Expressive vs. Strategic Voters: An Empirical Assessment*

JÖRG L. SPENKUCH
Northwestern University

February 2018

Abstract
Leading theories of how voters choose between candidates are rooted in two very different paradigms, with starkly different behavioral implications. Exploiting the incentive structure of Germany’s electoral system, I develop a novel set of empirical tests that pit the canonical pivotal voter model against alternative accounts according to which individuals derive expressive utility from supporting their most preferred candidate. The results show that neither paradigm can explain the most-salient features of the data. In addition, the evidence suggests that voters cannot be neatly categorized into sincere and strategic “types.”

*Previous versions of this paper circulated under the titles “(Ir)rational Voters?” and “On the Extent of Strategic Voting.” I have benefitted from comments by Daron Acemoglu, Scott Ashworth, Gary Becker, Daniel Benjamin, Eric Budish, Ethan Bueno de Mesquita, Steve Coate, Tony Cookson, Daniel Diermeier, Tim Feddersen, Alex Frankel, Richard Holden, Antonio Merlo, Matthew Notowidigdo, Nicola Persico, Jesse Shapiro, Yasutora Watanabe, and Richard Van Weelden. I am also indebted to Gabriele Schömel at the office of the Bundeswahlleiter for assistance in acquiring the data used in this paper. Steven Castongia provided excellent research assistance. Financial support from the Beryl W. Sprinkel Fund at the University of Chicago is gratefully acknowledged. All views expressed in this paper as well as any errors are solely my responsibility. Correspondence can be addressed to the author at MEDS Department, Kellogg School of Management, 2211 Campus Dr, Evanston, IL 60208, or by e-mail: j-spenkuch@kellogg.northwestern.edu.
1. Introduction

Understanding how voters choose between different candidates is essential for understanding the democratic process. One influential school of thought casts voters as strategic agents whose choices are driven by the possibility of casting the decisive vote (e.g., Arrow 1951; Austen-Smith and Banks 1999, 2005; Satterthwaite 1975). Others, however, argue that pivot probabilities in large elections are generally so small that tactical considerations cannot possibly affect voters’ decisions (see, e.g., Downs 1957; Green and Shapiro 1994). According to the leading alternative theory, individuals derive expressive utility directly from how they vote (see Brennan and Lomasky 1993; Hamlin and Jennings 2011). Hence, voters sincerely support their most preferred candidate, irrespective of her chances of winning.

How voters behave is ultimately an empirical question. Yet, there exists no consensus. Even the best journals regularly publish articles based on either paradigm (see Gerber et al. 2017 for a similar point). Whether voters are strategic is not only interesting in and of itself, but it is also important for developing accurate theories of electoral politics. Any model in which voters face more than two alternatives requires an assumption about the tactical sophistication of agents, and the conclusions from otherwise identical theories may depend critically on whether voters are taken to be strategic or sincere.

I address the gap in knowledge by devising a novel set of empirical tests that pit the canonical pivotal voter model against its most prominent alternative. The difficulty in disentangling expressive and strategic behavior is that individuals’ true preferences are unobserved. Without any additional structure it is impossible to know whether voters simply selected their most preferred candidate (see Degan and Merlo 2009). Even if one is willing to postulate that some voters did or did not act strategically—think, for instance, of Floridians voting for Ralph Nader in the 2000 presidential election—short of knowing the number of individuals who had an incentive to cast tactical ballots in the first place, it is unclear whether the observed behavior is quantitatively important. Yet, measuring the extent to which actual conduct violates a particular model’s predictions is necessary for assessing the positive content of any theory. All formal models are abstractions from reality and will, therefore, mispredict the behavior of some individuals. Only if the deviations are large would we want to reject a paradigm.

To quantify contradictions of either of the two leading theories of voter behavior, I exploit the incentive structure of parliamentary elections in Germany, where individuals have two votes that are cast simultaneously but counted under different electoral rules. As explained below, these votes need to satisfy basic consistency properties in order to conform to the predictions of either the pivotal voter model or its expressive counterpart.

---

1Compare, for instance, Besley and Coate (1997) with Osborne and Slivinski (1996).
My results imply that neither paradigm provides an adequate description of reality. Specifically, I estimate that, on average, almost two thirds of individuals violate the predictions of the pivotal voter model, while about one third does not behave expressively. Based on this evidence, I conclude that the two most prominent theories of how voters choose between candidates should both be rejected. I also present evidence to suggest that voters cannot be neatly classified into strategic and sincere “types.”

2. Related Literature

There exists a large literature concerned with detecting either expressive or strategic voting. Within this literature, laboratory experiments provide typically convincing evidence of tactical behavior by some, but not all, individuals (e.g., Bouton et al. 2016, 2017; Duffy and Tavits 2008; Eckel and Holt 1989; Esponda and Vespa 2014). How well existing results generalize to large, real-world elections, however, remains unknown.

Coate et al. (2008), for instance, argue that the pivotal voter model is unable to replicate winning margins in Texas liquor referenda. Yet, Reed (1990) and Cox (1994) find that the aggregate distribution of votes in Japan’s multimember districts does conform to the predictions of canonical rational choice theory. The results of Cox (1997) are suggestive of strategic behavior in some electoral systems but not in others. More recently, Fujiwara (2011) shows that in Brazil third-place candidates are more likely to be deserted in races under simple plurality rule than in runoff elections. Pons and Tricaud (2018) document that, in French parliamentary elections, the presence of a third candidate reduces the vote shares of the two front-runners. Their results support the view that many voters have expressive concerns.

Even less is known about the actual extent of expressive and strategic voting. Degan and Merlo (2009) study under what assumptions strategic voting can be detected in observational data. They conclude that the behavior of voters in U.S. national elections is, for the most part, consistent with sincere voting. Spenkuch (2015) exploits a highly unusual by-election in Germany, which allowed a party to gain one seat by receiving fewer votes, to show that about 9% of voters did not behave expressively. Kawai and Watanabe (2013) estimate a fully structural model of voting decisions in Japan’s general election, concluding that between 63% and 85% of voters are strategic. Quantifications like these are necessary to truly evaluate the strategic and expressive paradigms.

Recall, the fundamental difficulty in inferring (non)strategic behavior from naturally occurring data is that voters’ preferences are not observed. A separate strand of the literature tries to circumvent this problem by using survey data on voting decisions and political orientations (see, e.g., Abramson et al. 1992; Blais et al. 2001; Niemi et al. 1993). Estimates of tactical voting in this tradition are often very low—a few percentage points. Wright (1990,
1992), however, points to important survey biases and raises serious doubts about conclusions based on self-reported votes. Alvarez and Nagler (2000) even show that, depending on the survey design, results differ by as much as a factor of seven. As pointed out by Kawai and Watanabe (2013), another reason for why estimates of strategic voting tend to be low is that most analyses do not account for the fact that the vast majority of voters has no incentive to cast strategic ballots. Scholars may, therefore, falsely conclude that strategic behavior is substantively unimportant, even if many individuals would vote tactically if they had an incentive to do so.

3. Germany’s Electoral System

The political landscape in Germany used to be dominated by five major parties: CDU/CSU (conservative), SPD (center-left), FDP (libertarian), Green Party (green/left-of-center), and The Left (far left). The CDU/CSU and the SPD each had nearly as many supporters as the three smaller parties combined. Neither party, however, could govern on the federal level without a coalition partner. Since the mid-1980s, the CDU/CSU’s traditional partner has been the FDP, whereas the SPD has typically entered into coalitions with the Green Party. These “preferences” are well-known to voters.

My empirical strategy exploits the incentive structure of elections to the Bundestag, the lower house of the German legislature. Elections are held every four years according to a mixed-member system with approximately proportional representation. Except for minor modifications, the same system has been in place since 1953. In what follows, I describe the exact set rules as of the 2005 and 2009 parliamentary elections, which are the focus of the analysis below.\(^2\)

As mentioned in the introduction, each voter casts two different votes. The first vote, or candidate vote (Erststimme), is used to elect a constituency representative in each of 299 single-member districts. District representatives are determined in a first-past-the-post system. That is, whoever achieves the plurality of candidate votes in a given district is automatically awarded a seat in the Bundestag. Winners are said to hold direct mandates, and votes cast for any other candidate are discarded.\(^4\)

The arguably more important vote is the list vote (Zweitstimme). It is cast for a party, and the total number of party members who enter the Bundestag is roughly proportional to a party’s share of the national list vote among parties clearing a 5%-threshold. To achieve approximately proportional representation despite potentially lopsided outcomes in the candidate vote, the German electoral system awards list mandates. First, all list votes are ag-

\(^2\)Kiewiet (2013) is an important exception

\(^3\)The description borrows heavily from Spenkuch (2015).

\(^4\)Since the introduction of the two-ballot system in 1953, no independent candidate has won a district.
aggregated up to the national level, and a total of 598 preliminary seats are distributed on a proportional basis. Each party’s allotment is then broken down to the state level and compared with its number of direct mandates in the same state. Whichever number is greater determines how many seats the party will actually receive.

More formally, let $d_{p,s}$ denote the number of districts that party $p$ won in state $s$, and let $l_{p,s}$ be the number of mandates it would have received in the same state under proportional representation. The final number of seats that $p$ retains in $s$ equals $n_{p,s} = \max\{d_{p,s}, l_{p,s}\}$, and its total in the Bundestag is given by $n_p = \sum_s n_{p,s}$ (cf. Appendix B).

If $d_{p,s} < l_{p,s}$, then, in addition to the district winners, the first $l_{p,s} - d_{p,s}$ candidates on $p$’s list are elected as well. Otherwise, only holders of direct mandates receive a seat. Parties are said to win overhang mandates (Überhangmandate) whenever $d_{p,s} > l_{p,s}$. In such cases, the total number of seats in the Bundestag increases beyond 598. Since the total number of mandates awarded under proportional representation, i.e., $\sum_p \sum_s l_{p,s}$, exceeds the number of districts, $\sum_p \sum_s d_{p,s}$, by a factor of two, situations in which $d_{p,s} > l_{p,s}$ are not as common as one might imagine. For instance, relative to its share of the list vote, the CDU/CSU received an additional seven mandates in 2005, whereas the SPD secured nine extra seats. In 2009, there were 24 overhang mandates, 21 of which accrued to the CDU.\footnote{Starting with the 2013 election, the number of list mandates also increases when a party wins more direct than list mandates in a particular state. Given that my analysis focuses on the 2005 and 2009 elections, the 2013 electoral reform has no bearing on the results.}

4. Disentangling Strategic and Expressive Behavior

To see why the German context is useful, first consider the null hypothesis of expressive voting. Under the null, list and candidate votes must reveal voters’ true preferences over parties and candidates, respectively. Thus, if candidates were perfect representatives of their parties, then a candidate’s own vote share in a given district or municipality should be equal to that of her party. Given that preferences over candidates and parties are unlikely to be perfectly correlated, it is unreasonable to expect an exact correspondence between candidate- and party-vote shares in every district. However, after carefully controlling for all candidate characteristics, we should see a relationship that is nearly one-for-one on average. Importantly for my purposes, under the null, the degree to which party and candidate votes track each other cannot be a function of the electoral incentives. After all, the expressive voter model stipulates that, in large elections, individuals simply choose their most preferred candidates and parties, regardless of strategic considerations.

On a basic level, Germany’s electoral system provides straightforward incentives. In years in which all major parties are expected to clear the 5%-threshold, it is optimal for instrumentally rational agents to cast their list vote for whichever party they would like to gain the marginal
seat in parliament. The key incentive associated with the candidate vote is common to all elections under plurality rule. If a particular candidate is known to be out of the race, then instrumentally rational agents can generally do better by voting for somebody else. This is because, in large elections, a tie is orders of magnitudes more likely to involve candidates believed to be front-runners than an underdog (Myerson 2000). As a result, only a subset of candidates can be in the race, and pivotal voters should behave as if their choice set consisted only of serious contenders. Supporting anybody else would amount to a wasted vote, which, by definition, pivotal voters do not do.

It is thus possible to quantify deviations from the canonical pivotal voter model by determining the share of voters who stick with a candidate who is out of the race, conditional on voting for the associated party. Simply put, agents who—for whatever reason—vote for a party whose direct candidate in a particular district is not in contention for victory violate the predictions of the theory if they also vote for said candidate. They play a dominated strategy.

Note, the contrapositive does not necessarily hold. Individuals who cast split tickets may, but need not, be strategic. Thus, without imposing further assumptions, my empirical strategy recovers a lower bound on the extent to which individuals’ choices contradict the predictions of the pivotal voter model.

Figure 1 provides a graphical, high-level summary of my approach. The panels on the left entertain the possibility that all voters cast sincere ballots, as predicted by models of expressive voting. If this were indeed the case, then, after controlling for candidates’ personal appeal, there should be a very close correspondence between party and candidate votes, regardless of whether a particular candidate is viable.

By contrast, the middle panels assume that all voters behave tactically. If individuals’ behavior conformed to the predictions of the pivotal voter model, then only candidates believed to be in contention for victory would receive any votes. For noncontenders, the curve representing the relationship between party and candidate votes ought to be flat.

The panels on the right consider the case in which the electorate is composed of a mixture of expressive and pivotal voters. As before, expressive agents support their preferred candidate, but no pivotal voter chooses a nonviable nominee. Hence, the line relating noncontenders’ share of the candidate vote to their parties’ list votes must have a slope between zero and one. It is this slope that identifies the fraction of individuals who fail to cast a tactical ballot despite having an incentive to do so. In sticking with a nonviable candidate, they deviate

---

6Note, if voters take post-election coalition formation into account or if they live in a state in which overhang mandates may occur, then this party need not be the one they prefer based on their ideological convictions.
from the core prediction of the pivotal voter model. Conversely, one minus the slope estimate measures the share of voters whose observed choices are sensitive to the strategic incentives, and, therefore, violate the expressive paradigm.

