How Likely Is It that Courts Will Select the US President?

The Probability of Narrow, Reversible Election Results in the Electoral College versus a National Popular Vote

Abstract: Extremely narrow election outcomes—such as could be reversed by rejecting a few thousand ballots—are likely to trigger dispute over the results. Narrow vote tallies may generate recounts and litigation; they may be resolved by courts or elections administrators (e.g., Secretaries of State disqualifying ballots) rather than by voters; and they may reduce the peacefulness, perceived legitimacy, or predictability of the transfer of political power. In this paper we evaluate the probability of such disputable US presidential elections under a hypothetical National Popular Vote versus the current Electoral College system. Starting from probabilistic simulations of likely presidential election outcomes that are similar to the output from election forecasting models, we calculate the likelihood of disputable, narrow outcomes under the Electoral College. We find that the probability that the Electoral College is decided by 20,000 ballots or fewer in a single, pivotal state is greater than 1-in-10. Although it is possible in principle for either system to generate more risk of an election outcome narrow enough to dispute, the Electoral College today is about 40 times as likely as a National Popular Vote to generate scenarios in which a small number of ballots in a pivotal voting unit determines the presidency. This disputed-election risk is asymmetric across political parties. It is about twice as likely that a Democratic (rather than Republican) Electoral College victory in a close election could be overturned by a judicial decision affecting fewer than 1,000, 5,000, or 10,000 ballots in a single, pivotal state.

You know the old saying: you win some, you lose some ...And then there’s that little-known third category.”
– Al Gore at the 2004 Democratic Convention

Authors:
Michael Geruso (UT-Austin & NBER)
mike.geruso@utexas.edu
Dean Spears (UT-Austin & IZA)
dspears@utexas.edu
How Likely Is It that Courts Will Select the US President?

The Probability of Narrow, Reversible Election Results
in the Electoral College versus a National Popular Vote

Michael Geruso*  Dean Spears†

October 2020

Abstract

Extremely narrow election outcomes—such as could be reversed by rejecting a few thousand ballots—are likely to trigger dispute over the results. Narrow vote tallies may generate recounts and litigation; they may be resolved by courts or elections administrators (e.g., Secretaries of State disqualifying ballots) rather than by voters; and they may reduce the peacefulness, perceived legitimacy, or predictability of the transfer of political power. In this paper we evaluate the probability of such disputable US presidential elections under a hypothetical National Popular Vote versus the current Electoral College system. Starting from probabilistic simulations of likely presidential election outcomes that are similar to the output from election forecasting models, we calculate the likelihood of disputable, narrow outcomes under the Electoral College. We find that the probability that the Electoral College is decided by 20,000 ballots or fewer in a single, pivotal state is greater than 1-in-10. Although it is possible in principle for either system to generate more risk of an election outcome narrow enough to dispute, the Electoral College today is about 40 times as likely as a National Popular Vote to generate scenarios in which a small number of ballots in a pivotal voting unit determines the presidency. This disputed-election risk is asymmetric across political parties. It is about twice as likely that a Democratic (rather than Republican) Electoral College victory in a close election could be overturned by a judicial decision affecting less than 1,000, 5,000, or 10,000 ballots in a single, pivotal state.

“You know the old saying: you win some, you lose some...And then there’s that little-known third category.”
— Al Gore at the 2004 Democratic Convention

*University of Texas at Austin and the National Bureau of Economic Research
†University of Texas at Austin; Indian Statistical Institute - Delhi Centre; IZA; IFFS
1 Introduction

In the 2000 US presidential election, a change of 537 votes in Florida would have reversed the outcome. This was true even though half a million votes separated the two candidates nationally. The result was uncertainty, litigation, and an outcome that is widely regarded as having been influenced by judicial decisions of the US Supreme Court (rather than solely by the votes cast). Beyond courts, an administrative decision by an election official—such as disqualifying punched ballots for hanging chads or mailed ballots for improper enveloping—could be pivotal in deciding a state outcome and ultimately a presidential race.

Large literatures in the economics of social choice and in quantitative political science seek to quantify the probability that a single citizen’s ballot is pivotal in an election \cite{Gelman1998}. This literature ignores the experience of the 2000 election: once an election outcome is narrow enough, there is an incentive for candidates to contest the outcome in court and in other post-election proceedings \cite{Posner2001, Miller2012}. Narrow outcomes—and their subsequent disputation—have a number of well-known and hypothetical consequences. From the standpoint of political theory, they may reduce the actual or apparent legitimacy of election. They undercut the value of elections in decisively transferring political power. From the standpoint of practical economics and politics, they have been argued to be bad for financial markets \cite{Nippani2002, Mattozzi2008}, for the foreseeability\footnote{In the weeks before an election, polling aggregators like fivethirtyeight.com report thousands of voter-observations being collected every day, because forecasting the EC outcome requires knowing state-level means and variances, the variance of a national shock, and correlations in errors across states. Estimating the NPV outcome, in contrast, would require merely a much simpler random sample for a proportion (bracketing uncertainty in asymmetric turnout) and could be done with statistical precision using a well-designed survey of only a few thousand voters.} of investment climates, and for the stability and effectiveness of presidential transition periods. All of these consequences are conceptually distinct from any normative argument that the candidate who receives the most citizen votes (under a National Popular Vote) or electoral votes (under the Electoral College system) “should” win the office.

