Inferring Strategic Voting*

Kei Kawai† Northwestern University

Yasutora Watanabe‡ Northwestern University

May 2009

Abstract

We estimate a model of strategic voting by adopting a recently developed inequality-based estimator in a discrete-choice framework. The difficulty of identification comes from the fact that preference and voting behavior do not necessarily have a one-to-one correspondence for strategic voters. 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 variation generated by multiple equilibria. Using Japanese general-election data, we find a large fraction (68.2%, 82.7%) of *strategic voters*, only a small fraction (2.2%, 7.4%) 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, discrete-choice models

*We thank Kenichi Ariga, David Austen-Smith, Ivan Canay, Amrita Dhillon, Tim Feddersen, 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. Aggregate municipality level data from the Japanese general election is used for estimation. 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.

In our paper, we make a clear distinction between *strategic voting,* as defined above (this is the standard definition in the theoretical literature[1]), 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 all the more 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.

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

Our model is an adaptation of Myerson and Weber (1993) and Myerson (2002) with the addition of sincere voters.[2] We employ iterated deletion of weakly dominated strategies as the solution concept of the voting game. We use this solution concept so that the outcome of the model is robust to different assumptions regarding voter beliefs and is able to account for diverse patterns of outcome as observed in the data.

The presence of strategic voters poses a challenge to the identification of the model. 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, using particular features found in typical general-election data, we relate aggregate variation in the vote share to demographic characteristics. Typically, general-election data contains data from hundreds of elections where each electoral district is our unit of observation. Moreover, an electoral district is usually comprised of several sub-districts. For example, each electoral district, such as a U.S. State (for Senate elections) or a Congressional District (for House elections), is comprised of several sub-districts, such as counties or municipalities. Usually, the breakdown of vote counts and demographic characteristics is available at the sub-district (hereinafter referred to as municipality) level. This data structure allows us to use variation in the vote share and 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.

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

[2] Our model can be naturally extended to elections with N candidates competing for N_S ($N_S < N$) seats under single non-transferrable voting as in Cox (1994).

outcome in municipalities with a given characteristic vis-à-vis the variation in the 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. We use the difference between the predicted vote share and the actual vote share caused by multiplicity of equilibria to partially identify the fraction of strategic voters.

The following thought experiment illustrates our argument. Consider many municipalities *across* electoral districts. For expositional ease, let the municipalities have the same demographic and candidate characteristics. Moreover, the municipalities are chosen so that if all voters vote according to their preferences, a candidate from any one party is equally likely to obtain the highest number of votes as any other candidate. In principle, it is possible to choose the municipalities this way once the preference parameters have been identified. Now suppose that candidates from one party are in close contention less often than candidates from other parties.

Strategic voters who prefer candidates from this party will then strategically vote for a candidate from another party who has a greater chance of winning. As a result, the frequency with which each candidate obtains the highest number of votes in the municipalities may not be equal. As the fraction of strategic voters increases, the effect of the multiplicity of equilibria becomes more severe and introduces the possibility of more skew in the frequency with which each candidate obtains the highest number of votes. Hence, the degree of this skew identifies the fraction of strategic voters.[3]

Our identification argument is similar to that of Sweeting (2009), who uses the multiplicity of equilibria for identification. In a study of radio-commercial timing, he uses commercial-timing dispersion across markets, caused by the multiplicity of equilibria, in order to identify coordination benefits.

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. Our construction of inequality restrictions is similar to that of Aradillas-Lopez and Tamer (2008) in that we both use a weaker solution concept than Nash Equilibrium, such as rationalizability and iterated deletion of dominated strategies, to obtain restrictions.

We use data on the Japanese House of Representatives elections for estimation.[4] Once the primitives 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 of voters are strategic voters. The mean of the estimated proportion of strategic voters is (68.2%, 82.7%). We can also recover the fraction of misaligned voting once we estimate the model, by simulating the equilibrium behavior. Our results show that (1.9%, 5.0%) of the voters engage in misaligned voting, or (2.2%,7.4%) of the strategic voters. A counterfactual experiment that in-

[3] Note that the identification argument above involves comparing municipalities with very particular features. However, this was only for expositional ease and the argument applies equally well to any frequency of obtaining the highest number of votes and to any set of demographic characteristics. For instance, we could easily have considered any arbitrary frequency and examined the deviation from it. Differences in demographics can also be controlled for because the preference parameters have already been identified.

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

troduces 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. A second counterfactual experiment, which assumes sincere voting by all voters, shows a significant increase in the vote share for candidates of a minor party. The difference between the actual outcome and the outcome in our first counterfactual is composed of two effects. Our second counterfactual can isolate and quantify one of the effects.

There are both experimental and 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.[5] 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 actual vote to his/her 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 (1.9%, 5.0%) is comparable to the estimates of strategic voting reported in the previous literature, which ranges from 3% to 17%.[6]

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 preference and the fraction of strategic voters is then possible without requiring data on individuals' repeated voting records. This is analogous to estimating individual preferences from aggregate market level data as in Berry, Levinsohn, and Pakes (1995).

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

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

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 appear in Section 6. Finally, we close the paper with concluding remarks in Section 7.

2 Model

2.1 Model Set-up

Our model is an adaptation of Myerson and Weber (1993) [hereafter denoted as MW] and Myerson (2002). We model plurality-rule elections in which N candidates compete for one seat. Voters cast a vote for one candidate,[7] 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 $N \geq 3$ since strategic voting is otherwise not an issue. There are L municipalities in an electoral district, and we use subscript $l \in \{1, 2, ..., L\}$ to denote a municipality. There are a finite number of voters, $\sum_{l=1}^{L} K_l < \infty$, who are the players of the game (K_l is the number of voters in municipality l). Voter k's utility from having candidate i in office is

$$u_{ki} = u(x_k, z_i) + \varepsilon_{ki}, \tag{1}$$

where x_k are voter characteristics, z_i are candidate characteristics, and ε_{ki} is a preference shock.

We consider two types of voters, sincere (behavioral) and strategic (rational). Sincere voters cast their votes for the candidate they most prefer, i.e., a sincere voter votes for candidate i if and only if $u_{ki} \geq u_{kj}$ $\forall j$. On the other hand, strategic voters cast their votes considering the probability that their votes are pivotal. Hence, if we let $T_{k,ij}$ denote voter k's beliefs that candidates i and j will be tied for first place (or that i will be one vote behind j), the expected utility from voting for candidate i is

[7] We abstract from the issue of voter abstention.

given by[8]

$$\bar{u}_{ki} = \sum_{j\in\{1,..,N\}} T_{k,ij}(u_{ki} - u_{kj}),$$

and strategic voters vote for candidate i if and only if $\bar{u}_{ki} \geq \bar{u}_{kj}$ $\forall j$. 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. Being a strategic voter is a necessary condition for misaligned voting, but not a sufficient condition.

We denote the type of voter by $\alpha_{kl} \in \{0, 1\}$ where $\alpha_{kl} = 0$ denotes the sincere voter and $\alpha_{kl} = 1$ denotes the strategic voter. We write the probability that voter k in municipality l is a strategic voter as

$$\Pr(\alpha_{kl} = 1) = \alpha_l \eta_k,$$

where α_l is a municipality-level random term and η_k is an individual-voter-level shock with $E[\eta_k] = 1$. The probability that the voter is sincere is $\Pr(\alpha_{kl} = 0) = 1 - \alpha_l \eta_k$.

We make the following assumption on beliefs $\{T_{k,ij}\}$ following MW.