Importantly, the results from my approach are informative even if people do not fully understand the algorithm that determines parties’ final number of seats in parliament. The two key requirements for my empirical strategy to go through are that: (i) voters know the district winner is determined by plurality rule, and (ii) voters are not perfectly indifferent between the candidates who remain in the running. If correct, then a pivotal voter would never support a nonviable candidate, but an expressive voter might.

5. Testing the Null Hypothesis of Expressive Voting

To test the two leading theories of how voters choose between candidate, I rely on official results of the 2005 and 2009 federal elections. Restricting attention to 2005 and 2009 is useful because all important parties were widely expected to clear the national 5%-threshold. Critical for my approach is that the data are disaggregated by polling precinct (Wahlbezirk). In Germany, precincts are the smallest administrative units in which votes are counted. Each precinct is fully contained within an electoral district and associated with one polling station. As of 2009, there were 299 electoral districts and almost 89,000 precincts, which handled about 615 votes on average. Since races take place at the district level, these data allow me to exploit within-candidate variation, thereby conditioning on all observable as well as unobservable candidate characteristics.

Recall, under the null hypothesis of expressive voting, list and candidate votes must reveal voters’ true preferences over parties and candidates, respectively. Thus, after carefully controlling for candidate quality, precinct-level candidate- and party-vote shares should track each other almost one-for-one. The results in Table 1 show that this prediction holds in situations in which voters have no strategic incentive to abandon their favored party’s nominee, but not when tactical voting might be beneficial.

---

7Karp (2006) reports that the most common source of confusion among voters is that they overestimate the importance of the candidate vote for a party’s total number of mandates.

8Arguably, a small preference for one candidate over another is realistic. After all, there is a small chance that the outcome in a particular district will affect the aggregate distribution of seats. Even conditional on the aggregate distribution of seats, voters may care about sending “good” local representatives to parliament, especially since representatives elected via the candidate vote are likely to become members of committees that allow them to serve their geographically based constituency (Stratmann and Baur 2002). Consistent with this argument, in the 2009 German Longitudinal Election Survey (GLES), almost three out of four respondents said that it is either “important” or “very important” that candidates represent the interests of their home districts.

9For instance, more than 90% of adults sampled in the 2009 pre-election survey of the German Longitudinal Election Study expected the FDP and Green Party to receive more than five percent of the list vote.

10For descriptive statistics, see Appendix Table A.1.
The ordinary least squares estimates in the table are based on the following econometric model

\begin{equation}
\begin{aligned}
v_{k,r,t}^C = \chi_{m,k,t} + \phi v_{k,r,t}^L + \epsilon_{k,r,t},
\end{aligned}
\end{equation}

where \(v_{k,r,t}^C\) denotes contestant \(k\)'s share of the candidate vote in precinct \(r\) during election year \(t\), and \(v_{k,r,t}^L\) is her party's share of the list vote in the same precinct. To allow for almost arbitrary forms of autocorrelation in the residuals as well as for correlation within and across districts, standard errors are clustered by state. Going from the left of the table to the right, the set of fixed effects grows steadily. The most inclusive specification contains \(\chi_{m,k,t}\), a municipality- and year-specific candidate fixed effect. I, therefore, control nonparametrically for the appeal of individual candidates as perceived by the voters in a given town or village. The parameter of interest in equation (1) is \(\phi\). It measures the share of party supporters who also vote for the associated candidate.

The upper panel of Table 1 restricts attention to the eventual winner and runner-up of each race. Voters who support the parties of these candidates have no strategic reason to cast split ballots. After all, surprises in large-scale elections are very rare, and partisans have no incentive to desert someone they should have believed to be in contention for victory. Consistent with the idea that party votes are heavily correlated with individuals' preferences over candidates, the results in column (4) show that an extra list vote results in about .989 additional candidate votes. Although the point estimate is quite precise, it is not possible to statistically rule out that it is exactly equal to one \((p = .544)\).

By contrast, the middle panel focuses on candidates who finished in third place or worse. At least some individuals who supported the parties associated with these candidates had a strategic incentive to vote for someone else; and my estimates show that about one in three did so.

In the bottom panel of Table 1, I test the null hypothesis of expressive voting by estimating the difference between the settings above, i.e., between situations in which voters do and do not have a strategic incentive to cast split ballots. Specifically, I estimate the augmented model

\begin{equation}
\begin{aligned}
v_{k,r,t}^C = \chi_{m,k,t} + \phi v_{k,r,t}^L + \delta v_{k,r,t}^L \times 1 \{k \geq 3\} + \epsilon_{k,r,t},
\end{aligned}
\end{equation}

where \(1 \{k \geq 3\}\) is an indicator variable for whether candidate \(k\) finished third or worse. One can dismiss the null that all voters behave expressively if it is possible to statistically reject \(H_0 : \delta = 0\). Clearly, \(\delta\) is negative and statistically significant in all specifications \((p < .001)\). In fact, taking the coefficient in column (12) at face value, the choices of almost a third of
individuals are inconsistent with the expressive paradigm. Based on this evidence, I conclude that the expressive voter model should be rejected.

6. Quantifying Deviations from the Pivotal Voter Model

6.1. Econometric Details

In order to shed light on how frequently the predictions of the canonical pivotal voter model are violated, I estimate two related empirical specifications. The first one identifies the share of voters whose choices deviate from the theory by restricting attention candidates who, by any reasonable standard, were clearly not in contention for victory. For this set of candidates, I estimate

\[
v_{k,r,t}^C = \alpha_{m,k,t} + \lambda v_{k,r,t}^L + \epsilon_{k,r,t},
\]

where all symbols are as defined above. The parameter of interest is \( \lambda \). Since I restrict attention to nonviable candidates only, \( \lambda \) denotes the fraction of party supporters who stick with the associated candidate despite her being out of the race. As long as there is no heterogeneity in \( \lambda \), it is irrelevant if the set of candidates who are included in the sample used to estimate equation (3) is chosen too conservatively, i.e., if one fails to include some candidates who were also believed to be nonviable. Settling on a too narrowly defined set of noncontenders would only come at a loss of statistical power, but it would not prevent consistent estimation of \( \lambda \).

If there is heterogeneity in \( \lambda \) and if this heterogeneity is systematically correlated with who remains in contention for victory, then restricting attention to supporters of parties that field candidates who trail far behind might lead to unrepresentative estimates. The second (and, therefore, preferred) empirical specification addresses this issue by adopting a data-driven approach to classifying contestants.\(^{11}\)

Specifically, drawing from the literature on structural breaks in time series data, I estimate a cutoff value, \( \kappa \), separating candidates into contenders and noncontenders. Since theory predicts that there are always at least two candidates in the race, even if one of them trails far behind (Myerson 2000; Myerson and Weber 1993), I classify candidate \( k \) as a contender if, and only if, the difference in support for her party and the district’s second-most-popular one is less than \( \kappa \) percentage points. I, therefore, allow for equilibria in which voters perceive three or more candidates to be in contention for victory, as in Cox (1994).

\(^{11}\)Pre-election surveys in Germany are too small to derive reliable estimates of voters’ expectations. For instance, in only 50 electoral districts did the German Longitudinal Election Study—the best available data source—survey more than 15 adults prior to the 2009 elections.
In symbols, the second empirical model is given by

\[ v_{C}^{k;r;t} = m_{k;t} + v_{L}^{k;r;t} - h v_{L;2nd}\text{d}t + \gamma v_{L}^{k;r;t} - \epsilon_{k;r;t}. \]

Here, \( v_{L}^{k;d;t} \) denotes the list-vote share of candidate \( k \)'s party in district \( d \), and \( v_{L;2nd}\text{d}t \) is that of the second-most-popular one. If (4) is correctly specified, then searching for the value of \( \kappa \) that maximizes the \( R^2 \) yields a super-consistent estimate of the true break point (Hansen 2000). Moreover, under the null hypothesis that such a point exists, estimates of the model’s other parameters are normally distributed, and standard errors need not be adjusted for sampling variability in the location of the break (Bai 1997).

Although intuitively appealing, there is no guarantee that this method classifies all candidates correctly. For this reason, in Appendix C, I report results from a series of robustness checks, demonstrating that my conclusions are qualitatively robust to more than 25 alternative assumptions on how voters form beliefs about which candidates are in contention for victory.

6.2. Results

Focusing on nominees of the five major parties, Table 2 presents results for both empirical specifications. The upper panel is based on the first one, restricting attention to candidates who trailed the runner-up by more than ten percentage points. The lower panel is based on equation (4). Within each panel, the first row presents estimates of the share of voters who stick with a party’s candidate despite her having virtually no chance of winning.

Controlling for the appeal of individual candidates, estimates of \( \lambda \) range from .613 to .696 and are quite precise.\(^{12}\) Moreover, it is worth noting that the evidence from both specifications lines up remarkably well. Taken at face value, the results indicate that (at least) 61% of voters do not behave in accordance with the canonical pivotal voter model. Thus, as a general, positive theory of voter behavior, it should be dismissed.\(^{13}\)

7. Comparative Statics and Learning

In sum, neither of the two leading theories of how voters choose between different candidates accords well with real-world behavior. An important question is, therefore, whether an alter-

---

\(^{12}\) In Appendix C, I show that local party strength is essentially uncorrelated with the estimated share of voters who stick with the respective candidate. It is, therefore, not the case that the results in Table 2 are driven by parties fielding weak candidates in districts where they expect to perform poorly.

\(^{13}\) In Appendix F, I estimate a structural model of voting decisions. At the cost of imposing additional assumptions, this model allows me to estimate the share of instrumentally rational voters, rather than obtaining bound on the fraction of individuals whose conduct is (in)consistent with the pivotal voter model. The results accord well with the reduced-form evidence. I also assess the impact of strategic voting on the distribution of seats in parliament.
native theory is needed to rationalize the most-salient features of the data, or whether each model explains the behavior of some voters, but not others. That is, there may be different “types” of agents, which are well-described by the pivotal and the expressive voter model, respectively.

One suggestive piece of evidence against a theory with immutable types is that voters appear to learn to cast strategic candidate votes over time. To provide evidence of learning, I rely on the German Reunification as a natural experiment.\textsuperscript{14} Although the German Democratic Republic (GDR) held regular, formal elections to the \textit{Volkskammer} (People’s Chamber), they were effectively meaningless. East Germans could only choose from candidates on a single list controlled by the Socialist Unity Party (SED), and it was customary to cast one’s ballot in public, simply accepting all nominated candidates. In stark contrast, citizens of the Federal Republic of Germany had the opportunity to participate in free elections since 1949, and, from 1953 on, under a two-ballot system almost identical to the current one. They had thus more than 40 years of democratic experience by the time the GDR joined the West.

The first parliamentary elections in unified Germany were held in December of 1990 and were subject to (essentially) the same rules that had previously been used in the West and that continued to be in place thereafter. If experience does, indeed, lead voters to behave more strategically, then we would expect large initial differences in the share of agents whose choices are at odds with the pivotal voter model, which should disappear over time.

This prediction is borne out in Figure 2. For each election since 1990, the figure plots the estimated difference between East and West German voters. Negative values indicate more violations of the pivotal voter model among residents of the former GDR.\textsuperscript{15} The results show that just two months after reunification, East Germans were almost 16 percentage points more likely to stick with a noncontender than their Western counterparts. By 2005, however, the gap had vanished. Although none of the point estimates is very precise, one can nevertheless reject the null hypothesis of a constant difference ($p < .01$).

In order to speak more directly to the possibility that there may be sincere and strategic types of voters, I present comparative statics based on the regression models in equations (2) and (4). If there are, indeed, strategic and expressive types, then we would expect to find an approximately constant share of individuals who violate each theory, irrespective of other circumstances. The reason is that strategic (expressive) types would \textit{always (never)} abandon their favorite candidate if she happened to be nonviable. By contrast, systematic

\textsuperscript{14}For related evidence on learning to vote strategically among U.S. Senators, see Spenkuch et al. (2018).
\textsuperscript{15}The specification on which the estimates are based is akin to equation (4) but allows for different slopes and cutoff values in East and West Germany. A qualitatively similar picture would emerge if one were to restrict the cutoff to be the same in both regions.
heterogeneity in the extent of (non)strategic behavior rules out a type-based explanation and provides new evidence that helps to narrow down the set of alternative theories.

Table 3 compares estimates across a number of different settings. The first set of results demonstrates that the extent to which observed behavior violates the pivotal voter model depends on who remains in contention for victory. That is, conditional on voting for a party whose candidate is nonviable, voters are about 25 percentage points less likely to stick with a noncontender when the candidate of an allied party is still in the race than when faced with the choice between two evils, i.e., less palatable alternatives.\textsuperscript{16} A test for equality of coefficients rejects the null hypothesis of equal point estimates at the 1\%-level.

Moreover, distinguishing between districts in which the race for the direct mandate ended up being close and those in which it was not, violations of the pivotal voter model appear to have been less prevalent in the former—though the difference is not statistically significant—and disaggregating the data by election year shows that desertion of noncontenders was significantly more common in 2005 than in 2009 ($p < .001$).

