Inferring Strategic Voting*

Kei Kawai†
Northwestern University

Yasutora Watanabe‡
Northwestern University

January 2010

Abstract

We estimate a model of strategic voting and quantify the impact it has on election outcomes. Because the model exhibits multiplicity of outcomes, we adopt a set estimator. Using Japanese general-election data, we find a large fraction [75.3%, 80.3%] of strategic voters, only a small fraction [2.4%, 5.5%] of whom *voted for a candidate other than the one they most preferred* (*misaligned voting*). Existing empirical literature has not distinguished between the two, estimating misaligned voting instead of strategic voting. Accordingly, while our estimate of strategic voting is high, our estimate of misaligned voting is comparable to previous studies.

Keywords: strategic voting, set estimation, partially identified models, election

*We thank Kenichi Ariga, David Austen-Smith, Ivan Canay, Amrita Dhillon, Tim Feddersen, Marc Henry, Igal Hendel, Antonio Merlo, Aviv Nevo, Mark Satterthwaite, and Elie Tamer for helpful comments and suggestions.

†Department of Economics, Northwestern University, 2003 Sheridan Road, Evanston, IL 60208. Email: k-kawai@northwestern.edu

‡Department of Management and Strategy, Kellogg School of Management, Northwestern University, 2001 Sheridan Road, Evanston, IL 60208. Email: y-watanabe@kellogg.northwestern.edu

1 Introduction

Strategic voting in elections has been of interest to researchers since Duverger (1954) and Downs (1957). Models of strategic voting are fundamental to the study of political economy, and have been used to investigate topics ranging from performance of different electoral rules to information aggregation in elections. Whether voters actually behave strategically, however, is an empirical question.

Strategic voting is also of interest to politicians and voters. It is widely believed that if Ralph Nader had not run in the 2000 U.S. Presidential election, Al Gore would have won the election. The presence of minor candidates and third parties affects election outcomes, and the extent of that effect depends heavily on the fraction and behavior of strategic voters.

In this paper, we study how to identify and estimate a model of strategic voting and quantify the impact strategic voting has on election outcomes by adopting an inequality-based estimator. We estimate the model using aggregate municipality level data from the Japanese general election which uses plurality rule. In our counterfactual policy experiments, we investigate election outcomes under alternative electoral rules. Strategic voters are defined as those who make voting decisions conditioning on the event that their votes are pivotal. Unlike sincere voters who always vote according to their preferences, strategic voters do not necessarily vote for their most preferred candidate in plurality-rule elections with three or more candidates.[1]

In our paper, we make a clear distinction between *strategic voting,* as defined above (this is the standard definition in the theoretical literature[2]), and *voting for a candidate other than the one the voter most prefers* (hereafter referred to as *misaligned voting*). Strategic voters may vote for their most preferred candidate or they may not. Hence, the set of voters who engage in misaligned voting is only a subset of the set of strategic voters. Existing empirical literature has not distinguished between the two. In fact, previous attempts at estimating strategic voting have estimated misaligned voting instead of strategic voting. This distinction is important because the fraction of strategic voters is a model primitive while misaligned voting is an equilibrium object. In our paper we recover the extent of strategic voting, which allows us to conduct counterfactual policy experiments.

Our model is an adaptation of Myerson and Weber (1993) and Myerson (2002) with the addition of sincere voters.[3] We relax the equilibrium requirement that Myerson and Weber

[1] There are other behavioral models of voting, such as expressive voting (voters may vote for a candidate to send a signal). We focus on sincere voting and strategic voting, which have been the main focus of the emipirical literature.

[2] See, e.g., the entry of "strategic voting" in *The New Palgrave Dictionary of Economics* by Feddersen (2008).

[3] Our model can be naturally extended to elections with N candidates competing for N_S ($N_S < N$) seats

place on voters' beliefs on pivot probabilities. We use a weaker solution concept so that the outcome of the model is robust to different assumptions regarding voter beliefs and can better account for diverse patterns of outcome as observed in the data.[4] The consequence of adopting a weaker solution concept is that we have to deal with the issue of multiplicity of solution outcomes in identification and estimation.

The key source of the multiplicity of the solution outcomes - and hence the key source of difficulty in the identification of the model - is the presence of strategic voters. The difficulty stems from the fact that preference and voting behavior do not necessarily have a one-to-one correspondence for strategic voters. Our identification argument proceeds in three steps.

First we derive restrictions in terms of how preferences, which we write as a function of demographic characteristics, relate to voting behavior at the individual level. Unlike in other applications of discrete-choice models, the fact that a voter votes for candidate A does not imply that the voter preferred candidate A most. It could well be that the voter preferred candidate B over A, but voted for A instead because the voter believed that candidate B had little chance of winning. However, we can infer from the voter's behavior that the voter did not rank candidate A last in his order of preference. It is a weakly dominated strategy for all voters, sincere and strategic, to vote for their least preferred candidate.

Second, we relate aggregate variation in the vote share to demographic characteristics using two particular features often found in general-election data. The first feature is that general-election data typically consists of data from many elections taking place simultaneously (e.g., 646 elections for House of Commons in U.K., 435 elections for U.S. House of Representatives). This feature is essential for identification and estimation because we take each election to be our unit of observation. The second feature is that the breakdown of votes and demographic characteristics *within* each electoral district is available. (e.g., county-level breakdown of votes for U.S. Congressional Elections). This data structure allows us to relate variation in the vote share to variation in the demographic characteristics within a single electoral district, holding constant common components such as beliefs over tie probabilities and candidate characteristics. This partially identifies the preference parameters. (For the rest of the paper, we use the term "municipality" to denote the sub-district within an electoral district, such as counties. Note that several municipalities comprise one "district", which in turn corresponds to one election. See Figure 1.)

Lastly, we consider identification of the extent of strategic voting. Intuitively, the variation in the data that we would like to exploit is the variation in the voting outcome among municipalities (in different districts) with similar characteristics vis-à-vis the variation in the

under single non-transferrable voting as in Cox (1994).

[4]See footnote 12 for details.

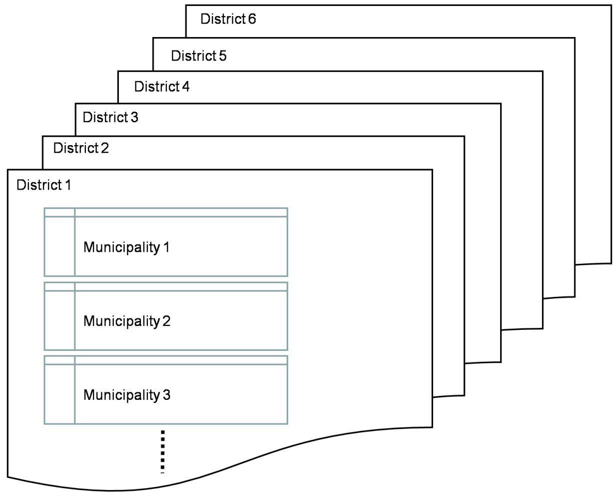

Figure 1: Data Structure. The district is our unit of observation, each of which is comprised of multiple municipalities. Breakdown of data is available at the municipality level.

vote shares and characteristics of other municipalities in the same district. For example, consider two liberal municipalities, one in a generally conservative electoral district and the other in a generally liberal district. Suppose that there are three candidates, a liberal, a centrist and a conservative candidate in both districts. If there are no strategic voters, we would not expect the voting outcome to differ across the two municipalities. However, in the presence of strategic voters, the voting outcome in these two municipalities could differ. If the strategic voters of the municipality in the conservative district believe that the liberal candidate has little chance of winning, those voters would vote for the centrist candidate, while strategic voters in the other municipality (in the liberal district) would vote for the liberal candidate according to their preferences (if they believe that the liberal candidate has a high chance of winning).

More generally, given the preference parameters, the model can predict what the vote share would be in each municipality if all of the voters voted according to their preferences. If there were no strategic voters, the difference between the actual outcome and the predicted sincere-voting outcome would only be due to random shocks. However, when there is a large number of strategic voters, the actual vote share can systematically diverge from the predicted outcome. This is due to the multiplicity of equilibria induced by strategic voters. Recall that strategic voters make voting decisions conditional on the event that their votes are pivotal. If the beliefs regarding the probability of being pivotal differ across electoral districts – and we have no reason to believe that they do not – the behavior of strategic voters will also differ across districts. This corresponds to different equilibria being played in different districts. Of course it is impossible to directly test for the relationship between voter behavior and voter beliefs regarding tie probabilities as beliefs are unobservable. However, we can still use the systematic difference between the predicted vote share and the actual vote share to partially identify the fraction of strategic voters.

Our estimation applies an estimator based on moment inequalities developed by Pakes, Porter, Ho and Ishii (2007). We use a bounds estimator because our voting model does not yield a unique outcome and we may only be able to set-identify the model parameters.

We use data on the Japanese House of Representatives elections for estimation.[5] Once the primitives of the model have been estimated, we investigate the extent of strategic voting using the estimated model. In our counterfactual policy experiments, we study how the outcome would change under proportional representation and under the assumption that all voters vote sincerely.

We find that a large proportion [75.3%, 80.3%] of voters are strategic voters. We also recover the extent of misaligned voting once we estimate the model, by simulating the equilibrium behavior. Our results show that [2.4%, 5.5%] of the voters engage in misaligned voting, or [3.0%, 7.3%] of the strategic voters. In our first counterfactual experiment, in which we introduce proportional representation, we find that the number of votes for major parties decreases by a large margin, and the number of seats decreases by an even greater margin.

In our second counterfactual experiment, we investigate what the outcome would be if all voters vote sincerely under plurality rule. We find that the number of seats for the parties would change significantly: one party would add [17, 40] seats while another would lose [20, 45] seats out of a total of 175 seats. Even though the extent of misaligned voting is small [2.4%, 5.5%], the impact on the number of seats is considerable because the winning margin is often small.

Related Literature There are both an experimental and an empirical literature on strategic voting in elections. In small-scale laboratory experiments with three candidates under plurality rule, Forsythe, Myerson, Rietz, and Weber (1993, 1996) find evidence of strategic voting.[6] They also find that strategic voting is more likely to occur if pre-election coordination devices such as polls and shared voting histories are available.

There is also a large empirical literature on strategic voting (see, e.g., Alvarez and Nagler (2000), Blais, Nadeau, Gidengil, and Nevitte (2001) and papers cited therein). Previous work in the literature has attempted to identify strategic voting by comparing each voter's

[5] Our implementation does not depend on any specific institutional feature of the Japanese election. Our approach can be applied to any election with plurality rule or single non-transferrable voting.

[6] See Holt and Smith (2005), Morton and Williams (2006), Palfrey (2006), and Rietz (2008) for a survey of the literature on experiments.

actual vote to his preferences. Voter preferences are proxied by measures such as voting behavior in previous elections and surveys eliciting voter preferences. However, as pointed out earlier, the difference between voting and preferences is a measure of misaligned voting rather than that of strategic voting. Accordingly, our estimate of misaligned voting [2.4%, 5.5%] is comparable to the estimates of strategic voting reported in the previous literature, which ranges from 3% to 17%.[7]

One closely related paper is Degan and Merlo (2007). They consider the falsifiability of sincere voting, and show that individual-level observations of voting in at least two elections are required to falsify sincere voting. They examine whether there exists a preference profile that is consistent with the observed election outcome without imposing any relationship between preferences and observable covariates. Our approach relates preferences to voter covariates within a standard discrete-choice framework. Identification of voter preferences and the fraction of strategic voters is then possible without requiring data on repeated voting records. This is analogous to papers such as Berry, Levinsohn, and Pakes (1995) which estimate individual preferences using aggregate data.[8]

Our paper is also related to the literature on strategic voter turnout. Shachar and Nalebuff (1999) and Coate, Conlin, and Moro (2008) estimate a model of voter turnout in which voter turnout is a function of the expected closeness of the election. These papers study turnout focusing on two candidate elections, a setting in which the issue of strategic voting does not arise. Our paper focuses on the issue of strategic voting instead of strategic turnout, although it is conceptually straightforward to extend our approach to a model of elections with both strategic voting *and* strategic turnout. We discuss this extension at the end of Section 4.

We describe the model in the next section, and explain the data in Section 3. Details on identification and estimation are provided in Section 4. Section 5 presents the results and the counterfactual experiments. Finally, we close the paper with concluding remarks in Section 6.

[7] See Alvarez and Nagler (2000), Blais, Nadeau, Gidengil, and Nevitte (2001) and papers cited therein.

[8] Regarding the use of aggregate data, the political science literature has been concerend about the issue of ecological inference (See, e.g., King, 1997). King (1997) proposes a solution to this problem by assuming a random coefficients type model with a particular functional form. Our approach can be thought of as microfounding the distribution of the random coefficients in his statistical model. We do so by considering a game theoretic model of voting.

2 Model

2.1 Model Set-up

Our model is an adaptation of Myerson and Weber (1993) [hereinafter denoted as MW] and Myerson (2002). We model plurality-rule elections in which K candidates compete for one seat. Voters cast a vote for one candidate,[9] and the candidate receiving the highest number of votes is elected to office (ties are broken with equal probability). We restrict attention to the case when $K \geq 3$ since strategic voting is otherwise not an issue. There are M municipalities in an electoral district, and we use subscript $m \in \{1, 2, ..., M\}$ to denote a municipality. There are a finite number of voters, $\sum_{m=1}^{M} N_m < \infty$, who are the players of the game (N_m is the number of voters in municipality m). Voter n's utility from having candidate k in office is

$$u_{nk} = u(\mathbf{x}_n, \mathbf{z}_k) + \xi_{km} + \varepsilon_{nk},$$

where $\mathbf{x}_n$ are voter characteristics, $\mathbf{z}_k$ are candidate characteristics, ξ_{km} is a candidate-municipality shock, such as the ability of a candidate to bring pork to municipality m, and ε_{nk} is an i.i.d. preference shock.

We consider two types of voters, sincere (behavioral) and strategic (rational). A sincere voter casts his vote for the candidate he prefers most, i.e., a sincere voter votes for candidate k if and only if $u_{nk} \geq u_{nl}$, $\forall l$. On the other hand, a strategic voter casts his vote taking into consideration that the only events in which his vote is pivotal are when the election is exactly tied or when the second place candidate is one vote behind. When voter n is pivotal and he casts the decisive vote between k and l, he changes the outcome of the election. In this situation, voting for candidate k gives utility $\frac{1}{2}(u_{nk} - u_{nl})$.[10] Hence, if we let $T_n = \{T_{n,kl}\}_{kl}$ denote voter n's beliefs that candidates k and l will be tied for first place or that k will be one vote behind l, the expected utility from voting for candidate k is given by[11]

$$\bar{u}_{nk}(T_n) = \frac{1}{2} \sum_{l \in \{1,..,K\}} T_{n,kl}(u_{nk} - u_{nl}),$$

[9] We abstract from the issue of voter abstention. We discuss the issue of turnout at the end of Section 4.

[10] Voter n's vote is pivotal in two cases. First, consider the case when candidates k and l are exactly tied without voter n's vote. In this case, candidate k wins if voter n votes for k. Because ties are broken with equal probability for each candidate, the utility from voting for candidate k is $u_{nk} - \frac{1}{2}(u_{nk} + u_{nl})$. Second, consider the case when candidate k is one vote behind candidate l without voter n's vote. The two candidates will tie if voter n votes for candidate k, while candidate l wins if voter n does not. Thus, the utility from voting for k is $\frac{1}{2}(u_{nk} + u_{nl}) - u_{nl}$. Therefore, in both cases, the utility from voting for candidate k is $\frac{1}{2}(u_{nk} - u_{nl})$.