\footnote{1Although litigation from the 2000 election reached the Supreme Court, the Constitution provides for the House to choose the President—one vote per state—if the Electoral College does not produce a majority winner.}

\footnote{2For example, Przeworski (2018) argues that “the greatest value of elections, for me by itself sufficient to cherish them, is that at least under some conditions they allow us to process in relative liberty and civic peace whatever conflicts arise in society, that they prevent violence” (p. 4). A tendency to extremely narrow outcomes, in practice, is likely to be a condition that interrupts such processing.}

\footnote{3For example, Przeworski (2018) argues that “the greatest value of elections, for me by itself sufficient to cherish them, is that at least under some conditions they allow us to process in relative liberty and civic peace whatever conflicts arise in society, that they prevent violence” (p. 4). A tendency to extremely narrow outcomes, in practice, is likely to be a condition that interrupts such processing.}
Table 1: Example counts of rejected ballots from swing states

<table>
<thead>
<tr>
<th></th>
<th>Absentee Ballots Rejected</th>
<th>Mail-In Ballots Rejected</th>
</tr>
</thead>
<tbody>
<tr>
<td>2016 Presidential General</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Florida</td>
<td>21,973</td>
<td>18,000</td>
</tr>
<tr>
<td>Pennsylvania</td>
<td>2,534</td>
<td>37,119</td>
</tr>
<tr>
<td>Ohio</td>
<td>10,189</td>
<td>21,154</td>
</tr>
<tr>
<td>Michigan</td>
<td>6,171</td>
<td>7,997</td>
</tr>
<tr>
<td>Wisconsin</td>
<td>284</td>
<td>23,196</td>
</tr>
<tr>
<td>Arizona</td>
<td>10,769</td>
<td>2,573</td>
</tr>
</tbody>
</table>


Administrative and judicial decisions that occur before ballots are cast or before the initial tallies are made could also be decisive in tipping the presidency. Table 1 shows that state and local election administrators typically reject thousands or tens of thousands of voter ballots. We ask whether very narrow, disputable outcomes—outcomes that could be reversed by judges or administrators rejecting or invalidating a thousand votes, or another small number—are more likely under the Electoral College (EC) system, or would be more likely under a National Popular Vote (NPV). The answer is theoretically ambiguous. Sometimes, disqualifying a few thousand ballots in one state would be enough to reverse that state’s Electoral College winner. Sometimes, flipping that one state in the Electoral College would be enough to change who is elected President. Is that more or less likely to occur than a close vote nationally?

To see why theory cannot resolve this question, it is useful to consider the tipping point state. The tipping point state is the state that, if one ordered the states by their realized partisan margin, contains the pivotal EC vote that confers a majority of the electoral votes (today, 270). Such a state could host a close election or a blowout. Intuitively, the EC is relatively likely to generate the conditions for a disputed election if the state likely to be a tipping point state is also likely to have a close outcome (e.g., Florida in 2000, won by 537 votes by Bush). In contrast, the state likely to be the tipping point state may be less likely to have a close outcome (e.g., Ohio in 2004, was the tipping point state and was won by Bush by over 100,000 votes, although other states had narrower outcomes). If the Electoral College tipping point state does not host

---

4 According to U.S. Election Assistance Commission (2017), the most common reason given by states for rejecting 2016 absentee ballots, reflecting 27.5% of rejections, was the judgment that “the signature on the ballot [was] not matching the signature on the state’s records” (p. 10).
a narrow election but the national election is nonetheless close, then deciding the presidency according to the NPV would be more likely to generate a winner reversible by few enough votes to face serious disputation risk. (For simplicity, we call such elections disputable below.)

We begin in the next section by formalizing this intuition in a simple mathematical model that shows that either case is theoretically possible. The toy model makes clear that which electoral system—the Electoral College or the NPV—is more likely to produce an election outcome close enough to be disputable depends on the expected closeness of the election in the state likely to be the EC tipping point. This, in turn, depends on the facts of political alignment across states and is ultimately an empirical question.

The main contribution of this paper is to answer that question, characterizing the empirical probability of a disputable (i.e., narrow) election under the EC system and under a hypothetical National Popular Vote system. Our approach uses the tools of statistical research, in contrast with prior studies of these risks in legal scholarship. Without specifying exactly how disputation or manipulation would occur, we compute the chances that a presidential election turns out to be vulnerable to officials’ disqualifying, reversing, or simply preventing a small number of votes. In other words, whereas other researchers have outlined the facts and circumstances that may justify or otherwise trigger election litigation or administrative interventions (Posner, 2001; Hasen, 2005), we ask simply how frequently voting outcomes are likely to be close enough that such events could decide who becomes President.

Our analysis takes as a starting point the election simulations in Geruso et al. (2020). These simulated election outcomes are generated from various election models estimated over the past two centuries of presidential contests. The models nest the methods of professional forecasters (Silver, 2016) and the standard approaches in positive political science (Katz et al., 2004). Because Geruso et al. (2020) produce simulated elections over the history of presidential voting, we are able to investigate our question for politics not only in the Modern era of the late 20th and early 21st century, but also in the post-Reconstruction and Antebellum periods.

We find that it is much more likely under the Electoral College system than under a

---

5Geruso et al. (2020) study presidential inversions—mismatches between the EC and NPV. In contrast with the present paper, but like others in the literature, they ignore post-election disputation and assume that the quantitative winner of the Electoral College wins the presidency.
hypothetical National Popular Vote that the election outcome will be narrow enough to be reversible by judicial or administrative processes. For example, in 100,000 simulation runs of the election model that describes the period 1988–2016, it is about 40 times more likely that a few thousand votes—if discarded by courts or election administrators—could reverse the outcome of the election.\(^6\) In particular, the chance that the election is reversible by disqualifying 1,000 votes or fewer is 0.465% in the EC versus 0.008% in an NPV system.\(^7\) The chance that the election is reversible by disqualifying 10,000 votes or less is 4.0% in the EC versus 0.1% in an NPV system. These probabilities are much greater in elections that are likely to be close in terms of the national vote outcome. As we show in robustness checks, the large differences in probabilities hold not only for a range of alternative models fit to actual election data, but also across thousands of hypothetical models of how politics could change under a National Popular Vote—with races becoming tighter, races becoming looser, states becoming more Democrat-leaning, states becoming more Republican-leaning, etc.