The latter finding may not be surprising. The 2005 election followed a failed motion of confidence that triggered the dissolution of the Bundestag and was widely perceived to be a “critical election,” in which differences between parties and, therefore, the stakes were significantly higher than usual (see, e.g., Korte 2009).\textsuperscript{17} In line with the results in Table 3, official statistics show a substantially larger fraction of split tickets in 2005 and an approximately 7 percentage points higher turnout than in 2009 (Bundeswahlleiter 2006, 2010). The change in turnout, however, is too small to account for the entire difference in $\lambda$. Estimating the share of individuals who did not behave according to the pivotal voter model for each municipality-year combination separately and regressing the resulting $\lambda_{m,t}$ on turnout in the respective village in the same year yields a point estimate of $-.698$ (with a standard error of $.173$). A 7 percentage point increase in turnout would, therefore, be predicted to lead to an approximately 4.9 percentage points lower fraction of nonstrategic agents. Although the available evidence does suggest that inframarginal voters are considerably more likely to deviate from the pivotal voter model than marginal ones, the difference in turnout is far too small to cause a near 50\% change in the estimated extent to which the theory is violated. Some simple back-of-the-envelope calculations show that this conclusion holds even if every

\textsuperscript{16}The following parties are defined as allies: CDU and FDP, as well as SPD and Green Party. The results are qualitatively similar if supporters of The Left are assumed to consider SPD candidates to be close substitutes.

\textsuperscript{17}Campaigning to stay in office, Chancellor Schröder and his SPD–Green coalition promised to undo some of their unpopular labor market and welfare reforms while raising taxes on the rich. In stark contrast, led by Angela Merkel, the conservative–libertarian bloc sought to further increase the pace and scope of deregulation, slashing income taxes and public spending in the process.
single marginal voter were strategic.\footnote{In 2005, about 13.3 million voters chose a party whose direct candidate is estimated to be “out of the race,” and almost half of them also abandoned the respective nominees. Suppose that every single one of the approximately 4 million additional voters in 2005 chose a party whose direct candidate was not in contention for victory and deserted the respective direct candidate. If this were, indeed, the case, then about 70% of the inframarginal voters, i.e. 6.5 out of 9.3 million, would not have behaved in accordance with the pivotal voter model. Even under these extreme assumptions, the difference in turnout cannot account for the entire change in $\lambda$.}

Importantly, the results in Table 3 are at odds with a type-based explanation. A theory in which a particular voter is either strategic or sincere can neither explain why defection of weak candidates is more common when the stakes are higher nor why it depends on which other candidates remain in contention for victory. The evidence in this section suggests that individuals can learn to vote strategically and that they trade off tactical and expressive considerations according to the electoral circumstances.

8. Conclusion

The scientific method requires that formal theories be rigorously tested and, if necessary, rejected. This paper develops a novel set of empirical tests in order to directly pit the canonical pivotal voter model against the most prominent alternative according to which individuals derive expressive utility from supporting their most preferred candidate. The results indicate that about two-thirds of individuals deviate from the predictions of the pivotal voter model, while almost one third do not behave expressively.

In light of a plethora of anecdotal evidence, one might not have expected literally all voters to abide by a particular theory. Nevertheless, the results above are noteworthy because they carefully measure the extent to which voter behavior violates either of two leading paradigms. Without proper quantification, there would be no basis to conclude that the gap between actual conduct and theory is so large that neither paradigm is empirically tenable. The evidence further suggests that voters cannot be neatly categorized into strategic and sincere types. Instead, individuals’ tendency to vote tactically varies substantially with the electoral circumstances. A promising alternative theory of voter behavior may be one in which individuals \textit{endogenously} decide whether to act strategically.

References


Figure 1: Theoretical Predictions under Expressive and Strategic Voting

A. Expressive Voters

B. Strategic Voters

C. Mixture of Expressive and Strategic Voters

I. Noncontender

II. Contender
Figure 2: Difference in the Incidence of "Wasted Votes" between East and West Germany, 1990–2009

Notes: Figure shows the percentage point difference in the incidence of nonstrategic voting between East and West Germany for each federal election from 1990 to 2009 as well as the associated 95%-confidence intervals. Negative values indicate more ballot combinations that violate the pivotal voter model among residents of the former GDR. The null hypothesis of a constant difference across all years can be rejected at the 1%-significance level, and that of an equal difference in 1990 and 2009 is rejected at the 1%-level as well.
Table 1: Testing the Null Hypothesis of Expressive Voting

A. Voters with No Strategic Incentives to Cast Split Ballots

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Share of Candidate Vote</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
</tr>
<tr>
<td>Share of Party Vote (φ)</td>
<td>1.061</td>
</tr>
<tr>
<td></td>
<td>(.010)</td>
</tr>
<tr>
<td>H₀: φ=1 [p-value]</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>

Fixed Effects:
- Party: Yes No No No
- Candidate: No Yes No No
- Candidate × Year: No No Yes No
- Candidate × Municipality × Year: No No No Yes

R-Squared: .903 .946 .950 .968
Number of Observations: 354,642 354,642 354,642 354,642

B. Voters with Strategic Incentives to Cast Split Ballots

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Share of Candidate Vote</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(5)</td>
</tr>
<tr>
<td>Share of Party Vote (φ)</td>
<td>.798</td>
</tr>
<tr>
<td></td>
<td>(.026)</td>
</tr>
<tr>
<td>H₀: φ=1 [p-value]</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>

Fixed Effects:
- Party: Yes No No No
- Candidate: No Yes No No
- Candidate × Year: No No Yes No
- Candidate × Municipality × Year: No No No Yes

R-Squared: .813 .888 .897 .934
Number of Observations: 527,419 527,419 527,419 527,419

C. Difference Between Settings with and without Strategic Incentives

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Share of Candidate Vote</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(9)</td>
</tr>
<tr>
<td>Share of Party Vote (φ)</td>
<td>1.040</td>
</tr>
<tr>
<td></td>
<td>(.010)</td>
</tr>
<tr>
<td>Share of Party Vote × Incentive to Cast Split Ballot (δ)</td>
<td>-.221</td>
</tr>
<tr>
<td></td>
<td>(.032)</td>
</tr>
<tr>
<td>H₀: δ=0 [p-value]</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>

Fixed Effects:
- Party: Yes No No No
- Candidate: No Yes No No
- Candidate × Year: No No Yes No
- Candidate × Municipality × Year: No No No Yes

R-Squared: .965 .980 .982 .989
Number of Observations: 882,061 882,061 882,061 882,061

Notes: Entries are coefficients and standard errors from estimating equations (1) (upper two panels) and (2) (lower panel) by ordinary least squares. The upper panel restricts the sample to candidates who finished first or second, giving supporters of the associated parties no strategic incentives to cast split ballots. The middle panel considers only candidates who finished third or worse, meaning that at least some supporters of the associated parties had a strategic incentive to cast split ballots. The lower panel pools the data from both settings. Heteroskedasticity robust standard errors are clustered by state and reported in parentheses. To account for the small number of clusters, reported p-values are based on the wild bootstrap procedure suggested by Cameron et al. (2008) with 10,000 iterations. See the Data Appendix for the precise definition and source of each variable.
### Table 2: Quantifying Deviations from the Pivotal Voter Model

#### A. Candidates Trailing Far Behind the Runner-Up

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Share of Candidate Vote</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Share of Party Vote (λ)</td>
<td>0.682</td>
<td>0.670</td>
<td>0.632</td>
<td>0.613</td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td>(0.010)</td>
<td>(0.014)</td>
<td>(0.016)</td>
</tr>
</tbody>
</table>

\[ H_0: \lambda = 1 \quad [p\text{-value}] \quad < 0.001 \quad < 0.001 \quad < 0.001 \quad < 0.001 \]

\[ H_0: \lambda = 0 \quad [p\text{-value}] \quad < 0.001 \quad < 0.001 \quad < 0.001 \quad < 0.001 \]

**Fixed Effects:**
- **Party:** Yes No No No
- **Candidate:** No Yes No No
- **Candidate × Year:** No No Yes No
- **Candidate × Municipality × Year:** No No No Yes

**R-Squared:** 0.717 0.816 0.832 0.885

**Number of Observations:** 463,544 463,544 463,544 463,544

#### B. All Candidates

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Share of Candidate Vote</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Share of Party Vote</td>
<td>0.765</td>
<td>0.696</td>
<td>0.657</td>
<td>0.656</td>
</tr>
<tr>
<td>× Noncontender (λ)</td>
<td>(0.022)</td>
<td>(0.021)</td>
<td>(0.019)</td>
<td>(0.026)</td>
</tr>
</tbody>
</table>

\[ Share of Party Vote \times Contender (γ) \]

\[ (0.010) \quad (0.010) \quad (0.012) \quad (0.021) \]

**Noncontender**

\[ 3.664 \quad -3.887 \]

\[ (0.433) \quad (0.614) \]

**Contender**

\[ 6.477 \quad -7.42 \]

\[ (0.717) \quad (0.140) \]

**Structural Break**

\[ 0.021 \quad 0.065 \quad 0.064 \quad 0.023 \]

\[ H_0: \lambda = 1 \quad [p\text{-value}] \quad < 0.001 \quad < 0.001 \quad < 0.001 \quad < 0.001 \]

\[ H_0: \lambda = 0 \quad [p\text{-value}] \quad < 0.001 \quad < 0.001 \quad < 0.001 \quad < 0.001 \]

**Fixed Effects:**
- **Party:** Yes No No No
- **Candidate:** No Yes No No
- **Candidate × Year:** No No Yes No
- **Candidate × Municipality × Year:** No No No Yes

**R-Squared:** 0.965 0.980 0.982 0.989

**Number of Observations:** 882,061 882,061 882,061 882,061

**Notes:** Entries are coefficients and standard errors from estimating equations (3) (upper panel) and (4) (lower panel) by ordinary least squares. The upper panel restricts the sample to candidates who finished more than 10 percentage points behind the one in second place, whereas the lower panel includes all candidates. Heteroskedasticity robust standard errors are clustered by state and reported in parentheses. To account for the small number of clusters, reported \( p \)-values are based on the wild bootstrap procedure suggested by Cameron et al. (2008) with 10,000 iterations. See the Data Appendix for the precise definition and source of each variable.
Deviations from Violations of the Sample Restriction Expressive Voting Pivotal Voter Model

<table>
<thead>
<tr>
<th>Sample Restriction</th>
<th>Deviations from Expressive Voting</th>
<th>Violations of the Pivotal Voter Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>.326</td>
<td>.656</td>
</tr>
<tr>
<td></td>
<td>(.018)</td>
<td>(.026)</td>
</tr>
<tr>
<td>By Availability of Close Substitute:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Allied Party's Candidate in the Race</td>
<td>.399</td>
<td>.556</td>
</tr>
<tr>
<td></td>
<td>(.039)</td>
<td>(.024)</td>
</tr>
<tr>
<td>Only Rival Parties' Candidates in the Race</td>
<td>.123</td>
<td>.817</td>
</tr>
<tr>
<td></td>
<td>(.012)</td>
<td>(.014)</td>
</tr>
<tr>
<td>By Difference between Winner and Runner-Up:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 1%</td>
<td>.342</td>
<td>.606</td>
</tr>
<tr>
<td></td>
<td>(.056)</td>
<td>(.034)</td>
</tr>
<tr>
<td>1% and 5%</td>
<td>.312</td>
<td>.621</td>
</tr>
<tr>
<td></td>
<td>(.048)</td>
<td>(.028)</td>
</tr>
<tr>
<td>&gt; 5%</td>
<td>.332</td>
<td>.662</td>
</tr>
<tr>
<td></td>
<td>(.037)</td>
<td>(.026)</td>
</tr>
<tr>
<td>By Year:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005</td>
<td>.399</td>
<td>.488</td>
</tr>
<tr>
<td></td>
<td>(.018)</td>
<td>(.016)</td>
</tr>
<tr>
<td>2009</td>
<td>.271</td>
<td>.726</td>
</tr>
<tr>
<td></td>
<td>(.029)</td>
<td>(.021)</td>
</tr>
</tbody>
</table>

Notes: Entries in the are coefficients and standard errors on $\delta$ in equation (2) (left column) and $\lambda$ in equation (4) (right column), using different subsamples of the data. The respective restriction is indicated on the left of each row. All specifications control for candidate-municipality-year fixed effects. See the Data Appendix for the precise definition and source of each variable.
Online Appendix to
“Expressive vs. Strategic Voters: An Empirical Assessment”

Contents

A Theoretical Underpinnings 3
  A.1 A Formal Model with Expressive and Strategic Voters . . . . . . . . . . . . . 3
  A.2 Equilibrium . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 4
  A.3 Mapping Theory into Data . . . . . . . . . . . . . . . . . . . . . . . . . . . 5

B Calculating a Party’s Number of Seats 6

C Robustness and Sensitivity Checks 7

D Strategic Voting under Proportionality Rule? 9

E Data Appendix 9

F Structural Analysis 10
  F.1 Adding Structure . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 11
  F.2 Results and Counterfactual Experiments . . . . . . . . . . . . . . . . . . . 14
  F.3 Numerical Methods . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 15

References 17

List of Figures
A.1 Distribution of Direct Mandates in the 2005 and 2009 Federal Elections . . . 19
A.2 Relationship between List and Candidate Votes for Candidates Trailing Far
    Behind . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 20
A.3 Observed vs. Predicted Distribution of Votes based on Structural Analysis,
    2009 Federal Election . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 21
A.4 Counterfactual Seat Distributions in the 17th Bundestag . . . . . . . . . . 22

List of Tables
A.1 Summary Statistics for Electoral Precincts . . . . . . . . . . . . . . . . . . . 23
A.2 Sensitivity Analysis Using Alternative Classifications of Contenders . . . . 24
A.3 Additional Sensitivity and Robustness Checks . . . . . . . . . . . . . . . . . 25
A.4 Candidate–List–Vote Gradient among Candidates Trailing Far Behind, by Distance between the First- and Second- as well as Second- and Third-Ranked Candidates .................................................. 26
A.5 Initial Distribution of Fractional Mandates by Party, 2005 & 2009 Federal Elections ....................................................................................................................... 27
A.6 Actual vs. Predicted Ranking of Candidates, Structural Analysis of 2009 Federal Elections .............................................................................................................. 28
A.7 Voters’ Partial Preference Orderings ..................................................................... 29
A.8 Distribution of District Winners and Expressive and Strategic Voting, Structural Analysis ................................................................................................................. 30
Appendix A: Theoretical Underpinnings

To show standard theory maps into my data and, thereby, frame the empirical approach in the main text, I introduce a simple model of voting under plurality rule. Instead of considering the complete decision problem associated with list and candidate votes in Germany, I focus on the race for a direct mandate in one electoral district, i.e., on a single subgame. The model is a straightforward extension of Myerson and Weber (1993) with the addition of expressive voters, stochastic turnout, and endogenous pivot probabilities.