[11] We assume that voter beliefs over three-way ties are infinitesimal compared to two-way ties, as is commonly assumed in the literature.

as in MW. Strategic voters vote for candidate k if and only if $\bar{u}_{nk}(T_n) \geq \bar{u}_{nl}(T_n)$, $\forall l$. Depending on the value of T_n, strategic voters may choose to vote for any candidate other than the one he prefers the least (i.e. the candidate k with the lowest value of u_{nk}). We come back to this fact when we discuss identification.

Note that we distinguish strategic voting and misaligned voting as discussed in the Introduction. We define misaligned voting as casting a vote for a candidate other than the one the voter most prefers. Hence, only strategic voters engage in misaligned voting, but a strategic voter may or may not engage in misaligned voting. In other words, being a strategic voter is a necessary condition for misaligned voting, but not a sufficient condition.

We assume that for at least some candidate pair $\{k, l\}$, beliefs over pivot probability, $T_{n,kl}$, is non-zero. Even if there is an obvious frontrunner, there is always some chance that a vote will be pivotal although it may be very small. As long as some $T_{n,kl}$ is always non-zero, we can normalize $T_{n,kl}$ so that $\sum_k \sum_{l>k} T_{n,kl} = 1$. This normalization is possible because a voter's decision is determined by the relative size of $\bar{u}_{nk}(T_n)$, which is not affected by rescaling $T_{n,kl}$ by a constant factor.

We denote the type of voter n in municipality m by a random variable $\alpha_{nm} \in \{0, 1\}$ drawn from a binomial distribution, where $\alpha_{nm} = 0$ denotes the sincere voter and $\alpha_{nm} = 1$ denotes the strategic voter. We also let the mean of the binomial distribution to be a random variable drawn for each municipality from some distribution F_α. Then the probability that voter n in municipality m is a strategic voter can be written as

$$\Pr(\alpha_{nm} = 1|\alpha_m) = \alpha_m,$$

where α_m is the municipality-level random term drawn from F_a and we assume that $\alpha_{nm} \perp \alpha_{n'm}$ $\forall n, n'$ conditional on α_m. The probability that the voter is sincere is $\Pr(\alpha_{nm} = 0|\alpha_m) = 1 - \alpha_m$.

We make the following assumption on beliefs T_n following MW.

Assumption *Beliefs over tie probabilities T_n are common across all voters in the same electoral district, i.e., $T_n = T$, $\forall\ n \in \{1, ..., N_1\} \cup ... \cup \{1, ..., N_M\}$.*

This assumption simply imposes voters in the same electoral district to have common beliefs over pivot probabilities, T. The assumption reflects the fact that information regarding the expected outcome of the election is widely available from news reports and poll results. By gaining access to this kind of information, voters in the same electoral district can form similar beliefs regarding the outcome.

Let $V_{k,m}^{SIN}$ be the fraction of votes cast by sincere voters for candidate k in municipality m, and let $V_{k,m}^{STR}(T)$ be the fraction of votes cast by strategic voters for candidate k. Note that $V^{STR}_{k,m}(T)$ is a function of beliefs, T. We can write these fractions as

$$V^{SIN}_{k,m} = \frac{\sum_{n=1}^{N_m}(1-\alpha_{nm})\cdot \mathbf{1}\{u_{nk}\geq u_{nl},\ \forall l\}}{\sum_{n=1}^{N_m}(1-\alpha_{nm})}, \tag{1}$$

$$V^{STR}_{k,m}(T) = \frac{\sum_{n=1}^{N_m}\alpha_{nm}\cdot \mathbf{1}\{\overline{u}_{nk}(T)\geq \overline{u}_{nl}(T),\ \forall l\}}{\sum_{n=1}^{N_m}\alpha_{nm}}. \tag{2}$$

The total vote share for candidate k in municipality m is then

$$V_{k,m}(T) = \frac{\sum_{n=1}^{N_m}(1-\alpha_{nm})}{N_m}V^{SIN}_{k,m} + \frac{\sum_{n=1}^{N_m}\alpha_{nm}}{N_m}V^{STR}_{k,m}(T).$$

Note that these expressions are approximated by their expectation as the number of voters, N_m, becomes large;

$$V^{SIN}_{k,m} \underset{p}{\rightarrow} \quad v^{SIN}_{k,m} \equiv \iint \mathbf{1}\{u_{nk}\geq u_{nl},\ \forall l\}]g(\boldsymbol{\varepsilon})d\boldsymbol{\varepsilon} f_m(\mathbf{x})d\mathbf{x}, \text{ and}$$

$$V^{STR}_{k,m}(T) \underset{p}{\rightarrow} \quad v^{STR}_{k,m}(T) \equiv \iint \mathbf{1}\{\overline{u}_{nk}(T)\geq \overline{u}_{nl}(T),\ \forall l\}g(\boldsymbol{\varepsilon})d\boldsymbol{\varepsilon} f_m(\mathbf{x})d\mathbf{x},$$

where f_m denotes the distribution of the demographic characteristics, $\mathbf{x}$, in municipality m, and g denotes the distribution of idiosyncratic shock, $\boldsymbol{\varepsilon}_n = (\varepsilon_{n1},...,\varepsilon_{nK})$. We obtain these expressions by computing the vote share for candidate k among voters of a given demographic characteristics $\mathbf{x}$, and then integrating this vote share with respect to characteristics $\mathbf{x}$ using its distribution f_m. We obtain a similar expression for the total vote share as N_m becomes large:

$$V_{k,m}(T) \underset{p}{\rightarrow} v_{k,m}(T) \equiv (1-\alpha_m)v^{SIN}_{k,m} + \alpha_m v^{STR}_{k,m}(T). \tag{3}$$

2.2 Solution Outcome

Until now, our model has been the same as the one considered in MW with the only difference being the presence of sincere voters. In order to take the model to the data, we relax the consistency requirement on beliefs, T, that MW place in equilibrium. The equilibrium requirement on voters' beliefs imposed by MW results in outcomes that may not rationalize diverse patterns of actual election data even when we add sincere voters to their model.[12]

[12] In an election with three candidates, the original equilibrium of MW predicts that either (i) the first place candidate wins, and the second and third place candidates receive exactly the same number of votes (with corresponding beliefs $\{T_{12}, T_{13}, T_{23}\} = \{p, 1-p, 0\}$ for some $p \in [0,1]$) or (ii) the third place candidate

To better account for the variation in the data and be robust to alternative specifications regarding beliefs, we weaken MW's consistency requirement on beliefs. Hence, our set of solution outcomes is a superset of the set of MW equilibria.

Let us denote the district level vote share, which is the total number of votes obtained by a candidate divided by the total number of votes cast in the election, by $V_k \equiv \sum_{m=1}^{M} N_m V_{k,m} \Big/ \sum_{m=1}^{M} N_m$. MW imposes the following consistency requirement in equilibrium: $V_k > V_l \Rightarrow T_{kj} \geq \varepsilon T_{lj}$, $\forall \varepsilon \in [0,1)$, $\forall k,l,j$. This implies that pivot probabilities involving candidates with low vote shares are zero. The consistency requirement (**C1**) we impose between beliefs, T, and the election outcomes is a weaker version of MW's ordering condition:

C1 : For an election with K candidates,

$$V_k > V_l \ \Rightarrow T_{kj} \geq T_{lj} \ \ \forall k,l,j \in \{1,...,K\}.$$

This condition implies that pivot probabilities involving candidates with high vote shares are larger than those with low vote shares. For the case of $K=3$ with vote shares $V_1 > V_2 > V_3$, **C1** implies that $T_{12} \geq T_{13} \geq T_{23}$, i.e., beliefs on the pivot probability between candidates 1 and 2, T_{12}, is higher than those between candidates 1 and 3, T_{13}, and so on.

Our second condition, **C2**, simply requires that given beliefs T, strategic voters vote optimally (and sincere voters vote for their most preferred candidate). Now we define the solution outcome of the voting game.

Definition *A set of solution outcomes $W \subseteq \Delta^{{}_K C_2} \times \left(\times_{m=1}^{M} \Delta^K\right)$ is defined as the set* $W = \left\{T, \left\{\{V_{k,m}\}_{k=1}^{K}\right\}_{m=1}^{M}\right\}$ such that **C1** and **C2** are satisfied.

$$\mathbf{C1} : V_k > V_l \ \Rightarrow T_{kj} \geq T_{lj} \ \ \forall k,l,j \in \{1,...,K\}.$$

$$\mathbf{C2} : V_{k,m} = \frac{\sum_{n=1}^{N_m}(1-\alpha_{nm})}{N_m} V_{k,m}^{SIN} + \frac{\sum_{n=1}^{N_m}\alpha_{nm}}{N_m} V_{k,m}^{STR}(T)$$

A few comments are in order. First, the set of solution outcomes, W, is not empty: That is, a solution outcome exists. This can be shown in a similar way as in the proof of Theorem

receives zero votes (with beliefs $\{T_{12}, T_{13}, T_{23}\} = \{1,0,0\}$).

Even if we (1) introduce sincere voters, (2) add shocks to voter preferences or (3) introduce randomness to the fraction of strategic voters (or any combination of (1), (2), and (3)) to MW, there would still only be two types of equilibria: One with beliefs $\{T_{12}, T_{13}, T_{23}\} = \{p, 1-p, 0\}$ and the other with $\{T_{12}, T_{13}, T_{23}\} = \{1,0,0\}$. As before, Equilibrium (i) has the undesirable property that the second and third candidates receive exactly the same number of votes. In equilibrium (ii), all three candidates can receive a positive and different number of votes, but the only beliefs that can support the equilibrium is $\{T_{12}, T_{13}, T_{23}\} = \{1,0,0\}$, which is a strong assumption to impose, unlikely to hold in many races.

1 in MW. The proof is in Appendix A. Second, W is not a singleton in general. In order to cope with the issue of multiplicity of solution outcomes, we adopt an inequality-based estimator in our estimation. Third, W is a superset of the set of equilibria considered in MW. This is because condition **C1** is weaker than that of MW. Finally, note that W does not depend on the information structure of the model, i.e., whether we assume that the voters know the realization α_{nm} and ε_{nk} of other voters, or only their distributions.

Finally, we remark on the empirical restriction implied by our solution outcome. Note that **C2** embodies the restriction that no voter votes for his least preferred candidate through equations (1) and (2), which give the expressions for vote shares of the sincere and strategic voters. However, beyond this restriction, the model leaves considerable freedom in how $V_{k,m}^{STR}(T)$ is linked to voter preferences. This is because the solution outcome does not pin down T (only a weak restriction is imposed via **C1**), nor do we observe the value of T. Hence, the empirical content of our solution outcome would be similar if we had instead adopted rationalizability[13] as our solution concept (See Bernheim, 1984, Pearce, 1984).

3 Data

We use data from the Japanese House of Representatives election held on September 11, 2005. Out of a total number of 480 Representatives, 300 members were elected by plurality rule. We use the data from these 300 plurality-rule elections.[14] For each electoral district, the breakdown of vote-share data is available by municipality as shown in Figure 1. An electoral district is usually comprised of several municipalities (9.26 on average, in our sample).[15] This particular data structure plays an important role in our identification.

We obtained the data on the vote shares and candidate characteristics from *Yomiuri Shimbun*, a national newspaper publisher. The demographic characteristics we use are obtained from the *Social and Demographic Statistics of Japan* published by the Statistics Bureau of the Japanese Ministry of Internal Affairs and Communications.[16] We match these two data sets at the municipality level.

[13] To be more precise, *perfect rationalizability* of Bernheim (1984) or *cautious rationalizability* of Pearce (1984).

[14] An additional 180 Representatives were elected by proportional representation from 11 regional electoral districts. In proportional representation, voters cast ballots for parties, and a closed list is used to determine the winner. It is possible for a person to be a candidate in both plurality and proportional elections. When two candidates are ranked equally on the party list, the results of the plurality rule election affect the relative rank of the two candidates. Only the LDP and the DPJ ranked more than two candidates equally in this election.

[15] In the vast majority of cases, municipal borders do not cross electoral districts.

[16] The basic information for the data is available at http://www.stat.go.jp/english/data/ssds/outline.htm and http://www.stat.go.jp/english/data/zensho/intex.html.

Out of a total of 300 districts, we keep the districts that satisfy the following criteria.

(i) There are three or four candidates,[17] and the composition of the candidates' parties in the district is any three or four of the following four parties; the Liberal Democratic Party (LDP), the Democratic Party of Japan (DPJ), the Japan Communist Party (JCP), or the Yusei (YUS). Technically, the Yusei is not a single party, but we grouped former LDP candidates who split away from the LDP and ran on a common platform against postal privatization.

(ii) There are at least two municipalities within the electoral district.

(iii) There are no mergers of municipalities within the electoral district during the period from April 1, 2004 to the day of the election.

We are left with 175 electoral districts. We drop samples that do not satisfy criterion (i) because we treat party affiliation as a candidate characteristic, and we cannot precisely estimate the coefficients on parties that only fielded a very small number of candidates. Criterion (i) ensures that we have enough elections with the same combination of parties fielding candidates to construct our moment inequalities.[18] We need criterion (ii) because our estimation requires at least two municipalities in each electoral district. Criterion (iii) is required to deal with an issue that arises when merging two data sets. Because the demographics data and the vote share data are collected on different dates (April 1, 2004 and September 11, 2005), municipalities that merged with others between these dates are dropped from the sample. In some cases, however, we are able to match the data properly. When this is possible, we keep the merging municipalities in the sample.

We report the descriptive statistics of electoral-district vote shares in Table 1. There are 9.26 municipalities per electoral district on average. The average winner's vote share is about 52% and the winning margin is about 14%. The mean vote share of the winner is higher in three-candidate districts (52.9%) than in four-candidate districts (41.2%). The mean winning margin is also higher in three-candidate districts (14.2%) than in four-candidate districts (9.4%). Similarly, the margin between the second- and third-place candidates is

[17] We do not include 15 observations in which there are only two candidates for technical reasons. We use an estimator of Pakes, Porter, Ho and Ishii (2007) in our estimation, but it is not clear whether their method of inference can be applied when some of the parameters are point-identified. While two candidate districts contain no information about the extent of strategic voting, they point-identify some of the preference parameters of the voters. For our estimation, this is problematic. Alternatively, we can use other inequality based estimators (e.g. Chernozhkov, Hong and Tamer (2007)), which give consistent estimates even when parameters are point identified. However, this comes at a very high computational cost in our application.

