Delaware | 45.4 | 9.7 | 0.41% | 0.04% | 0.00% | Delaware | 46.4 | 12.9 | 0.31% | 0.03% | 0.00% |
| Florida | 47.3 | 11.2 | 0.36% | 0.04% | 0.00% | Florida | 47.7 | 6.7 | 0.60% | 0.06% | 0.01% |
| Georgia | 39.6 | 20.5 | 0.19% | 0.02% | 0.00% | Georgia | 41.4 | 15.9 | 0.25% | 0.03% | 0.00% |
| Hawaii | 27.3 | 8.3 | 0.48% | 0.05% | 0.00% | Hawaii | 45.0 | 9.0 | 0.44% | 0.04% | 0.00% |
| Idaho | 62.4 | 14.3 | 0.28% | 0.03% | 0.00% | Idaho | 55.0 | 10.3 | 0.39% | 0.04% | 0.00% |
| Illinois | 44.4 | 8.7 | 0.46% | 0.05% | 0.00% | Illinois | 53.6 | 11.6 | 0.35% | 0.03% | 0.00% |
| Indiana | 56.3 | 13.6 | 0.29% | 0.03% | 0.00% | Indiana | 52.4 | 7.0 | 0.57% | 0.06% | 0.01% |
| Iowa | 54.7 | 10.0 | 0.40% | 0.04% | 0.00% | Iowa | 52.0 | 7.3 | 0.54% | 0.05% | 0.01% |
| Kansas | 69.0 | 13.2 | 0.30% | 0.03% | 0.00% | Kansas | 51.4 | 10.6 | 0.38% | 0.04% | 0.00% |
| Kentucky | 49.4 | 10.3 | 0.39% | 0.04% | 0.00% | Kentucky | 42.3 | 8.5 | 0.47% | 0.05% | 0.00% |
| Louisiana | 45.0 | 19.5 | 0.20% | 0.02% | 0.00% | Louisiana | 54.5 | 15.5 | 0.26% | 0.03% | 0.00% |
| Maine | 54.7 | 15.2 | 0.26% | 0.03% | 0.00% | Maine | 49.5 | 7.2 | 0.56% | 0.06% | 0.01% |
| Maryland | 40.6 | 9.8 | 0.41% | 0.04% | 0.00% | Maryland | 42.5 | 10.5 | 0.38% | 0.04% | 0.00% |
| Massachusetts | 38.4 | 14.2 | 0.28% | 0.03% | 0.00% | Massachusetts | 50.8 | 10.1 | 0.39% | 0.04% | 0.00% |
| Michigan | 44.2 | 6.5 | 0.62% | 0.06% | 0.01% | Michigan | 51.6 | 7.8 | 0.51% | 0.05% | 0.01% |
| Minnesota | 45.8 | 8.0 | 0.50% | 0.05% | 0.00% | Minnesota | 48.8 | 7.5 | 0.54% | 0.05% | 0.01% |
| Mississippi | 62.4 | 20.4 | 0.20% | 0.02% | 0.00% | Mississippi | 51.2 | 8.9 | 0.45% | 0.04% | 0.00% |
| Missouri | 51.0 | 7.6 | 0.53% | 0.05% | 0.01% | Missouri | 48.6 | 8.0 | 0.50% | 0.05% | 0.00% |
| Montana | 45.2 | 9.1 | 0.44% | 0.04% | 0.00% | Montana | 48.5 | 11.4 | 0.35% | 0.03% | 0.00% |
| Nebraska | 52.4 | 14.1 | 0.28% | 0.03% | 0.00% | Nebraska | 53.3 | 13.0 | 0.31% | 0.03% | 0.00% |
| Nevada | 47.3 | 7.2 | 0.55% | 0.06% | 0.01% | Nevada | 47.9 | 15.9 | 0.25% | 0.03% | 0.00% |
| New Hampshire | 55.0 | 8.3 | 0.48% | 0.05% | 0.00% | New Hampshire | 50.0 | 11.1 | 0.36% | 0.04% | 0.00% |
| New Jersey | 44.7 | 5.0 | 0.81% | 0.08% | 0.01% | New Jersey | 49.0 | 10.1 | 0.40% | 0.04% | 0.00% |
| New Mexico | 49.6 | 12.2 | 0.33% | 0.03% | 0.00% | New Mexico | 48.0 | 6.6 | 0.61% | 0.06% | 0.01% |
| New York | 39.4 | 11.2 | 0.36% | 0.04% | 0.00% | New York | 46.5 | 12.3 | 0.32% | 0.03% | 0.00% |
| North Carolina | 49.5 | 5.3 | 0.75% | 0.07% | 0.01% | North Carolina | 46.8 | 5.9 | 0.68% | 0.07% | 0.01% |
| North Dakota | 47.9 | 15.3 | 0.26% | 0.03% | 0.00% | North Dakota | 56.8 | 12.7 | 0.31% | 0.03% | 0.00% |
| Ohio | 48.7 | 10.1 | 0.40% | 0.04% | 0.00% | Ohio | 53.0 | 10.3 | 0.39% | 0.04% | 0.00% |
| Oklahoma | 56.2 | 16.1 | 0.25% | 0.02% | 0.00% | Oklahoma | 50.5 | 9.8 | 0.41% | 0.04% | 0.00% |
| Oregon | 48.2 | 9.9 | 0.40% | 0.04% | 0.00% | Oregon | 47.1 | 7.1 | 0.56% | 0.06% | 0.01% |
| Pennsylvania | 52.8 | 6.2 | 0.65% | 0.06% | 0.01% | Pennsylvania | 48.5 | 7.9 | 0.50% | 0.05% | 0.01% |
| Rhode Island | 38.2 | 13.1 | 0.30% | 0.03% | 0.00% | Rhode Island | 45.8 | 12.8 | 0.31% | 0.03% | 0.00% |
| South Carolina | 53.2 | 12.2 | 0.33% | 0.03% | 0.00% | South Carolina | 50.7 | 9.7 | 0.41% | 0.04% | 0.00% |
| South Dakota | 54.8 | 15.6 | 0.26% | 0.03% | 0.00% | South Dakota | 58.8 | 7.7 | 0.52% | 0.05% | 0.01% |
| Tennessee | 56.9 | 12.0 | 0.33% | 0.03% | 0.00% | Tennessee | 54.1 | 13.4 | 0.30% | 0.03% | 0.00% |
| Texas | 56.4 | 7.4 | 0.54% | 0.05% | 0.01% | Texas | 53.7 | 11.2 | 0.36% | 0.04% | 0.00% |
| Utah | 62.7 | 9.7 | 0.41% | 0.04% | 0.00% | Utah | 56.9 | 14.6 | 0.27% | 0.03% | 0.00% |
| Vermont | 51.4 | 19.7 | 0.20% | 0.02% | 0.00% | Vermont | 50.2 | 12.6 | 0.32% | 0.03% | 0.00% |
| Virginia | 50.5 | 22.9 | 0.17% | 0.02% | 0.00% | Virginia | 52.9 | 14.5 | 0.28% | 0.03% | 0.00% |
| Washington | 41.6 | 10.6 | 0.37% | 0.04% | 0.00% | Washington | 47.9 | 5.6 | 0.71% | 0.07% | 0.01% |
| West Virginia | 37.2 | 15.0 | 0.27% | 0.03% | 0.00% | West Virginia | 46.0 | 9.3 | 0.43% | 0.04% | 0.00% |
| Wisconsin | 42.5 | 8.1 | 0.49% | 0.05% | 0.00% | Wisconsin | 51.9 | 6.1 | 0.65% | 0.07% | 0.01% |
| Wyoming | 69.2 | 9.1 | 0.44% | 0.04% | 0.00% | Wyoming | 51.7 | 14.6 | 0.27% | 0.03% | 0.00% |