Assumption *Beliefs over tie probabilities $\{T_{k,ij}\}$ are common across all voters in the same electoral district, i.e., $\{T_{k,ij}\} = \{T_{ij}\}$, $\forall\ k \in \{1, ..., K_1\} \cup ... \cup \{1, ..., K_L\}$.*

This assumption simply imposes common beliefs over tie probabilities among candidates on voters in the same electoral district . The assumption reflects the fact that information regarding the expected outcome of the election is widely available from news reports and poll results. Based on this information, voters form similar beliefs regarding the outcome.

Let P_{il}^{SIN} be the fraction of votes cast by sincere voters for candidate i in municipality l, and let $P_{il}^{STR}(\{T_{ij}\})$ be the fraction of votes cast by strategic voters for candidate i. Note that $P_{il}^{STR}(\{T_{ij}\})$ is a function of beliefs, $\{T_{ij}\}$. We can write these

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

fractions as

$$
\begin{aligned}
P_{il}^{SIN} &= \frac{\sum_{k=1}^{K_l}(1-\alpha_{kl})\cdot \mathbf{1}\{u_{ki}\geq u_{kj}\forall j\}}{\sum_{k=1}^{K_l}(1-\alpha_{kl})}, \\
P_{il}^{STR}(\{T_{ij}\}) &= \frac{\sum_{k=1}^{K_l}\alpha_{kl}\cdot \mathbf{1}\{\overline{u}_{ki}\geq \overline{u}_{kj}\ \forall j,\{T_{ij}\}\}}{\sum_{k=1}^{K_l}\alpha_{kl}}.
\end{aligned}
$$

The total vote share for candidate i in municipality l is then

$$
P_{il}^{TOT}(\{T_{ij}\}) = \frac{\sum_{k=1}^{K_l}(1-\alpha_{kl})}{K_l}P_{il}^{SIN} + \frac{\sum_{k=1}^{K_l}\alpha_{kl}}{K_l}P_{il}^{STR}(\{T_{ij}\}). \tag{2}
$$

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

$$
\begin{aligned}
P_{il}^{SIN} \underset{p}{\rightarrow} \iint \mathbf{1}\{u_{ki} &\geq u_{kj}\ \forall j\}]dF_{x,l}dF_{\varepsilon} \\
&\equiv v_{il}^{SIN},
\end{aligned}
$$

$$
\begin{aligned}
P_{il}^{STR}(\{T_{ij}\}) \underset{p}{\rightarrow} \iint \mathbf{1}\{\overline{u}_{ki} &\geq \overline{u}_{kj}\ \forall j,\{T_{ij}\}\}dF_{x,l}dF_{\varepsilon} \\
&\equiv v_{il}^{STR}(\{T_{ij}\}),
\end{aligned}
$$

where $F_{x,l}$ denotes the distribution of the demographic characteristics, x, in municipality l, F_{ε} denotes the distribution of idiosyncratic shock, ε, $\int dF_{\varepsilon}$ denotes integration over N random variables ε_{ki}, and $\int dF_{x,l}$ denotes integration over all dimensions of demographic characteristics x. Analogously, we obtain similar expression for the total vote share as K_l becomes large;

$$
\begin{aligned}
P_{il}^{TOT}(\{T_{ij}\}) &\underset{p}{\rightarrow} (1-\alpha_l)v_{il}^{SIN} + \alpha_l v_{il}^{STR}(\{T_{ij}\}) \\
&\equiv v_{il}(\{T_{ij}\}).
\end{aligned} \tag{3}
$$

2.2 Solution Concept and 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. We diverge in the solution concept, however. In addition to the common belief assumption, MW further place strong consistency requirements on beliefs $\{T_{ij}\}$ in equilibrium, which results in outcomes that may not rationalize diverse patterns of actual election data even when we add sincere voters to their model.[9] Thus, in order to fit the data and be robust to alternative specifications regarding beliefs, we adopt a weaker solution concept.

We take an approach that is closely related to that of Dhillon and Lockwood (2004) and Buenrostro, Dhillon and Vida (2007), who study dominance solvability of election models. Our solution concept combines iterated deletion of weakly dominated strategies and common beliefs across voters. We require that players have common beliefs $\{T_{ij}\}$ and only play actions that survive iterated deletion of weakly dominated strategies. This does not impose strong consistency requirements on beliefs $\{T_{ij}\}$ such as rational expectations.

Let S^n denote the set of strategy profiles that survive the n-th iterated deletion of weakly dominated strategies via maximal simultaneous deletion. S^1 is the set of strategy profiles in which there does not exist any strategy profile such that a voter votes for the least-preferred candidate. Let $T(S^\infty)$ be the set of tie probabilities that can be rationalized by S^∞.

[9] In an election with three candidates, the original equilibrium of MW predicts that either (i) the second and third 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 candidate receives zero votes (with beliefs $\{T_{12}, T_{13}, T_{23}\} = \{1, 0, 0\}$). Even if we introduce sincere voters in the MW model, there would still be two types of equilibria, one with beliefs $\{T_{12}, T_{13}, T_{23}\} = \{p, 1-p, 0\}$ and the other with $\{T_{12}, T_{13}, T_{23}\} = \{1, 0, 0\}$. Equilibrium (i) still has the undesirable property that the second and third candidates receive the exact same number of votes, which is not supported by the data. In equilibrium (ii), all three candidates can receive a positive and different number of votes, but the only belief that can support the equilibrium is $\{T_{12}, T_{13}, T_{23}\} = \{1, 0, 0\}$, which is a very strong assumption to impose, unlikely to be true in many races.

Definition *A set of solution outcomes $W \subseteq \Delta^N \times L$ is defined as the set W s.t.*

$$W = \left\{ \{W_l = (P_{1l}^{TOT}, P_{2l}^{TOT}, ..., P_{Nl}^{TOT})\}_{l=1}^{L} \middle| P_{il}^{TOT}(\{T_{ij}\}) = \frac{\sum_{k=1}^{K_l}(1-\alpha_{kl})}{K_l} P_{il}^{SIN} \right.$$

$$\left. + \frac{\sum_{k=1}^{K_l}\alpha_{kl}}{K_l} P_{il}^{TOT}(\{T_{ij}\}), \quad \{T_{ij}\} \in T(S^{\infty}) \right\}. \qquad (4)$$

This solution outcome first applies iterated deletion of a weakly dominated strategy, and then requires common beliefs over the set of fully reduced strategies, S^{∞}. Note that we do not impose the common belief assumption at the time of deleting weakly dominated strategies.

It is well known that the iterated deletion of weakly dominated strategies depends on the order of elimination, in general. In our case, the order of deletion does not affect the solution outcome because voters have strict preferences with probability one due to idiosyncratic preference shocks. This implies that condition (A1) in Dhillon and Lockwood (2004) is satisfied in our model, which, in turn, is a sufficient condition for transference of decision-maker indifference, (TDI*), a condition identified by Marx and Swinkels (1997), who guarantee that the solution is order-independent.

Note that W depends on the informational structure we assume, i.e., whether we assume that the voters know the realization of α_{kl} and ε_{ki}, or only their distribution. Under both structures, a strategy profile in which voters vote for their least-preferred candidate can be deleted immediately. In our particular application, we base our estimation only on this restriction, which holds under both informational structures.

Finally, note that the set of outcomes, W, is not a singleton in general, and each outcome in W has a corresponding belief $\{T_{ij}\} \in T(S^{\infty})$. Thus, we have a multiple solution outcome (we use the term equilibrium interchangeably with the term solution outcome hereafter) in the model.[10]

[10] With complete information, Dhillon and Lockwood (2004) have shown that the set of outcomes, W, is a singleton when a large proportion of the population agrees on who the worst candidate is at every round of deletion.