[18] The Kagoshima 5th District is dropped from the sample because no other district had the same combination of parties fielding candidates (LDP, JCP, YUS) as this district. This is the only district we dropped that satisfied all three criteria.

	mean	st. dev.	min	max	# obs
# of municipalities per district	9.26	7.14	2	36	175
3-candidate district	8.67	6.82	2	36	158
4-candidate district	14.71	7.67	3	36	17
winner's vote share (%)	51.74	6.69	28.98	73.61	175
3-candidate district	52.87	5.60	36.03	73.61	158
4-candidate district	41.23	6.84	28.98	55.89	17
winning margin (%)	13.71	10.15	0.06	53.91	175
3-candidate district	14.17	10.09	0.16	53.91	158
4-candidate district	9.40	9.71	0.06	35.50	17
margin between 2nd and 3rd (%)	28.47	9.46	0.57	23.32	175
3-candidate district	30.37	7.40	4.74	43.32	158
4-candidate district	10.71	8.04	0.57	23.32	17
vote share – JCP	7.74	3.00	2.77	23.30	170
vote share – DPJ	38.37	8.82	10.78	60.10	175
vote share – LDP	49.71	8.89	22.00	73.62	175
vote share – YUS	35.02	8.87	14.50	49.58	22

Table 1: Descriptive Statistics of Electoral Districts – Vote Shares

significantly lower in four-candidate districts than in three-candidate districts. The last four rows report the vote-share breakdown for the four political parties. The mean vote share of the LDP is 49.7%, the highest among all parties. It is followed by the DPJ with 38.4%, the YUS with 35.0% and the JCP with 7.7%.[19]

Table 2 reports the descriptive statistics of candidate characteristics. The first three rows contain information on the candidates' hometowns.[20] The next three rows provide descriptive statistics on the candidates' political experience. An average of 1.32 (in three-candidate districts) and 1.47 (in four-candidate districts) candidates are incumbents. Note that the number of incumbents is higher than 1 because some candidates who had previously been elected to the House of Representatives in a proportional-rule election ran in the plurality election. Less than 0.51 candidates on average have previously held public office.[21]

Table 3 reports the descriptive statistics of the municipalities' demographic characteristics. The mean income per capita is about 3.16 million yen (about $35,000), and the mean

[19] Note that the sum of these percentages is greater than 100%. This is because not all parties field candidates in every district.

[20] In case a candidate has a hometown in his/her electoral district (as reported in the first row), we have additional information on candidates' hometowns that identifies exactly which municipality the candidate's hometown is in. We do not report it here, but use it in our estimation.

[21] This includes former and current municipality councillors, mayors, members of a prefectural assembly, prefectural governors, and the Members of the Houses of Councillors, as well as former Members of the House of Representatives.

	3 cand. district	4 cand. district
# of candidates w/ hometown in district	1.01 (0.96)	1.71 (1.05)
# of candidates w/ hometown in prefecture	0.95 (0.86)	0.71 (0.92)
# of candidates w/ hometown in another pref.	1.04 (0.82)	1.58 (1.23)
# of incumbents	1.32 (0.53)	1.47 (0.51)
# of candidates who previously held public office	0.51 (0.62)	0.35 (0.49)
# of candidates with no exp. in public office	1.16 (0.67)	2.18 (0.73)
# of observations	158	17

Table 2: Descriptive Statistics of Electoral Districts – Candidate Characteristics. The mean of each variable is reported. Standard errors are in parenthesis.

		mean	st. dev.	min	max	# obs
income per capita (in million yen)		3.16	0.42	2.27	6.47	1,621
years of schooling	$\leq$ 11 years (%)	35.00	12.37	7.16	71.08	1,621
	12-14 years (%)	45.41	6.37	20.09	62.59	1,621
	15-16 years (%)	9.83	3.34	2.86	19.41	1,621
	$\geq$ 16 years (%)	9.76	5.86	1.51	39.38	1,621
population above age 65 (%)		22.45	7.16	8.06	49.71	1,621

Table 3: Descriptive Statistics of Municipalities

length of schooling is about 12 years on average. The mean fraction of the population above age 65 is 22.5 percent. In the estimation, we use the distribution of demographic characteristics, which is readily available for years of schooling and age. Regarding income, only the mean of the distribution was available at the municipality level. We use the prefectural Gini coefficients as well as the average income to construct the distribution.[22]

[22] We have data on the total taxable income and the total number of taxpayers for each municipality. The mean income for each municipality can be computed from these numbers. We compute the quantiles of the income distribution by assuming a log-normal distribution where the variance is calculated by fitting the prefecture-level income distribution. Data on the prefecture-level income distritubtion is obtained from *the 2004 National Survey of Family Income and Expenditure* published by the Statistics Bureau of the Japanese Ministry of Internal Affairs and Communications.

4 Identification and Estimation

We first describe the econometric specification of the model we have presented in Section 2 in order to facilitate our identification and estimation arguments. Then, we discuss identification of the model and estimation.

4.1 Specification

We specify the utility function of voter n in municipality m with candidate k elected to office as

$$u_{nmk} = u(\mathbf{x}_n, \mathbf{z}_{km}; \theta^{PREF}) + \xi_{km} + \varepsilon_{nk},$$

where ξ_{km} is an i.i.d. idiosyncratic candidate-municipality level shock which follows a normal distribution, $N(0, \theta_\xi)$, denoted as F_ξ, and ε_{nk} is an i.i.d. idiosyncratic voter-candidate level shock which follows a Type-I extreme value distribution. An example of ξ_{km} is the candidate's ability to bring pork spending to municipality m. θ^{PREF} is a vector of preference parameters. $\mathbf{x}_n$ denotes the characteristics of voter n, including years of education, income level, and an indicator of whether or not the voter is above age 65. $\mathbf{z}_{km} = \{\mathbf{z}_k^{POS}, \mathbf{z}_{km}^{QLTY}\}$ is a vector of observable attributes of candidate k in municipality m. We partition $\mathbf{z}_{km}$ depending on how it interacts with voter characteristics. Let $\mathbf{z}_k^{POS}$ be the attributes of candidate k which are related to his ideological position such as his party affiliation. Let $\mathbf{z}_{km}^{QLTY}$ be other non-ideological attributes of candidate k such as the candidate's previous political experience and an indicator of whether municipality m is the candidate's hometown (which is why $\mathbf{z}_{km}$ is indexed by m). As for $u(\mathbf{x}_n, \mathbf{z}_{km}; \theta^{PREF})$, we assume a functional form with a quadratic loss term in the distance between the voter's and the candidate's ideological positions:

$$u(\mathbf{x}_n, \mathbf{z}_{km}; \theta^{PREF}) = -(\theta^{ID}\mathbf{x}_n - \theta^{POS}\mathbf{z}_k^{POS})^2 + \theta^{QLTY}\mathbf{z}_{km}^{QLTY},$$

where $\theta^{PREF} = \{\theta^{ID}, \theta^{POS}, \theta^{QLTY}\}$. We consider a unidimensional ideological space, and let the ideology of the voter be a function of his demographics, $\theta^{ID}\mathbf{x}_n$, and the ideology of the candidate be $\theta^{POS}\mathbf{z}_k^{POS}$. The utility of the voter depends on the distance between his ideology, $\theta^{ID}\mathbf{x}_n$, and that of the candidate, $\theta^{POS}\mathbf{z}_k^{POS}$, which is captured by the quadratic term. The additive term captures the non-ideological component of utility, which we write as $\theta^{QLTY}\mathbf{z}_{km}^{QLTY}$.

As described in the model section, the objective of a sincere voter is to vote for candidate k, who gives the highest value of u_{nmk}, while the objective of a strategic voter is to vote for candidate k, who gives the highest value of $\bar{u}_{nmk}(T)$, where

$$\bar{u}_{nmk}(T) = \sum_{l \in \{1,..,K\}} T_{kl}(u_{nmk} - u_{nml}).$$

As we discussed in Section 2, we assume that for at least some candidate pair $\{k, l\}$, T_{kl} is positive, no matter how small. This allows us to normalize T so that $\sum_k \sum_{l>k} T_{kl} = 1$, because utility representation is invariant to multiplication by a constant factor.

Recall that we denote the type of voter n in municipality m by a random variable $\alpha_{nm} \in \{0, 1\}$ drawn from a binomial distribution, where $\alpha_{nm} = 0$ denotes the sincere voter and $\alpha_{nm} = 1$ denotes the strategic voter. We also let the mean of the binomial distribution be a random variable drawn for each municipality from some distribution F_α. Then the probability that voter n in municipality m is a strategic voter can be written as

$$\Pr(\alpha_{nm} = 1 | \alpha_m) = \alpha_m,$$

where α_m is the municipality-level random term drawn from a Beta distribution, $Beta(\theta_{\alpha 1}, \theta_{\alpha 2})$, denoted as F_α.

4.2 Identification

In this subsection, we discuss the identification of the model when we let the number of districts (denoted as D) go to infinity. As described in the Data Section, our election data includes observations from many districts, for each of which we have a municipality-level breakdown of vote-share data and demographic characteristics. In terms of our notation, the number of districts is large ($D \to \infty$), but the number of municipalities per electoral district, denoted by M^d, is small ($M^d < \infty$, $\forall d \in \{1, ..., D\}$). We assume that voting games (i.e., elections) are played in D districts independently of each other, and we treat each district as a unit of observation.

Our identification argument proceeds in two steps. We first discuss partial identification of preference parameters. Then, given partial identification of preference parameters, we discuss partial identification of the fraction of strategic voters.

4.2.1 Partial Identification of Preference Parameters

Preference parameters are (partially) identified by the relationship between demographic and vote-share variation within each electoral district that we observe in the data. In order to exploit this variation for identification of preference parameters, the main restriction we use is that voters do not vote for their least preferred candidate. This restriction, however, does not give us point identification: The restriction only implies whom a voter will *not* vote for, but it does not imply whom a voter *will* vote for. The question of whom a voter will vote for is determined by T^d (beliefs over tie probabilities in district d) which is not observable to the econometrician nor uniquely determined by our solution concept.

If T^d were known (either observed or uniquely pinned down by our solution concept), the data would point-identify the preference parameters. As T^d is neither uniquely determined nor observable, the identified set of preference parameters is the union of parameter values, each corresponding to a value of $T^d \in \Delta^{{}_K C_2}$. Notice how our consistency requirement on the beliefs **C1** constrains the identified set by putting a restriction on the set of values T^d can take. Without **C1**, T^d can take any value as long as it adds up to one. Below, we illustrate how parameters are partially identified (and not point identified).

Consider the preference parameter, $\theta^{above65}$, which captures the effect of age (being 65 or older) on ideology, for the case of $K = 3$. Take two municipalities ($\{m^d_{old}, m^d_{young}\}$) in the same district, one with a high proportion of voters above age 65 (say 30%) and the other with a low proportion (say 20%), but otherwise with similar demographic characteristics. Now take pairs of municipalities in other districts ($\{m^{d'}_{old}, m^{d'}_{young}\}$, $\{m^{d''}_{old}, m^{d''}_{young}\}$, ...), that have similar demographic characteristics as the first pair (i.e., one with 30% of voters above age 65 and the other with 20%. Other demographic characteristics are similar to the first pair). It is important to note that each pair belongs to the same district and that we have many such pairs (from many different districts). Suppose that the vote share for Party A in the "old" municipalities are 5% higher than in the "young" municipalities on average. Then this implies that being older makes the voters become ideologically closer to Party A: But how much closer depends on the beliefs T^d.

In order to exhibit how T^d affects identification of $\theta^{above65}$, consider two polar cases as in Figure 2:

> Case 1: T^d is such that the tie probability between candidates from Parties B and C is close to one, and that the other two tie probabilities are close to zero, for all d ($T^d_{BC} \approx 1$, $T^d_{AB} \approx T^d_{AC} \approx 0$, $\forall d$).
>
> Case 2: T^d is such that the tie probability between candidates from Parties A and C is close to zero, and that the other two tie probabilities are near 0.5, for all d ($T^d_{AB} \approx T^d_{BC} \approx 0.5$, $T^d_{AC} \approx 0$, $\forall d$).

In Case 1, no strategic voter votes for the Party A candidate; hence, the 5% increase in the vote shares of Party A candidates in the "old" municipalities must be attributed to the difference in the sincere voters' behavior alone. Because the 5% increase must be explained

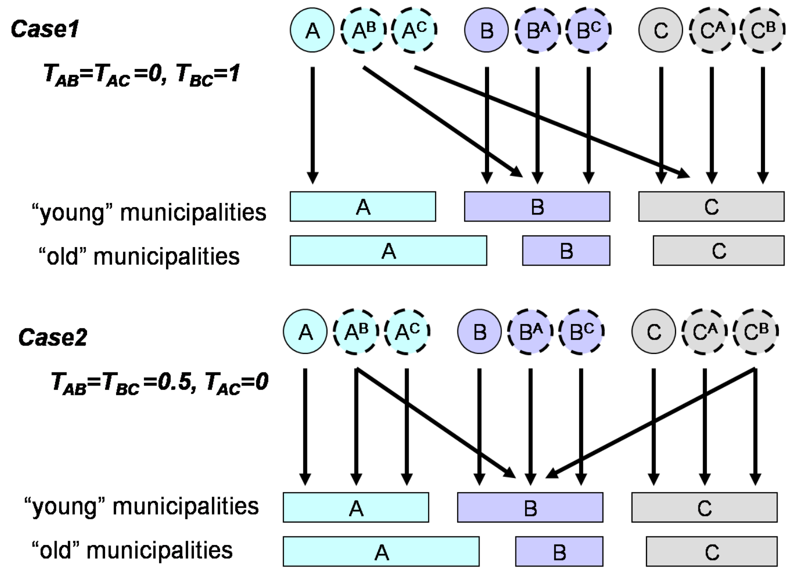

Figure 2: Identification of Preference. Sincere voters are illustrated with regular circles, and strategic voters, with dotted circles. The letters inside the circle indicate the most-preferred candidate and the superscripts indicate the second preferred candidate for strategic voters. The rectangles indicate the respective vote shares. In Case 1, only sincere voters who prefer candidate A the most vote for A. In Case 2, both sincere and strategic voters who prefer candidate A the most vote for A. In Case 1, the 5% difference in the vote share is then attributed to the difference in the behavior of only sincere voters, while it is attributed to the difference in the behavior of both types of voters in Case 2. Thus, the effect of demographic characteristics on utility depends on T.

only by the fraction of the population that is sincere, the effect of the parameter $\theta^{above65}$ must be quite large. In Case 2, the votes for Party A candidates come not only from sincere voters, but also from strategic voters. The 5% increase in the vote share for Party A candidates can then be accounted for by the difference in the behavior of both sincere *and* strategic voters. Thus, compared to Case 1, the value of $\theta^{above65}$ will be relatively small in Case 2 because we can attribute the 5% increase to the difference in the behavior of both types of voters. As T^d is unobservable, we cannot rule out Case 1 nor Case 2. Thus the identified set for $\theta^{above65}$ will be a set that includes the values implied by Case 1and by Case 2.

The parameters on candidate characteristics, θ^{POS} and θ^{QLTY}, can similarly be (partially) identified by taking municipalities across districts and relating the variation in the vote share and candidate characteristics. For example, the effect on utility of electing a candidate with no experience is identified by the difference in the vote shares between candidates with no experience and those with experience, controlling for other candidate and demographic characteristics.

4.2.2 Partial Identification of the Fraction of Strategic Voters

Second, we discuss the identification of the average fraction of strategic voters. In the following discussion, we fix the preference parameters, θ^{PREF}, and consider the identification of the extent of strategic voting given θ^{PREF}. Once this is accomplished, we can vary θ^{PREF} in the identified set of θ^{PREF} to trace out the identified set of the parameters that determine the extent of strategic voting.[23]

Intuitively, the variation in the data that we would like to exploit is the variation in the voting outcome among municipalities (in different districts) with similar characteristics vis-à-vis the variation in the vote shares and characteristics of other municipalities in the same district. For example, consider two districts, one that is generally conservative and another that is liberal. Suppose that we can find a liberal municipality from each district. Suppose also that there are three candidates, a liberal, a centrist and a conservative candidate in both districts. If there are no strategic voters, we would not expect the voting outcome to differ across the two municipalities. However, in the presence of strategic voters, the voting outcome in these two municipalities could differ. If the strategic voters of the municipality in the conservative district believe that the liberal candidate has little chance of winning, those voters would vote for the centrist candidate, while the strategic voters in the other municipality (in the liberal district) would vote for the liberal candidate according to their preferences (if they believe that the liberal candidate has a high chance of winning).