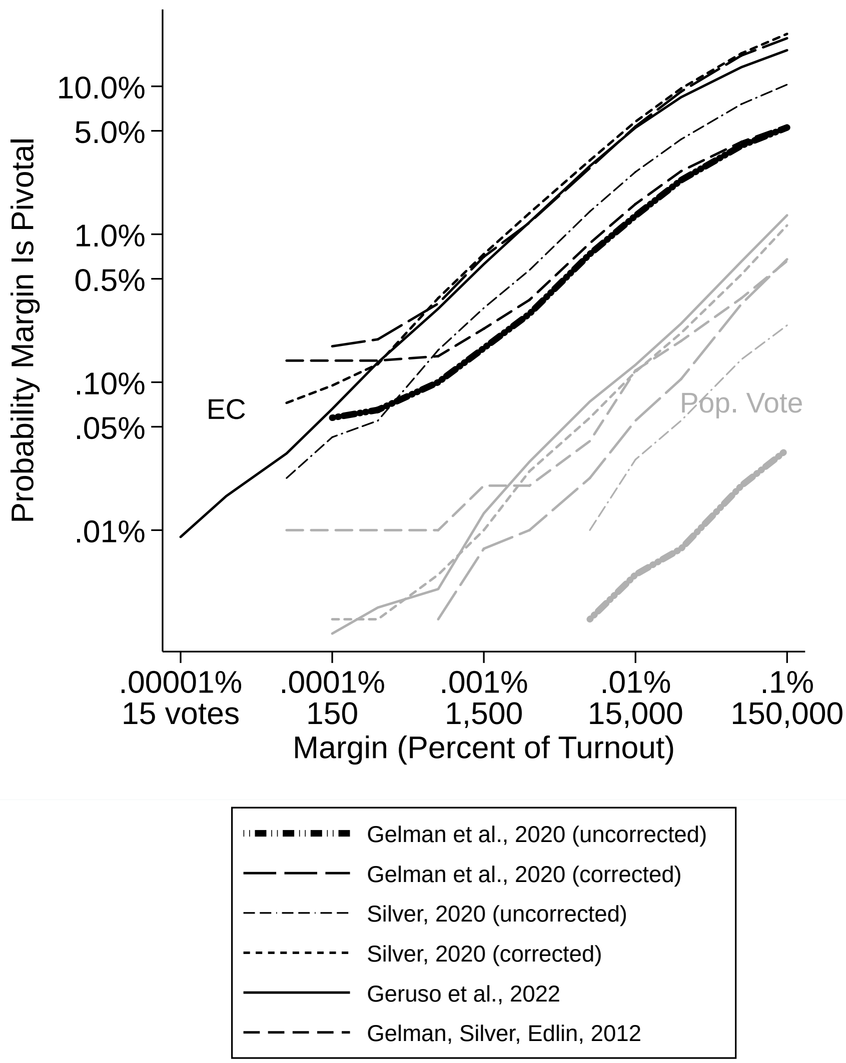
**Note:** Table reports summary statistics that describe US governor and US Senate elections, along with the implied probabilities of a close race, by election type and state. Within a state, data on the two Senate seat races are combined to estimate a single statistic. The probability of a close race is calculated by fitting the empirical standard deviation to a normal distribution centered on a 50-50 vote, as described in the text. Statistics related to the US Electoral College are repeated from Table 2, Panel C for reference. The rows labelled Senate Average and Governor Average report the unweighted means of the indicated statistics across all states. Data for this table come from the Leip Election Atlas (Leip, 2019) and include all available data on races since 1960.