We also show that the disputed-election risk is asymmetric across political parties, with the probabilities favoring Republicans over the past 30 years: Our election model covering the period 1988-2016 predicts that it has been about twice as likely that a Democratic Electoral College victory in a close election would be within a disputable 1,000 vote margin in a pivotal state than that a Republican Electoral College victory would be.

To understand the mechanism behind our main finding of vastly different risk of narrow outcomes between the EC and NPV systems, we investigate which counterfactual patterns of US political geography could reverse our main finding. The exercise shows that if the US were much more geographically polarized than it currently is or ever was, then it is possible that the EC system could result in fewer narrowly decided elections than an NPV system.\(^8\) This

---

\(^6\)In other words, that a few thousand votes or fewer separate the presidential winner and loser in the pivotal voting unit. The pivotal voting unit under a hypothetical NPV is the entire US. Under the EC system it is a state whose outcome, if changed, could reverse the Electoral College winner.

\(^7\)Throughout the paper, we typically report the pivotal vote thresholds in terms of ballot *reversals*—e.g., replacing a ballot cast for a Democratic with a ballot cast for a Republican. Here, in contrast, we are reporting vote thresholds corresponding to ballot *disqualifications*. 10,000 reversals from party A to party B have the same effect on the outcome as 20,000 disqualifications of party-A ballots.

\(^8\)We show that disputed election risk could in principle be higher in an NPV system if the variance in across-state average election outcomes were much greater—for example if California were even further from Oklahoma in partisan lean (leaving the national mean and distribution of election outcomes unchanged).
hypothetical reversibility emerges only under large counterfactual changes to the electoral map. The actual patterns of partisan geographies today—and the partisan geographies across US history, which has contained different states, different dominant parties, and different groups of enfranchised voters—all indicate that very narrow outcomes are likely to be a perennial feature of future elections, if held under the Electoral College. Our findings indicate that a National Popular Vote would be significantly less likely to generate disputed-election risk and the apparent illegitimacy of a very narrow outcome.

2 Illustrative model

In this section, we present a toy model of two-tier voting that motivates our empirical analysis beginning in the next section. The model illustrates two facts. First, either the EC or the NPV could be more likely to generate a disputable, narrow-margin election outcome. Second, an important factor shaping these probabilities is the expected closeness of the vote in the tipping point state—the state likely to be pivotal in generating the minimum number of votes needed to win the EC. More precisely, if one sorted all states by partisan alignment (such as by ex post Republican vote shares) and tallied the cumulative total of the EC votes at each ranking, the tipping point state is the state that would contain the pivotal EC vote (today, the 270th vote). The tipping point state need not be close to 50-50 partisan alignment in principle, even though it often has been in recent politics.

For the purposes of the stylized model, assume that there are three equal-sized states ($s \in \{1, 2, 3\}$) with realized Republican vote share $x_s \in [0, 1]$ and expected Republican vote share $\mu_s$. We assume that states are numbered according to increasing $\mu_s$ and that the probability that state 2 is not pivotal for the EC is small enough to ignore. So the Republican wins the EC if and only if the Republican wins the tipping point state—that is, if and only if the realized $x_2$ is greater than 0.5. The Republican wins the NPV if and only if $\bar{x} (\equiv \frac{1}{3} \sum_s x_s)$ is greater than 0.5.

Let $x_s$ be normally distributed for each state, and assume a homoscedastic variance ($\sigma^2_s = \sigma^2$) and that $x_s$ are independent. Assuming independence ignores the empirical reality of
Table 2: Disputed Election Risk in the EC and NPV: An Illustrative Model

<table>
<thead>
<tr>
<th>map</th>
<th>states’ map sizes</th>
<th>state expectations</th>
<th>NPV, $E[\bar{x}]$</th>
<th>expected number of votes needed to reverse outcome if</th>
<th>expected number of votes needed to reverse outcome if</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1,000</td>
<td>35% 53% 65%</td>
<td>53%</td>
<td><strong>47</strong></td>
<td><strong>73</strong></td>
</tr>
<tr>
<td>B</td>
<td>1,000</td>
<td>20% 63% 70%</td>
<td>53%</td>
<td><strong>130</strong></td>
<td><strong>73</strong></td>
</tr>
</tbody>
</table>

**Note:** Table shows how the number of votes that could overturn the presidential winner in an EC or NPV system depends on the expected vote share in the tipping point state. The table considers two alternative “maps” of where Democratic and Republican voters reside. In both maps, we hold fixed the Republican share of the national popular vote, $E[\bar{x}]$ at 53%. In both maps, states each contain 1,000 voters. Map B is more geographically polarized and the tipping point state is less close than in map A. As a result, in map A, it is more likely that a small number of votes could reverse the outcome in an EC system than in an NPV system. In map B, it is more likely that a small number of votes could reverse the outcome in an NPV system than in an EC system.

correlated national election-year shocks.\(^9\) Such an assumption clarifies this toy mathematical model. In contrast, the empirical model of the data-generating process of election outcomes that we use below to establish our main result incorporates correlated national election-year shocks (as well as regional shocks, states of heterogenous sizes and apportionment representation, etc.). Because the sum of independent normal random variables is normal, the NPV is normal with an expectation of $\frac{1}{3} \sum s \mu_s$ and a variance of $\frac{1}{3} \sigma^2$.

Let each state contain $n$ voters. Under the EC, the number of votes, $v_{EC}$, that would reverse the outcome is:

$$v_{EC} = n |0.5 - x_2| \text{ where } x_2 \sim \mathcal{N}(\mu_2, \sigma^2). \quad \text{(EC)}$$

Under the NPV, the number of votes, $v_{NPV}$ that would reverse the outcome is:

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

Taking expectations of $v_{NPV}$ and $v_{EC}$ shows that which is likely to be greater 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 $\frac{1}{3} \sigma^2$.\(^{10}\)

Table 2 illustrates this numerically, in an example in which we assume states of 1,000 voters,

\(^9\)One can read this model as conditional on national shocks, which are embedded in $\mu_s$.