More generally, given the preference parameters, the model can predict what the vote share would be in each municipality if all of the voters voted according to their preferences. If there were no strategic voters, the difference between the actual outcome and the predicted sincere-voting outcome would only be due to random shocks. However, when there is a large number of strategic voters, the actual vote share can systematically diverge from the predicted outcome. This is due to the multiplicity of solution outcomes induced by strategic

[23] Our two-step identification strategy can be schematically described as follows. Let Θ^{PREF} and Θ^{α} be the parameter spaces for θ^{PREF} and $\theta^{\alpha}(\equiv \{\theta_{\alpha 1}, \theta_{\alpha 2}, \theta_{\xi}\})$. First, we consider $I_1(\Theta^{\alpha}) \subset \Theta^{PREF}$, the identified set of θ^{PREF}, given that we may allow θ^{α} to take any value in Θ^{α}. We then consider $I_2(I_1(\Theta^{\alpha})) \subseteq \Theta^{\alpha}$, the identified set of θ^{α} given that we allow θ^{PREF} to take any value in $I_1(\Theta^{\alpha})$. We do not know whether $I_2(I_1(\Theta^{\alpha})) \subsetneq I_2(\Theta^{PREF})$, but the important fact is that $I_2(I_1(\Theta^{\alpha})) \subsetneq \Theta^{\alpha}$. This would be the case if $\exists\ \theta^{\alpha}$, $\nexists\ \theta^{PREF} \in \Theta^{PREF}$ such that $I_2(\theta^{PREF}) = \theta^{\alpha}$. Here, we illustrate this point by example. Let θ^{α_1} and θ^{α_2} be such that $\theta_{\alpha 1}/(\theta_{\alpha 1} + \theta_{\alpha 2}) \approx 0$. In this case, almost every voter votes according to his preferences. Thus, we would not expect the vote share of a municipality to be correlated with the demographic characteristics of other municipalities within the same electoral district. But it could well be the case that voting behavior in a very liberal municipality in a generally conservative electoral district is systematically different from the voting behavior in a very liberal municpality in a generally liberal district. There are no preference parameters that can rationalize such data patterns. Thus, $I_2(I_1(\Theta^{\alpha})) \subsetneq \Theta^{\alpha}$.

Our two-step procedure has empirical content because preferences are partly identified by demographic and vote-share variation *within* districts, while the parameters concerning the distribution of α are identified by variation *across* districts.

voters. Recall that strategic voters make voting decisions conditional on the event that their votes are pivotal. If the beliefs regarding the probability of being pivotal differ across electoral districts – and we have no reason to believe that they do not – the behavior of strategic voters will also differ across districts. This corresponds to different outcomes being played in different districts. The example in the previous paragraph is a manifestation of this. We use the difference between the predicted vote share and the actual vote share to partially identify the fraction of strategic voters.

To further illustrate our identification argument, consider the case of three candidates. In this case, the vote shares in municipality m can be drawn as a point in a simplex. Recall that given a particular value of α_m (the fraction of strategic voters in municipality m) and T, the vote shares can be written as a convex combination of the vote shares of sincere and strategic voters;

$$\mathbf{v}_m(T, \alpha_m) = (1 - \alpha_m)\mathbf{v}_m^{SIN} + \alpha_m \mathbf{v}_m^{STR}(T).$$

where $\mathbf{v}_m$ is the vector of vote shares of the three candidates (v_{1m}, v_{2m}, v_{3m}) and similarly for $\mathbf{v}_m^{SIN}$ and $\mathbf{v}_m^{STR}$. Notice that here, we have made the dependence of $\mathbf{v}_m$ on α_m explicit. Now define $\Delta_m(\alpha_m)$ as the set of all possible vote shares when we vary T in $\mathfrak{T}$ (We denote the set of T satisfying **C1** by $\mathfrak{T}$),

$$\Delta_m(\alpha_m) = \bigcup_{T \in \mathfrak{T}} \mathbf{v}_m(T, \alpha_m).$$

Note that $\Delta_m(\alpha_m)$ and $\Delta_m(1)$ are similar, by a factor of α_m around the singleton $\Delta_m(0) = \mathbf{v}_m^{SIN}$ because α_m is the weight of the convex combination. The dotted circle in Figure 3 corresponds to $\Delta_m(1)$.

For expositional purposes, we first present our identification argument when we can take the number of municipalities to go to infinity and the municipality level shock $\boldsymbol{\xi}_m$ is close to zero. Consider a subset of municipalities in a single electoral district which all have the same demographic characteristics (Note that this does not literally have to be the case because we can control for demographic characteristics once preference parameters are known). In this case, the vote share observations should all lie on the line segment between $\Delta_m(0) = \mathbf{v}_m^{SIN}$ and $\mathbf{v}_m^{STR}(T)$ because these two endpoints are the same in all municipalities[24] and only the realizations of α_m vary across municipalities. Denote this support of the observed distribution as L and the endpoint of L as $\bar{L}$ (the other endpoint is $\mathbf{v}_m^{SIN} = \Delta_m(0)$). We also define the point L' where the extension of L intersects the boundary $\Delta_m(1)$ (See Figure 4). Note that

[24] To see this, recall that $\Delta_m(0)$ is a function of demographic characteristics, and $v_m^{STR}(T)$ is a function of demographic characteristics and T. As the municipalities belong to the same district they share the same T and they share the same demographic characteristics because of the way in which we selected them.

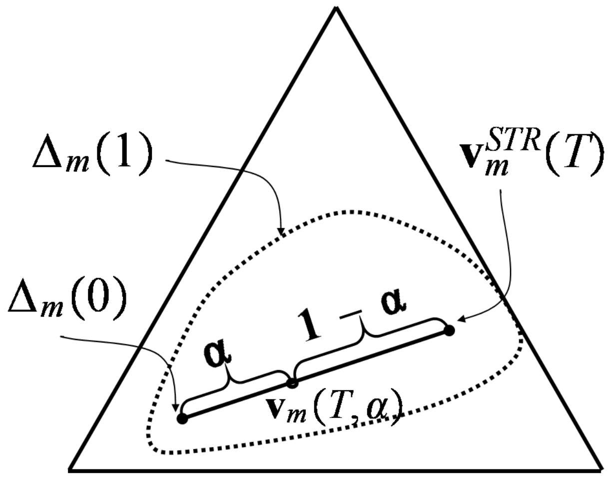

Figure 3: Vote Shares for the Case of $N = 3$. Vote shares $\mathbf{v}_m(T, \alpha)$ is a mixture of sincere votes ($\Delta_m(0) = \mathbf{v}_m^{SIN}$) with fraction $1 - \alpha$, and strategic votes ($\mathbf{v}_m^{STR}(T)$) with fraction α.

L, $\Delta_m(0)$, and L' are all identified. L is just the support of the observed vote shares, and $\Delta_m(0)$ and L' are identified once preferences are identified. On the other hand, the exact position of $\mathbf{v}_m^{STR}(T)$ cannot be determined as T is unknown. The only thing that we know about its location is that it lies somewhere on the dashed line segment between $\bar{L}$ and L'.[25]

Consider two polar cases, Case A and Case B in Figure 4. Case A depicts the situation where $\mathbf{v}_m^{STR}(T)$ is at $\overline{L}$ and Case B depicts the situation where $\mathbf{v}_m^{STR}(T)$ is at L'. For each of the two cases, observations of vote shares can be mapped into realizations of $\alpha_m \in$[0,1]. This mapping is different in Case A and Case B and results in different distributions of α as can be seen in Figure 4. Case A corresponds to the upper bound of the extent of strategic voting, and Case B provides the lower bound. We therefore can partially identify the distribution of α_m as well as the upper and lower bounds of its mean.

Now we discuss how we can modify this discussion to the case where the number of municipalities are finite but the number of districts goes to infinity. Parallel to the previous argument, consider subsets of municipalities from each district with the same demographic characteristics. The key differences from the previous situation are that (1) even if we condition on the same demographics, $\mathbf{v}_m^{STR}(T)$ differs across districts because T is not the same across districts, and (2) we can only take a finite number of municipalities from the same district. Figure 5 illustrates the case where we have three municipalities from two districts. Notice that $\Delta_m(0)$ is the same across these municipalities because the demographics are the same. However, as municipalities in different districts have different T^d, the vote share

[25]This is because vote shares are given by $\mathbf{v}_m(T, \alpha_m) = (1 - \alpha_m)\mathbf{v}_m^{SIN} + \alpha_m \mathbf{v}_m^{STR}(T)$, so that any point on L must lie between $\Delta_m(0)$ and $\mathbf{v}_m^{STR}(T, \alpha_m)$.

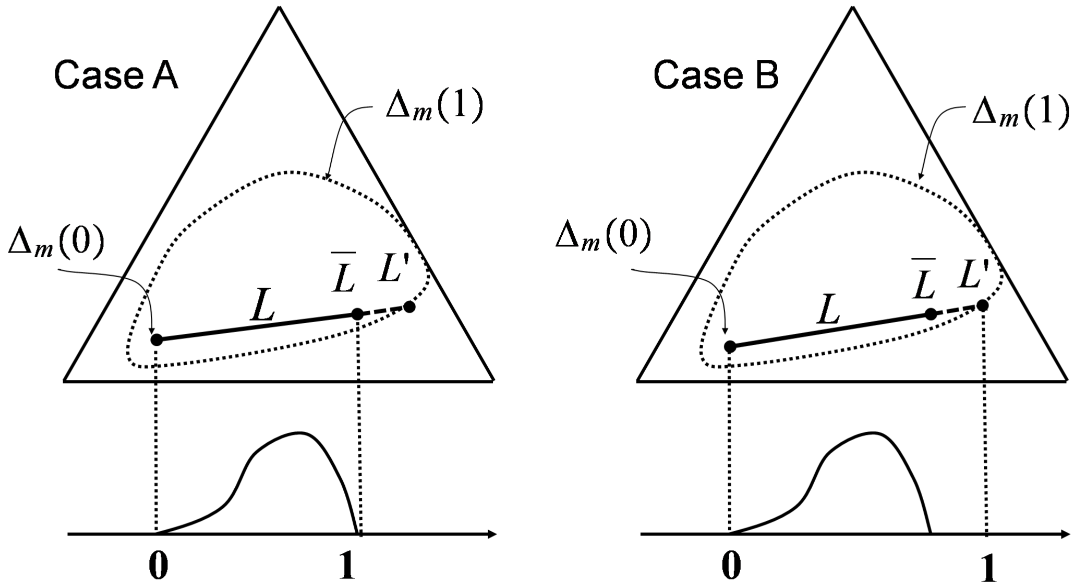

Figure 4: Partial Identification of the Extent of Strategic Voting When $D = 1$ and $M^d \to \infty$. Vote share observations map differently into different values of α depending on the position of $v_m^{STR}(T)$. Case A corresponds to the upper bound of the distribuion, and Case B to the lower bound.

data will be on different line segments for different districts. As in the previous argument, consider two polar cases, Case A′ and Case B′ in Figure 5. Case A′ is the situation where $\mathbf{v}_m^{STR}(T)$ is at $\overline{L}_m$ and Case B′ corresponds to the situation where $\mathbf{v}_m^{STR}(T)$ is at L'_m. For each of the two cases, we can map the vote share observations into realization of $\alpha_m \in [0, 1]$. Note that even though the number of municipalities in a given district is finite, by taking the number of districts to infinity, we can obtain an infinite number of α_ms on $[0, 1]$ that are transformed from the vote share observations. Note that Case A′ gives the upper bound of the distribution of α_m, and Case B′ gives the lower bound. Thus, we set-identify the distribution of α_m.

In the actual data, the vote shares may not lie on the same line segment as in Figure 5, even when we take observations from municipalities with the same demographics. Recall that $\boldsymbol{\xi}_m$ is the municipality level shock that accounts for this kind of variation. It is true that if we do not restrict the distribution of $\boldsymbol{\xi}_m$ in any way, it may not be possible to separately identify the distribution of $\boldsymbol{\xi}_m$ and α_m nonparametrically. However, it should be intuitive from Figure 5 that if restrict the distribution of $\boldsymbol{\xi}_m$ to well-behaved distributions which are mean-zero and unimodal, the same intuition would carry through. We assume that the distribution of the random shock $\boldsymbol{\xi}_m$ follows a Normal distribution with mean zero. Then, we can parametrically account for the dispersion of vote shares around the line segment and

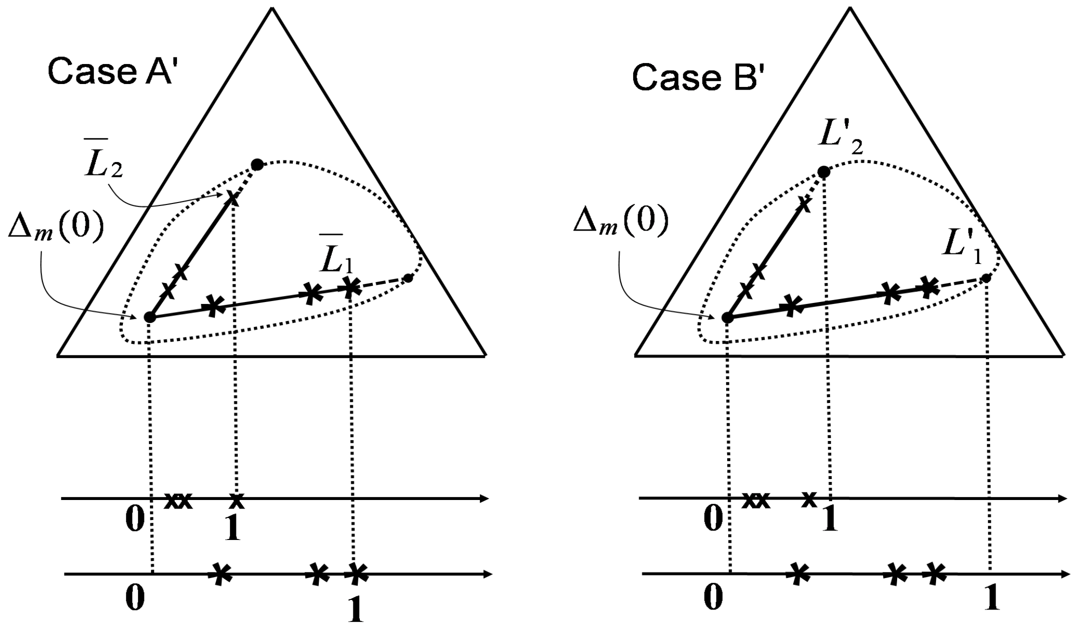

Figure 5: Partial Identification of the Extent of Strategic Voting When $D \to \infty$, but $M^d < \infty$. The figure illustrates the situation when there are two districts with three municipalities each. Case A$'$ corresponds to the upper bound, and Case B$'$ to the lower bound.

the above identification discussion remains valid.

Finally we describe how to extend our argument when preference parameters are only partially identified. For each θ^{PREF} in the identified set, we can partially identify the extent of strategic voting by following our previous argument. To the extent that preference parameters are only partially identified, we can vary θ^{PREF} in the identified set: This allows us to trace out the identified set of the extent of strategic voting.