**Table A5:** EC *Closeness* Compared to OECD Single-Tier Presidential Elections

| | Winning Party Share Statistics | | Probability of a Race Decided by Less Than: | | |
|---|---|---|---|---|---|
| | Empirical Mean (ignored) | Empirical Standard Deviation | 0.1 points | 0.01 points | 0.001 points |
| **OECD Average** | **61.3** | **9.3** | **0.64%** | **0.06%** | **0.006%** |
| Austria | 65.9 | 15.8 | 0.25% | 0.03% | 0.003% |
| Chile | 58.4 | 7.5 | 0.53% | 0.05% | 0.005% |
| Czech Republic | 53.1 | 2.4 | 1.64% | 0.16% | 0.016% |
| Finland | 60.7 | 13.5 | 0.30% | 0.03% | 0.003% |
| France | 60.0 | 12.1 | 0.33% | 0.03% | 0.003% |
| Iceland | 75.3 | 17.7 | 0.23% | 0.02% | 0.002% |
| Ireland | 60.9 | 7.4 | 0.54% | 0.05% | 0.005% |
| Lithuania | 61.4 | 11.9 | 0.34% | 0.03% | 0.003% |
| Mexico | 57.4 | 8.9 | 0.45% | 0.04% | 0.004% |
| Poland | 58.8 | 11.1 | 0.36% | 0.04% | 0.004% |
| Portugal | 68.6 | 10.0 | 0.40% | 0.04% | 0.004% |
| Slovakia | 58.1 | 1.8 | 2.26% | 0.23% | 0.023% |
| Slovenia | 64.8 | 8.9 | 0.45% | 0.04% | 0.004% |
| South Korea | 56.2 | 6.3 | 0.63% | 0.06% | 0.006% |
| Turkey | 60.3 | 4.1 | 0.97% | 0.10% | 0.010% |