District	Municipality	Candidate 1	Candidate 2	Candidate 3	Candidate 4
		Ibuki	Tamaki	Kokuta	
Kyoto 1	Kita-ku	28,018	17,468	15,384	
	Kamigyo-ku	20,090	12,179	9,754	
	⋮	⋮	⋮	⋮	
⋮	⋮	⋮	⋮	⋮	
		Narumiya	Kitagami	Nakagawa	Tanaka
Kyoto 4	Ukyo-ku	16,929	30,967	31,021	17,945
	Miyama-cho	581	498	808	1,554
	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮

Table 1: Data Structure. This is an example of vote counts. Demographics are recoreded similarly. Each district is a unit of observation.

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.[11] For each electoral district, the breakdown of vote-share data is available by municipality as shown in Table 1. An electoral district is usually comprised of several municipalities (9.55 on average in our sample).[12] This particular data structure plays an important role in our identification.

[11] An additional 180 Representatives were elected by proportional representation from 11 regional electoral districts. In proportional representation, voters cast ballots for parties, and 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 such candidates are ranked equally on the party list, the result of the plurality rule affects the relative rank of the two candidates. Only the LDP and the DPJ ranked more than two candidates equally in this election.

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

The data on the vote share and candidate characteristics were collected by *Yomiuri Shimbun*, a national newspaper publisher. The demographic characteristics we use are obtained from *Social and Demographic Statistics of Japan* published by the Statistics Bureau of the Japanese Ministry of Internal Affairs and Communications.[13] We match these two data sets at the municipality level.

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

(i) There are three or four candidates, and the composition of the candidates' parties in the district is any three 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 that satisfy the three criteria.[14] We drop samples that do not satisfy criterion (i) because we treat party affiliation as a candidate characteristic, and our moment inequalities are constructed for each combination of parties. Criterion (i) ensures that we have enough elections with the same combination of candidate parties to construct our moment inequalities. We need criterion (ii) because our estimation requires at least two municipalities in each electoral district. Criteria (iii) is required to deal with an issue that arises when merging two data sets. Because the data sets 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 properly the data and, thus, keep some of these merging municipalities in the sample.

We report the descriptive statistics of electoral district vote shares in Table 2. 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

[13] 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.

[14] The Kagoshima 5th District is the only district we dropped that satisfied all three criteria. We excluded this district because there was no other district in which the combination of candidate parties was the same as in this district (LDP, JCP, YUS).

	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 2: Descriptive Statistics of Electoral Districts – Vote Shares

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 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%.[15]

Table 3 reports the descriptive statistics of candidate characteristics. The first three rows contain information on the candidates' hometowns.[16] 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

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

[16]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.

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

Table 3: Descriptive Statistics of Electoral Districts – Candidate Characteristics. The mean of each variable is reported. Standard errors are in parenthesis.

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.[17]

Table 4 reports the descriptive statistics of the municipalities' demographic characteristics. The mean income per capita is about 3.16 million yen, and the mean of municipality average years of schooling is about 12 years. 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.[18]

[17] 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.

[18] 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, but we do not have data that permits us to compute the dispersion. We compute the quantiles of the income

	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 ≤ 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
≥ 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 4: Descriptive Statistics of Municipalities

4 Identification and Estimation

Our identification strategy is to use the solution concept introduced in the model section to relate variation in the vote share to variation in the characteristics of candidates and demographics. As described in the previous section, our election data includes observations from many districts, and for each electoral district we have a municipality-level breakdown of vote-share data and demographic characteristics. In terms of model notation, the number of elections, denoted by M, is large ($M \to \infty$), but the number of municipalities per electoral district, denoted by L^m, is small ($L^m < \infty, \forall m \in \{1, ..., M\}$). We assume that voting games are played in M districts independently of each other, and we treat each district as a unit of observation. In terms of estimation, we need to consider the issue of multiplicity of equilibria because our solution concept does not provide a unique outcome. We deal with the issue of multiplicity by using an inequality-based estimator developed by Pakes, Porter, Ho and Ishii (2007).

4.1 Specification

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

$$u_{kli} = u(x_k, z_{il}; \theta^{PREF}) + \epsilon_{kli},$$

distribution by assuming a log-normal distribution where the variance is calculated by fitting the prefecture-level income distribution. Data on 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.

where $\epsilon_{kli} = \xi_{il} + \varepsilon_{ki}$. In our specification, we let the unobservable error ϵ_{kli} be composed of an idiosyncratic candidate-municipality level shock, ξ_{il}, which follows i.i.d. normal distribution, $F_\xi = N(0, \theta_\xi)$, as well as an idiosyncratic voter-candidate shock, ε_{ki}, which is distributed Type I extreme value. An example of shock ξ_{il} is the candidate's ability to bring pork spending to municipality l. θ^{PREF} is a vector of the preference parameters of the model. x_k denotes the characteristics of voter k, including years of education, income level, and an indicator of whether or not the voter is above 65. z_{il} is a vector of observable attributes of candidate i at municipality l. z_{il} includes the candidate's previous political experience, political party and an indicator of whether municipality l is the candidate's hometown, which is why z_{il} is indexed by l. For $u(x_k, z_{il}; \theta^{PREF})$, we assume a quadratic loss function as follows:

$$u(x_k, z_{il}; \theta^{PREF}) = -(\theta^{ID} x_k - \theta^{POS} z_i^{POS})^2 + \theta^{QLTY} z_{il}^{QLTY}, \tag{5}$$

where $z_{il} = \{z_i^{POS}, z_{il}^{QLTY}\}$ are two separate types of candidate attributes and $\theta^{PREF} = \{\theta^{ID}, \theta^{POS}, \theta^{QLTY}\}$. We let the ideology of the voter be a function of his demographics, $\theta^{ID} x_k$, and the ideology of the candidate be $\theta^{POS} z_i^{POS}$. The utility of the voter depends on the distance between his ideology, $\theta^{ID} x_k$, and that of the candidate, $\theta^{POS} z_i^{POS}$, which is captured by the quadratic term. The additive term captures the non-ideological component of utility, which we write as $\theta^{QLTY} z_{il}^{QLTY}$.

As described in the model section, the objective of a sincere voter is to vote for a candidate i, who maximizes u_{kli}, while the objective of a strategic voter is to vote for a candidate i, who maximizes

$$\bar{u}_{kli} = \sum_{j \in \{1,..,N\}} T_{ij} (u_{kli} - u_{klj}).$$

We assume that for at least some pair $\{i, j\}$, T_{ij} is positive, no matter how small. Then, as utility representation is invariant to a constant scaling factor, we can assume $\sum_{i,j} T_{ij} = 1$ for convenience.

Recall that the probability that voter k in municipality l is a strategic voter is written as

$$\Pr(\alpha_{kl} = 1) = \alpha_l \cdot \eta_k,$$

where α_l is the term common to all voters in municipality l, and η_k is the individual-voter-level shock with $E[\eta_k] = 1$. We assume that α_l follows a Beta distribution,

$F_{\alpha} = Beta(\theta_{\alpha 1}, \theta_{\alpha 2})$.