A.1. A Formal Model with Expressive and Strategic Voters

Let the set of candidates be denoted by $K = \{1, 2, \ldots, k\}$. Members of the electorate (simultaneously) cast single nontransferable votes, and the contestant with the highest vote total is declared the winner of the election. Ties are broken by the flip of a fair coin.

For simplicity, assume that voters behave either expressively or tactically, $\theta \in \{e, t\}$. Expressive voters always choose their most preferred candidate, whereas tactical agents act based on personal preferences as well as their beliefs about the actions of other players in the game. Although the data suggest that voters choose whether to act strategically, I abstract from this complication in order to focus on how to quantify the extent of each behavior in a given election. The share of agents who behave expressively is given by $\lambda \in [0, 1]$.

Each voter has strict preferences over candidates summarized by a vector $u = (u_1, \ldots, u_k)$ in some finite set $U \subset \mathbb{R}^k$, where $u_i$ is the expected utility from candidate $i$ winning the district (conditional on the expected outcomes in all other districts as well as the expected realization of the national list vote). $f(u)$ denotes the fraction of individuals with a particular preference profile. That is, $f$ is a probability distribution over $U$. A voter’s type is defined as the tuple $(u, \theta) \in I \equiv U \times \{e, t\}$. For simplicity, $u$ and $\theta$ are assumed to be independent random variables.

Agents know their own type, but are uncertain about the number of other players in the game. This captures the idea that real world elections are characterized by substantial uncertainty about turnout, and that voters are typically not aware of everybody else’s identity. Following Myerson (2000, 2002), assume that the total number of voters is a random variable drawn from a Poisson distribution with mean $n < \infty$. $n$, $f$, as well as $\lambda$ are common knowledge.

As mentioned above, strategic agents maximize expected utility taking the behavior of others into account. More specifically, tactical voters choose candidate $k$ only if doing so maximizes

$$\pi(k, \pi|u, t) = \frac{1}{2} \sum_{k' \in K \setminus \{k\}} \pi(k, k') [u_k - u_{k'}],$$

where $\pi = (\pi(k, k'))_{k, k' \in K}$ denotes players’ common beliefs about the probability of casting a pivotal vote.\(^1\)

\(^1\)To see that (1) follows from expected utility maximization note that an individual’s vote affects his payoff
By contrast, expressive players always select their most preferred contestant. They maximize the utility function:

$$u(k, \pi|u,e) = u_k.$$ 

A.2. Equilibrium

Let $\sigma(k|u, \theta)$ denote voters’ strategies. That is, $\sigma : \mathcal{I} \rightarrow \Delta(K)$ specifies the probability that a type $(u, \theta)$ voter casts a ballot for candidate $k$. In equilibrium it must be the case that, for all $(u, \theta) \in \mathcal{I}$,

$$\sigma(k|u, \theta) > 0 \text{ only if } k \in \arg\max_{k' \in K} \mu(k, \pi|u, \theta).$$

Given $\sigma$, realized vote totals, $v = (v(k))_{k \in K}$, are random variables with means, $\mu = (\mu(k))_{k \in K}$, equal to

$$\mu(k) = n \sum_{u \in U} [\lambda \sigma(k|u,e) + (1 - \lambda) \sigma(k|u,t)] f(u).$$

From the Poisson assumption it follows that the elements of $v$ are independently distributed (see Myerson 1998 for a proof), which allows the probability of casting the pivotal vote to be expressed in closed form. More specifically, not knowing the exact number of players in the game, the ex ante probability of candidate $k$ being tied for first or one vote behind $k_0$ is given by

$$\pi(k, k_0) = \lambda \sum_{\tau=1}^{\infty} \left[ \left( \sum_{\tau'=\tau}^{\tau+1} \psi(v = \tau|\mu(k)) \right) \prod_{k'' \in K \setminus \{k, k_0\}} \sum_{\tau''=0}^{\tau-1} \psi(v = \tau''|\mu(k'')) \right],$$

where $\psi(v = \tau|\mu(k))$ denotes the probability of a Poisson random variable $v$ with parameter $\mu(k)$ being equal to $\tau$.\(^2\)

**Definition:** Given the Poisson game $\Gamma(K, \mathcal{I}, n, f, \lambda)$, a voting equilibrium consists of a strategy function $\sigma$ satisfying, for all $(u, \theta) \in \mathcal{I}$,

(i) $\sigma(k|u, \theta) \geq 0 \quad \forall k \in K,$
(ii) $\sum_{k \in K} \sigma(k|u, \theta) = 1,$ and
(iii) $\sigma(k|u, \theta) > 0$ only if $k \in \arg\max_{k' \in K} \mu(k', \pi|u, \theta);$  

as well as a set of beliefs such that

only if it changes the outcome of the election, i.e. if two candidates are either tied for first or one vote apart. If candidate $k$ and $k'$ are tied, then voting for the former results in an expected utility gain of $u_k - \frac{1}{2}(u_k + u_{k'})$. If $k$ is one vote behind $k'$, then choosing $k$ changes payoffs by $\frac{1}{2}(u_k + u_{k'}) - u_{k'}$, which is the same as the previous expression. Summing over all candidate pairs and weighting by $\pi$ gives (1).\(^2\)

As is typical in the literature on strategic voting, the probability of three-way ties is assumed to be negligible.
(iv)  \( \bar{\pi}(k, k') = \pi(k, k') \quad \forall k, k' \in K. \)

**Proposition 1:** The set of voting equilibria is always non-empty.

**Proof:** See Spenkuch (2013).

To get a sense of what equilibrium play looks like note that strategic voters’ utility function is homogenous in \( e \). Hence, tactical voting decisions are determined by the relative—not absolute—size of perceived pivot probabilities. From the magnitude theorem in Myerson (2000) it follows that some pivot probabilities are going to be several orders of magnitude larger than others; although for large electorates all elements of \( \pi \) will be very close to zero. That is, as \( n \to \infty \) most pivot probabilities become infinitesimal relative to, at most, a few remaining ones. Intuitively, this is because homogeneity of the utility function implies that \( \bar{\pi}(k, k') \) can be rewritten as the probability of \( k \) and \( k' \) running neck-and-neck ahead of all other contestants, conditional on the election being tied in the first place. Such a tie, however, is substantially more likely to involve the two front-runners than an underdog. Hence, almost all of the probability mass must be concentrated in one or two candidate pairs, which gives rise to the following corollary.

**Corollary 1:** In large elections only a subset of candidates will be in the race, and strategic voters behave as if choosing only among those candidates who are believed to be serious contenders.

Since tactical agents become more inclined to select a particular candidate as they form favorable beliefs about her being in contention for victory—say, because her standing in pre-election polls improves, or due to campaign activities that manipulate voters’ perception of candidate viability—the model above exhibits the potential for bandwagon effects and self-fulfilling prophecies (Simon 1954). In general there may be multiple equilibria, and any candidate that is not a Condorcet loser may be the sole likely winner under plurality rule (cf. Myerson and Weber 1993). Thus, without further refinement the model makes no prediction about the set of candidates who will be “in the race,” i.e., which of the possible equilibria is being observed by the econometrician.

A.3. **Mapping Theory into Data**

For identifying the share of strategic voters, however, this is inconsequential. The key takeaway from the discussion above is that it is not optimal for strategic agents to vote for a candidate who is out of the race, i.e., for whom tie probabilities are orders of magnitude smaller than for other candidates. It must, therefore, be the case that \( \sigma(k|u, t) = 0 \) for all candidates who are noncontenders, whereas \( \sigma(k|u, e) \) equals either 0 or 1, depending on whether type \( (u, e) \) agents prefer \( k \) over every other contestant. The model above, therefore, directly maps into Figure 1 in the main text.

To see this, note that given these strategies and focusing on noncontenders, i.e., on candidates believed to be “out of the race,” equation (2) simplifies to

\[
\mu(k) / n = \lambda \sum_{\bar{u} \in \{ u \in U | u_k \geq u_{k'}, \forall k' \}} f(\bar{u}).
\]
The left-hand side of this expression denotes \( k \)'s share of the candidate vote, whereas the right-hand side equals the share of expressive voters multiplied by the fraction of individuals who actually had a strategic incentive to abandon \( k \).\(^3\)

I observe \( \mu(k)/n \) in the data and I am interested in estimating \( \lambda \). If \( \sum_{\tilde{u} \in \{u \in U| u_k > u_{k'} \}} f(\tilde{u}) \) was also observeable, then I could infer \( \lambda \) simply by regressing \( \mu(k)/n \) on \( \sum_{\tilde{u} \in \{u \in U| u_k > u_{k'} \}} f(\tilde{u}) \). In practice, I use the list vote share of \( k \)'s party instead of \( \sum_{\tilde{u} \in \{u \in U| u_k > u_{k'} \}} f(\tilde{u}) \) (and I control for candidates’ valence using fixed effects). Voters supporting the party of candidate \( k \) can, conditional on \( k \) being out of the race, always do better by voting for somebody else. They have, therefore, an incentive to cast split ballots, and \( \lambda \) measures the share of individuals who fail to do so and thereby violate the pivotal voter model.

Under the additional assumption that preferences over parties and candidates are heavily correlated (conditional on candidates’ idiosyncratic appeal), the list vote is, in fact, a direct measure of \( \sum_{\tilde{u} \in \{u \in U| u_k > u_{k'} \}} f(\tilde{u}) \), which implies that my empirical approach identifies \( \lambda \) exactly. Although I find such an assumption plausible, I emphasize that it is not required. Without it, my approach recovers a lower bound on the extent of nonstrategic voting. The reason is that some supporters of \( k \)'s party will cast split ballots for nonstrategic reason. These vote combinations are consistent with the pivotal voter model and, therefore, included in \( 1 - \lambda \).

**Appendix B: Calculating a Party’s Number of Seats**

Following Spenkuch (2015), this appendix explains the exact algorithm that was used to calculate a party’s number of seats in the 16th and 17th Bundestag. Let \( d_{p,s} \) denote the number of direct mandates accruing to party \( p \) in state \( s \). \( v_{p,s} \) is the number of list votes that \( p \) received in \( s \), with the equivalent number on the national level given by \( \overline{v}_p = \sum_s v_{p,s} \). With this notation in hand, party \( p \)'s seat total is calculated in three steps:

**Step 1: Proportional Allocation of List Mandates to Parties.** Absent overhang mandates, there are 598 seats in the Bundestag. These are allocated by proportionality rule to the set of parties clearing the 5%-threshold or winning at least three direct mandates. That is, the number of list mandates of party \( p \) equals

\[
\overline{l}_p \approx \begin{cases} 
598 \frac{\overline{v}_p}{\sum_{p' \in \tilde{P}} \overline{v}_{p'}} & \text{if } p \in \tilde{P} \\
0 & \text{otherwise}
\end{cases},
\]

where \( \tilde{P} = \left\{ p | \frac{\overline{v}_p}{\sum_{p' \in \tilde{P}} \overline{v}_{p'}} \geq .05 \lor \sum_s d_{p,s} \geq 3 \right\} \) and \( \approx \) represents equality after rounding according to the Sainte-Laguë method, which ensures that \( \sum_p \overline{l}_p = 598 \).\(^4\)

**Step 2: Proportional Allocation of Mandates to State Lists.** German electoral law requires parties to compete with different lists in each state. Therefore, list mandates need to be allocated to

\(^3\)Note, voters for whom \( u_k < u_{k'} \) for some \( k' \neq k \) would never vote for \( k \) anyway. The choice of whether or not to stick with \( k \) in light of her being “out of the race” is irrelevant for these agents.

\(^4\)In 2005 the method of Hare-Niemeyer was used instead.
the respective state lists. In practice, the number of mandates awarded to a party’s state list is proportional to the list’s contribution to the party’s vote total. More precisely, for all $s$ and all $p$,

$$l_{p,s} = \begin{cases} \tilde{l}_p \frac{\pi_p}{\pi_P} & \text{if } p \in \tilde{P} \\ 0 & \text{otherwise} \end{cases},$$

where $\equiv$ is defined as above.

**Step 3: Determination of the Actual Number of Seats.** However, the actual number of seats that party $p$ receives in state $s$ is given by

$$n_{p,s} = \max \{d_{p,s}, l_{p,s} \}.$$

If $d_{p,s} < l_{p,s}$ then, in addition to the district winners, the first $l_{p,s} - d_{p,s}$ candidates on $p$’s list in $s$ are elected to the Bundestag as well. Otherwise, only holders of direct mandates receive a seat.

Note that only if $d_{p,s} \leq l_{p,s}$ for all $s$, will party $p$’s seat total, $\pi_p = \sum_s n_{p,s}$, be equal to the number of seats it would be assigned under proportional representation, i.e. $\tilde{l}_p$.

**Appendix C: Robustness and Sensitivity Checks**

**Functional Form** The results in the main text are based on linear regression models. In Appendix A, I show that the linear model specification is theoretically grounded, and that the slope parameter recovers the relevant mixture weights, i.e., the share of voters whose behavior violates the pivotal voter model. Nonetheless, in reality, the relationship between list- and candidate-vote shares need not necessarily be linear. In order to show that linearity provides an excellent approximation to the true functional relationship in the data, Appendix Figure A.2 depicts a nonparametric estimate of said relationship among candidates who trail far behind the runner-up, i.e., the same set of candidates used in the upper panel of Table 2 in the main text. As is apparent from Figure A.2, a linear regression model appears to be reasonable.