**Note:** Table reports summary statistics that describe chief of state elections by popular vote in OECD countries, along with the implied probabilities of a close race for each country. The probability of a close race is calculated by fitting the empirical standard deviation of the winning party's vote share to a normal distribution centered on a 50-50 vote, as described in the text. Statistics related to the US Electoral College are repeated from Table 2, Panel C for reference. OECD Average reports the unweighted means of the indicated statistics across the countries in the table.

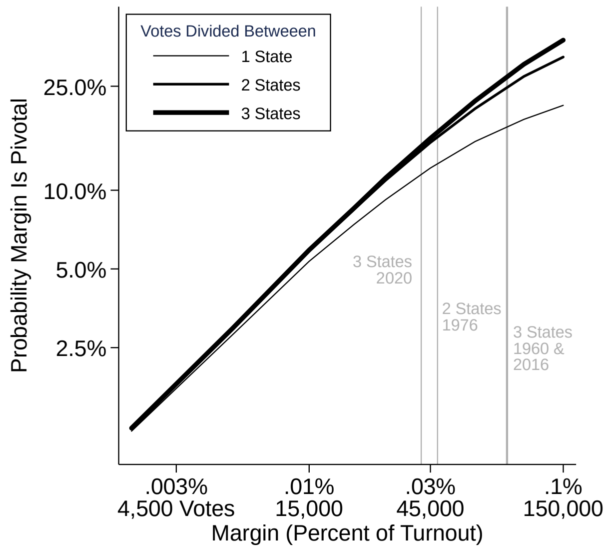
**Figure A1:** Main Result Replicated in 2008 Forecast Model and Uncorrected 2020 Forecast Models



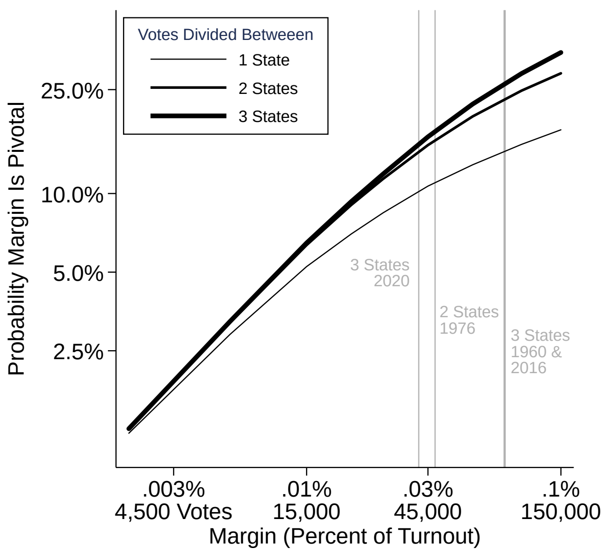

**Note:** Figure replicates the main result (Figure 2) for an extended set of models of the modern period. We plot both corrected and uncorrected versions of the 2020 forecasting models. The corrected versions subtract a 2 percentage point ex-post model bias favoring Democrats from every state-level simulated outcome in the *Economist* (Gelman, Heidemanns and Morris, 2020) and *FiveThirtyEight* (Silver, 2020) forecasts. The Gelman, Silver and Edlin (2012) forecast model characterizes the 2008 election. See Figure A4 for descriptive statistics of these models.