4.2 Identification

We informally discuss identification in this section. Beliefs $\{T_{ij}^m\}$ (hereafter we use superscripts to index electoral districts), are different across electoral districts simply because the candidates and the electorate are different. Moreover, we cannot treat the beliefs, $\{T_{ij}^m\}$, in each district m, as parameters because the number of districts is large ($M \to \infty$) while the number of municipalities within each district is small ($L^m < \infty$). Because we cannot treat $\{T_{ij}^m\}$ as parameters, we need restrictions that do not involve $\{T_{ij}^m\}$.

First, we discuss the identification of the preference parameters that relate a voter's demographic characteristics to his ideological position, θ^{ID}. Preference parameters, θ^{ID}, are identified by demographic and vote-share variation within each electoral district. Consider two municipalities, l_1^{m0} and l_2^{m0}, from the same electoral district, $m0$, whose distribution of demographic characteristics differs only along one dimension (we call this characteristics x^n and we call the corresponding preference parameter θ^n). As $M \to \infty$, we can find districts $m1, m2, ...$ so that pairs of municipalities, $\{(l_1^{m1}, l_2^{m1}), (l_1^{m2}, l_2^{m2}), ...\}$, have the same distribution of demographic characteristics as the pair (l_1^{m0}, l_2^{m0}). Suppose that we can observe $\{T_{ij}^m\}$. Then, we can point identify parameter θ^n from the difference in the vote share and the corresponding difference in the demographic characteristics between observations $\{l_1^{m0}, l_1^{m1}, l_1^{m2}, ...\}$ and $\{l_2^{m0}, l_2^{m1}, l_2^{m2}, ...\}$.[19] The preceding argument assumed that the values of $\{T_{ij}^m\}$ are known. Although we cannot observe $\{T_{ij}^m\}_{m=1}^M$, $\{T_{ij}^m\}$ is restricted to lie on a N-dimensional simplex, where N is the number of candidates. By assigning arbitrary values of $\{T_{ij}^m\} \in \Delta^N$ to each m, we can obtain different parameter values for θ^n denoted by $\theta^n(\{T_{ij}^m\}_{m=1}^M)$. We identify the upper and lower bounds of θ^n by taking the maximum and minimum of $\theta^n(\{T_{ij}^m\}_{m=1}^M)$ with respect to $\{T_{ij}^m\}_{m=1}^M$.

For example, consider the identification of a preference parameter, $\theta^{above65}$, which captures the effect of age (being 65 or older) on ideology, for the case of $N = 3$.[20] Let l_{old}^{m0} and l_{young}^{m0} be two municipalities in district $m0$, with 30% and 20% of the population above age 65, respectively, but with otherwise the same demographic char-

[19] Note that effects of α and ξ will be averaged out when we have many municipalities.

[20] In this example, we use the case of $N = 3$ candidates, but our discussion applies to the case of $N > 3$ as well.

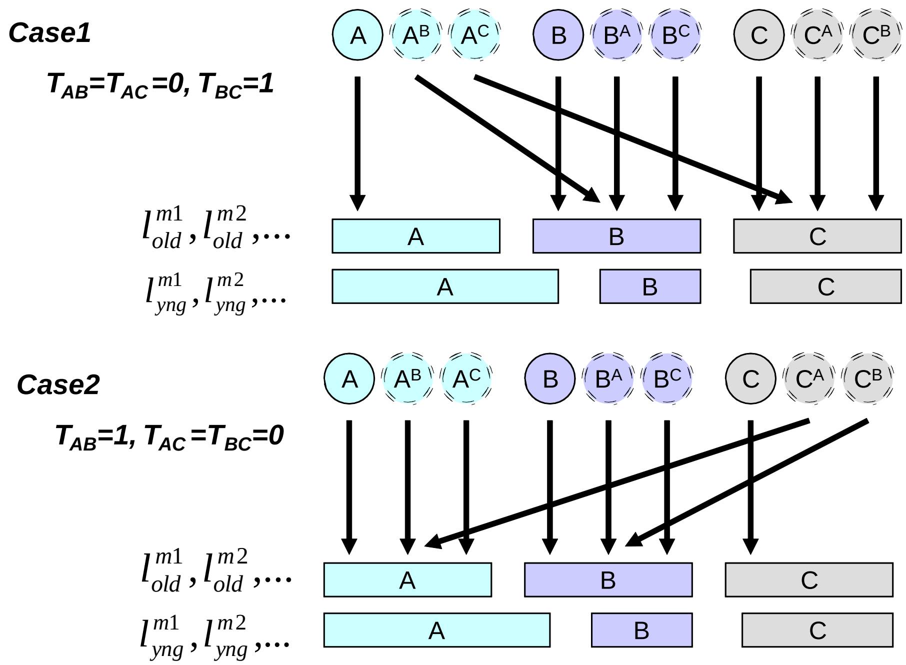

Figure 1: Identification of Preference. Sincere voters are illustrated with regular circles, and strategic voters, with dotted circles. The letters inside the circle indicate the most-preferred candidates and the superscripts indicate the second preferred candidates for strategic voters. The rectangles indicate respective vote shares. In Case (1), only sincere voters who rank candidate A first vote for A. In Case (2), both sincere and strategic voters who rank candidate A first vote for A as do strategic voters who rank candiate C first and A second. The 5% difference in the vote share is then attributed to the difference in the behavior of only sincere voters in Case (1), 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 $\{T_{ij}\}$.

acteristics. We can find municipalities in other districts $\{(l_{old}^{m1}, l_{young}^{m1})$, $(l_{old}^{m2}, l_{young}^{m2})$, $...\}$ that have the same demographic characteristics as the pair, $(l_{old}^{m0}, l_{young}^{m0})$. Suppose that the average vote share for Party A in municipalities $\{l_{old}^{m0}, l_{old}^{m1}, l_{old}^{m2}, ...\}$ is 5% lower than in municipalities $\{l_{young}^{m0}, l_{young}^{m1}, l_{young}^{m2}, ...\}$. In order to exhibit how $\{T_{ij}^m\}$ affects identification of $\theta^{above65}$, we now consider two polar cases as illustrated in Figure 1: (1) the belief that the tie probability between candidates from Parties B and C is one, and that the other two tie probabilities are zero ($T_{BC}^m = 1$, $T_{AB}^m = T_{AC}^m = 0$, $\forall m \in \{m0, m1, ...\}$), and (2) the belief that the tie probability between candidates from Parties A and B is one, and that the other two tie probabilities are zero ($T_{AB}^m = 1$,

$T^m_{AC} = T^m_{BC} = 0$, $\forall m \in \{m0, m1, ...\}$). In case (1), no strategic voter votes for the Party A candidate; hence, the 5% difference in the vote share must be attributed to the difference in the sincere voters' behavior alone. Because the 5% difference must be explained only by the fraction of the population that is sincere, the effect of parameter $\theta^{above65}$ must be quite large. Next, consider case (2). In this case, the votes for Party A candidate come not only from sincere voters, but also from strategic voters. The 5% difference in the vote share for the Party A candidate can then be accounted for due to 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 because we can attribute the 5% difference to the difference in the behavior of both types of voters.

The parameters on candidate characteristics, θ^{POS} and θ^{QLTY}, can similarly be 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 vote shares between the candidates with no experience and those with experience, controlling for other candidate and demographic characteristics.