**Misclassification of Contenders** For the point estimates in the main text to correctly identify deviations from the pivotal voter model, it must be the case that the regressors are uncorrelated with the error term. One obvious source of bias may be systematic misclassification of contenders. While it is unproblematic to falsely classify some candidates whom voters believed to be out of the race as contenders—at least as long as $\lambda$ is not heterogeneously distributed—making the opposite mistake would lead to upward bias in $\lambda$ and, therefore, to an overstatement of the extent to which observed behavior violates the theory. To ameliorate this concern, Appendix Table A.2 presents estimates of $\lambda$ using twenty-six alternative definitions of contenders. For each definition, the table shows two estimates: one based on candidate-year fixed effects, and another using candidate-year fixed effects that are specific to individual municipalities. For comparison, the top row displays the main results from the lower panel of Table 2.

Although individual point estimates do, of course, vary, the majority of them are very close to
their baseline values. For instance, assuming that voters have perfect foresight regarding the winner and runner-up of the election, one would estimate the fraction of nonstrategic votes to equal 66.3% instead of 65.6%, whereas adaptive expectations based on the outcome of the last election (i.e., the winner and runner-up in the previous federal election are believed to be in the race) would lead to point estimates ranging from 67.8% to 71.3%. Of the fifty-two additional estimates in Table A.2, the lowest one is 58.9% and the highest one equals 71.6%. Slightly more than 90% of coefficients fall within the original 95%-confidence intervals. The evidence, therefore, suggests that misclassification of contestants is not a first-order problem.

**Exact Indifference** Some individuals could be exactly indifferent about who carries their district, and might therefore stick with a candidate who is out of the race. The empirical strategy in this paper would classify these agents as nonstrategic, leading to estimates of $\lambda$ that include indifferent voters.

One piece of evidence suggesting that the vast majority of voters are not indifferent to who represents their home district in parliament comes from the fact that less than 2% of those going to the polls cast invalid or no candidate votes (despite the fact that it is possible to cast a valid list vote while leaving the candidate vote blank). For the U.S., for instance, it has been argued that ballot roll-off (i.e., voters not completing one of several sections on the ballot) is a sign of voters not caring “enough” about a particular race (e.g., Bullock and Dunn 1996; Burnham 1965). If Germans were exactly indifferent about district-level races, then one would not expect them to be willing to incur even a small “hassle cost” to cast their candidate vote. The fact that more than 98% of voters do cast valid candidate votes suggests that the potential bias from exact indifference is small.

**Endogenous Nomination of Candidates** One may be worried that parties field better candidates in districts in which they have more supporters and that this may lead to biased point estimates. However, estimating $\lambda$ for each candidate-year combination and regressing the resulting $\lambda_{k,t}$ on the district-wide list vote as a measure of party strength yields a point estimate of 0.001 with a standard error of 0.003, which is not only economically small but also statistically indistinguishable from zero. Put differently, local party strength is essentially uncorrelated with the estimated share of voters who stick with the respective candidate.

**Strategic List Votes** As explained in the main text, interpreting $\lambda$ as a lower bound on the share of nonstrategic voters does not require an assumption as to whether list votes accurately reveal voters’ preferences over parties and candidates. This is because it is never optimal for supporters of parties fielding nonviable candidates to vote for the respective nominees, regardless of why these voters cast their list vote for the party. The clear benefit of imposing such an assumption would be that $\lambda$ need not be regarded as a lower bound anymore. In order to provide additional evidence consistent with voters choosing their favorite party according to their preferences, Appendix D presents an explicit (though imperfect) test of strategic voting in the PR part of the German system. Intuitively, if voters cast strategic list votes, one would expect parties to “bunch”
near thresholds where they gain (or lose) a seat. In reality, however, fractional mandates are
approximately uniformly distributed on the unit interval, as one would expect if strategic list votes
were quantitatively unimportant.

Additional Robustness Checks Appendix Table A.3 demonstrates that the results do not
depend on the weighting scheme, whether overhang mandates occurred, or whether one also includes
candidates of “micro-parties.” Appendix Table A.4 further shows that my substantive results do
not depend on the distance between the two frontrunners or the gap between the second and third
candidate.

Appendix D: Strategic Voting under Proportionality Rule?
Given that my main results focus on the 2005 and 2009 elections, in which all major parties were
widely expected to clear the 5%-threshold, voters should have no theoretical incentive to cast
strategic list votes if the party they would like to gain the marginal seat in parliament could,
indeed, be awarded the fractional mandate associated with an additional vote. In reality, however,
parties can only be awarded whole mandates, which means that some may be closer to thresholds
where they gain (or lose) a seat. Thus, if voters cast strategic list votes one would expect parties
to “bunch” near the endogenously determined cutoff levels.5 By contrast, if voters cast sincere
list votes one would expect parties’ number of fractional mandates to be approximately uniformly
distributed on the unit interval. Below, I test this prediction.

The upper panel of Appendix Table A.5 shows the initial distribution of fractional mandates
according to the list vote on the national level (i.e. before applying the rounding methods of Hare-
Niemeier or Sainte-Lagué). The lower panel displays parties’ initial number of fractional mandates
by state. While the former distribution determines the total number of list mandates a given party
receives in parliament, the latter one governs how a party’s number of seats are allocated across
states (cf. Appendix B). The p-values below each panel refer to Kolmogorov-Smirnov tests of the
null hypothesis that the distribution of fractional mandates is uniform on the unit interval.

Based on this approach it is not possible—neither on the national nor on the state level—to reject
the null and, therefore, the assumption that individuals cast list votes that reveal their (induced)
preferences over which party wins the marginal seat in parliament.

Appendix E: Data Appendix
This appendix provides a description of all data used in the paper, as well as precise definitions
together with the sources of all variables. Data containing the official results of the 1980, 1983, 1987,
1990, 1994, and 1998 federal elections by municipality (Gemeinde) as well as the 2002, 2005, and
2009 elections by polling precinct (Wahlbezirk) have been purchased from the Federal Returning

5In 2005 the method of Hare-Niemeyer was used for “rounding”, whereas the Sainte-Lagué method was
used in 2009. It is important to note that whether a party’s number of seats in parliament is adjusted upwards
or downwards depends in both of these methods not just on its own (fractional) vote share, but also on that
of other parties.
These data include information on the number of list and candidate votes for each party and each candidate, the number of eligible voters, as well as the number of valid and invalid votes. In 2009 there were approximately 89,000 precincts. Whenever necessary precinct level numbers are aggregated using the municipality identifiers contained in the raw data. Municipalities spanning multiple districts are discarded. Throughout the analysis the following variables are used:

**Number of Eligible Voters** is defined as the number of residents of each precinct that were allowed to vote in the particular year. In general this encompasses all German citizens over the age of 18, who have not been declared mentally unfit, or whose voting rights have not been suspended due to criminal behavior.

**Turnout** is defined as the number of actual voters over the number of eligible voters. This number cannot be calculated for absentee precincts, as absentee voters are included in the number of eligible voters in their district of residence. Hence, in-person turnout in each district needs to be adjusted for absentee voters. In practice, this is done by multiplying the number of issued absentee ballots by .95 (which corresponds to the empirical frequency with which they are cast) and adding them to the ballots that are cast in person.

**Share of List Vote** is defined as the portion of all valid list votes (in %) that are cast for a particular party. “Micro parties”, i.e., those not clearing the 5%-threshold, are grouped together.

**Share of Candidate Vote** is defined as the portion of all valid candidate votes (in %) that are cast for the candidate of a particular party. Votes for candidates of “micro parties” are pooled.

**Absentee Precinct** is an indicator variable equal to one if a given precinct handles only absentee ballots.

Prior to every election to the Bundestag the Federal Returning Officer publishes information on all official candidates (Bundeswahlleiter 2005c, 2009b). I have manually transcribed these lists and use them to identify individual candidates.

**Appendix F: Structural Analysis**

Although the reduced form results in the main text provide evidence of expressive as well as strategic voting, they are subject to some limitations. For instance, the assumption that candidate quality enters equations (1)–(4) in the main text linearly might be overly restrictive. To rigorously account for the drawbacks of the reduced form analysis and to be able to construct counterfactuals, this section replicates the reduced-form results by estimating a structural model of voting decisions in the 2009 federal election.⁶

---

⁶Results are qualitatively similar when looking at the 2005 election instead.
F.1. Adding Structure

In order to replicate the main results about the average extent of non-instrumentally rational voting, it is convenient to group voters into two sets: strategic agents and expressive ones. Doing so comes at the cost of ignoring a voters’ choice to act strategically, but it simplifies the analysis considerably. Given the very limited variation in district size, and therefore pivot probabilities, it would be extremely challenging to leverage this type of variation.

The Magnitude Theorem in Myerson (2000) shows that voters will generally group contestants into two categories: candidates who are in the race and those who are not. It is, therefore, appealing to model agents’ decisions as a discrete choice problem in which expressive and strategic voters face different equilibrium choice sets. The former choose among all contestants in a particular district, whereas the latter consider only candidates who are in contention for victory. When it comes to the list vote, however, all voters pick from the set of parties.

In order to represent agents’ (induced) preference profiles in a tractable yet flexible fashion, I assume that individual $i$ receives utility

$$u_{i,p}^L = \xi_{p,m} + \zeta_{i,p} + \eta_{i,p}$$

from voting for party $p$’s list. Here, $\xi_{p,m}$ denotes the average utility that agents living in municipality $m$ derive from voting for $p$, and $\zeta_{i,p}$ are individual specific deviations from the mean. $\eta_{i,p}$ is an i.i.d. type-I extreme value (T1EV) taste shock. Any potential strategic considerations with respect to the list vote are assumed to enter via this error term.

Moreover, let the underlying utility from casting one’s candidate vote for the nominee of party $p$ to equal

$$u_{i,k}^C = \xi_{p,m} + \zeta_{i,p} + \chi_k + \epsilon_{i,k},$$

where $k$ indexes candidates, and $\chi_k$ is voters’ assessment of $k$ relative to that of her party. $\chi_k$ plays, therefore, a very similar role as the candidate fixed effect in the reduced form part of the analysis. $\epsilon_{i,k}$ denotes another i.i.d. T1EV shock.\(^7\)

It is important to note that $\xi_{p,m}$ and $\zeta_{i,p}$ appear in both (3) and (4), implying that official party positions influence not only voters’ perceptions of the respective organizations, but also that with respect to their candidates. This assumption captures the fact that German politicians campaign heavily on their own party’s platform and it introduces the correlation between list and candidate votes that has been the identifying source of variation in the reduced-form part of the analysis.

To allow individuals’ preferences to systematically deviate from the average in their municipality, $\zeta_{i,p}$ is assumed to follow a multivariate normal distribution with an unrestricted covariance matrix.

---

\(^7\)The mean utility from abstaining is normalized to zero. Since the available data do not allow turnout to be calculated for individual precincts, the analysis in this section is conducted at the municipality level instead (restricting attention to the set of municipalities that are fully contained within an electoral district).
That is, \( (\zeta_{i,p})_{p \in P} \sim N(0, \Sigma) \). Hence, supporters of the conservative CDU may, for example, also have a taste for the FDP, while holding more negative views of the communist Left.

While \( \xi_{p,m} \) and \( \zeta_{i,p} \) model commonalities in voters’ assessments of parties and the respective contestants, \( \eta_{k,p} \) and \( \varepsilon_{i,k} \) allow for differences in tastes that go beyond the common perception of candidate quality, i.e., \( \chi_k \). The TIEV assumption is convenient because it results in a smooth closed form representation of individual choice probabilities.

Given the structure of preferences, the expected share of the list vote of party \( p \in P \) in municipality \( m \) equals

\[
\hat{v}^{L}_{p,m} = \frac{\exp(\xi_{p,m} + \zeta_{i,p})}{1 + \sum_{p' \in P} \exp(\xi_{p',m} + \zeta_{i,p'})} d\Phi(\zeta),
\]

and that of the associated candidate is

\[
\hat{v}^{C}_{k,m} = \lambda \hat{v}^{C,E}_{k,m} + (1 - \lambda) \hat{v}^{C,T}_{k,m}.
\]

Here, \( \hat{v}^{C,E}_{k,m} \) denotes candidate \( k \)'s share among expressive voters, and \( \hat{v}^{C,T}_{k,m} \) that among tactical ones. As before, \( \lambda \) is the fraction of agents who are sincere.

Since expressive voters consider every candidate, \( \hat{v}^{C,E}_{k,m} \) is given by

\[
\hat{v}^{C,E}_{k,m} = \frac{\exp(\xi_{p,m} + \zeta_{i,p} + \chi_k)}{1 + \sum_{k' \in K(d)} \exp(\xi_{p',m} + \zeta_{i,p'} + \chi_{k'})} d\Phi(\zeta),
\]

where \( K(d) \) marks the set of all contestants in district \( d \). Tactical agents, however, behave as if they are choosing only among the set of serious contenders, \( C(d) \). That is, irrespective of the underlying utility in (4), strategic voters disregard all candidates that are not “in the race.” Consequently, \( k \)'s share among strategic individuals equals

\[
\hat{v}^{C,T}_{k,m} = \begin{cases} 
\int \frac{\exp(\xi_{p,m} + \zeta_{i,p} + \chi_k)}{1 + \sum_{k' \in C(d)} \exp(\xi_{p',m} + \zeta_{i,p'} + \chi_{k'})} d\Phi(\zeta) & \text{if } k \in C(d) \\
0 & \text{otherwise}
\end{cases}.
\]

A seemingly natural way to estimate \( (\xi, \chi, \Sigma, \lambda) \) would be to find the parameter combination that produces the best fit between predicted vote shares and the data. This, however, entails that preferences would be partially identified from candidate votes, which may confound strategic desertion with simple distaste. In order to avoid this problem, I infer electorates’ average tastes solely from list votes.