**Figure A2:** Close Election Risk, Allowing that Multiple States Could Be Pivotal if Combined

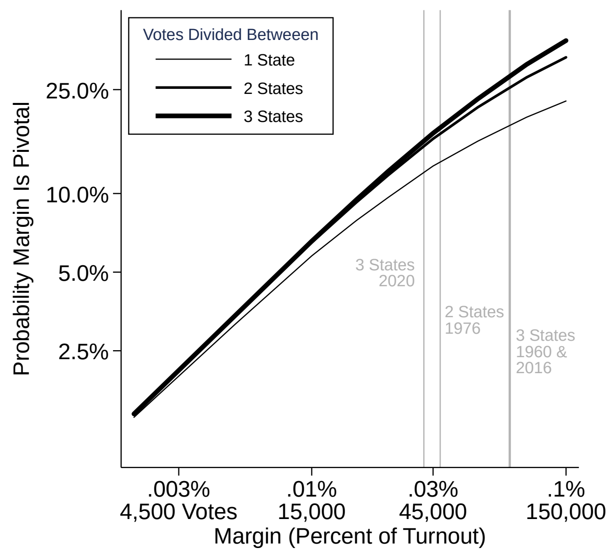
(A) Gelman et al. (2020)




(B) Geruso et al. (2022)



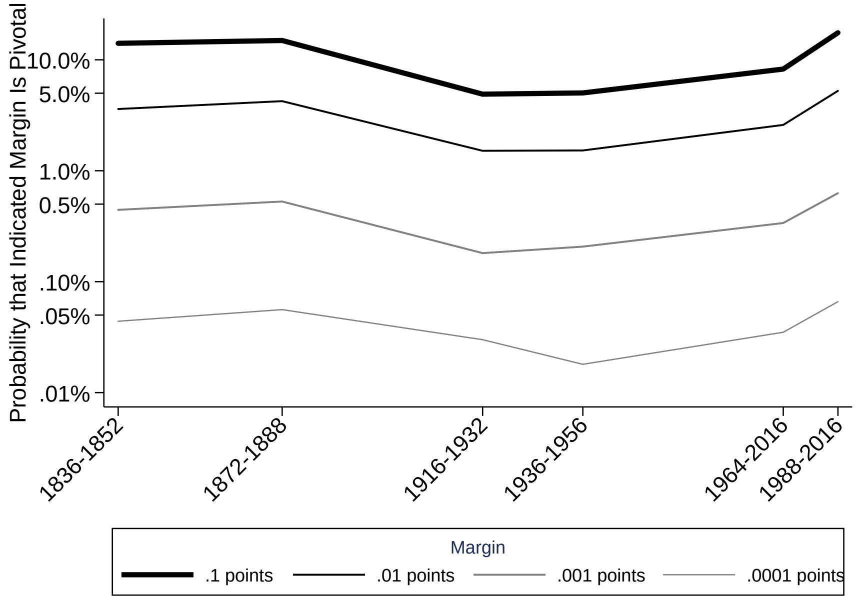

(C) Silver (2020)