Second, we discuss the identification of the average fraction of strategic voters, i.e., the mean of the Beta distribution, μ $(= \theta_{\alpha 1}/(\theta_{\alpha 1} + \theta_{\alpha 2}))$. In the following discussion, we fix a particular preference parameter, $\theta^{PREF} = \{\theta^{ID}, \theta^{POS}, \theta^{QLTY}\}$, and consider the identification of μ 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 μ.[21] Consider districts $\{m1, m2, ...\}$ and corresponding municipalities $\{l^{m1}, l^{m2}, ...\}$ such that their demographics are the same. Given θ^{PREF}, the model can predict the distribution of

[21] Our two-step identification strategy can be schematically described as follows. Let Θ^{PREF} and Θ^{α} be the space in which the parameters θ^{PREF} and $\theta^{\alpha} = (\theta_{\alpha 1}, \theta_{\alpha 2}, \theta_{\xi})$ lie. 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 claim that this inclusion can be strict, $I_2(I_1(\Theta^{\alpha})) \subsetneq \Theta^{\alpha}$. This would be the case if for some θ^{α}, $I_1(\{\theta^{\alpha}\}) = \phi$. Here, we will give an example. Let $\theta^{\alpha_1}$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 preference. 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 that of a very liberal municpality in a generally liberal district. This is because liberal candidates have little chance of winning in a consevative district, so strategic voters in thosel districts vote for other candidates.

Our second step in 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.

vote shares for these municipalities if all voters are sincere. The aggregate distribution of predicted vote shares across these municipalities will be close to the observed aggregate distribution of actual vote shares if μ is close to zero. On the other hand, if μ is large, the vote shares across these municipalities can be far from the predicted vote shares depending on the beliefs, $\{T_{ij}^m\}$, in each district. This is due to the multiplicity of equilibria induced by strategic voters. 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 simply corresponds to different equilibria being played in different districts. Hence, the larger is the μ, the more divergent is the observed outcome from the predicted outcome. The extent to which the predicted outcome differs from the actual outcome identifies μ. Note that, similar to Sweeting (2009), we use multiplicity of equilibria for the identification of μ.[22]

For illustration, consider districts $\{m1, m2, m3, ...\}$ and municipalities $\{l^{m1}, l^{m2}, l^{m3}, ...\}$ so that $\{l^{m1}, l^{m2}, l^{m3}, ...\}$ have the same demographic characteristics. Moreover, we choose these municipalities in such a way that $(h^A, h^B, h^C) = (1/3, 1/3, 1/3)$, where h^i is the frequency with which the candidate from party i wins the most votes in the municipality (but not necessarily in the entire electoral district) if all voters are sincere. That is, in the absence of strategic voters, the Party A candidate obtains the highest number of votes in the municipality one-third of the time on average, as do the other two candidates. Now suppose that there is a large number of strategic voters.

[22] In addtion to this argument, identification is also facilitated by the variation *within* electoral districts generated by the multiplicity of equilibria as in Sweeting (2009). The particular structure of the data allows us to compare varinace in vote shares within districts and variance across all districts. A higher proportion of strategic voters is consistent with a higher level of variance in vote shares across electoral districts than within electoral districts. Thus, the variance ratio partially identifies the proportion of strategic voters.

For illustration, consider two electoral districts where the candidates have exactly the same characteristics and all the municipalities have the same demographic characteristics. If there are no strategic voters, the vote share across municipalities would differ only due to random shocks. Hence, the variance in the vote share within each electoral district should not be different from the variance computed by pooling the municipalities across the two districts. However, if there are strategic voters, the pooled variance may be larger than the variance in each electoral district. This is again due to the multiplicity of equilibria induced by strategic voters. If the beliefs on pivot probabilities are different in the two electoral districts, the behavior of strategic voters will be different across the two districts. Hence, the variation within each electoral district is still only induced by random shocks, but now there can be large differences in the vote share across the two electoral districts, which results in an increase in the pooled variance relative to the variance within each electoral district. Thus, the variance ratio helps identify the fraction of strategic voters.

h^A decreases as the frequency of close elections involving candidates from Party A decreases. If candidates from Party A are seldom in close contention, strategic voters who most prefer Party A will vote for candidates from either Party B or Party C. Hence, the distribution (h^A, h^B, h^C) will be skewed in favor of the candidates from Parties B and C, and against candidates from Party A. The degree of this skewness (partially) identifies the degree of strategic voting.

The identification argument for μ compared municipalities with very particular features only for expositional purposes. However, the argument applies equally well to any frequency $(h^1, ..., h^N)$ and to any demographic characteristics because preference parameters are already identified and differences in demographics can be controlled for. We can also easily consider any arbitrary frequency and examine the deviation from it.

Finally, we discuss the identification of $\theta_{\alpha 1}$, $\theta_{\alpha 2}$, and θ_{ξ}. Once the preference parameters, θ^{PREF}, are identified, we know from the above discussion that μ can also be partially identified. Note that after controlling for μ and for the preferences, the randomness in the vote share across municipalities within an electoral district can only be accounted for by the randomness in ξ and α. As the observed variance of vote shares is attributed either to the variance of ξ, θ_{ξ}, or to the variance of the Beta distribution, σ, we can obtain bounds for these two parameters. As $\theta_{\alpha 1}$ and θ_{a2} corresponds one-to-one to μ and σ, partial identification of μ and σ implies that $\theta_{\alpha 1}$ and θ_{a2} are partially identified.

4.3 Estimation

We estimate the model using inequality-based estimator developed by Pakes, Porter, Ho, and Ishii (2007). If voter beliefs, $\{T_{ij}\}$, were known, a unique outcome would correspond to every realization of the unobserved error terms (ξ, α). In such a case, we could employ estimation procedures such as GMM or MLE. However, the multiplicity of outcomes induced by the presence of strategic voters, together with the fact that we cannot observe voter beliefs, $\{T_{ij}\}$, make the inequality-based estimator appropriate.

We construct the moment inequalities using the following idea, which is somewhat similar to indirect inference (Smith (1993) and Gouriéroux, Monfort, and Renault (1993)). First, we fix some parameter θ. If we knew the beliefs of voters ($\{T_{ij}\}$), we can simulate the vote share outcome. By regressing the simulated vote share on the demographic characteristics in each municipality, we can capture the relationship between the vote share and the demographic characteristics as coefficients of the regression. Because we cannot observe $\{T_{ij}\}$, we vary $\{T_{ij}\} \in \Delta^N$ to obtain the minimum and maximum values of the regression coefficients.[23] At the true parameter θ_0, the coefficients obtained by regressing the actual vote share on demographics should fall between the minimum and maximum values. If for some θ, the regression coefficients obtained using the actual vote share does not lie between the minimum and maximum values of the regression coefficients computed using θ, then θ is not in the identified set. This is how we construct one set of moment inequalities. We also use the mean vote share conditional on candidate characteristics to construct another set of moment inequalities in a similar fashion.

We use municipality-level aggregate data for our estimation. We denote the vote-share data of candidate i in municipality l by v_{il}^{data}, and the distribution of demographic characteristics, x_k, in municipality l by $f_{x,l}$. We let $(\varepsilon_{ki})_{i=1}^{N}$ denote the N draws of individual-candidate-specific shock, and we let $f_{(\varepsilon)}$ denote the distribution of (ε_{ki}). Lastly, candidate i's characteristics are denoted by z_{il}.