Accordingly, with data on \( C(d) \) and actual vote shares in hand, estimates of \( (\xi, \chi, \Sigma, \lambda) \) could be
obtained by minimizing the objective function:

\[
SSR(\xi, \chi, \Sigma, \lambda|v^C, v^L) = \sum_{d \in D} \sum_{m \in M(d)} \sum_{k \in K(d)} (\hat{v}_{k,m}^C - v_{k,m}^C)^2
\]

subject to the set of constraints

\[
\hat{v}_{p,m}^L = v_{p,m}^L \quad \forall p, m, d.
\]

Yet, as \(C(d)\) is not observed, it needs to be estimated as well. I assume that a candidate is believed to be a contender if, and only if, her party trails the district’s second most popular one by less than \(\kappa\) percentage points, and recover \(\kappa\) from the data.\(^9\) Thus, estimating \(C(d)\) adds the following set of equilibrium constraints

\[
\begin{align*}
\tau_{d}^{L,2nd} & - \tau_{k,d}^{L} \leq \kappa & \forall k \in C(d), \forall d \\
\tau_{d}^{L,2nd} & - \tau_{k,d}^{L} > \kappa & \forall k \notin C(d), \forall d.
\end{align*}
\]

Given the granularity of the data, the optimization problem defined by equations (8)–(11) is extremely large. Finding the solution involves optimizing over more than 63,000 parameters, solving about 61,500 non-linear constraints, and approximating roughly 120,000 different five dimensional integrals. To keep the computational burden manageable without compromising the quality of the solution, the analysis relies on recent advances in numerical methods, such as integration on sparse-grids (Heiss and Wünschel 2008) and mathematical programming with equality constraints (Dube et al. 2012; Su and Judd 2012).

Before proceeding to the results it is useful to provide some intuition on how the parameters are identified. Identification of \(\xi_{p,m}\) is straightforward. From Berry (1994) it follows that, for every \(\Sigma\), there exists a unique vector \(\xi\) which solves (9). Economically, this means that the list vote pins down the average taste in different markets.

Akin to the analysis in the main text, identification of \(\lambda\) is based on the intuition in Figure 1. That is, the share of sincere voters can be inferred from the ratio of noncontenders’ observed vote shares (depicted on the \(y\)-axis) to those they would receive if all agents acted solely based on their preferences (proxied by the position on the \(x\)-axis).

Candidate quality, i.e., \(\chi_k\), can be gleaned by comparing contestant’s actual performance in different municipalities with predictions thereof based on party preferences and \(\lambda\). \(\chi_k\) will be positive for candidates whose vote shares systematically exceed their predicted values, and negative for those who underperform.

Lastly, \(\Sigma\) is identified from the empirical covariance between noncontenders’ list and contenders’ candidate votes. Take, for instance, a district in which the FDP candidate is out of the race, while

\(^9\)Experimentation with a subset of the contender classifications in Appendix Table A.2 yielded qualitatively similar results.
the nominee of the CDU is a contender. If the latter receives, on average, more votes in villages that have a greater taste for the FDP, then the respective parameter in the covariance matrix will be positive. Analogous arguments apply to the remaining elements of $\Sigma$.

F.2. Results and Counterfactual Experiments

With 73.7% (and a standard error of 7.8%) the estimated share of nonstrategic voters, i.e. $\lambda$, is strikingly close to the corresponding reduced-form results in Tables 2 and A.2. Unfortunately, few of the model’s other parameters are easily interpretable by themselves. Thus, instead of listing parameter estimates, the following discussion presents results in a way that relates straightforwardly to common intuition.\(^{10}\)

In order to judge the model’s fit consider Appendix Figure A.3. The upper two panels contrast the true marginal distributions of candidate and list votes (dark bars) with those predicted by the model (light bars). Given that $(\xi, \chi, \Sigma, \lambda, \kappa)$ have been chosen to mimic these data, there are practically no discernible differences.

The lower panel depicts the frequency of valid list and candidate vote combinations. It is important to note that information on the joint distribution of votes come from an independent source (Bundeswahlleiter 2009a, 2010) and were not used to fit the model. Thus, the lower panel of Figure A.3 provides a strong quasi-out-of-sample test of whether the estimation results are reasonable. Although there do exist differences, on the whole the predicted distribution matches the qualitative features of its real world counterpart fairly well, lending credibility to the results.

Appendix Table A.6 compares actual and simulated outcomes of district level races. As can be seen from the entries on the diagonal, the model does an excellent job at ranking candidates. In particular, it predicts almost 95% of winners correctly.

While Figure A.3 and Table A.6 are useful in evaluating the goodness of fit, a more interesting question might be for whom supporters of different parties would vote if their preferred candidate was out of the race. In order to shed light on the ordering of preferences, Appendix Table A.7 shows the frequency with which voters would substitute toward the candidate of any other party, assuming that all but their preferred contestant were still in the race. Thus, the entries correspond to the probability of some other party’s candidate being “the next best choice.” The model predicts FDP adherents to substitute toward candidates of the CDU, whereas most supporters of the Green Party and The Left would choose SPD contenders instead. Given parties’ ideological positions, these patterns conform exactly to what one would expect.

Based on the structural estimates, Appendix Figure A.4 presents several counterfactual election results by which to judge the impact of strategic voting.\(^{11}\) The top left panel shows the actual distribution of seats in the Bundestag, whereas the panel on the right displays the distribution that would prevail if mandates were awarded based solely on a single vote counted under proportionality

\(^{10}\)A list of all estimates is available from the author upon request.

\(^{11}\)For details on the computation of these counterfactuals see below.
rule with a 5%-threshold, i.e., the list vote. Evidently, the current Bundestag mirrors a parliament formed under proportional representation fairly closely: all five major parties are represented, with more than 60% of seats accruing to the CDU and the SPD. In the current equilibrium, distortions introduced through strategic candidate votes are very small.

The remaining two panels assume a single vote counted under plurality rule on the district level (akin to the candidate vote, or elections to the House of Representatives in the U.S.) The counterfactual on the bottom left shows the model’s predictions for such a first-past-the-post scheme with 26.3% of voters behaving strategically and the current set of candidates. In the panel on the bottom right all voters are expressive.

In line with common intuition, relative to proportional representation a “winner-take-all” system would result in dramatic losses for small parties. However, as comparing the panels on the right shows, these losses are due to the way different electoral rules map vote shares into mandates and not to tactical voting.

The impact of strategic behavior can be gleaned from comparing the two counterfactuals on the bottom. Given its estimated extent, tactical voting has only a modest effect on the overall allocation of seats. Not a single party’s share of seats would change by more than 5 percentage points, often substantially less. The evidence in Table A.8 further indicates that, compared to the current equilibrium, about one in ten districts would change hands if all voters were to cast sincere ballots.

F.3. Numerical Methods

This appendix describes the numerical methods used to solve the optimization problem defined by equations (8)–(11) as well as the construction of counterfactual election results above.

Mathematical Programming with Equality Constraints Typically, to recover mean utilities in models of discrete choice (i.e. $\xi_{p,m}$) researchers turn to inverting the system of non-linear markets share equations via the nested fixed point (NFP) algorithm in Berry (1994) and Berry et al. (1995). Recently, however, Su and Judd (2012) and Dube et al. (2012) have shown how to recast extremum estimators in general, and the one in the Berry et al. (1995) in particular, as a mathematical programming problem with equality constraints (MPEC).

Key to the MPEC approach is the insight that the inner loop can be eliminated entirely by recasting the estimator as an optimization problem subject to a set of non-linear constraints, i.e., (9), which require predicted market shares to equal observed ones.

Since objective function and market share equations are usually smooth, one can rely on state-of-art optimization software to find candidate solutions. Moreover, dispensing with the inner loops avoids numerical problems associated with loose inner loop error tolerances (see Dube et al. 2012 for a discussion of the NFP algorithm’s numerical properties), and it may significantly increase computational speed because the system of market share equations does not have to be solved exactly at each iteration. (The constraints have to be satisfied only at the solution.) Importantly,
Su and Judd (2012) prove that MPEC and NFP solve the same problem, yielding the same estimates with the same statistical properties.

The implementation of MPEC in this paper is based on the MATLAB code of Dube et al. (2012), using both of the KNITRO solver’s interior-point and active set algorithms (Byrd et al. 1999, 2004, 2006). To improve numerical accuracy as well as computational performance, KNITRO is provided with hand-coded first-order analytical derivatives of the objective function and the constraints, second order derivatives, as well as the sparsity patterns of the constraint Jacobian and the Hessian. Since the Hessian contains almost $4 \times 10^9$ elements of which only about $1.8 \times 10^6$ are non-zero, supplying the solver with the sparsity pattern is critical in order to economize on memory usage and time. To increase the likelihood of finding the global optimum five different starting points are used. Relative optimality and feasibility error tolerances, i.e. the maximum violation of the first order conditions and the constraints, have each been set to $10^{-6}$. Reported standard errors are based on the block-bootstrap with 100 iterations.

In order to provide the solver with a completely smooth optimization problem, the constraints in (10)–(11) have been rewritten as an indicator function for each candidate, $c_k (\kappa)$, and are numerically approximated by the hyperbolic tangent. That is,

$$c_k (\kappa) = \frac{1}{2} + \frac{1}{2} \tanh \left( \rho \left( \kappa + \bar{\eta}_{k,d} - \bar{\eta}_{d}^{L, a} \right) \right)$$

for $\rho = 5,000$. Thus, equation (7) becomes

$$\widehat{v}_{k,m}^{C,T} = \int \frac{c_k \exp \left( \xi_{p,m} + \eta_{i,p} + \chi_k \right)}{1 + \sum_{k' \in K} c_{k'} \exp \left( \xi_{p',m} + \eta_{i,p'} + \chi_{k'} \right)} d\Phi (\zeta).$$

**Sparse Grid Integration** Instead of solving the approximately 120,000 five dimensional integrals in equations (5), (6), and (7) using simulation methods, I rely on sparse grid integration (SGI), introduced into economics by Heiss and Winschel (2008). SGI provides a way to approximate integrals numerically avoiding the curse of dimensionality associated with conventional quadrature rules (see Judd 1998). Monte Carlo evidence by Skrainka and Judd (2011) indicates that SGI imposes a significantly lower computational burden than simulation methods achieving the same level of accuracy.

SGI is closely related to conventional Gaussian quadrature rules, but by exploiting symmetry properties it relies only on a small subset of nodes and (appropriately rescaled) weights. My analysis uses a Konrad-Patterson rule with Gaussian kernel for choosing nodes, as explained in Heiss and Winschel (2008). This particular rule has only 151 nodes; yet it exactly integrates (over five dimensions) all complete polynomials of total order less than 7. Experimentation with more accurate rules yielded essentially the same point estimates, but required significantly more CPU time.

**Construction of Counterfactuals** The counterfactual election results in Section F.2 of the appendix have been constructed by simulation. More specifically, for each municipality in the data 100 times its actual number of voters have been simulated by randomly drawing $\zeta$, $\eta$, and $\varepsilon$ from
the respective (estimated) distributions. A fraction $\lambda$ of simulated voters (rounded to the nearest integer) are designated to behave sincerely. Next, each voter’s candidate and party specific utilities are calculated and his (partial) preference orderings for the list and candidate vote are determined. Naturally, sincere voters consider all candidates, whereas tactical voters choose only among those contestants who are estimated to be contenders. Election results are then constructed by aggregating votes to the appropriate level, and applying the specified electoral rule.

References


Figure A.1: Distribution of Direct Mandates in the 2005 and 2009 Federal Elections

**Notes:** Figure depicts the winner of the candidate vote by electoral district and candidates’ party affiliation in the 2005 (left) and 2009 (right) federal elections. In the 2005 (2009) election, candidates running for the CDU/CSU won the plurality of votes in 150 (218) out of 299 electoral districts. SPD candidates gained 145 (64) direct mandates. Candidates of The Left won 3 (16) districts, and the Green Party achieved 1 (1) direct mandate. No FDP contestant won a district race.