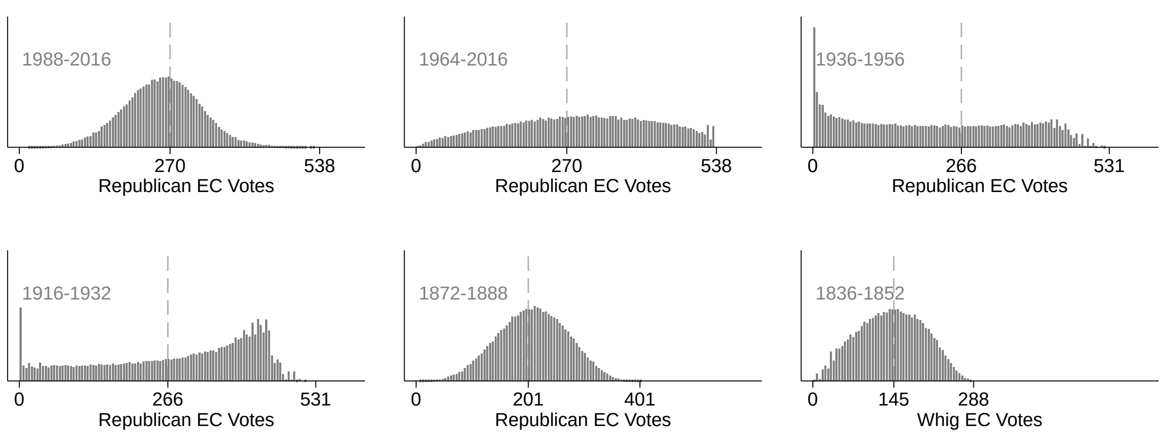

**Note:** Figure plots the probability that a national election outcome would be reversible by changing a small number of votes. The construction follows Figure 2 but allows for the indicated vote margin to be split across 2 or 3 states that together could be pivotal in determining the EC outcome. The horizontal axis indicates the closeness margin being evaluated, labelled in the top line as a percent of overall turnout, and labelled in the bottom line as vote counts, assuming turnout of 150 million voters (roughly 2020 turnout). The vertical axis indicates the probability that the indicated margin is pivotal. Panel (A) plots the probability of a close EC outcome in the models from Gelman, Heidemanns and Morris (2020); Panel (B) uses Geruso, Spears and Talesara (2022) (modern period, 1988–2016); and Panel (C) uses Silver (2020). The 1960, 1976, 2016, 2020 outcomes in which multiple states were jointly pivotal are plotted for reference, and set along the horizontal axis according to the margin in those races expressed as a fraction of the turnout in those election years. We do not plot reference cases like 1876, 1884, 1916, or 2000 in which a single state was pivotal.