\(^{10}\)In particular, the expectations $E[v_{EC}]$ and $E[v_{NPV}]$ can be calculated as the means of folded normals:

$$\sqrt{\frac{2}{\pi}} \sigma \exp \left(-\frac{\mu^2}{2\sigma^2}\right) + \mu \left(1 - 2 \Phi\left(-\frac{\mu}{\sigma}\right)\right)$$
expected vote shares $\mu_s$ as indicated in the table, and a standard deviation of 5% ($\sigma = 50$) for all states. Both electoral maps A and B have the same expected Republican share of the NPV (indeed, they have the same distribution of the NPV), but the location of the Republican votes across states differs in the two maps. In map A, the expectation for the decisive number of votes in the EC ($E[v^{\text{EC}}]$) is smaller than in the NPV ($E[v^{\text{NPV}}]$) because the tipping point state is close to 50% in expectation. Of the two maps, A is the scenario that better corresponds to the actual geographic alignment in US politics.\(^{11}\) In map B, the decisive number of votes in the EC is expected to be larger than in the NPV. Because the facts of geographic partisan alignment matter, we now turn to simulations of likely presidential election outcomes, estimated from voting 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, turnout, and so on.

3 Method

3.1 Election simulation method of Geruso et al. (2020)

We build upon the models and simulations of Geruso et al. (2020), to which we refer the reader for complete details. Their approach proceeds in two stages: First they estimate the statistical model (i.e., the data-generating process) for presidential elections at the level of the states based on historical voting data, and then they sample from the estimated model to build distributions of likely outcomes.\(^{12}\) We use the same set of 100,000 simulated election draws that they produce from each estimated model. We focus on two of their preferred models (within each time period), which track the typical approaches to election modeling in positive political science and election forecasting. Each simulation draw provides a Republican (or Whig) vote share for each of the states and DC in a given simulated election year outcome. The simulations reveal the

\(^{11}\)In the empirical configuration of the US electoral map, the expected outcome of the states that are likely to be pivotal in the EC — here represented by $\mu_2$ — is close to 0.5, so the EC is more likely to have a very small number of decisive votes. But, as we show in a counterfactual computation, if the map changed such that each state’s $\mu_s$ moved away from 0.5, then $\mu_2$ could be much farther from 0.5 than $\frac{1}{2} \sum_s \mu_s$ would be (because polarization of Republican and Democratic states would balance each other out) so a small number of decisive votes would be more likely under the NPV.

\(^{12}\)Other aspects of two-tier election systems have been studied with Monte Carlo simulation by Feix et al. (2004), Esposito and Merlin (2010), and others, although such studies typically simulate stylized systems (such as equal population across “states”).
Monte Carlo probability distributions of presidential election outcomes for a generic Republican (or Whig) and Democratic candidate for the time period over which each model is estimated.

Following Geruso et al. (2020), we study three time periods during which US presidential politics produced close elections: 1988–2016, which we call the Modern era; 1872–1888, which we label post-Reconstruction; and 1836–1852, Antebellum. We ignore third parties throughout; when we write of the “Republican vote share” we mean the two-party share, that is, the share of votes for for the Republican (or Whig) versus the Democratic candidate.

Geruso et al. flexibly model the data-generating process for a state-by-election-year \((st)\) outcome as consisting of a state expectation, \(\pi_s\), and a mean-zero shock \((\epsilon_{st})\), which may be correlated across states in an election:

\[
\logit(V_{st}) = \pi_s + \epsilon_{st} 
\]

\[
\epsilon_{st} = \gamma_t + \phi_{st} + X_s \delta_t 
\]

The outcome variable of interest is \(V_{st}\), the two-party vote share for the indexed party (normalized to be Whigs before the Civil War and Republicans afterward) in the state-year.

In one model, they simplify to \(\epsilon_{st} = \gamma_t + \phi_{st}\), where \(\gamma_t\) is a national shock specific to an election year and \(\phi_{st}\) is a state-year shock, both normally distributed. State-year shocks are drawn independently with a common variance, which is different from the variance of the national shock. These two variances and state-specific means are estimated via maximum likelihood with a structural equation model of this data-generating process. This model is an application of the “unified method of evaluating electoral systems” of Gelman and King (1994). We follow Geruso et al. (2020) in describing this as “model 1” for each time period.

In robustness checks, we present further models. Model 2 parameterizes \(X_s \delta_t\), which allows for correlated shocks across states. The national shock \(\gamma\) also introduces correlation across states, but this set of shocks allows the strength of the correlation to depend on state-level observable \(X\). Following Silver (2016), the state-level observables that Geruso, et al use for the Modern period are race, education, and census region (only census region is plausible among these for the earlier periods). They use maximum likelihood to jointly estimate the variances of \(\gamma, \phi\) and...
the set of $\delta$. Also following Silver (2016), they draw shocks from a $t$ distribution with degrees of freedom one less than the number of years used to estimate that model. Simulations methods, with descriptive statistics, are summarized in Appendix Table A1.

### 3.2 Method for identifying narrow outcomes

The Geruso et al. (2020) method produces 100,000 simulated election outcomes for a given model, parameterized to a given time period. We use these existing simulated outcomes directly to ask a different question than that study.\(^{13}\) For each draw, for a given narrow margin $m$, such as 1,000 votes, we ask whether a change in that number of votes (from Republican to Democrat or vice versa) would change the outcome of the election.

For the NPV this trivially requires observing the national popular vote share, and whether transferring $m$ votes from the candidate who received the most popular votes would result in the other candidate receiving the most popular votes. If so, that simulation would be reversed by that number of votes. We estimate the probability as the fraction of the 100,000 simulations for which this is true. Note that, throughout our paper, we report these pivotal vote margins $m$ as thresholds of vote reversals that could alter the winner of the presidency---e.g., replacing a ballot counted for a Democratic with a ballot counted for a Republican. To translate any threshold $m$ that we discuss or compute into the corresponding threshold number of ballot disqualifications that could alter the winner of the presidency, multiply $m$ by two. Disqualifications could include ballot-level judgements that signatures, envelopes, postmarks, or chad-punches are unclear or inadequate; locality-level judicial decisions to halt counting; administrative decisions to remove absentee ballot drop-boxes or otherwise change early-voting procedures; or any act that prevents a voter’s effort to vote for a candidate from counting as a vote for any candidate (cf. Alvarez et al., 2008).