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

$$\begin{aligned} v_{il}^{data}(\{T_{ij}^m\}, f_x, z_i; \theta_0, \{\xi_{il}\}, \alpha_l) &= (1-\alpha_l) v_{il}^{SIN}(\theta_0, \{\xi_{il}\}) \\ &\quad + \alpha_l v_{il}^{STR}(\{T_{ij}^m\}; \theta_0, \{\xi_{il}\}), \\ \forall l &\in \{1, ..., L^m\}, \forall\, m \in \{1, ..., M\}, \end{aligned} \tag{6}$$

where

$$v_{il}^{SIN}(\theta_0, \{\xi_{il}\}) = \int_{(\varepsilon_{ki})} \int_{x} \mathbf{1}\{u_{ki} \geq u_{kj}\ \forall j\} f_{x,l} dx f_{(\varepsilon)} d(\varepsilon_{ki})$$

[23] In the actual implementation with observations such that , we assume that $T_{ij} \geq T_{ik} \Rightarrow T_{lj} \geq T_{lk}$ (which only has bite for $N \geq 4$). Intuitively, this assumption introduces a natural ordering in terms of who is most likely to win. For example, if candidates $1, 2, 3, 4$ are likely to win in decreasing order, it is natural to expect $T_{12} \geq T_{13} \geq T_{14}$, and $T_{13} \geq T_{23} \geq T_{24}$ and $T_{14} \geq T_{24}$. Of course our assumption is less restrictive than this, but our assumption is similar in spirit.

In principle, we do not require this assumption for identification nor estimation, but we require it when we construct our measure of misaligned voting after the estimation. The assumption ensures the following: If a voter who prefers candidate i most votes strategically for candidate j, then no voter who prefers candidate j most will vote for candidate i.

is the expression for the vote share for candidate i among sincere voters, and

$$v_{il}^{STR}(\{T_{ij}^m\};\theta_0,\{\xi_{il}\}) = \\ \int\limits_{(\varepsilon_{ki})}\int\limits_{x} \mathbf{1}\{\overline{u}_{ki} \geq \overline{u}_{kj}\ \forall j, \{T_{ij}^m\}\} f_{x,l} dx f_{(\varepsilon)} d(\varepsilon_{ki}),$$

is the corresponding expression for the strategic voters. Note that $\int\limits_{(\varepsilon_{ki})} f_{(\varepsilon)} d(\varepsilon_{ki})$ denotes integration over N random variables (ε_{ki}) and $\int\limits_{x} f_{x,l} dx$ denotes integration over all dimensions of the demographic characteristics. This expression is obtained by assuming that the vote share for each municipality is generated by an infinite number of voters in each municipality. Of course, the number of voters in each municipality is finite, but we assume that 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.[24]

Now, we construct moment inequalities based on the regression coefficients in each electoral district. Given some parameter θ, beliefs $\{T_{ij}^m\}$, and realizations $\{\xi_l\}_l$, $\{\alpha_l\}_l$, we can predict the vote share for the candidates in each of the municipalities by using expression (6) and replacing θ_0 with θ, which we denote as $v_{il}(\{T_{ij}^m\},\theta,\xi_l,\alpha_l)$. Now consider regressing L^m predicted vote shares on demographic characteristics for each of the L^m municipalities. We let $\beta_{i,m}$ denote the coefficient obtained by regressing $v_{il}(\{T_{ij}^m\},\theta,\xi_l,\alpha_l)$ onto $f_{x,m,l}$ for each candidate in district m, i.e.,

$$\beta_{i,m}(\{f_{x,m,l}\}_l,\{z_{il}\}_l,\theta,\{T_{ij}^m\},\{\xi_l\}_l,\{\alpha_l\}_l) = \\ \arg\min_{\beta}\left[\sum_{l=1}^{L^m}\left(v_{il}(\{T_{ij}^m\},\theta,\xi_l,\alpha_l) - \beta\cdot f_{x,m,l}\right)^2\right].$$

[24] Correspondingly, we only use restrictions that are robust to the realization of individual voter level draws of $(\varepsilon_{ki},\eta_k)$, to the precise informational structure, and also to the number of voters in the municipality. We define W' as

$$W' = \left\{\{W_l' = \{v_l(\{T_{ij}\})\}_{l=1}^L \;\middle|\; \begin{array}{c} v_l(\{T_{ij}\}) = (1-\alpha_l)\Delta_{l0} + \alpha_l\Delta_l(\{T_{ij}\}), \\ \text{for some } \{T_{ij}\} \in T(S^1) \end{array}\right\}.$$

The difference between W and W' is that P_{il}^0 has been replaced by its infinite counterpart, and S^∞ has been replaced by S^1. This means that we use the fact that voters eliminate voting for their least-preferred candidate. Note that $w \in W'$ is a necessary condition for $w \in W$ and, hence, is a weaker restriction on the possible set of outcomes than W. However, this gives us testable parameter restrictions, which we use for the estimation of parameters.

The details on the construction of $\beta_{i,m}$ are provided in Appendix A. Similarly, we let $\gamma_{i,m}$ denote the average vote share conditional on candidate characteristics for each candidate characteristic z_{il}, i.e.,

$$\gamma_{i,m}(\{f_{x,m,l}\}_l, \{z_{il}\}_l, \theta, \{T_{ij}^m\}, \{\xi_l\}_l, \{\alpha_l\}_l) = \sum_{l=1}^{L^m} \mathbf{1}\{z = z_{il}\} \cdot v_{il}(\{T_{ij}^m\}, \theta, \xi_l, \alpha_l),$$

where $\mathbf{1}\{\cdot\}$ denotes an indicator function. The details on the construction of $\gamma_{i,m}$ are provided in Appendix A. Note that we obtain $\beta_{i,m}$ and $\gamma_{i,m}$ for each district.

Corresponding to different values of $\{T_{ij}^m\}$,[25] which cannot be observed by the econometrician, the coefficient $\beta_{i,m}(\{f_{x,m,l}\}_l, \{z_{il}\}_l, \theta, \{T_{ij}^m\}, \{\xi_l\}_l, \{\alpha_l\}_l)$ takes values between $\overline{\beta}_{i,m}(\{f_{x,m,l}\}_l,\{z_{il}\}_l,\theta,\{\xi_l\}_l,\{\alpha_l\}_l)$ and $\underline{\beta}_{i,m}(\{f_{x,m,l}\}_l,\{z_{il}\}_l,\theta,\{\xi_l\}_l,\{\alpha_l\}_l)$, defined as[26]

$$\begin{aligned}
\overline{\beta}_{i,m}(\{f_{x,m,l}\}_l, \{z_{il}\}_l, \theta, \{\xi_l\}_l, \{\alpha_l\}_l) &= \max_{\{T_{ij}^m\}\in\Delta^N} \beta_{i,m}(\{f_{x,m,l}\}_l, \{z_{il}\}_l, \theta, \{T_{ij}^m\}, \{\xi_l\}_l, \{\alpha_l\}_l), \\
\underline{\beta}_{i,m}(\{f_{x,m,l}\}_l, \{z_{il}\}_l, \theta, \{\xi_l\}_l, \{\alpha_l\}_l) &= \min_{\{T_{ij}^m\}\in\Delta^N} \beta_{i,m}(\{f_{x,m,l}\}_l, \{z_{il}\}_l, \theta, \{T_{ij}^m\}, \{\xi_l\}_l, \{\alpha_l\}_l).
\end{aligned}$$

We obtain the similar expression for γ. Now we integrate over two sets of random

[25] In our estimation, we place the following restriction on $\{T_{ij}\}$: In addition to the restriction on $\{T_{ij}\}$ we mentioned in footnote 23 ($T_{ij} \geq T_{ik} \Rightarrow T_{lj} \geq T_{lk}$), we restrict $\{T_{ij}\}$ to be such that if $v_i > v_j > v_k$, where v_l is the vote share of candidate l, $T_{ij} \geq T_{ik} \geq T_{jk}$ and similarly for four candidate districts.