**Figure A3:** Close Election Risk Has Been High over the History of Presidential Voting

(A) Probabilities of Election Closeness over US History



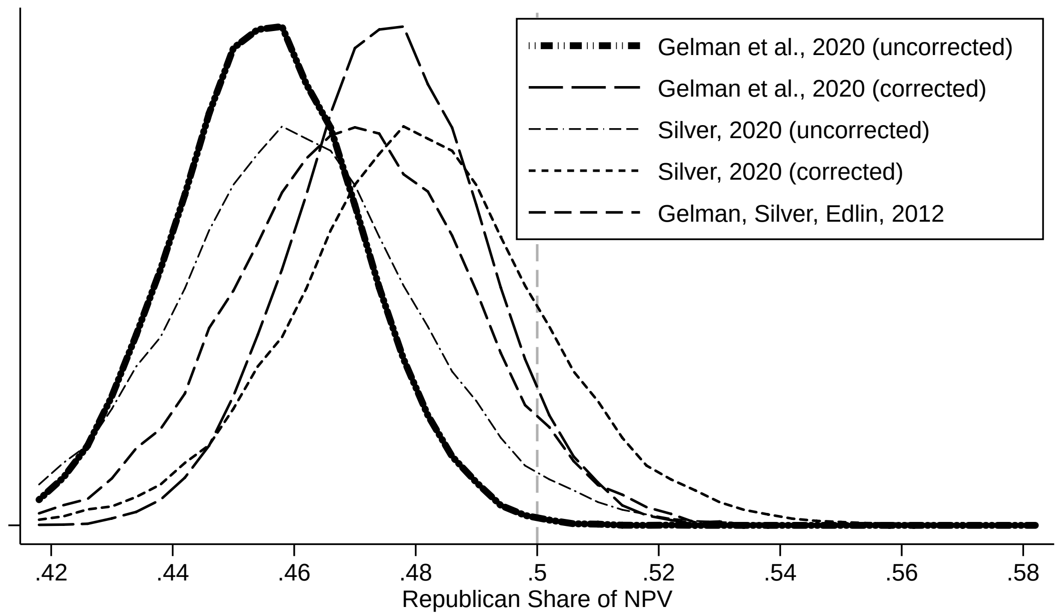

(B) Electoral Vote Probability Distributions over US History



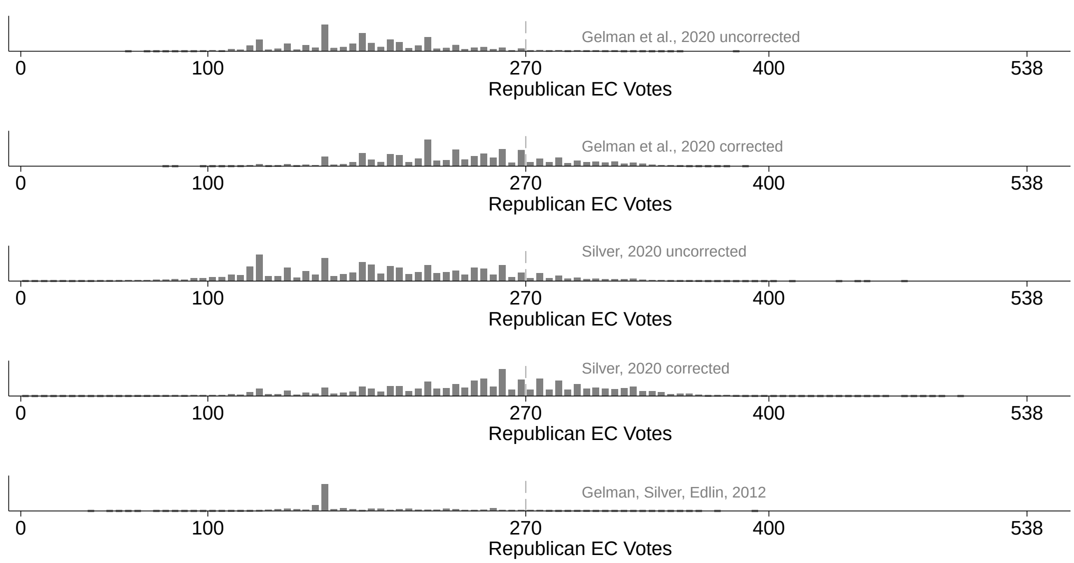

**Note:** Figure plots the probability of a close outcome for different historical periods and different thresholds of closeness. For the same set of models used to construct panel (A), panel (B) plots the distribution of the expected Republican (or, for the earliest period, Whig) electoral vote total. For the 1836–1852 panel, the GST dataset includes a total of 288 EC votes at stake. This excludes South Carolina's 8 ballots, because during this period there was no citizen vote for president in South Carolina. See also Figure A5 for popular vote distributions corresponding to these periods.

**Figure A4:** Data and Descriptive Statistics for 2008 Forecast Model and Uncorrected 2020 Forecast Models

(A) Implied Distributions of the National Popular Vote

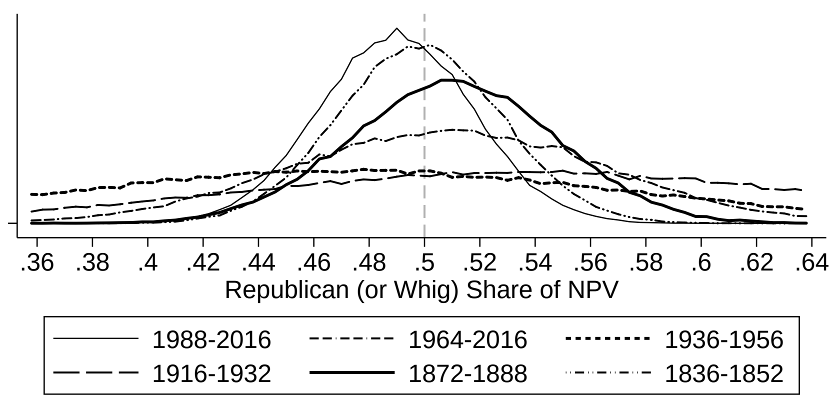



(B) Implied Distributions of the EC Vote




**Note:** Figure shows the probability distributions over electoral outcomes implied by three datasources, constructed from the 10,000 or 40,000 simulation draws in each datasource. Panel (A) plots the distribution of popular vote outcomes. Panel (B) plots the distribution of electoral vote outcomes. The corrected versions subtract a 2 percentage point ex-post model bias favoring Democrats from every state-level simulated outcome in the *Economist* (Gelman, Heidemanns and Morris, 2020) and *FiveThirtyEight* (Silver, 2020) forecasts. The Gelman, Silver and Edlin (2012) forecast model characterizes the 2008 election.

**Figure A5:** Popular Vote Distributions in the Historical Models in Geruso, Spears and Talesara (2022)




**Note:** Each trace represents a model from Geruso, Spears and Talesara (2022) describing a different time period, as indicated. See Figure A3 notes for additional detail.