For the EC, we repeat this procedure for each of the states, including Washington DC in the Modern period (DC receives three EC ballots). If transferring $m$ votes would change the winner of a state, then we additionally compute whether changing that state’s outcome would change the overall winner of the Electoral College. For a simulation, we count it as reversed by that

---

\(^{13}\)So, in the present paper, we neither estimate model parameters in equation 3.1 nor draw random outcomes.
number of votes if this is true for any state capable of conferring an electoral vote majority—that is, if that number of votes could be pivotal in reversing which candidate wins the Electoral College. We estimate the probability as the fraction of simulations for which this is true (for one or more states). Note that we therefore ignore the possibility that \( m \) votes could change the Electoral College outcome if split over two or more carefully chosen states; we do this because our motivating issues of legitimacy and disputation are most salient in one-state cases, which are the level at which election administrators and courts may act. But we note that this is a reason that the probabilities that we compute for the Electoral College will be, if anything, underestimates.

4 Results

4.1 Main Result: Probabilities of Narrow Outcomes

Figure 1 presents our main result: a disputably narrow outcome, in which a small number of votes would change the winner of the presidency, is more likely under the EC than under the NPV. This is true at all narrow margins considered—from 500 up to 10,000 votes. The scales in the figure are logged because the differences between the probabilities corresponding to the EC and NPV systems are so large.

The difference in probabilities of narrow, disputable election outcomes in the Modern period is about 40 times larger under the Electoral College system than if the winner were chosen by popular vote in the same simulated elections. For example, the probability that an election would be reversed by changing 1,000 votes in an NPV system in Figure 1 is about one-in-4,500. The probability that an election would be reversed by changing 1,000 votes in our current EC system is about one-in-100. That approximately 40-times difference is fairly constant across models and across the vote thresholds considered.\(^{14}\) The chance of an election outcome being reversible by an administrative or judicial ruling that swung the tallies by 1,000 votes.

---

\(^{14}\)Specifically, the chance that the election is reversible by reversing 1,000 votes or fewer is 0.90% in the EC versus 0.02% in the NPV system. The chance that the election is reversible by reversing 10,000 votes or fewer is 6.9% in the EC versus 0.2% in the NPV system. At very large thresholds, the difference shrinks, but an order-of-magnitude gap remains: The chance that the election is reversible by reversing 100,000 votes or fewer is 19.7% in the EC versus 1.9% in an NPV system.
Figure 1: Main Result: The Probability that a Few Votes Decides the Election is Higher in the Electoral College than under an NPV

Note: Figure shows probabilities that election outcomes could be reversible by a small number of votes. Probabilities are computed as fractions of 100,000 Monte Carlo simulations. For example, 2,153 out of 100,000 simulated election outcomes would be reversed in the EC by a change in 2,500 votes in a state, so the EC has a 2.2% chance of being reversible at a 2,500 vote threshold of contestability. In contrast, the NPV has only a 0.05% chance of being reversible by a 2,500 vote change.
votes is much smaller than a reversal due to changing 10,000 votes. But at both thresholds and in the range between, the disputed-election risk is an order of magnitude higher in the EC system than under a National Popular Vote. Considering a slightly larger threshold, we find that the probability that the Electoral College is decided by 20,000 ballots or fewer in a single, pivotal state is greater than 1-in-10. The corresponding probability for the NPV is less than 1-in-250. We show below that order of magnitude difference isn’t a mathematical necessity or in any way linked to the number of US states: If one altered the underlying model of voting so that geographic polarization across the fifty states and DC was different than is reflected in the estimated election models, the gap between the EC and NPV disputed-election risk could close or even reverse.

In Figure 2 we repeat the analysis of Figure 1 for different modeling approaches and for earlier time periods. For the Antebellum and post-Reconstruction periods, we likewise find an order of magnitude separating the probability of a narrow outcome that could be reversed by changing a small number of votes when comparing the EC and NPV. As in the Modern period, this holds regardless of the reversal threshold considered. It is important to note that for the earlier periods, the margins in levels of votes represented a larger share of the electorate. In the Antebellum period for example, some states had less than 100,000 popular votes and national popular vote turnout was 2 to 3 million.

4.2 Robustness to Alternative Parameterizations of the National Popular Vote

Perhaps the distribution of likely election outcomes would change if the rules of electing a president changed to become an NPV system. Perhaps campaigns, parties, and voters would change their strategies or other behaviors, and different election outcomes would be likely. If so, we cannot learn about such possible futures from our main statistical models, because they are informed by data about the past, under the EC.

Would our main result (that narrow, disputable elections are relatively unlikely under an NPV system) still be true if the distribution of likely NPV outcomes changes? To answer this
Figure 2: Robustness: Repeating Analysis for Other Election Modeling Choices and for Elections in Other Time Periods

(A) Modern (1988–2016) Model M1

(B) Modern (1988–2016) Model M2

(C) Post-Reconstruction (1872–1888) Model R1

(D) Post-Reconstruction (1872–1888) Model R2

(E) Antebellum (1836–1852) Model A1

(F) Antebellum (1888–2016) Model A2

Note: Figure shows probabilities that election outcomes could be reversible by a small number of votes. Panel (A) reprints Figure 1 for comparison. Probabilities are computed as fractions of 100,000 Monte Carlo simulations. For example, in Panel (A), 2,153 out of 100,000 simulated election outcomes would be reversed in the EC by a change in 2,500 votes in a state, so the EC has a 2.2% chance of being reversible at a 2,500 vote threshold of contestability. In contrast, the same panel shows that the NPV has only a 0.05% chance of being reversible by a 2,500 vote change.
**Figure 3:** The Main Result is Robust to Many Counterfactual Scenarios of How Politics Might Change Under a National Popular Vote System