4.3 Estimation

At the outset, it is useful to clarify the set of parameters that we estimate: They are the preference parameters, θ^{PREF}, the distribution of strategic voters, $(\theta_{\alpha 1}, \theta_{\alpha 2})$, and the variance of ξ, θ_ξ. It is important to note that we do not estimate the beliefs T. This is because our unit of observation is the district, and as the number of districts increases, so does the number of tie beliefs T. Because we cannot treat T as parameters, we need restrictions that do not involve T.

We estimate the model using an inequality-based estimator developed by Pakes, Porter, Ho, and Ishii (2007). If voter beliefs, T, were known (either observed, or uniquely determined by the model), a single outcome would correspond to one realization of the unobserved error terms (ξ, α). In such a case, we could employ estimation procedures such as GMM or MLE. However, as discussed in Section 4.2, the multiplicity of outcomes induced by the presence of strategic voters, together with the fact that we cannot observe voter beliefs, T, imply that the model parameters are only partially identified: This makes the use of set-based estimator appropriate.

We construct the moment inequalities using an idea which is somewhat similar to indirect inference (Smith (1993) and Gouriéroux, Monfort, and Renault (1993)). The following explains the steps we take to construct the moment inequalities. A more detailed description of each step is found in Appendix B.

1. Take some district d and denote the municipalities that belong to this district as $\{1, 2, ...M^d\}$. Regress the vote share data of candidate k in each municipality, $v_{k,m}^{data}$, on the demographics of each municipality, f_m,[26] to obtain the regression coefficient $\beta_{k.d}^{data} = (f_d' f_d)^{-1} f_d' v_{k,d}^{data}$, where $v_{k,d}^{data} = (v_{k,1}^{data}, ..., v_{k,M^d}^{data})'$ and $f_d = (f_1, ...f_{M^d})'$. Note that we obtain K coefficients for each district.

2. Fix some parameter θ and beliefs of voters, T^d. Also fix particular values of $\boldsymbol{\alpha}_d = \{\alpha_m\}_{m=1}^{M^d}$ and $\boldsymbol{\xi}_d = \{\boldsymbol{\xi}_m\}_{m=1}^{M^d}$, which are the fractions of strategic voters and the candidate-municipality shocks, respectively. Given θ, T^d, $\boldsymbol{\alpha}_d$ and $\boldsymbol{\xi}_d$, compute the predicted vote share outcome for each municipality in the district, $(v_{k,1}^{PRED}(T^d, \alpha_1, \boldsymbol{\xi}_1; \theta)$, ..., $v_{k,M^d}^{PRED}(T^d, \alpha_{M^d}, \boldsymbol{\xi}_{M^d}; \theta))$.

3. Parallel to step 1, regress the simulated vote share, $v_{k,m}^{PRED}(T^d, \alpha_m, \boldsymbol{\xi}_m; \theta)$, on the demographic characteristics in each municipality, f_m, to obtain the regression coefficient $\beta_{k,d}(T^d, \boldsymbol{\alpha}_d, \boldsymbol{\xi}_d; \theta) = (f_d' f_d)^{-1} f_d' v_d^{PRED}(T^d)$, where $v_d^{PRED}(T^d) = (v_{k,1}^{PRED}(T^d, \alpha_1, \boldsymbol{\xi}_1; \theta)$, ..., $v_{k,M^d}^{PRED}(T^d, \alpha_{M^d}, \boldsymbol{\xi}_{M^d}; \theta))'$.

4. Because we do not know T^d, we vary $T^d \in \mathfrak{T}(v_d^{data})$ to obtain the minimum and maximum values of the regression coefficients as

$$\begin{aligned}\underline{\beta}_{k,d}(\boldsymbol{\alpha}_d, \boldsymbol{\xi}_d; \theta) &= \min_{T^d \in \mathfrak{T}(v_d^{data})} \beta_{k,d}(T^d, \boldsymbol{\alpha}_d, \boldsymbol{\xi}_d; \theta), \text{ and} \\ \overline{\beta}_{k,d}(\boldsymbol{\alpha}_d, \boldsymbol{\xi}_d; \theta) &= \max_{T^d \in \mathfrak{T}(v_d^{data})} \beta_{k,d}(T^d, \boldsymbol{\alpha}_d, \boldsymbol{\xi}_d; \theta),\end{aligned}$$

[26] We used f_m to denote the distribution of demographic characteristics $\mathbf{x}$ in municipality m in Section 2. If we discretize f_m, we can identify f_m with a vector of probabilities. We use the same notation f_m to denote the distribution and the vector of probabilities. The vector f_m contains, for example, the fraction of the population above 65, the fraction of population in different income ranges, etc.

where $v_d^{data} = \left(\sum_{m=1}^{M^d} v_{k,m}^{data} N_m \Big/ \sum_{m=1}^{M^d} N_m \right)_{k=1}^{K^d}$ is the *district* level vote share data and $\mathfrak{T}(v_d^{data})$ is defined as the set of beliefs that is consistent with condition **C1**. Recall that **C1** requires that beliefs be consistent with the vote share outcome.

5. Integrate out $\boldsymbol{\alpha}_d$ and $\boldsymbol{\xi}_d$ by simulating values of $\boldsymbol{\alpha}_d$ and $\boldsymbol{\xi}_d$ from F_α and F_ξ, and obtain $\overline{\beta}_{k,d}(\theta) = \iint \overline{\beta}_{k,d}(\boldsymbol{\alpha}_d, \boldsymbol{\xi}_d; \theta) dF_{\boldsymbol{\alpha}} dF_{\boldsymbol{\xi}}$ and $\underline{\beta}_{k,d}(\theta) = \iint \underline{\beta}_{k,d}(\boldsymbol{\alpha}_d, \boldsymbol{\xi}_d; \theta) dF_{\boldsymbol{\alpha}} dF_{\boldsymbol{\xi}}$.

6. Then, by construction, we have $E[\underline{\beta}_{k,d}(\theta_0)] \leq E[\beta_{k,d}^{data}] \leq E[\overline{\beta}_{k,d}(\theta_0)]$ at the true parameter θ_0. Thus, we obtain the following moment inequalities;

$$\begin{aligned} E[\underline{\beta}_{k,d}(\theta_0) - \beta_{k,d}^{data}] &\leq 0, \text{ and} \\ E[\overline{\beta}_{k,d}(\theta_0) - \beta_{k,d}^{data}] &\geq 0. \end{aligned}$$

   Moreover, we can construct moment inequalities conditioning on candidate characteristics z (z only takes discrete values).[27] We can do so by running the regressions in steps 1 and 3 only on a subset of the sample for which candidate characteristics z takes a particular value:

$$\begin{aligned} E[\underline{\beta}_{k,d}(\theta_0) - \beta_{k,d}^{data} | z] &\leq 0, \text{ and} \\ E[\overline{\beta}_{k,d}(\theta_0) - \beta_{k,d}^{data} | z] &\geq 0. \end{aligned}$$

   The identified set is the set of θ that satisfy the above equations.

   We base our estimation on the conditional moment inequalities. We take the sample analog of the conditional moment inequalities by repeating steps 1 through 5 for each district. Then, by taking the average, we obtain the criterion function

$$\begin{aligned} Q^+(\theta) &= \sum_{\zeta,k} \left\| \frac{1}{D} \sum_d 1_{\{z=\zeta\}} \left[\overline{\beta}_{k,d}(\theta) - \beta_{k,d}^{data} \right] \right\|_- , \\ Q^-(\theta) &= \sum_{\zeta,k} \left\| \frac{1}{D} \sum_d 1_{\{z=\zeta\}} \left[\beta_{k,d}^{data} - \underline{\beta}_{k,d}(\theta) \right] \right\|_+ , \end{aligned}$$

   where $\|a\|_+ = \max\{0, a\}$, and $\|a\|_- = \min\{0, a\}$. We then apply Pakes, Porter, Ho, and Ishii (2007).

Note that in computing the predicted vote shares in Step 3, we use $v_{k,m}(T)$ in equation (3). $v_{k,m}(T)$ is the infinite counterpart of the vote share $V_{k,m}(T)$ in equation (3); that is,

[27] z only includes variables such as indicators for party affiliation and hometown as described in Section 4.1.

the probability limit of $V_{k,m}(T)$ when the number of voters tends to infinity. Of course, the number of voters in each municipality is finite,[28] but this is not a problem as long as the error from approximating the vote share by its infinite counterpart is sufficiently small compared to the variance of other error terms in the model.

Extending the Model to Include Voter Turnout Our approach can be extended to include the voter's turnout decision. We can, for example, introduce a cost of voting (or a consumption value of voting) into our model, and allow the voters to abstain. In terms of the standard discrete choice model, this would be analogous to the inclusion of an outside option (e.g., not buying a good). Of course, with this modification, we would no longer be able to normalize T to sum up to 1 (i.e., $\sum_k \sum_{l>k} T_{kl} = 1$) as we do in our paper. The scale of T matters for turnout. However, it should be straightforward in principle to identify and estimate a model with voter turnout. The scale of T would be identified by the level of turnout. Then, the identification of the model parameters would follow similarly as the discussion in Section 4.2. Estimation would proceed by simulating the vote shares and turnout for all possible values of T including those that do not add up to 1.

In this paper, we only focus on the issue of strategic voting for computational reasons. In the standard pivotal voter model, turnout is sensitive to small changes in T. For example, a change in T from 10^{-11} to 10^{-10} increases the voter's utility of turning out by ten-fold. This means that we would need to simulate the outcome on a grid in the space of pivot probability that is fine enough to clearly differentiate values 10^{-11}, 10^{-10} (and in between). Hence, the computational cost of implementing this approach could be very high.

5 Results and Counterfactual Experiments

5.1 Parameter Estimates

The confidence intervals for the parameters are reported in Table 4. The exact specification of the utility function we estimate is

$$\begin{aligned}
&u(\mathbf{x}_n, \mathbf{z}_{km}; \theta^{PREF}) = \\
&-\left\{[\theta^{const}, \theta^{income}, \theta^{education}, \theta^{above65}, \theta^{below65}]\mathbf{x}_n - [\theta^{LDP}, \theta^{JCP}, \theta^{DPJ}, \theta^{YUS}]\mathbf{z}_k^{POS}\right\}^2 \\
&+[\theta^{incumbent}, \theta^{previous}, \theta^{no_experience}, \theta^{hometown1}, \theta^{hometown2}, \theta^{hometown3}, \theta^{hometown4}]\mathbf{z}_{km}^{QLTY} \\
&+\xi_{km} + \varepsilon_{kn},
\end{aligned}$$

[28] The average number of voters in a municipality is more than 43,000.

	Confidence Interval
$\theta_{\alpha 1}$	[5.210, 6.005]
$\theta_{\alpha 2}$	[1.473, 1.706]
θ_{ξ}	[0.373, 0.385]
$\theta^{hometown1}$	[0.437, 0.444]
$\theta^{hometown2}$	[0.180, 0.187]
$\theta^{hometown3}$	[0.038, 0.041]
θ^{const}	$[-1.420, -1.418]$
θ^{income}	$[-0.164, -0.162]$
$\theta^{education}$	[0.177, 0.179]
$\theta^{above65}$	$[-0.003, -0.001]$
θ^{YUS}	$[-0.068, -0.065]$
θ^{JCP}	$[-3.467, -3.448]$
θ^{DPJ}	$[-2.998, -2.990]$
$\theta^{previous}$	$[-0.204, -0.199]$
$\theta^{no\_experiecne}$	[0.080, 0.083]

Table 4: Confidence Intervals. Confidence intervals reported are asymptotically more conservative than 95%. These confidence intervals are calculated following Pakes, Porter, Ho, and Ishii (2007).

where we use normalizations $\theta^{below65}$=0, $\theta^{incumbent}$=0, $\theta^{hometown4}$=0, and θ^{LDP}=0.[29]

First, we discuss our parameter estimates for the first term of the utility function. This term captures the ideological component of the voter's utility and it is written as a function of the distance between the voter's ideological position and the candidates' ideological positions. We have estimated the ideological positions of the candidates' parties as, θ^{JCP}=[-3.467, -3.448], θ^{DPJ}=[-2.998, -2.990], and θ^{YUS}=[-0.068,-0.065], where $\theta^{LDP} = 0$, by normalization. We can interpret this result as follows. The JCP and the DPJ are close in ideological space relative to the position of the LDP and the YUS, but compared with the JCP, the position of the DPJ is slightly closer to the LDP and the YUS. This is consistent with the general understanding that on the left-right spectrum, the JCP is very liberal, the DPJ is moderately liberal, and the LDP and the YUS are moderately conservative. Regarding voter positions,

[29] If we let the first three elements of the vector $\mathbf{z}_{km}^{QLTY}$ be dummy variables for whether (1) candidate k has been an incumbent, (2) has had previous political experience, or (3) has had no political experience, then the first three elements of $\mathbf{z}_{km}^{QLTY}$ add up to 1: $z_{km}^{QLTY}(1)+ z_{km}^{QLTY}(2) + z_{km}^{QLTY}(3) = 1$. Thus we need to normalize one of the coefficients (The fact that we are dealing with a discrete choice model precludes us from including a constant term as well.). For the same reason, $\theta^{below65}$ and $\theta^{hometown4}$ are normalized to 0. As for θ^{LDP}, this is normalized to 0 because only the difference between the candidate's ideology, $\theta^{POS}\mathbf{z}_k^{POS}$, and the voter's ideology, $\theta^{ID}\mathbf{x}_n$ matter. Note that because we include a constant term in $\mathbf{z}_k^{POS}$, one of the elements in θ^{ID} can be normalized to zero.

a voter with a lower income, a longer years of schooling, and younger than 65 is ideologically closer to candidates from the LDP and the YUS than to candidates from the DPJ and the JCP.

The estimates of the parameters on candidate experience are $\theta^{previous}$=[-0.204, -0.199], and $\theta^{no\_experience}$= [0.080, 0.083], where $\theta^{incumbent} = 0$, by normalization. $\theta^{previous}$ measures the effect of previously having held public office and $\theta^{no-experience}$ measures the effect of not having had any experience in public office. We have estimated $\theta^{previous}$ to be [-0.204, -0.199], which means that incumbents have an advantage over non-incumbent candidates with previous political experience. We have estimated $\theta^{no-experience}$ to be [0.080, 0.083], which implies that candidates with no prior experience do slightly better than incumbents. This may seem somewhat surprising, but the biggest issue in this election was about postal reform, pitting old guard politicians against new challengers. Our result can be interpreted as voters preferring fresh candidates to both incumbents and candidates with previous experience.

Hometown effects are estimated as $\theta^{hometown1}$=[0.437, 0.444], $\theta^{hometown2}$=[0.180, 0.187], and $\theta^{hometown3}$=[0.038, 0.041], where $\theta^{hometown4}$=0, by normalization. The parameter $\theta^{hometown1}$ captures the effect of having a hometown in the same municipality as the voter, and $\theta^{hometown2}$ is the effect of having a hometown in the same electoral district but in a different municipality. $\theta^{hometown3}$ is the effect of having a hometown in the same prefecture as the voter but not in the same electoral district, and lastly, $\theta^{hometown4} = 0$ is the effect of having a hometown in a different prefecture. The results show that voters receive the highest utility from a candidate whose hometown is in the same municipality as theirs, and the utility decreases as the distance between the candidate's hometown and the voters' municipality increases.