[26] For any matrix A, we can consider $\overline{A\beta}_{i,m} \equiv \max_{\{T_{ij}^m\}} A\beta_{i,m}$ and $\underline{A\beta}_{i,m} \equiv \min_{\{T_{ij}^m\}} A\beta_{i,m}$, and construct moment inequalities by following the argument presented in the main text. In our implementation, we use the identity matrix for A regarding $\beta_{i,m}$. For $\gamma_{i,m}$, we use $A^T = \begin{pmatrix} & 1 & \cdots & \cdots & 1 & 0 & \cdots & 0 & \cdots & \cdots \\ & -1 & 0 & \cdots & 0 & 1 & \cdots & 1 & 0 & \cdots \\ I_\gamma & 0 & -1 & \ddots & \vdots & -1 & 0 & 0 & 1 & \cdots \\ & \vdots & \ddots & \ddots & 0 & 0 & \ddots & 0 & -1 & 0 \\ & 0 & \cdots & 0 & -1 & \vdots & \ddots & -1 & 0 & \ddots \end{pmatrix}$, where I_γ is the identity matrix.

This allows us to add restrictions on the pairwise difference of the effects of candidates' hometowns on the vote share, e.g., the difference in the vote share for a candidate whose hometown is out of state compared to a candidate whose hometown is within state.

variables $\{\xi_l\}_l$ and $\{\alpha_l\}_l$, to obtain expected upper and lower bounds:

$$\begin{aligned}
\overline{\beta}_{i,m}(\{f_{x,m,l}\}_l, \{z_{il}\}_l, \theta) &= \int_{(\xi_l)} \int_{(\alpha_l)} \overline{\beta}_{i,m}(\{f_{x,m,l}\}_l, \{z_{il}\}_l, \theta, \{\xi_l\}_l, \{\alpha_l\}_l) dF_{(\alpha)} dF_{(\xi)}, \\
\underline{\beta}_{i,m}(\{f_{x,m,l}\}_l, \{z_{il}\}_l, \theta) &= \int_{(\xi_l)} \int_{(\alpha_l)} \underline{\beta}_{i,m}(\{f_{x,m,l}\}_l, \{z_{il}\}_l, \theta, \{\xi_l\}_l, \{\alpha_l\}_l) dF_{(\alpha)} dF_{(\xi)},
\end{aligned}$$

where $\int_{(\xi_l)} dF_{(\xi)}$ denotes integration over $N \times L_m$ random variables $\{\xi_l\}_l$, and similarly, $\int_{(\alpha_l)} dF_{(\alpha)}$ denotes integration over L_m random variables $\{\alpha_l\}_l$. Again, we obtain parallel expressions for γ. Then, we obtain the following inequalities that must be satisfied at the true θ_0:

$$\begin{aligned}
E[\overline{\beta}_m(\{f_{x,m,l}\}_l, \{z_{il}\}, \theta_0) - \beta_m^{data} | \{z_i^{POS}\}] &\geq 0, \\
E[\underline{\beta}_m(\{f_{x,m,l}\}_l, \{z_{il}\}_l, \theta_0) - \beta_m^{data} | \{z_i^{POS}\}] &\leq 0,
\end{aligned}$$

where β_m^{data} is the coefficient obtained by regressing the actual vote-share data $\{v_{m,l}^{data}\}_l$ onto $\{f_{x,m,l}\}_l$. The expectation in the last expression is taken with respect to $\{\xi_l\}_l$,$\{\alpha_l\}_l$, $\{f_{x,m,l}\}_l$ as well as the number of municipalities (L^m), which we take to be a random variable. Note also that the expectation is conditional on the party affiliation, z_i^{POS}. For $\gamma_{i,m}$, we take the expectation conditional on z_{il}, i.e.,

$$\begin{aligned}
E[\bar{\gamma}_m(\{f_{x,m,l}\}_l, \{z_{il}\}_l, \theta_0) - \gamma_m^{data} | \{z_{il}\}] &\geq 0, \\
E[\underline{\gamma}_m(\{f_{x,m,l}\}_l, \{z_{il}\}_l, \theta_0) - \gamma_m^{data} | \{z_{il}\}] &\leq 0.
\end{aligned}$$

We take the sample analog of these moment inequalities, which are

$$\begin{aligned}
Q^+(\theta, z_i^{POS}) &= \left\| \sum_m \overline{\beta}_m(\{f_{x,m,l}\}_l, \{z_{il}\}_l, \theta) - \sum_m \beta_m(\{v_{m,l}^{data}\}_l, \{f_{x,m,l}\}_l, \{z_{il}\}_l,) \right\|_- \\
Q^-(\theta, z_i^{POS}) &= \left\| \sum_m \beta_m(\{v_{m,l}^{data}\}_l, \{f_{x,m,l}\}_l, \{z_{il}\}_l,) - \sum_m \overline{\beta}_m(\{f_{x,m,l}\}_l, \{z_{il}\}_l, \theta) \right\|_+ ,
\end{aligned}$$

where $\|a\|_+ = \max\{0, a\}$, and $\|a\|_- = \min\{0, a\}$. We obtain corresponding expressions for γ. We then apply Pakes, Porter, Ho, and Ishii (2007) for these moment

	Confidence Interval
θ^{const}	$(-1.494, -1.428)$
θ^{income}	$(-0.156, -0.152)$
$\theta^{education}$	$(0.200, 0.208)$
$\theta^{above65}$	$(-0.025, -0.021)$
θ^{JCP}	$(-3.494, -3.479)$
θ^{DPJ}	$(-3.031, -2.985)$
θ^{YUS}	$(-0.391, -0.052)$
$\theta^{previous}$	$(-0.211, -0.203)$
$\theta^{no\_experiecne}$	$(0.084, 0.090)$
$\theta^{hometown1}$	$(0.355, 0.434)$
$\theta^{hometown2}$	$(0.177, 0.232)$
$\theta^{hometown3}$	$(0.045, 0.052)$
$\theta_{\alpha 1}$	$(5.984, 6.109)$
$\theta_{\alpha 2}$	$(1.456, 1.482)$
θ_{ξ}	$(0.374, 0.393)$

Table 5: Confidence Intervals. Confidence intervals reported are asymptotically more conservative than 95%. These confidence intervals are calculated following Pakes, Porter, Ho, and Ishii (2007).

inequalities.

5 Results and Counterfactual Experiments

5.1 Parameter Estimates

The confidence intervals for the parameters are reported in Table 5. The exact specification of utility function we estimate is

$$
\begin{aligned}
& u(x_k, z_{il}; \theta^{PREF}) = \\
& -\left\{[\theta^{const}, \theta^{income}, \theta^{education}, \theta^{above65}, \theta^{below65}]x_k - [\theta^{LDP}, \theta^{JCP}, \theta^{DPJ}, \theta^{YUS}]z_i^{POS}\right\}^2 \\
& +[\theta^{incumbent}, \theta^{previous}, \theta^{no\_experience}, \theta^{hometown1}, \theta^{hometown2}, \theta^{hometown3}, \theta^{hometown4}]z_{il}^{QLTY} \\
& +\xi_{il} + \varepsilon_{ik},
\end{aligned}
$$

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

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 ideal point and the candidates' ideological positions. We have estimated the ideological positions of the candidates' parties as, θ^{JCP}=(-3.494,-3.479), θ^{DPJ}=(-3.031,-2.985), and θ^{YUS}=(-0.391,-0.052), 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 YUS are moderately conservative. Regarding voter positions, a voter with a lower income, 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 $\theta^{previous}$=(-0.211,-0.203), $\theta^{no\_experience}$= (0.084,0.090), and $\theta^{incumbent}$ normalized to 0. $\theta^{previous}$ 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 (-0.211,-0.203), which means that incumbents have an advantage over non-incumbent candidates with previous political experience. We have estimated $\theta^{no\_experience}$ to be (0.084,0.090), 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. Voters in general supported the reform, resulting in many newly elected Representatives by historical standards.