(A) Main Estimated Election Models (M1 and M2) and Many Alternative Counterfactual Models

(B) Histograms of 1,000-Vote Reversal Probabilities Under Counterfactuals

(C) Histograms of 5,000-Vote Reversal Probabilities Under Counterfactuals

**Note:** Figure examines whether a changing national electorate (a changing national partisan lean and a changing spread of potential outcomes) would overturn the finding that close races are unlikely under a National Popular Vote system. Results for the EC are Models 1 and 2 of Figure 2 for the Modern period (1988–2016). For visibility, only a subset of the 3,295 alternative parameterizations of the National Popular Vote are plotted in Panel A. Histogram points in Panels B and C are results from all 3,295 alternative parameterizations of the National Popular Vote, described in detail in Section 4.2.
question, we replicated our main result with 3,295 alternative models for the NPV.\textsuperscript{15} These include close elections, Democratic landslides, and Republican landslides; electoral systems with low variation in outcomes from year to year and high variation in outcomes; elections with fat tails and narrow tails; and elections with larger and smaller turnout than in 2016. These include models in which there is a greater than 99.99\% chance that the NPV will be within 0.49 to 0.51 and a greater than 95\% chance that it will be within 0.495 to 0.505. (In fact, only 6 out of 48 elections since the US adopted citizen voting for president in the early 1800s have been within a margin of one percentage point.) So, if an NPV system would incentivize narrow national margins, then our simulations include that possibility—in addition to including wide national margins, too.

Figure 3 plots the result of this search over thousands of models. The top panel plots the NPV density functions of the models considered to highlight the range of diversity. The remaining panels calculate summary statistics of the probability of a disputable election at thresholds of 1,000 or 5,000 votes. In these histograms there is one observation per counterfactual model. In none of these models is the probability of a very narrow outcome greater under the NPV than under the EC. For narrow outcomes in the hundreds or thousands, a result close enough to risk litigation and administrative intervention is more likely under the Electoral College than under the National Popular Vote for \textit{any} plausible model (and many implausible models) of how the National Popular Vote outcome might endogenously change under a National Popular Vote system of electing presidents.
Figure 4: Asymmetry: Minimum Votes Needed to Reverse Electoral College Victory

Note: Figure plots all of the simulated election draws for which the Electoral College outcome could be reversed (i.e., the alternate candidate would win the presidency) by an administrative or judicial ruling that reversed 10,000 votes or fewer. Simulated elections are computed using the “Model 1” variant for the Modern era, as in Figure 1. Of these 100,000 simulated election draws from model M1, 6,897 of these probable elections could be reversed by exchanging 10,000 votes or fewer. Several happen to be reversible by a handful of votes in a pivotal state.
4.3 Asymmetry

In Figure 4 we focus on the probability of a disputable, narrow election outcome under the EC system and investigate whether the odds are asymmetrical across parties in the Modern period. Of 100,000 simulated election draws from model M1, 6,897 of these probable elections could be reversed by exchanging 10,000 votes or fewer. Each such outcome is plotted in the top panel of the figure according to the number of votes needed to swing the election and the NPV in that simulation. The figure shows that even with just 100,000 simulated elections, there are cases where the EC is decided by a mere handful of votes, while the same simulated elections have the candidate totals differing by a few hundred thousand to several million votes nationally. The figure also shows that disputed-election risk is greater when the Democratic candidate gains a larger share of the popular vote. Of these 6,897 disputable election outcomes, 71% are cases where Democrats won the NPV. This statistic is complicated to interpret because in many simulations in which the Democratic candidate would “win” the national popular vote, they would nonetheless lose the Electoral College and therefore the presidency.

In Figure 5, we take a different look at partisan asymmetry, this time conditioning on which candidate wins the EC, and further zooming in on close elections. Here, we use the same 100,000 simulated elections as in Figure 4 (which is the same as in Figure 2). In races that are close at the national level—decided by less than one, two, three, or four percentage points of the NPV—it is more likely that an Electoral College victory by a Democratic candidate could be overturned by reversing small number of votes. This is true whether one considers a 1,000 or 5,000 vote reversal threshold. It is also true at larger thresholds (not pictured). By this measure, the disputed-election risk is more favorable to Republicans than Democrats. At wider NPV

---

15In particular, we model the NPV directly (not state outcomes), using every combination of:
- NPV expectation in 0.5 percentage point steps from 0.45 to 0.55.
- NPV standard deviations of 0.25, 0.5, and 0.75 percentage points, then 0.5 point steps from 1.0 to 7.5.
- National turnout at 90%, 95%, 100%, 105%, and 110% of 2016 levels.
- A statistical distribution that is either normal or $t$ with 7 degrees of freedom (because there are 8 observations in our Modern period).

Note that it is not necessary to model state-level outcomes to model the NPV, if a distribution such as $t$ or normal can be assumed. Unlike in our main results, which are derived from fitting a structural model to actual election outcomes, for these 3,295 models we do not compute simulations but instead compute results directly from a continuous normal or $t$ distribution.
Figure 5: Asymmetry: Probability of a Narrow, Disputable Electoral College Outcome by Party when the National Popular Vote is Close

(A) Probabilities that 1,000 Votes Could Reverse Outcome

<table>
<thead>
<tr>
<th>Probability Election Is Reversible by 1,000 Votes or Fewer</th>
<th>Close Races, Nationally</th>
</tr>
</thead>
<tbody>
<tr>
<td>Republican</td>
<td>Democrat</td>
</tr>
<tr>
<td>&lt; 1 point</td>
<td>&lt; 2 points</td>
</tr>
<tr>
<td>1.7%</td>
<td>3.2%</td>
</tr>
<tr>
<td>&lt; 3 points</td>
<td>&lt; 4 points</td>
</tr>
<tr>
<td>1.8%</td>
<td>1.7%</td>
</tr>
<tr>
<td>2.7%</td>
<td>2.2%</td>
</tr>
<tr>
<td>1.5%</td>
<td>1.9%</td>
</tr>
</tbody>
</table>

Note: Tight Electoral College wins by Democrats are more likely to be reversed by a ruling that disqualifies a small number of votes in a key state.