Finally, the mean of the distribution of strategic voters $(\theta_{\alpha 1}/(\theta_{\alpha 1} + \theta_{\alpha 2}))$ is estimated to be between 0.753 and 0.803, that is, [75.3%, 80.3%] of voters are strategic voters on average. This may sound surprising given the fact that the fraction of strategic voting reported in previous studies is between 3% and 17%. However, note that the fraction of "strategic voting" reported in previous studies is in fact the fraction of misaligned voting, as discussed in the Introduction, and not the standard definition of strategic voting (See, e.g., the entry of "strategic voting" in *The New Palgrave Dictionary of Economics* by Feddersen (2008).). Misaligned voting is an equilibrium behavior of strategic voters, and strategic voters may or may not vote for their most preferred candidate. In order to compare our result with the previous studies, we use the estimated model to compute the extent of misaligned voting in the next subsection.

5.2 Extent of *Misaligned Voting*

The extent of misaligned voting is given by the fraction of voters who do not vote for the most preferred candidate. Because we do not have any individual voting records (we only observe vote shares at the municipality level), we still face the task of identifying the extent of misaligned voting from aggregate data; that is, from the difference in the actual vote shares and the counterfactual vote shares we simulate using the estimated model, under the assumption that all voters vote sincerely. Identifying the extent of misaligned voting is not straightforward because there could be misaligned voting at the individual level, but the inflow of misaligned votes to candidate k (i.e., votes cast for candidate k by voters who do not prefer k the most) and the outflow of misaligned votes from candidate k may cancel each other out in the aggregate at the municipality level. Additionally, computing what the outcome would have been if all voters voted sincerely is itself not a simple task. This is because (1) the realization of municipality level shocks ($\boldsymbol{\xi}$) cannot be uniquely recovered and (2) the model parameters are set identified. We describe how to deal with these issues in Appendix C.

We obtained the upper and lower bounds of misaligned voting as 2.4% and 5.5%, that is, about [2.4%, 5.5%] of all voters voted for a candidate that they did not prefer most. Our estimates of misaligned voting are comparable to the numbers reported in the existing literature, ranging from 3% to 17%. Also, given that the estimated fraction of strategic voters is about [75.3%, 80.3%] of the population on average, the fraction of strategic voters who did not vote for their most preferred candidate is [3.0%, 7.3%].

5.3 Counterfactual Experiments

5.3.1 Proportional Representation

In our first counterfactual experiment, we consider what the election outcome would have been under proportional representation instead of plurality rule. In a typical election under proportional representation, voters cast ballots for parties rather than for individual candidates and parties are allotted seats in proportion to the vote share. As votes would not be wasted under proportional representation, there is little incentive for voters to vote strategically. Thus, minor parties generally gain more votes and seats than they would under plurality rule.

We computed the counterfactual vote share by assuming that all voters vote for the party whose ideological position is closest to their own.[30] We also allowed the voters to vote for

[30] We only used the party position to compute the counterfactual outcome because candidate-specific characteristics do not play role in proportional representation.

	JCP	DPJ	LDP	YUS
Actual (Plurality)				
Vote Share (%)	7.8	38.4	50.0	3.9
Number of Seats	0	35	131	9
Counterfactual (PR)				
Vote Share (%)	[7.40, 8.29]	[31.82, 33.55]	[26.56, 27.50]	[31.87, 33.02]
Number of Seats	[13.61, 14.43]	[55.69, 58.72]	[46.47, 48.13]	[55.77, 57.79]

Number of Seats is calculated as (vote share)×175.

Table 5: Counterfactual Experiment – Proportional Representation. Acutual vote share is computed by aggregating the number of votes for a party across all of the 175 districts and dividing it by the total number of votes cast in the 175 districts. Thus they add up to 100% (c.f., Table 6).

any of the four parties regardless of whether a party actually fielded a candidate in the voter's district or not. Hence, there are two effects that account for the difference in the vote shares between the actual election and the counterfactual experiment. One effect is the change in the behavior of strategic voters (sincere-voting effect). The second is the effect of expanding the choice set (choice-expansion effect). The second effect is present because in the counterfactual experiment, we let the voters vote for parties regardless of whether a party fielded a candidate in the voter's district. In our next counterfactual experiment, we will try to isolate and quantify each of the two effects.

Table 5 compares the vote shares and the number of seats each party obtains in the experiment with the actual data under plurality rule. Firstly, the vote share for the DPJ and the LDP would be smaller under proportional representation. As we will confirm in the next counterfactual experiment, a large part of the decrease can be explained by the choice-expansion effect. Secondly, the vote share for the YUS would be larger in the counterfactual experiment. The fact that the YUS did not field candidates in many districts increased its vote share under the counterfactual through the choice-expansion effect (We find almost no sincere-voting effect in the next experiment for the YUS).

As for the number of seats in the counterfactual experiment, we simply multiplied the vote shares of each party by the number of total districts (175). The difference between the actual and the counterfactual is even greater for the number of seats than for vote shares because votes are translated very differently into seats under plurality and proportionality.

	JCP	DPJ	LDP	YUS
Actual				
Vote Share (%)	7.7	38.4	49.7	35.0
Number of Seats	0	35	131	9
Counterfactual				
Vote Share (%)	[8.39, 10.19]	[40.60, 43.77]	[42.63, 45.73]	[33.93, 38.77]
Number of Seats	[0, 0]	[52, 75]	[86, 111]	[11, 18]

Table 6: Counterfactual Experiment – Sincere Voting under Plurality Rule. Acutual vote share is computed by taking the average of the vote shares only over districts in which the party fielded a candidate. Thus, they do *not* add up to 100% (c.f., Table 5).

5.3.2 Sincere Voting under Plurality Rule

In our second counterfactual experiment, we investigate what the outcome would have been if all voters had voted sincerely under plurality rule. It is well known from Gibbard (1973) and Satterthwaite (1975) that there does not exist a strategy-proof voting mechanism (except for a dictatorial mechanism or a mechanism in which a particular candidate is never chosen under any circumstances). Even though a strategy-proof voting mechanism does not exist, we can simulate the sincere-voting outcome under any mechanism because we have recovered the primitives of the model. In this experiment, we compute the sincere-voting outcome under plurality rule. This exercise also enables us to isolate the sincere-voting effect as we discussed in the previous subsection.

Table 6 compares the actual vote shares and the number of seats with those of the sincere-voting experiment (Note that the vote shares do not add up to 100% because the vote shares are computed by taking the average of the vote shares only over districts in which the party fielded a candidate). The details on how we obtained Table 6 are provided in Appendix D.

We find that the number of seats for the DPJ and the LDP change significantly in spite of the fact that the extent of misaligned voting is small [2.4%, 5.5%]. The DPJ would add [17, 40] seats and the LDP would lose [20, 45] seats. Compared to the relatively small change in the vote share, the change in the number of seats is considerable. Note that this difference in the number of seats is accounted for by misaligned voting. Even though the extent of misaligned voting is small, the impact on the number of seats is large because the winning margin is often small.

With respect to vote shares, we find that the vote shares for the JCP and the DPJ increase while the vote share for the LDP decreases in our experiment. This is what we would expect given that the LDP candidates tended to be strong while some fraction of DPJ candidates and even a greater fraction of the JCP candidates were not. On the other hand, we find that the sincere-voting effect for the YUS is nearly zero. This implies that the gain in the vote share for the YUS in the previous counterfactual experiment is due mostly to the choice-expansion effect. Our findings also suggest that a large part of the decrease in the vote shares in the previous experiment for the LDP and the DPJ are due to the choice-expansion effect. Lastly, given that vote share for the JCP remains almost unchanged in the previous experiment, the choice-expansion effect and the strategic-voting effect for the JCP were of similar magnitude, but worked in opposite directions.

6 Concluding Remarks

In this paper, we study how to identify and estimate a model of strategic voting and quantify its impact on election outcomes by adopting an inequality-based estimator. Preference and voting behavior do not necessarily have a one-to-one correspondence for strategic voters, and we obtain partial identification of preference parameters from the restriction that voting for the least preferred candidate is a weakly dominated strategy. The extent of strategic voting is identified using particular features of general-election data. We also make a clear distinction between strategic voting and misaligned voting.

By using aggregate data from the Japanese general election, we find that a large proportion of voters are strategic voters. We estimate the fraction of strategic voters to be [75.3%, 80.3%], on average. A counterfactual experiment that introduces proportional representation decreases the number of votes for major-party candidates by a large margin, and the number of seats by an even greater margin. In the second counterfactual experiment, which assumes sincere voting by all voters under plurality, we find that the number of seats for the parties change significantly. Even though the extent of misaligned voting is small [2.4%, 5.5%], the impact on the number of seats is considerable because the winning margin is often small.

One of the important issues that we did not deal with in this paper is voter turnout. Voters' beliefs on pivot events are also important for models of voter turnout, and it may be possible to extend our approach in this direction. We leave this for future research.

References

[1] Alvarez, Michael R., and Jonathan Nagler (2000) "A New Approach for Modelling Strategic Voting in Multiparty Elections," *British Journal of Political Science*, Vol. 30, pp. 57-75.

[2] Bernheim, Douglas B. (1984), "Rationalizable Strategic Behavior," *Econometrica*, 52(4), pp1007–1028

[3] Berry, Steven T., James A. Levinsohn and Ariel Pakes (1995), "Automobile Prices in Market Equilibrium," *Econometrica*, 60(4), pp. 889–917

[4] Blais, André, Richard Nadeau, Elisabeth Gidengil, and Neil Nevitte (2001) "Measuring Strategic Voting in Multiparty Plurality Elections," *Electoral Studies,* Vol. 20 (3), pp. 343-352.

[5] Chernozhukov, Victor, Han Hong and Elie Tamer (2007) "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica*, 75(5), pp. 1243-1284.

[6] Coate, Stephen, Michael Conlin, and Andrea Moro (2008) "The Performance of Pivotal-Voter Models in Small-Scale Elections: Evidence from Texas Liquor Referenda," *Journal of Public Economics*, Vol. 92(3), pp. 582-596.

[7] Cox, Gary W. (1994), "Strategic Voting Equilibria Under the Single Nontransferable Vote," *American Political Science Review*, Vol. 88, No. 3, pp. 608-621.

[8] Degan, Arianna, and Antonio M. Merlo (2007) "Do Voters Vote Sincerely?" *mimeo*, University of Pennsylvania

[9] Downs, Anthony (1957), *An Economic Theory of Democracy.* New York: Harper

[10] Duverger, Maurice (1954), *Political Parties: Their Organization and Activity in the Modern State*, translated by Barbara and Robert North, New York, Wiley

[11] Feddersen, Timothy (2008), "Strategic Voting," *The New Palgrave Dictionary of Economics.* Second Edition. Eds. Steven N. Durlauf and Lawrence E. Blume. Palgrave Macmillan

[12] Forsythe, Robert E., Roger B. Myerson, Thomas A. Rietz, and Robert J. Weber (1996) "An Experimental Study of Voting Rules and Polls in Three-Way Elections," *International Journal of Game Theory.* Vol 25, pp 355-83.

[13] Forsythe, Robert E., Roger B. Myerson, Thomas A. Rietz, and Robert J. Weber (1993) "An Experiment on Coordination in Multi-candidate Elections: The Importance of Polls and Election Histories," *Social Choice and Welfare.* Vol. 10, pp. 223-47.

[14] Gouriéroux Christian, Alain Monfort, and Eric Renault (1993) “Indirect Inferece,” *Journal of Applied Econometrics*, Vol. 8, pp.S85-S118.

[15] Gibbard, Allan (1973) “Manipulation of Voting Schemes: A General Result,” *Econometrica*, Vol. 41, pp 587-601.

[16] Holt, Charles A. and Angela M. Smith (2005) “A Selective Survey of Experiments in Political Science” *mimeo*, University of Virginia.

[17] King, Gary (1997), *A Solution to the Ecological Inference Problem: Reconstructing Individual Behavior from Aggregate Data*. Princeton University Press.

[18] Morton, Rebecca B. and Kenneth C. Williams (2006) ”Experimentation in Political Science” *The Oxford Handbook of Political Methodology,* ed. by Janet Box-Steffensmeier, David Collier, and Henry Brady, Oxford University Press

[19] Myerson, Roger, B (2002). “Comparison of Scoring Rules in Poisson Voting Games,” *Journal of Economic Theory*, Vol. 103, pp. 219-251

[20] Myerson, Roger, B. and Robert J. Weber (1993). “A Theory of Voting Equilibria,” *American Political Science Review* , Vol. 87, No. 1, pp. 102-114.

[21] Pakes, Ariel, Jack Porter, Katherine Ho, and Joy Ishii (2007) “Moment Inequalities and Their Application,” *mimeo*, Harvard University

[22] Palfrey, Thomas R. (2006) “Laboratory Experiments in Political Economy,” in *The Oxford Handbook of Political Economy*, ed. by Barry Weingast and Donald Wittman, Oxford University Press

[23] Pearce, David G. (1984), “Rationalizable Strategic Behavior and the Problem of Perfection,” *Econometrica*, 52(4), pp1029–1050

[24] Rietz, Thomas A. (2008) “Three-way Experimental Election Results: Strategic Voting, Coordinated Outcomes and Duverger’s Law.” Forthcoming in *The Handbook of Experimental Economics Results,* ed. by Charles Plott and Vernon Smith, Elsevier Science, Amsterdam.

[25] Satterthwaite, Mark A. (1975) “Strategy-Proofness and Arrow’s Condidtions: Existence and Correspondence Theorems for Voting Procedures and Social Welfare Functions,” *Journal of Economic Theory*, Vol. 10, pp 187-217.

[26] Shachar, Ron. and Nalebuff, Barry. (1999) "Follow the Leader: Theory and Evidence on Political Participation," *American Economic Review*, Vol. 89(3), pp. 525-49.

[27] Smith, Anthony A. (1993) "Estimating Nonlinear Time-Series Models Using Simulated Vector Autoregressions," *Journal of Applied Econometrics*, Vol. 8, pp.S63-S84.

[28] Yomiuri Shimbun (2005) *Shugiin-Senkyo 2005 CD-ROM [House of Representatives Election 2005 CD-ROM]*, Tokyo

7 Appendix

7.1 Appendix A: Existence of Solution Outcome

We provide a proof of the existence of the solution outcome. It is almost identical to the proof in MW. Take some $\varepsilon \in (0,1)$. We define a mapping Φ from the product space of vote shares $\mathcal{V}(=\Delta^{K\times M})$ and tie probability $\mathcal{T}(=\Delta^{_K C_2})$ to its power set $2^{\mathcal{V}\times\mathcal{T}}$ so that the fixed point of the mapping is an element in the solution outcome. Before we define Φ, let us first define Φ_1 to be a mapping from $\mathcal{V} \ni V = (V_1, ..., V_K)$ to $2^{\mathcal{T}}$: $\Phi_1(V) = \{T \in \mathcal{T} | V_k > V_l \Rightarrow T_{kn} \geq \varepsilon T_{ln} \forall k, l, n\}$. Φ_1 is the set of tie probability that satisfy a stronger version of **C1** (because $\varepsilon \in (0,1)$). Φ_1 is non-empty valued, convex-valued and upper-hemi continuous. Now define Φ_2 to be a mapping from $\mathcal{T}$ to $2^{\mathcal{V}}$ as $\Phi_2(T) = \{((V_{k,m}(T))_{k=1}^K)_{m=1}^M\}$ where $V_{k,m}(T)$ is defined by **C2**. $\Phi_2(T)$ is a singleton set. Φ_2 is also non-empty valued, convex valued and upper-hemi continuous. Now we define $\Phi : \mathcal{V}\times\mathcal{T} \ni (V,T) \mapsto \Phi(V,T) = (\Phi_2(T), \Phi_1(V)) \in 2^{\mathcal{V}\times\mathcal{T}}$. Then Φ is also non-empty, convex-valued, and upper-hemi continuous. By applying Kakutani's fixed point theorem to Φ, we know that there exists a fixed point of Φ. As the fixed point satisfies **C1** and **C2**, the solution outcome is nonempty.