**Sources:** Based on Bundeswahlleiter (2005a, 2005b, 2008, 2009).
Figure A.2: Relationship between List and Candidate Votes for Candidates Trailing Far Behind

Notes: Figure shows a semiparametric estimate of the relationship between list and candidate votes for candidates of the five major parties who trail the runner-up in their district by more than 10 percentage points as well as the associated asymptotic 95% confidence intervals. The estimating equation is $v_{k,r,t}^c = \chi_{m,k,t} + f(v_{k,r,t}^l) + \epsilon_{k,r,t}$, where $v_{k,r,t}^c$ denotes the vote share of candidate $k$ in precinct $r$ during election year $t$, $v_{k,r,t}^l$ is the list-vote share of the associated party, and $\chi_{m,k,t}$ is a municipality- and year-specific candidate fixed effect. $f(\cdot)$ is approximated by cubic B-splines with knots at every 1.5 percentage points. Standard errors account for clustering at the state level and have been calculated using the nonparametric bootstrap with 1,000 iterations.
Figure A.3: Observed vs Predicted Distribution of Votes based on Structural Analysis, 2009 Federal Elections

A. Marginal Distribution of Candidate Votes

B. Marginal Distribution of List Votes

C. Joint Distribution of Candidate and List Votes

Notes: Figure depicts actual and predicted vote shares in the 2009 federal election. Panel A shows the marginal distribution of candidate votes, and panel B that of list votes. Panel C depicts the frequency of valid list and candidate vote combinations, i.e. their joint distribution. Dark columns are based on official statistics by the Federal Returning Officer (Bundeswahlleiter 2009, 2010). Light columns corresponds to the predictions of the structural model in Appendix F.
Figure A.4: Counterfactual Seat Distributions in the 17th Bundestag

**A. Status Quo**

- CDU/CSU: 36.0%
- SPD: 24.5%
- FDP: 15.5%
- The Left: 12.7%
- Green Party: 11.4%

**B. Proportional Representation Based on Actual List Votes**

- CDU/CSU: 38.4%
- SPD: 23.5%
- The Left: 12.2%
- Green Party: 10.9%

**C. Plurality Rule Expressive and Strategic Voters**

- CDU/CSU: 77.6%
- SPD: 18.7%
- The Left: 3.1%
- Green Party: 0.7%

**D. Plurality Rule Expressive Voters Only**

- CDU/CSU: 73.8%
- SPD: 23.5%

Notes: Figure depicts counterfactual seat distributions in the Bundestag following the 2009 federal election. Results are based on the structural estimates in Appendix F. See the appendix for a description of the assumptions underlying each panel.
## Table A.1: Summary Statistics for Electoral Precincts

<table>
<thead>
<tr>
<th>Variable</th>
<th>Full Sample</th>
<th>West Germany 2005</th>
<th>West Germany 2009</th>
<th>East Germany 2005</th>
<th>East Germany 2009</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Eligible Voters</td>
<td>820.7</td>
<td>821.5</td>
<td>834.2</td>
<td>782.9</td>
<td>802.5</td>
</tr>
<tr>
<td></td>
<td>(406.1)</td>
<td>(385.4)</td>
<td>(387.6)</td>
<td>(460.3)</td>
<td>(487.7)</td>
</tr>
<tr>
<td>Turnout</td>
<td>.747</td>
<td>.789</td>
<td>.727</td>
<td>.751</td>
<td>.658</td>
</tr>
<tr>
<td></td>
<td>(.087)</td>
<td>(.071)</td>
<td>(.083)</td>
<td>(.069)</td>
<td>(.084)</td>
</tr>
<tr>
<td>Share of Candidate Vote (in %):</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CDU/CSU</td>
<td>41.07</td>
<td>44.81</td>
<td>41.94</td>
<td>29.65</td>
<td>32.94</td>
</tr>
<tr>
<td></td>
<td>(13.02)</td>
<td>(13.33)</td>
<td>(11.49)</td>
<td>(9.93)</td>
<td>(10.32)</td>
</tr>
<tr>
<td>SPD</td>
<td>32.23</td>
<td>38.73</td>
<td>28.80</td>
<td>31.35</td>
<td>20.03</td>
</tr>
<tr>
<td></td>
<td>(12.61)</td>
<td>(12.66)</td>
<td>(10.71)</td>
<td>(8.28)</td>
<td>(7.54)</td>
</tr>
<tr>
<td>FDP</td>
<td>7.04</td>
<td>4.59</td>
<td>9.66</td>
<td>5.16</td>
<td>8.17</td>
</tr>
<tr>
<td></td>
<td>(3.93)</td>
<td>(2.25)</td>
<td>(3.74)</td>
<td>(2.85)</td>
<td>(3.73)</td>
</tr>
<tr>
<td>The Left</td>
<td>9.66</td>
<td>3.95</td>
<td>7.35</td>
<td>24.86</td>
<td>28.61</td>
</tr>
<tr>
<td></td>
<td>(9.67)</td>
<td>(3.06)</td>
<td>(4.01)</td>
<td>(7.37)</td>
<td>(8.46)</td>
</tr>
<tr>
<td>Green Party</td>
<td>6.87</td>
<td>5.46</td>
<td>9.32</td>
<td>3.76</td>
<td>5.53</td>
</tr>
<tr>
<td></td>
<td>(5.29)</td>
<td>(3.78)</td>
<td>(5.23)</td>
<td>(5.80)</td>
<td>(6.08)</td>
</tr>
<tr>
<td>Others</td>
<td>3.08</td>
<td>2.38</td>
<td>2.88</td>
<td>5.22</td>
<td>4.72</td>
</tr>
<tr>
<td></td>
<td>(2.88)</td>
<td>(2.75)</td>
<td>(2.59)</td>
<td>(3.10)</td>
<td>(2.70)</td>
</tr>
<tr>
<td>Share of List Vote (in %):</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CDU/CSU</td>
<td>35.47</td>
<td>38.67</td>
<td>35.59</td>
<td>26.21</td>
<td>30.65</td>
</tr>
<tr>
<td></td>
<td>(11.60)</td>
<td>(12.48)</td>
<td>(10.25)</td>
<td>(8.59)</td>
<td>(8.92)</td>
</tr>
<tr>
<td>SPD</td>
<td>27.98</td>
<td>34.22</td>
<td>23.62</td>
<td>29.96</td>
<td>17.95</td>
</tr>
<tr>
<td></td>
<td>(10.91)</td>
<td>(10.84)</td>
<td>(8.63)</td>
<td>(7.16)</td>
<td>(5.94)</td>
</tr>
<tr>
<td>FDP</td>
<td>12.01</td>
<td>10.10</td>
<td>15.18</td>
<td>8.00</td>
<td>10.57</td>
</tr>
<tr>
<td></td>
<td>(4.83)</td>
<td>(3.56)</td>
<td>(4.61)</td>
<td>(3.25)</td>
<td>(3.76)</td>
</tr>
<tr>
<td>The Left</td>
<td>10.43</td>
<td>4.83</td>
<td>8.40</td>
<td>25.05</td>
<td>28.27</td>
</tr>
<tr>
<td></td>
<td>(9.23)</td>
<td>(3.09)</td>
<td>(4.26)</td>
<td>(6.26)</td>
<td>(7.56)</td>
</tr>
<tr>
<td>Green Party</td>
<td>8.83</td>
<td>8.38</td>
<td>10.93</td>
<td>4.78</td>
<td>5.90</td>
</tr>
<tr>
<td></td>
<td>(5.38)</td>
<td>(4.87)</td>
<td>(5.25)</td>
<td>(4.13)</td>
<td>(4.93)</td>
</tr>
<tr>
<td>Others</td>
<td>5.23</td>
<td>3.71</td>
<td>6.23</td>
<td>6.01</td>
<td>6.66</td>
</tr>
<tr>
<td></td>
<td>(2.85)</td>
<td>(2.03)</td>
<td>(2.76)</td>
<td>(3.03)</td>
<td>(3.03)</td>
</tr>
<tr>
<td>Absentee Precinct</td>
<td>.148</td>
<td>.155</td>
<td>.166</td>
<td>.090</td>
<td>.098</td>
</tr>
<tr>
<td></td>
<td>(.355)</td>
<td>(.362)</td>
<td>(.372)</td>
<td>(.286)</td>
<td>(.297)</td>
</tr>
<tr>
<td>Number of Observations</td>
<td>177,425</td>
<td>71,614</td>
<td>72,056</td>
<td>17,110</td>
<td>16,645</td>
</tr>
</tbody>
</table>

*Notes:* Entries are means and standard deviations for all precinct-level variables used in the analysis, differentiating between East and West Germany as well as election year. See the Data Appendix for a precise definition of each variable.
<table>
<thead>
<tr>
<th>Classification of Contenders</th>
<th>Fixed Effects:</th>
<th>Violations of Pivotal Voter Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Candidate × Year</td>
<td>Candidate × Municipality × Year</td>
</tr>
<tr>
<td>Baseline (Based on Party Vote, Original Cutoff)</td>
<td>.657</td>
<td>.656</td>
</tr>
<tr>
<td></td>
<td>(.019)</td>
<td>(.026)</td>
</tr>
<tr>
<td>Ex Post Outcome of Races (Original Cutoff)</td>
<td>.655</td>
<td>.651</td>
</tr>
<tr>
<td></td>
<td>(.018)</td>
<td>(.026)</td>
</tr>
<tr>
<td>Based on Party Vote, Using Different Cutoffs:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt; 1% behind Second-Ranked Candidate</td>
<td>.705</td>
<td>.668</td>
</tr>
<tr>
<td></td>
<td>(.029)</td>
<td>(.028)</td>
</tr>
<tr>
<td>&gt; 2% behind Second-Ranked Candidate</td>
<td>.693</td>
<td>.661</td>
</tr>
<tr>
<td></td>
<td>(.027)</td>
<td>(.027)</td>
</tr>
<tr>
<td>&gt; 5% behind Second-Ranked Candidate</td>
<td>.669</td>
<td>.641</td>
</tr>
<tr>
<td></td>
<td>(.021)</td>
<td>(.020)</td>
</tr>
<tr>
<td>&gt; 8% behind Second-Ranked Candidate</td>
<td>.648</td>
<td>.623</td>
</tr>
<tr>
<td></td>
<td>(.017)</td>
<td>(.018)</td>
</tr>
<tr>
<td>&gt; 10% behind Second-Ranked Candidate</td>
<td>.634</td>
<td>.609</td>
</tr>
<tr>
<td></td>
<td>(.015)</td>
<td>(.014)</td>
</tr>
<tr>
<td>&gt; 12% behind Second-Ranked Candidate</td>
<td>.615</td>
<td>.589</td>
</tr>
<tr>
<td></td>
<td>(.012)</td>
<td>(.009)</td>
</tr>
<tr>
<td>Ex Post Outcome of Races, Using Different Cutoffs:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt; 1% behind Second-Ranked Candidate</td>
<td>.690</td>
<td>.658</td>
</tr>
<tr>
<td></td>
<td>(.027)</td>
<td>(.028)</td>
</tr>
<tr>
<td>&gt; 2% behind Second-Ranked Candidate</td>
<td>.682</td>
<td>.652</td>
</tr>
<tr>
<td></td>
<td>(.026)</td>
<td>(.026)</td>
</tr>
<tr>
<td>&gt; 5% behind Second-Ranked Candidate</td>
<td>.663</td>
<td>.635</td>
</tr>
<tr>
<td></td>
<td>(.020)</td>
<td>(.021)</td>
</tr>
<tr>
<td>&gt; 8% behind Second-Ranked Candidate</td>
<td>.650</td>
<td>.626</td>
</tr>
<tr>
<td></td>
<td>(.018)</td>
<td>(.020)</td>
</tr>
<tr>
<td>&gt; 10% behind Second-Ranked Candidate</td>
<td>.632</td>
<td>.613</td>
</tr>
<tr>
<td></td>
<td>(.014)</td>
<td>(.016)</td>
</tr>
<tr>
<td>&gt; 12% behind Second-Ranked Candidate</td>
<td>.618</td>
<td>.597</td>
</tr>
<tr>
<td></td>
<td>(.011)</td>
<td>(.011)</td>
</tr>
<tr>
<td>Ranked First or Second Based on Party Vote</td>
<td>.716</td>
<td>.676</td>
</tr>
<tr>
<td></td>
<td>(.030)</td>
<td>(.030)</td>
</tr>
<tr>
<td>Ranked First, Second, or Third Based on Party Vote</td>
<td>.701</td>
<td>.665</td>
</tr>
<tr>
<td></td>
<td>(.023)</td>
<td>(.017)</td>
</tr>
<tr>
<td>Ranked First or Second Based on Ex Post Outcome</td>
<td>.695</td>
<td>.663</td>
</tr>
<tr>
<td></td>
<td>(.029)</td>
<td>(.029)</td>
</tr>
<tr>
<td>Ranked First, Second, or Third Based on Ex Post Outcome</td>
<td>.629</td>
<td>.600</td>
</tr>
<tr>
<td></td>
<td>(.013)</td>
<td>(.014)</td>
</tr>
<tr>
<td>Finished First or Second in Last Federal Election</td>
<td>.713</td>
<td>.678</td>
</tr>
<tr>
<td></td>
<td>(.032)</td>
<td>(.034)</td>
</tr>
<tr>
<td>Finished First, Second, or Third in Last Federal Election</td>
<td>.681</td>
<td>.643</td>
</tr>
<tr>
<td></td>
<td>(.015)</td>
<td>(.011)</td>
</tr>
<tr>
<td>Finish in Last Federal Election (Original Cutoff)</td>
<td>.684</td>
<td>.670</td>
</tr>
<tr>
<td></td>
<td>(.026)</td>
<td>(.031)</td>
</tr>
<tr>
<td>Finish in Last Federal Election Using Different Cutoffs:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt; 1% behind Second-Ranked Candidate</td>
<td>.709</td>
<td>.674</td>
</tr>
<tr>
<td></td>
<td>(.031)</td>
<td>(.032)</td>
</tr>
<tr>
<td>&gt; 2% behind Second-Ranked Candidate</td>
<td>.704</td>
<td>.670</td>
</tr>
<tr>
<td></td>
<td>(.030)</td>
<td>(.032)</td>
</tr>
<tr>
<td>&gt; 5% behind Second-Ranked Candidate</td>
<td>.687</td>
<td>.656</td>
</tr>
<tr>
<td></td>
<td>(.027)</td>
<td>(.029)</td>
</tr>
<tr>
<td>&gt; 8% behind Second-Ranked Candidate</td>
<td>.678</td>
<td>.648</td>
</tr>
<tr>
<td></td>
<td>(.024)</td>
<td>(.027)</td>
</tr>
<tr>
<td>&gt; 10% behind Second-Ranked Candidate</td>
<td>.671</td>
<td>.642</td>
</tr>
<tr>
<td></td>
<td>(.022)</td>
<td>(.023)</td>
</tr>
<tr>
<td>&gt; 12% behind Second-Ranked Candidate</td>
<td>.663</td>
<td>.634</td>
</tr>
<tr>
<td></td>
<td>(.019)</td>
<td>(.019)</td>
</tr>
</tbody>
</table>