(B) Probabilities that 5,000 Votes Could Reverse Outcome

<table>
<thead>
<tr>
<th>Probability Election Is Reversible by 5,000 Votes or Fewer</th>
<th>Close Races, Nationally</th>
</tr>
</thead>
<tbody>
<tr>
<td>Republican</td>
<td>Democrat</td>
</tr>
<tr>
<td>&lt; 1 point</td>
<td>&lt; 2 points</td>
</tr>
<tr>
<td>8.1%</td>
<td>14.0%</td>
</tr>
<tr>
<td>&lt; 3 points</td>
<td>&lt; 4 points</td>
</tr>
<tr>
<td>7.8%</td>
<td>7.2%</td>
</tr>
<tr>
<td>12.7%</td>
<td>10.4%</td>
</tr>
<tr>
<td>6.6%</td>
<td>8.6%</td>
</tr>
</tbody>
</table>

Note: Figure shows that Electoral College wins by Democratic candidates in close national elections (in terms of the national popular vote margin) are more likely to reversible by a small number of votes. The top panel plots the probabilities that Democratic and Republican EC victories could be reversible by an action or ruling that changed 1,000 votes in a key state. The bottom panel plots the probabilities that Democratic and Republican EC victories could be reversible by an action or ruling that changed 5,000 votes in a key state. Results are computed using the “Model 1” variant for the Modern era, as in Figure 1.
margins—races separated by larger national vote totals—the partisan gap shrinks and even reverses.

In contrast with this asymmetry in the EC, consider the NPV. Only 186 simulated elections have an NPV margin within 10,000 votes, compared with the 6,897 under the EC. Of these 186, 98 are Republican popular vote wins and 88 are Democratic wins, which is a 53%-47% split (compared with 29% and 71% for the EC). In other words, contestably narrow outcomes are much less likely under an NPV, and contestably narrow outcomes are more symmetrically distributed between parties in the rare chance that they occur under an NPV.

4.4 Counterfactual:

The Pattern Reverses as Cross-State Variance increases

Could the minimal vote margin to swing the presidency ever be greater for the NPV than under the EC? In other words, could a National Popular Vote possibly generate more disputed-election risk? Section 2 presents a simple numerical example that suggests the answer is yes (see electoral map B from that example). In that hypothetical case, the national popular vote was close, even though each state was lopsided. This possibility does not well describe the basic facts of American politics, where many states are competitive. In this section, we illustrate the very large change to the US electoral map that would be required for our result to reverse. That is, we ask how different US geographic polarization would have to be from the polarization at any time in the past 200 years for the disputed-election risk to be higher under an NPV system than under the EC system.

To do so, we construct counterfactuals that take as a starting point our simulated election outcomes in M1 (which are estimated on training data of the actual partisan geographies of the US over 1988–2016), but we step each state’s mean away from 50-50. For example, under “shift by 10” scenarios, states that typically vote Republican become 10 percentage points more Republican in expectation, and states that typically vote Democratic become 10 percentage points more Democratic. Figure 6 presents four cases (including replotting our main result as

---

16 A non-parametric Wilcoxon test can reject that elections with EC margins within 10,000 are equally likely to be Democratic and Republican NPV wins ($z = -35.6$) but cannot reject equal likelihood for elections with PV margins within 10,000 ($z = 0.7$).
Figure 6: Hypothetically, the Pattern Reverses if Polarization is Increased by Pushing Each State’s Mean Away from 50-50

(a) Estimated Model (No Shift to Partisan Lean)
- national mean = 48.97
- national s.d. = 2.77
- cross-state-mean s.d. = 9.80

(b) Increase Partisan Lean in Each State by 10 Points
- national mean = 48.96
- national s.d. = 2.77
- cross-state-mean s.d. = 18.46

(c) Increase Partisan Lean in Each State by 15 Points
- national mean = 48.96
- national s.d. = 2.77
- cross-state-mean s.d. = 23.02

(d) Increase Partisan Lean in Each State by 16 Points
- national mean = 48.96
- national s.d. = 2.77
- cross-state-mean s.d. = 23.95

Note: Figures are computed using the “Model 1” variant for the Modern era, as in Figure 1. Each state’s mean Republican share is moved away from 50-50 by the indicated number of percentage points (shares are bounded at 1% and 99%). The mean and variance of the national popular vote are held nearly constant, as shown in the summary statistics for each model, expressed in units of percentage points. “Cross-state-mean s.d.” is the standard deviation of the 51 state-level expected Republican shares.
a 0-point shift for comparison). The most extreme counterfactual plotted, which is different enough from the actual political map to reverse our main result, is a counterfactual in which each state is further polarized by 16 percentage points. In such a counterfactual world, California and Oklahoma would have a mean Republican vote share of 25.1% and 79.4%, instead of their true historical average of 41.8% and 62.1% in the Modern era. This finding reinforces our observation in the introduction that it is not logically necessary that an Electoral College be more likely to produce narrow-election risk than a National Popular Vote. Yet it is in fact true of the modern US electoral map and would be true even under extreme counterfactual changes. Indeed, a 15-point polarization shift, which would put California and Oklahoma at averages of 26.1% and 78.4%, would not be enough to reverse our conclusions, as Panel (c) of Figure 6 shows.