7.2 Appendix B: Estimation

We use municipality-level aggregate data for our estimation. We denote the vote-share data of candidate k in municipality m by $v_{k,m}^{data}$. We use f_m to denote the distribution of demographic characteristics $\mathbf{x}$ in municipality m. We let $\boldsymbol{\varepsilon}_n = (\varepsilon_{nk})_{k=1}^K$ denote the K draws of individual-candidate-specific shock, and we let g denote the distribution of $\boldsymbol{\varepsilon}_n$. Similarly, denote $\boldsymbol{\xi}_m = (\xi_{km})_{k=1}^K$. Lastly, candidate k's characteristics are denoted by $\mathbf{z}_{km}$.

Recall that as in equation (3) we can express the vote share for candidate k in municipality m as a composition of vote shares among strategic and sincere voters:

$$v_{k,m}^{data} \approx (1-\alpha_m) v_{k,m}^{SIN}(\boldsymbol{\xi}_m; \theta_0) + \alpha_m v_{k,m}^{STR}(T^d, \boldsymbol{\xi}_m, ; \theta_0) \quad (4)$$

where

$$
\begin{aligned}
v^{SIN}_{k,m}(\boldsymbol{\xi}_m;\theta_0) &= \iint \mathbf{1}\{u_{nk} \geq u_{nl},\ \forall l\} g(\boldsymbol{\varepsilon}) d\boldsymbol{\varepsilon} f_m(\mathbf{x}) d\mathbf{x} \\
v^{STR}_{k,m}(T^d,\boldsymbol{\xi}_m;\theta_0) &= \iint \mathbf{1}\{\overline{u}_{nk}(T^d) \geq \overline{u}_{nl}(T^d),\ \forall l\} g(\boldsymbol{\varepsilon}) d\boldsymbol{\varepsilon} f_m(\mathbf{x}) d\mathbf{x},
\end{aligned}
$$

are the expression for the vote share for candidate k among sincere and strategic voters.

Now, we construct moment inequalities based on the regression coefficients in each electoral district.

Step 1 Take some $\mathbf{z}$ and some district d. We obtain $\beta^{data}_{k,d}$ by regressing the vote share data $(v^{data}_{k,1}, ..., v^{data}_{k,M^d})$ on the demographics in each municipality $(f_1, ...f_{M^d})$,[31] i.e.,

$$
\beta^{data}_{k,d} = \arg\min_{\beta} \left[\sum_{m=1}^{M^d} 1_{\{z_{km}=z\}} (v^{data}_{km} - \beta \cdot f_m)^2 \right].
$$

Because M^d is not large, we cannot include many regressors. The number of regressors must be less than M^d. For this reason, we run 9 different types of regressions all involving just a constant or a constant and one component of f_m. For example, we run a regression of v^{data}_{km} on a constant and the fraction of population above 65 years old conditioned on $z_{km} = LDP$. The full set of regressions we use is in the Supplementary Material.

Step 2 Fix some parameter θ, beliefs T^d, and values of $\boldsymbol{\alpha}_d = \{\alpha_m\}^{M^d}_{m=1}$ and $\boldsymbol{\xi}_d = \{\boldsymbol{\xi}_m\}^{M^d}_{m=1}$. We can compute the vote shares for candidate k in each of the municipalities which we denote as $(v^{PRED}_{k,1}(T^d, \alpha_1, \boldsymbol{\xi}_1; \theta),\ ...,\ v^{PRED}_{k,M^d}(T^d, \alpha_{M^d}, \boldsymbol{\xi}_{M^d}; \theta))$. We can obtain a closed form solution for the predicted vote share of sincere voters because ε is distributed type 1 exteme value. Regarding strategic voters, the predicted vote share does not have a closed form solution, and we use Monte-Carlo integration. For Monte-Carlo integration, we take 10 draws of ε for each demographic characteristics, $\mathbf{x}$. As we group the voters into 32 types according to their characteristics $\mathbf{x}$,[32] we take 320 draws of ε for each municipality.

[31] As in footnote 28, we can identify the distribution of demographic characteristics f_m with a vector of probabilities. We use the same notation f_m to denote the distribution and the vector of probabilities. The vector f_m contains, for example, the fraction of the population above 65, the fraction of population in different income ranges, etc.

[32] We discretize income into four groups, age into two groups, and education into four groups. Thus, we have $4 \times 2 \times 4 = 32$ types.

Step 3 Parallel to Step 1, regress the simulated vote shares of candidate k, $(v_{k,1}^{PRED}(T^d, \alpha_1, \boldsymbol{\xi}_1; \theta)$ $, ..., v_{k,M^d}^{PRED}(T^d, \alpha_{M^d}, \boldsymbol{\xi}_{M^d}; \theta))$, on the demographic characteristics in each municipality $(f_1, ...f_{M^d})$, conditioning on a particular value of z. We obtain the regression coefficient as

$$\beta_{k,d}(T^d, \boldsymbol{\alpha}_d, \boldsymbol{\xi}_d; \theta) = \arg\min_{\beta} \left[\sum_{m=1}^{M^d} 1_{\{z_{km}=z\}} (v_{k,m}^{PRED}(T^d, \alpha_{M^d}, \boldsymbol{\xi}_{M^d}; \theta) - \beta \cdot f_m)^2 \right].$$

Step 4 Because we do not know T^d, we vary $T^d \in \mathfrak{T}(v^{data})$ to obtain the minimum and maximum values of the regression coefficients $\underline{\beta}_{k,d}(\boldsymbol{\alpha}_d, \boldsymbol{\xi}_d; \theta)$ and $\overline{\beta}_{k,d}(\boldsymbol{\alpha}_d, \boldsymbol{\xi}_d; \theta)$ as in the main text. In practice, we discretize $\mathfrak{T}(v^{data})$ with a grid size equal to 0.04.

Step 5 We integrate out $\boldsymbol{\alpha}_d$ and $\boldsymbol{\xi}_d$ by simulating values of $\boldsymbol{\alpha}_d$ and $\boldsymbol{\xi}_d$ from F_α and F_ξ, and obtain $\overline{\beta}_{k,d}(\theta)$ and $\underline{\beta}_{k,d}(\theta)$, as defined in the main text. We draw 10 realizations of α_m and $\boldsymbol{\xi}_m$ from F_α and F_ξ, hence we have $10 \times M^d$ draws for each district d.

Step 6 We take the average of $\overline{\beta}_{k,d}(\theta)$, $\underline{\beta}_{k,d}(\theta)$ and $\beta_{k,d}^{data}$ across d and obtain the empirical analog as in the main text.

Finally, to improve the sharpness of the identified set, we include another type of moment inequalities that harnesses the comovements in β that results from varying T. Notice that in Step 4, we have computed the maximum and the minimum values of β separately for each of the 9 types of regressions. But note that the coefficients from the regressions cannot move independently. Thus in an effort to use some of these restrictions, we can construct additional moment inequalities by taking linear combination of β. For example, let $\beta_{k,d}^{OLD}$ and $\beta_{k,d}^{RICH}$ be the regression coefficients that we obtain in Steps 1 and 4 when we regress vote shares on the proportion of the population above 65 and the proportion of the population in the highest income quartile, respectively. Then we can consider $\max_{\{T^d\}}(\beta_{k,d}^{OLD}(T^d) - \beta_{k,d}^{RICH}(T^d))$ and use this to form moment inequalities. More generally, for any matrix A, we can consider $\overline{A\beta}_{k,d} \equiv \max_{\{T^d\}} A\beta_{k,d}(T^d)$ and $\underline{A\beta}_{k,d} \equiv \min_{\{T\}} A\beta_{k,d}(T^d)$ and construct moment inequalities by following the same argument presented in the main text. We provide the exact form of matrix A that we use in our estimation in the Supplementary Material.

7.3 Appendix C: Comuptation of Misaligned Voting

The amount of misaligned voting is given by the fraction of voters who do not vote for the most preferred candidate. As we discussed in the main text, we do not have any individual voting records (we only observe vote shares at the municipality level), so we need to identify the extent of misaligned voting from aggregate data. In Step 1, we discuss issues arising from identifying the extent of misaligned voting from aggregated data, assuming that we can precisely recover the outcome when everyone votes sincerely. Then, in Steps 2 to 4, we will discuss issues related to recovering the sincere voting outcome from the estimated model.

Step 1

Let v_k^{data} denote the actual vote share for candidate k and let v_k^{sin} denote the vote share of candidate k when everyone votes sincerely (subscripts d, m are suppressed from now on). Also, let D_{kl} denote the total votes cast for candidate k by strategic voters who prefer candidate l most (inflow/outflow of misaligned votes from l to k). Then the object of interest, the amount of misaligned voting, can be expressed as $\sum_{k,l} D_{kl}$. On the other hand, the available information is summarized as $v_k^{data} - v_k^{sin} = \sum_l D_{kl} - \sum_l D_{lk}$, where $\sum_l D_{kl}$ is the inflow of misaligned votes into candidate k and $\sum_l D_{lk}$ is the outflow of misaligned votes from candidate k. (Note that **C1** implies that if $D_{kl} > 0$, then $D_{lk} = 0$.). The question we are concerned with is the following: What can we learn about $\sum_l D_{lk}$ given that we only know $v_k^{data} - v_k^{sin} (\equiv \Delta^k) = \sum_l D_{kl} - \sum_l D_{lk}$?

We can show that for $K = 3$, $\sum_l D_{lk}$ can be bounded below by

$$lb(\{\Delta^k\}) = \max_k \{|\Delta^k|\}$$

and above by

$$ub(\{\Delta^k\}) = \max_k \{\Delta^k\} - \min_k \{\Delta^k\}.$$

We provide an analogous expression for $K = 4$ in the Supplementary material. These bounds are also sharp among all bounds that can be obtained without imposing any distributional assumptions on the shocks in the utility function.[33] The proofs are provided in the Supplementary material.

Step 2 to Step 4

Now we discuss issues related to recovering the sincere voting outcome from the estimated model. Given preference parameters of the model, for any realization of $\boldsymbol{\xi}$, we can compute what the outcome would be if all voters vote sincerely. We denote this predicted sincere-voting outcome as $v^{sin}(\widehat{\theta}, \boldsymbol{\xi})$. Ideally, we would know the actual realization of $\boldsymbol{\xi}$, $\boldsymbol{\xi} = \boldsymbol{\xi}_0$ in each municipality, and compute the sincere voting outcome, $v^{sin}(\widehat{\theta}, \boldsymbol{\xi}_0)$, using this actual

[33] We do not know whether the bounds are sharp with regard to the class of DGPs that we considered in our estimation where we have imposed distributional assumptions on the unobservable shocks in the utility function. As our estimation bypasses inference on T, it is difficult to obtain bounds that are, at the same time, computable and sharp with regard to the DGPs we considered in the estimation.

realization of $\boldsymbol{\xi}_0$ and using a parameter value in the estimated set, $\widehat{\theta} \in \widehat{\Theta}_{CI}$. Then the difference between the observed vote share, v^{data} and $v^{sin}(\widehat{\theta}, \boldsymbol{\xi}_0)$, $(\Delta^k = v^{data} - v^{sin}(\widehat{\theta}, \boldsymbol{\xi}_0))$ would allow us to compute the lower and upper bounds, $lb(\{\Delta^k\})$ and $ub(\{\Delta^k\})$. However, $\boldsymbol{\xi}_0$ cannot be recovered uniquely. Also, the difference between $v^{data} = v(\boldsymbol{\xi}_0)$ and $v^{sin}(\widehat{\theta}, \boldsymbol{\xi})$ depends on $\widehat{\theta}$, which we have only set-identified. Hence, we compute the bounds on the extent of misaligned voting in the following three steps (Step 2 to Step 4).

In Step 2, fix $\widehat{\theta} \in \widehat{\Theta}_{CI}$. For any given draw of $\boldsymbol{\xi}$ from $\hat{F}_{\boldsymbol{\xi}}$, we compute $\widehat{\Delta}^k(\boldsymbol{\xi})$,

$$\widehat{\Delta}^k(\boldsymbol{\xi}) = v_k^{data} - v_k^{sin}(\widehat{\theta}, \boldsymbol{\xi})$$

and the corresponding bounds $lb(\{\widehat{\Delta}^k(\boldsymbol{\xi})\})$ and $ub(\{\widehat{\Delta}^k(\boldsymbol{\xi})\})$. By Monte Carlo, we then compute the expected value of the bounds where the expectation is taken with regard to the randomness in $\boldsymbol{\xi}$,

$$\begin{aligned} Lb_0 &= \int lb(\{\widehat{\Delta}^k(\boldsymbol{\xi})\}) d\hat{F}_{\boldsymbol{\xi}}(\boldsymbol{\xi}), \text{ and} \\ Ub_0 &= \int ub(\{\widehat{\Delta}^k(\boldsymbol{\xi})\}) d\hat{F}_{\boldsymbol{\xi}}(\boldsymbol{\xi}), \end{aligned}$$

for each municipality, where $\hat{F}_{\boldsymbol{\xi}}$ is the estimated distribution of $\boldsymbol{\xi}$. Note that Lb_0 and Ub_0 do not necessarily coincide with $lb(\{\widehat{\Delta}^k(\boldsymbol{\xi}_0)\})$ and $ub(\{\widehat{\Delta}^k(\boldsymbol{\xi}_0)\})$, which are the lower and upper bounds of the extent of misaligned voting we would obtain if we had full knowledge of the realizations of $\boldsymbol{\xi}$, $\boldsymbol{\xi} = \boldsymbol{\xi}_0$. Therefore, we need to account for the parts of Lb_0 and Ub_0 that are induced by the randomness in $\boldsymbol{\xi}$. We discuss this in Step 3.