Notes: Entries are coefficients and standard errors on $\lambda$ in equation (4) in the main text, using alternative classifications of “contender.” The respective definition is shown in the column on the left. Heteroskedasticity robust standard errors are clustered by state and reported in parentheses.
### Table A.3: Additional Sensitivity and Robustness Checks

<table>
<thead>
<tr>
<th>Restriction</th>
<th>Fixed Effects:</th>
<th>Violations of Pivotal Voter Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Candidate × Year</td>
<td>Candidate × Municipality × Year</td>
</tr>
<tr>
<td>Baseline</td>
<td>.657</td>
<td>.656 ( .019 )</td>
</tr>
<tr>
<td>Difference Estimator</td>
<td>.653</td>
<td>.678 ( .027 )</td>
</tr>
<tr>
<td>In States without Overhang Mandates</td>
<td>.624</td>
<td>.609 ( .029 )</td>
</tr>
<tr>
<td>Weighted by Number of Party Supporters</td>
<td>.678</td>
<td>.672 ( .029 )</td>
</tr>
<tr>
<td>Including &quot;Other&quot; Party Candidates</td>
<td>.659</td>
<td>.645 ( .020 )</td>
</tr>
</tbody>
</table>

Notes: Entries are coefficients and standard errors on the share of nonstrategic voters, i.e. $\lambda$ in Table 4, using different subsamples of the data and weighting schemes. The respective restriction is indicated on the left of each row. See the Data Appendix for the precise definition and source of each variable.
Table A.4: Candidate–List–Vote Gradient among Candidates Trailing Far Behind, by Distance between the First- and Second- as well as Second- and Third-Ranked Candidates

<table>
<thead>
<tr>
<th>Gap between First- and Second-Ranked Candidate</th>
<th>Gap between Second- and Third-Ranked Candidate</th>
<th>Cutoffs Based on Party Vote</th>
<th>Cutoffs Based on Ex Post Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Any</td>
<td>&gt; 10%</td>
<td>.609</td>
<td>.613</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.014)</td>
<td>(.016)</td>
</tr>
<tr>
<td>Any</td>
<td>&gt; 12%</td>
<td>.589</td>
<td>.597</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.009)</td>
<td>(.011)</td>
</tr>
<tr>
<td>&lt; 3%</td>
<td>&gt; 10%</td>
<td>.620</td>
<td>.646</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.023)</td>
<td>(.041)</td>
</tr>
<tr>
<td>&lt; 3%</td>
<td>&gt; 12%</td>
<td>.607</td>
<td>.610</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.028)</td>
<td>(.027)</td>
</tr>
<tr>
<td>&lt; 2%</td>
<td>&gt; 10%</td>
<td>.626</td>
<td>.659</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.026)</td>
<td>(.051)</td>
</tr>
<tr>
<td>&lt; 2%</td>
<td>&gt; 12%</td>
<td>.609</td>
<td>.606</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.034)</td>
<td>(.028)</td>
</tr>
<tr>
<td>&lt; 1%</td>
<td>&gt; 10%</td>
<td>.639</td>
<td>.592</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.029)</td>
<td>(.032)</td>
</tr>
<tr>
<td>&lt; 1%</td>
<td>&gt; 12%</td>
<td>.616</td>
<td>.592</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.042)</td>
<td>(.032)</td>
</tr>
</tbody>
</table>

Notes: Entries are coefficients and standard errors on the share of nonstrategic voters, i.e. λ in Table 4, using different subsamples of the data. The respective restriction is indicated on the left of each row. All regressions control for candidate × municipality × year fixed effects. Heteroskedasticity robust standard errors are clustered by state and reported in parentheses.
### Table A.5: Initial Distribution of Fractional Mandates by Party, 2005 & 2009 Federal Elections

#### A. National Level

<table>
<thead>
<tr>
<th>Election Year</th>
<th>CDU</th>
<th>SPD</th>
<th>FDP</th>
<th>The Left</th>
<th>Green Party</th>
<th>CSU</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005</td>
<td>.919</td>
<td>.170</td>
<td>.184</td>
<td>.208</td>
<td>.524</td>
<td>.996</td>
</tr>
<tr>
<td>2009</td>
<td>.517</td>
<td>.557</td>
<td>.655</td>
<td>.636</td>
<td>.115</td>
<td>.519</td>
</tr>
</tbody>
</table>

$H_o$: Fractional Mandates ~ $U[0,1]$

All Years: $p$-value = .721

2005: $p$-value = .542

2009: $p$-value = .310

#### B. State Level

<table>
<thead>
<tr>
<th>State &amp; Election Year</th>
<th>CDU</th>
<th>SPD</th>
<th>FDP</th>
<th>The Left</th>
<th>Green Party</th>
<th>CSU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bavaria, 2005</td>
<td>--</td>
<td>.761</td>
<td>.843</td>
<td>.209</td>
<td>.440</td>
<td>.996</td>
</tr>
<tr>
<td>Bavaria, 2009</td>
<td>--</td>
<td>.480</td>
<td>.377</td>
<td>.329</td>
<td>.534</td>
<td>.519</td>
</tr>
<tr>
<td>Baden-Württemberg, 2005</td>
<td>.066</td>
<td>.080</td>
<td>.106</td>
<td>.873</td>
<td>.279</td>
<td>--</td>
</tr>
<tr>
<td>Baden-Württemberg, 2009</td>
<td>.416</td>
<td>.467</td>
<td>.062</td>
<td>.743</td>
<td>.066</td>
<td>--</td>
</tr>
<tr>
<td>Brandenburg, 2005</td>
<td>.246</td>
<td>.388</td>
<td>.414</td>
<td>.460</td>
<td>.066</td>
<td>--</td>
</tr>
<tr>
<td>Brandenburg, 2009</td>
<td>.789</td>
<td>.124</td>
<td>.909</td>
<td>.831</td>
<td>.238</td>
<td>--</td>
</tr>
<tr>
<td>Berlin, 2005</td>
<td>.382</td>
<td>.387</td>
<td>.997</td>
<td>.981</td>
<td>.382</td>
<td>--</td>
</tr>
<tr>
<td>Berlin, 2009</td>
<td>.751</td>
<td>.122</td>
<td>.923</td>
<td>.139</td>
<td>.387</td>
<td>--</td>
</tr>
<tr>
<td>Bremen, 2005</td>
<td>.085</td>
<td>.043</td>
<td>.385</td>
<td>.401</td>
<td>.686</td>
<td>--</td>
</tr>
<tr>
<td>Bremen, 2009</td>
<td>.184</td>
<td>.507</td>
<td>.530</td>
<td>.713</td>
<td>.766</td>
<td>--</td>
</tr>
<tr>
<td>Hamburg, 2005</td>
<td>.588</td>
<td>.808</td>
<td>.110</td>
<td>.780</td>
<td>.870</td>
<td>--</td>
</tr>
<tr>
<td>Hamburg, 2009</td>
<td>.608</td>
<td>.575</td>
<td>.725</td>
<td>.461</td>
<td>.028</td>
<td>--</td>
</tr>
<tr>
<td>Hesse, 2005</td>
<td>.901</td>
<td>.754</td>
<td>.146</td>
<td>.346</td>
<td>.521</td>
<td>--</td>
</tr>
<tr>
<td>Hesse, 2009</td>
<td>.960</td>
<td>.958</td>
<td>.766</td>
<td>.001</td>
<td>.594</td>
<td>--</td>
</tr>
<tr>
<td>Lower Saxony, 2005</td>
<td>.070</td>
<td>.070</td>
<td>.595</td>
<td>.691</td>
<td>.715</td>
<td>--</td>
</tr>
<tr>
<td>Lower Saxony, 2009</td>
<td>.523</td>
<td>.098</td>
<td>.664</td>
<td>.713</td>
<td>.967</td>
<td>--</td>
</tr>
<tr>
<td>Mecklenburg-West Pomerania, 2005</td>
<td>.863</td>
<td>.141</td>
<td>.814</td>
<td>.078</td>
<td>.523</td>
<td>--</td>
</tr>
<tr>
<td>Mecklenburg-West Pomerania, 2009</td>
<td>.205</td>
<td>.113</td>
<td>.255</td>
<td>.708</td>
<td>.701</td>
<td>--</td>
</tr>
<tr>
<td>North Rhine-Westphalia, 2009</td>
<td>.508</td>
<td>.418</td>
<td>.534</td>
<td>.642</td>
<td>.852</td>
<td>--</td>
</tr>
<tr>
<td>Rhineland-Palatinate, 2005</td>
<td>.558</td>
<td>.812</td>
<td>.661</td>
<td>.733</td>
<td>.297</td>
<td>--</td>
</tr>
<tr>
<td>Rhineland-Palatinate, 2009</td>
<td>.225</td>
<td>.666</td>
<td>.370</td>
<td>.024</td>
<td>.104</td>
<td>--</td>
</tr>
<tr>
<td>Saarland, 2009</td>
<td>.622</td>
<td>.126</td>
<td>.022</td>
<td>.826</td>
<td>.579</td>
<td>--</td>
</tr>
<tr>
<td>Saxony, 2005</td>
<td>.474</td>
<td>.547</td>
<td>.538</td>
<td>.918</td>
<td>.685</td>
<td>--</td>
</tr>
<tr>
<td>Saxony, 2009</td>
<td>.714</td>
<td>.837</td>
<td>.405</td>
<td>.129</td>
<td>.216</td>
<td>--</td>
</tr>
<tr>
<td>Saxony-Anhalt, 2005</td>
<td>.710</td>
<td>.246</td>
<td>.537</td>
<td>.054</td>
<td>.786</td>
<td>--</td>
</tr>
<tr>
<td>Saxony-Anhalt, 2009</td>
<td>.299</td>
<td>.985</td>
<td>.829</td>
<td>.741</td>
<td>.904</td>
<td>--</td>
</tr>
<tr>
<td>Schleswig-Holstein, 2005</td>
<td>.224</td>
<td>.620</td>
<td>.275</td>
<td>.033</td>
<td>.923</td>
<td>--</td>
</tr>
<tr>
<td>Schleswig-Holstein, 2009</td>
<td>.583</td>
<td>.338</td>
<td>.854</td>
<td>.875</td>
<td>.984</td>
<td>--</td>
</tr>
<tr>
<td>Thuringia, 2005</td>
<td>.905</td>
<td>.692</td>
<td>.509</td>
<td>.961</td>
<td>.930</td>
<td>--</td>
</tr>
<tr>
<td>Thuringia, 2009</td>
<td>.613</td>
<td>.187</td>
<td>.776</td>
<td>.231</td>
<td>.081</td>
<td>--</td>
</tr>
</tbody>
</table>

$H_o$: Fractional Mandates ~ $U[0,1]$

All Years: $p$-value = .362

2005: $p$-value = .271

2009: $p$-value = .798

Notes: Entries denote the number of fractional mandates by party in the 2005 and 2009 federal elections, as explained in Appendix D. The upper panel does so for the national level, whereas the lower panel refers to the state level. $H_o$ refers to the null hypothesis that the number of fractional mandates is uniformly distributed on the unit interval. The respective $p$-values are based on Kolmogorov–Smirnov tests. For a detailed description of how mandates are allocated to parties, see Appendix B.
### Table A.6: Actual vs Predicted Ranking of Candidates, Structural Analysis of 2009 Federal Elections

<table>
<thead>
<tr>
<th>Actual Rank</th>
<th>Predicted Rank (as Fraction of Actual Rank)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>94.6%</td>
</tr>
<tr>
<td>2</td>
<td>5.1%</td>
</tr>
<tr>
<td>3</td>
<td>0.3%</td>
</tr>
<tr>
<td>4</td>
<td>0.0%</td>
</tr>
<tr>
<td>5</td>
<td>0.0%</td>
</tr>
</tbody>
</table>

**Notes:** Entries denote the frequency with which the predictions of the structural model in Appendix F coincide with observed outcomes, considering only candidates of the 5 major parties.

### Table A.7: Voters’ Partial Preference Orderings

<table>
<thead>
<tr>
<th>First-Choice Candidate</th>
<th>Second-Choice Candidate (as Fraction of First Choice)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CDU/CSU</td>
</tr>
<tr>
<td>CDU/CSU</td>
<td>--</td>
</tr>
<tr>
<td>SPD</td>
<td>16.1%</td>
</tr>
<tr>
<td>FDP</td>
<td>98.9%</td>
</tr>
<tr>
<td>The Left</td>
<td>18.0%</td>
</tr>
<tr>
<td>Green Party</td>
<td>14.5%</td>
</tr>
</tbody>
</table>

**Notes:** Entries denote the simulated relative frequency of voters' second-choice candidate, conditional on their first choice. See Appendix F for details.

### Table A.8: Distribution of District Winners under Expressive and Strategic Voting, Structural Analysis

<table>
<thead>
<tr>
<th>District Winner with Sincere Voters</th>
<th>District Winner with Sincere and Strategic Voters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CDU/CSU</td>
</tr>
<tr>
<td>CDU/CSU</td>
<td>70.8%</td>
</tr>
<tr>
<td>SPD</td>
<td>6.5%</td>
</tr>
<tr>
<td>FDP</td>
<td>0.0%</td>
</tr>
<tr>
<td>The Left</td>
<td>0.0%</td>
</tr>
<tr>
<td>Green Party</td>
<td>0.3%</td>
</tr>
</tbody>
</table>

**Notes:** Entries compare the simulated distribution of district winners in a first-past-the-post system with only sincere voters (left column) to the distribution that would obtain with a mixture of types (top row). Summing across columns gives the percentage of districts that would accrue to a particular party if all voters behaved sincerely, whereas summing across rows gives a party’s share of districts if 26.3% of voters behaved strategically. Consequently, adding the entries on the diagonal shows that about 90% of districts would accrue to the same party. See Appendix F for details on the simulation.