5 Conclusion

Extremely narrow election outcomes—such as would be reversed by a few thousand votes—have been claimed to have normatively disadvantageous properties. They may incentivize litigation or recounts; they may be resolved by courts or administrators rather than voters; and they may reduce the legitimacy, peacefulness, speed, or predictability of the transfer of political power. In principle, such narrow outcomes could be more likely under the Electoral College or under a National Popular Vote.

In fact, we show, they are more likely in US Presidential elections under the Electoral College than would be the case with a National Popular Vote. We demonstrate this with a new application of Geruso et al.’s (2020) statistical simulations of US presidential elections. Even in cases where the candidates are separated by up to a million votes in the National Popular Vote, we find that the chance is greater than 1-in-45 that the Electoral College outcome would be reversible by a number of votes in the hundreds, in one state. Our finding holds under a wide range of statistical models, hypothetical models for a National Popular Vote, and historical and modern time periods featuring close elections—diverse periods which span different sets of states, parties, and electoral maps. The findings here suggest that a National Popular Vote could significantly reduce the risks of extremely narrow, disputable election outcomes.
Only 8 of our 100,000 simulated outcomes have a popular vote tipping count as close as the 537 that actually occurred under the EC in 2000; at that rate, assuming Presidential elections continued for millennia to arrive once every four years, such an outcome would be expected under the popular vote once in every 50,000 years. Writing about the 2000 Presidential election, Lempert (2016) suggested that the possibility of recounts 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 there would not have been an NPV recount. In 2000, there were more than 500,000 popular votes separating the two candidates—not a hairsbreadth, and indeed a larger margin than any of our computations here consider. Richard Nixon, John F. Kennedy, and Abraham Lincoln were all elected in elections with a NPV margin larger than we plot in any of our figures, but less than the number of popular votes that separated Bush and Gore. Yet, in 2000, only a few hundred citizen votes divided the decisive state. A hypothetical national recount might have been difficult, but it would not have been needed and would not have occurred. In statistical expectation, only the Electoral College generates hairsbreadth outcomes.
References


<table>
<thead>
<tr>
<th>model name</th>
<th>parametric?</th>
<th>parties</th>
<th>states</th>
<th>shocks</th>
<th>covariates</th>
<th>mean NPV</th>
<th>NPV std. dev.</th>
<th>EC win</th>
<th>inversion rate</th>
<th>mean votes to change NPV outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modern Period: 1988-2016</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M1 52201</td>
<td>P (normal)</td>
<td>Republican Democratic</td>
<td>51</td>
<td>national</td>
<td>(w/DC)</td>
<td>52201 Democratic (w/DC) state</td>
<td>48.97%</td>
<td>2.77%</td>
<td>42.04%</td>
<td>12.28%</td>
</tr>
<tr>
<td>M2 52538</td>
<td>P (t)</td>
<td>Republican Democratic</td>
<td>51</td>
<td>national</td>
<td>(w/DC)</td>
<td>52538 Democratic (w/DC) state race education</td>
<td>48.98%</td>
<td>3.34%</td>
<td>43.40%</td>
<td>12.75%</td>
</tr>
<tr>
<td>NP 51222</td>
<td></td>
<td>Republican Democratic</td>
<td>51</td>
<td>national</td>
<td>(w/DC)</td>
<td>51222 Democratic (w/DC) state education</td>
<td>48.96%</td>
<td>2.85%</td>
<td>45.22%</td>
<td>7.39%</td>
</tr>
<tr>
<td>Post-Reconstruction: 1872-1888</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R1 22202</td>
<td>P (normal)</td>
<td>Republican Democratic</td>
<td>38</td>
<td>national</td>
<td>state</td>
<td>22202 Democratic state</td>
<td>51.03%</td>
<td>3.69%</td>
<td>55.21%</td>
<td>12.63%</td>
</tr>
<tr>
<td>R2 22538</td>
<td>P (t)</td>
<td>Republican Democratic</td>
<td>38</td>
<td>national</td>
<td>state</td>
<td>22538 Democratic state</td>
<td>51.00%</td>
<td>4.76%</td>
<td>54.38%</td>
<td>13.42%</td>
</tr>
<tr>
<td>NP 21222</td>
<td></td>
<td>Republican Democratic</td>
<td>38</td>
<td>national</td>
<td>state</td>
<td>21222 Democratic state</td>
<td>51.02%</td>
<td>3.86%</td>
<td>56.29%</td>
<td>12.61%</td>
</tr>
<tr>
<td>Antebellum: 1836-1852</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A1 12201</td>
<td>P (normal)</td>
<td>Whig Democratic</td>
<td>30</td>
<td>national</td>
<td>state</td>
<td>12201 Democratic state</td>
<td>49.81%</td>
<td>2.94%</td>
<td>47.62%</td>
<td>9.31%</td>
</tr>
<tr>
<td>A2 12538</td>
<td>P (t)</td>
<td>Whig Democratic</td>
<td>30</td>
<td>national</td>
<td>state</td>
<td>12538 Democratic state</td>
<td>49.81%</td>
<td>3.93%</td>
<td>47.65%</td>
<td>9.90%</td>
</tr>
<tr>
<td>NP 11222</td>
<td></td>
<td>Whig Democratic</td>
<td>30</td>
<td>national</td>
<td>state</td>
<td>11222 Democratic state</td>
<td>49.81%</td>
<td>2.99%</td>
<td>48.87%</td>
<td>5.59%</td>
</tr>
</tbody>
</table>

Note: P = parametric in logistic space, with draws from either a normal (after Gelman and King (1994)) or t (after Silver (2016)) distribution; NP = non-parametric (results are very similar to parametric models; details available on request). Mean and variance of National Popular Vote (NPV) in percentage points, for Whigs or Republicans. Model names are as given by Geruso et al. (2020). These non-parametric methods are not named by Geruso et al. (2020); they are Wild (× ± 1) with national and state shocks, homoscedastic with all observed state-year shocks pooled. Five-digit numbers under model names are Geruso et al.’s (2020) simulation code in publicly released replication do files.