In Step 3, we evaluate the components of Lb_0 and Ub_0 that are induced by the randomness in $\boldsymbol{\xi}$. To do so, we compute the mean effects of the randomness components by calculating (using Monte Carlo integration)

$$\begin{aligned} Lb_{\boldsymbol{\xi}} &= \int \int lb(\{\widetilde{\Delta}^k(\widetilde{\boldsymbol{\xi}}, \widetilde{\widetilde{\boldsymbol{\xi}}})\}) d\hat{F}_{\boldsymbol{\xi}}(\widetilde{\widetilde{\boldsymbol{\xi}}}) d\hat{F}_{\boldsymbol{\xi}}(\widetilde{\boldsymbol{\xi}}), \text{ and} \\ Ub_{\boldsymbol{\xi}} &= \int \int ub(\{\widetilde{\Delta}^k(\widetilde{\boldsymbol{\xi}}, \widetilde{\widetilde{\boldsymbol{\xi}}})\}) d\hat{F}_{\boldsymbol{\xi}}(\widetilde{\widetilde{\boldsymbol{\xi}}}) d\hat{F}_{\boldsymbol{\xi}}(\widetilde{\boldsymbol{\xi}}), \end{aligned}$$

where $\widetilde{\Delta}^k(\widetilde{\boldsymbol{\xi}}, \widetilde{\widetilde{\boldsymbol{\xi}}})$ is the difference in the vote share between two realizations of municipality-level shock, $\widetilde{\boldsymbol{\xi}}$ and $\widetilde{\widetilde{\boldsymbol{\xi}}}$, i.e.,

$$\widetilde{\Delta}^k(\widetilde{\boldsymbol{\xi}}, \widetilde{\widetilde{\boldsymbol{\xi}}}) = v_k^{sin}(\widehat{\theta}, \widetilde{\boldsymbol{\xi}}) - v_k^{sin}(\widehat{\theta}, \widetilde{\widetilde{\boldsymbol{\xi}}}).$$

We then compute the lower and upper bounds of misaligned voting at the municipality level as

$$
\begin{aligned}
LB &= Lb_0 - Lb_{\boldsymbol{\xi}}, \text{ and} \\
UB &= Ub_0 - Ub_{\boldsymbol{\xi}}.
\end{aligned}
$$

In Step 4, we account for the fact that θ is only set-identified. So far, we have been computing LB and UB implicitly treating θ as given. By denoting the dependence on θ more explicitly, LB and UB above can be written as $LB(\theta)$ and $UB(\theta)$. Because θ is partially identified, we need to compute $LB(\theta)$ and $UB(\theta)$ by allowing θ to move in the partially identified set Θ_{CI} in order to construct the most conservative bound on the extent of misaligned voting, $\underline{LB}$ and $\overline{UB}$, i.e.

$$
\begin{aligned}
\underline{LB} &= \min_{\theta \in \Theta_{CI}} LB(\theta), \text{ and} \\
\overline{UB} &= \max_{\theta \in \Theta_{CI}} UB(\theta).
\end{aligned}
$$

7.4 Appendix D: Comupation of Second Counterfactual

Computation of the second counterfactual proceeds in the same way as described in Steps 2 to 4 in Appendix B. This is because as was the case in our first counterfactual, we cannot recover the realization of the municipality level random shock $\boldsymbol{\xi}, \boldsymbol{\xi} = \boldsymbol{\xi}_0$. Denote the counterfactual vote share as $v^{sin}(\widehat{\theta}, \boldsymbol{\xi}_0)$. The problem is that we cannot compute this because $\boldsymbol{\xi}_0$ is unobserved. But we can obtain bounds for $v^{sin}(\widehat{\theta}, \boldsymbol{\xi}_0)$ by following the same procedure as in Appendix C. We can also compute the number of seats in the same way. Note that we do not need to take Step 1 in this case.

8 Supplementary Material

8.1 Supplementary Material A

In our estimation, we run regressions in Step 1 and Step 3 in order to obtain $\beta_{k,d}^{data}$ and $\beta_{k,d}(T^d, \boldsymbol{\alpha}_d, \boldsymbol{\xi}_d; \theta)$, which are

$$\begin{aligned}
\beta_{k,d}^{data}(T^d, \boldsymbol{\alpha}_d, \boldsymbol{\xi}_d; \theta) &= \arg\min_{\beta} \left[\sum_{m=1}^{M^d} 1_{\{z_{km}=z\}} (v_{k,m}^{data} - \beta \cdot f_m)^2 \right], \text{ and} \\
\beta_{k,d}(T^d, \boldsymbol{\alpha}_d, \boldsymbol{\xi}_d; \theta) &= \arg\min_{\beta} \left[\sum_{m=1}^{M^d} 1_{\{z_{km}=z\}} (v_{k,m}^{PRED}(T^d, \alpha_{M^d}, \boldsymbol{\xi}_{M^d}; \theta) - \beta \cdot f_m)^2 \right].
\end{aligned}$$

We run 9 different types of regressions (fourty eight regressions in total) for each district as follows.

1. Regressing the vote share onto a constant and the fraction of population above 65 years old, i.e. $f_m = (1,$ "fraction of population above 65"). If we let P denote $\{LDP, DPJ, JCP, YUS\}$, we run this regression for each combination of $(z_1^{POS}, ..., z_K^{POS}) \in P^K$.
2. f_m is a constant and the fraction of population with years of schooling between 12 to 14 years. Regression is run for each combination of $(z_1^{POS}, ..., z_K^{POS}) \in P^K$.
3. f_m is a constant and the fraction of population with years of schooling between 15 to 16 years. Regression is run for each combination of $(z_1^{POS}, ..., z_K^{POS}) \in P^K$.
4. f_m is a constant and the fraction of population with years of schooling over 16 years. Regression is run for each combination of $(z_1^{POS}, ..., z_K^{POS}) \in P^K$.
5. f_m is a constant and the fraction of population with income in the first quartile (lower than 1.870 million yen). Regression is run for each combination of $(z_1^{POS}, ..., z_K^{POS}) \in P^K$.
6. f_m is a constant and the fraction of population with income in the second quartile (between 1.870 million yen and 2.704 million yen). Regression is run for each combination of $(z_1^{POS}, ..., z_K^{POS}) \in P^K$.
7. f_m is a constant and the fraction of population with income in the third quartile (between 2.704 million yen and 3.911 million yen). Regression is run for each combination of $(z_1^{POS}, ..., z_K^{POS}) \in P^K$.
8. f_m is a constant. Regression is run for each combination of $(z_1^{POS}, ..., z_K^{POS}) \in P^K$.
9. f_m is a constant. Regression is run for each combination of $(z_1^{POS}, ..., z_K^{POS}) \in P^K$, and $(z_{km}^{EXPR}, z_{km}^{HOME})$ where $z_{km}^{EXPR} \in$ {incumbent, previous political experience, no previous political experience}, and $z_{km}^{HOME} \in$ {hometown of the candidate is outside the prefecture, hometown of the candidate is inside the prefecture (but outside the distrct), hometown of

the candidate is in the district (but outside municipality m), hometown of the candidate is in municipality $m\}$.

In order to improve the sharpness of the identified set, we include another type of moment inequalities that harnesses the comovements in β that results from changes in T as dissussed in Step 6 of Appendix B. We augment the moment conditions by using restrictions on the comovement of the coefficients for the 9th type of regressions. This allows us to add restrictions on the pairwise difference in the βs that relate to the effect of candidates' experience and hometowns, e.g., the difference in the vote share for a LDP candidate whose hometown is outside of the prefecture compared to a LDP candidate whose hometown is within the prefecture. In practice, the matrix A used in Step 6 in our estimation is $A^T = \begin{pmatrix} I_{60} & \begin{matrix} \mathbf{0} \\ B \end{matrix} \end{pmatrix}$ where $B = \begin{pmatrix} 1 & \cdots & \cdots & 1 & 0 & \cdots & 0 & \cdots & \cdots \\ -1 & 0 & \cdots & 0 & 1 & \cdots & 1 & 0 & \cdots \\ 0 & -1 & \ddots & \vdots & -1 & 0 & 0 & 1 & \cdots \\ \vdots & \ddots & \ddots & 0 & 0 & \ddots & 0 & -1 & 0 \\ 0 & \cdots & 0 & -1 & \vdots & \ddots & -1 & 0 & \ddots \end{pmatrix}$ and I_{60} is the identity matrix of size 60×60.

8.2 Supplementary Material B

In this Supplementary Material, we prove that the bounds $ub(\{\Delta^k\})$ and $lb(\{\Delta^k\})$ we have used in Appendix C in fact constitute bounds and that they are sharp. Because the bounds are different for $K = 3$ and $K = 4$, we prove each case in turn. We drop subscripts d and m for the rest of the section.

8.2.1 Case of $K = 3$

First, we prove that, for the case of $K = 3$, the extent of strategic voting is bound by $lb(\{\Delta^k\})$ and $ub(\{\Delta^k\})$, where

$$
\begin{aligned}
lb(\{\Delta^k\}) &= \max_k\{|\Delta^k|\}, \text{ and} \\
ub(\{\Delta^k\}) &= 1\{\#\{\Delta^k > 0\} = 2\}(\max_k\{\Delta^k|\Delta^k > 0\} - \min_k\{\Delta^k\}) \\
&\quad + 1\{\#\{\Delta^k > 0\} = 1\}(\max_k\{\Delta^k\} - \min_k\{\Delta^k|\Delta^k < 0\}) \\
&= \max_k\{\Delta^k\} - \min_k\{\Delta^k\},
\end{aligned}
$$

and $\#\{\Delta^k > 0\}$ indicates the number of Δ^ks that are positive, and $1\{\cdot\}$ is an indicator function. Let D_{kl} denote the votes cast for candidate k by strategic voters who prefers candidate l most. Then the amount of misaligned voting is $\sum_{kl} D_{kl}$ (Note that **C1** implies that if $D_{kl} > 0$, then $D_{lk} = 0$.).

First, we prove that the extent of strategic voting is bound by $lb(\{\Delta^k\})$ and $ub(\{\Delta^k\})$. Without loss of generality, index the candidates as 1, 2, and 3 such that the beliefs regarding the tie probabilities satisfy $T_{12} \geq T_{13} \geq T_{23}$. Then the amount of misaligned voting is $D = D_{12} + D_{13} + D_{23}$ (Note that $D_{21} = D_{31} = D_{32} = 0$.). Now, we can write

$$\begin{aligned} \Delta^1 &= D_{12} + D_{13}, & \text{(A1)} \\ \Delta^2 &= D_{23} - D_{12}, & \text{(A2)} \\ \Delta^3 &= -D_{13} - D_{23}. & \text{(A3)} \end{aligned}$$

Note that $|\Delta^1| + |\Delta^3| = D_{12} + 2D_{13} + D_{23} \geq D$, thus $|\Delta^1| + |\Delta^3|$ is an upper bound. We consider two cases; (i) $\{\#\{\Delta^k > 0\} = 1\}$, and (ii) $\{\#\{\Delta^k > 0\} = 2\}$. In case (i), we know that the positive number we observe is Δ^1, but cannot identify which of the two negative numbers correspond to Δ^2 or Δ^3. In case (ii), we know that the negative number we observe is Δ^3, but we cannot identify which of the two positive numbers correspond to Δ^1 or Δ^2. These two cases are exhaustive as $\Delta^1 + \Delta^2 + \Delta^3 = 0$. In case (i),

$$\begin{aligned} ub(\{\Delta^k\}) &= \max_k\{\Delta^k\} - \min_k\{\Delta^k|\Delta^k < 0\} = \Delta^1 - \min\{\Delta^2, \Delta^3\} \\ &= |\Delta^1| + \max\{|\Delta^2|, |\Delta^3|\} \\ &\geq |\Delta^1| + |\Delta^3|. \end{aligned}$$

In case (ii),

$$\begin{aligned} ub(\{\Delta^k\}) &= \max_k\{\Delta^k|\Delta^k > 0\} - \min_k\{\Delta^k\} = \max\{\Delta^1, \Delta^2\} - \Delta^3 \\ &= \max\{|\Delta^1|, |\Delta^2|\} + |\Delta^3| \\ &\geq |\Delta^1| + |\Delta^3|. \end{aligned}$$

We can also see that $\max_k\{|\Delta^k|\}$ is the lower bound because $|\Delta^1| = D_{12} + D_{13} \leq D$, $|\Delta^2| \leq D_{23} + D_{12} \leq D$, and $|\Delta^3| = D_{13} + D_{23} \leq D$.

Second, we prove by contradiction that the upper bound $ub(\{\Delta^k\})$ is sharp. Let $h(\Delta^1, \Delta^2, \Delta^3) \leq ub(\{\Delta^k\})$ for all $\{\Delta^k_{d,m}\}$, and moreover $h(\Delta^{1*}, \Delta^{2*}, \Delta^{3*}) < ub(\{\Delta^k\})$. Without loss of generality, consider the following two cases (i) $\Delta^{1*} > 0 > \max\{\Delta^{2*}, \Delta^{3*}\}$ and (ii) $\min\{\Delta^{1*}, \Delta^{2*}\} > 0 > \Delta^{3*}$. Note that we cannot identify whether the two negative numbers in case (i) correspond to Δ^{2*} or Δ^{3*}, and similarly, in case (ii), we cannot identify whether the two positive numbers correspond to Δ^{1*} or Δ^{2*}. This is the reason why we use the min and the max operators. In case (i), if we let $D_{12} = \Delta^{1*}$, $D_{23} = -\min\{\Delta^{2*}, \Delta^{3*}\}$ and $D_{13} = 0$, then the three equations (A1)-(A3) can be satisfied. In this instance, $D_{12}+D_{13}+D_{23} = \Delta^{1*} - \min\{\Delta^{2*}, \Delta^{3*}\} = ub(\{\Delta^{k*}\})$, achieving our bound. Hence, h cannot be an upper bound. In case (ii), let $D_{12} = \max\{\Delta^{1*}, \Delta^{2*}\}$, $D_{13} = 0$, $D_{23} = -\Delta^{3*}$. Then (A1)-(A3) are satisfied, and moreover, $D_{12} + D_{13} + D_{23} = \max\{\Delta^{1*}, \Delta^{2*}\} - \Delta^{3*} = ub(\{\Delta^{k*}\})$.

Third, we prove by contradiction that the lower bound $lb(\{\Delta^k\})$ is sharp. Let $h(\Delta^1, \Delta^2, \Delta^3) \geq lb(\{\Delta^k\})$ for all $\{\Delta^k_{d,m}\}$, and moreover $h(\Delta^{1*}, \Delta^{2*}, \Delta^{3*}) > lb(\{\Delta^k\})$. Without loss of generality, consider the following two cases (i) $\Delta^{1*} > 0 > \max\{\Delta^{2*}, \Delta^{3*}\}$ and (ii) $\min\{\Delta^{1*}, \Delta^{2*}\} > 0 > \Delta^{3*}$. In case (i), let $D_{12} = -\Delta^{2*}$, $D_{13} = -\Delta^{3*}$, and $D_{23} = 0$. This satisfies the three equations (A1)-(A3) and moreover, $D_{12} + D_{13} + D_{23} = -\Delta^{2*} - \Delta^{3*} = \Delta^{1*} = lb(\{\Delta^{k*}\})$. In case (ii) let $D_{12} = 0$ and $D_{23} = \Delta^{2*}$ and $D_{13} = -\Delta^{3*} - \Delta^{2*}$. This also satisfies equations (A1)-(A3), and implies $D_{12} + D_{13} + D_{23} = -\Delta^{3*} = lb(\{\Delta^{k*}\})$. Thus, h cannot be a lower bound.

8.2.2 Case of $K = 4$

For the case of $K = 4$, the lower and upper bounds $lb(\{\Delta^k\})$ and $ub(\{\Delta^k\})$ are written as

$$\begin{aligned} lb(\{\Delta^k\}) = {} & 1\{\#\{\Delta^k > 0\} = 3\} \max\left\{\min_{k,l\neq k}\{\Delta^k + \Delta^l | \Delta^k, \Delta^l > 0\}, -\min_k\{\Delta^k|\Delta^k < 0\}\right\} \\ & + 1\{\#\{\Delta^k > 0\} = 2\} \max\left\{\min_k\{\Delta^k|\Delta^k > 0\}, -\min_k\{\Delta^k|\Delta^k < 0\}\right\} \\ & + 1\{\#\{\Delta^k > 0\} = 1\} \max\left\{\max_k\{\Delta^k\}, -\max_k\{\Delta^k|\Delta^k < 0\}\right\}, \text{ and} \\ ub(\{\Delta^k\}) = {} & \max_{k,l\neq k}\{2\Delta^k + \Delta^l\} - \max_k\{\Delta^k|\Delta^k < 0\} \end{aligned}$$

The proof is similar to the case of $K = 3$.