Hometown effects are estimated as $\theta^{hometown1}$=(0.355,0.434), $\theta^{hometown2}$=(0.177,0.232), and $\theta^{hometown3}$=(0.045,0.052) with normalization $\theta^{hometown4}$=0. The parameter $\theta^{hometown1}$ captures the effects 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 the candidate 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 is estimated to be between 0.682 and 0.827, that is, (68.2%, 82.7%) of voters are strategic voters on average. This may sound surprising given the fact that the fraction of strategic voting reported in the literature is between 3% and 17%. However, note that the fraction of "strategic voting" in the previous literature is in fact the fraction of misaligned voting, as discussed in the Introduction, and not the fraction of strategic voting, as defined in the theoretical literature. 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 did not vote for the most preferred candidate. Because the vote share is only observed at the municipality level, we still have the task of identifying the extent of misaligned voting from aggregate data; that is, from the difference in the actual vote share and the vote share we would obtain if all voters voted sincerely. This is not straightforward because the inflow and the outflow of strategic votes into candidate i may cancel each other out. In addition, even if the parameters of the model are known, computing what the outcome would have been if all voters voted sincerely is itself not a simple task. This is because the realization of municipality level shocks (ξ) cannot be uniquely recovered. We describe how to deal with these issues in Appendix B.

We obtained the upper and lower bounds of misaligned voting as $\underline{LB} = 1.85\%$ and $\overline{UB} = 5.01\%$, that is, about (1.85%, 5.01%) 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 mean fraction of strategic voters is about (68.2%, 82.7%) of the population, the fraction of strategic voters who did not vote for their most preferred candidate is (2.24%, 7.35%).

	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 (%)	(3.7, 7.8)	(17.5, 34.2)	(18.9, 32.8)	(31.3, 56.0)
Number of Seats	(6.5, 13.6)	(30.7, 60.0)	(33.0, 57.5)	(54.9, 97.9)

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

Table 6: 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%.

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.[27] We also allowed the voters to vote for any of the four parties regardless of whether a party actually fielded a candidate in the 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). Under porportional representation, there is little incentive to vote strategically and hence there is no misaligned voting. 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 that did not field a can-

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

didate in the voter's district. In our next counterfactual experiment, we will try to isolate and quantify the sincere-voting effect.

Table 6 compares the vote share and the number of seats each party obtains in the experiment with the actual data under the plurality rule. Firstly, the vote share for the DPJ and the LDP would be smaller under proportional representation. This is presumably because both the sincere-voting effect and the choice-expansion effect work against the DPJ and the LDP. In many elections, the strong candidates come from the two parties, and also the two parties fielded candidates in all of the districts in our sample. 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. Lastly, the vote share for the JCP remains the same or decreases. The sincere-voting effect probably increasesd the votes for the JCP, but the choice-expansion effect worked against the JCP due to the fact that the JCP fielded candidates in many distrcicts.

As for the number of seats in the counterfactural experiment, we simply multiplied the vote share by the number of total districts in our sample. The difference in the number of seats between the actual and the counterfactual is even greater than for the vote shares.

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 the 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 candidate is never chosen. However, this counterfactual experiment enables us to isolate the sincere-voting effect as we discussed in the previous subsection.

Table 7 compares the actual vote share and the number of seats won for the four parties with those of the sincere-voting experiment. The details on how we obtained these figures are provided in Appendix C. The vote shares for the DPJ and the LDP in this counterfactual are not very different from the vote shares in the actual election. This suggests that the change in the vote shares for these two parties in our first counterfactual is driven mostly by the choice-expansion effect. The vote share and the number of seats for the YUS is likely to increase with the counterfactual experiment. This implies that some of the increase in the vote share of the YUS in

	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 (%)	(5.8, 9.7)	(31.7, 44.6)	(41.6, 55.9)	(33.2, 60.7)
Number of Seats	(0, 0)	(10, 20)	(95, 149)	(12, 20)

Table 7: Counterfactual Experiment – Sincere Voting under Plurality Rule. Acutual Vote Share is computed by taking the average of the vote share over all of the 175 districts. Thus, they do not add up to 100%.

the previous counterfactual is due to the sincere-voting effect, but also that the bulk of the increase came through the choice-expansion effect.

6 Concluding Remarks

In this paper, we study how to identify and estimate a model of strategic voting and quantify the impact it has 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 variation generated by multiple equilibria, similar to Sweeting (2009). 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. The mean of the estimated proportion of strategic voter is (68.2%, 82.7%). A counterfactual experiment that introduces proportional representation decreases the number of votes to major-party candidates by a large margin, and the number of seats by an even greater margin. A second counterfactual experiment, which assumes sincere voting by all voters, also shows a significant increase in the vote share for candidates of a minor party.

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] Aradillas-Lopez, Andres, and Elie Tamer (2008) “The Identification Power of Equilibrium in Simple Games,” *Journal of Business and Economic Statistics*, Vol. 26, pp. 261-283.

[3] Berry, Steven T., James A. Levinsohn and Ariel Pakes (1995), “Automobile Prices in Market Equilibrium,” *Econometrica*, 60(4), 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] Buenrostro, Lucia, Amrita Dhillon, and Peter Vida (2007), “Scoring Rule Voting Games and Dominance Solvability” mimeo, University of Warwick

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

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

[8] Dhillon, Amrita and Benjamin Lockwood (2004) “When are Plurality Rule Voting Games Dominance Solvable?” *Games and Economic Behavior*, Vol. 46, pp. 55-75

[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] Marx, Leslie M. and Jeroen M. Swinkels (1997) "Order Independence for Iterated Weak Dominance," *Games and Economic Behavior,* Vol. 18, pp. 219–245.

[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] 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.

[24] 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.

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

[26] Sweeting, Andrew (2009),"The Strategic Timing of Radio Commercials: An Empirical Analysis Using Multiple Equilibria," forthcoming, *Rand Journal of Economics.*

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

7 Appendix

7.1 Appendix A

We construct $\beta_{i,m}$ as follows:

$\beta^1_{i,m}, \beta^2_{i,m}$: coefficients of regressing $v_{il}(\{T^m_{ij}\}, \theta, \xi_l, \alpha_l)$ onto a constant and fraction of population above 65 years old.

$\beta^3_{i,m}, \beta^4_{i,m}$: coefficients on a constant and fraction of population with years of schooling between 12 to 14 years.

$\beta^5_{i,m}, \beta^6_{i,m}$: coefficients on a constant and fraction of population with years of schooling between 15 to 16 years.

$\beta^7_{i,m}, \beta^8_{i,m}$: coefficients on a constant and fraction of population with years of schooling above 16 years.

$\beta^{9}_{i,m}, \beta^{10}_{i,m}$: coefficients on a constant and fraction of population with income in the first quartile (lower than 1.870 million yen).

$\beta^{11}_{i,m}, \beta^{12}_{i,m}$: coefficients on a constant and fraction of population with income in the second quartile (higher than 1.870 million yen and lower than 2.704 million yen)

$\beta^{13}_{i,m}, \beta^{14}_{i,m}$: coefficients on a constant and fraction of population with income in the third quartile (higher than 2.704 million yen and lower than 3.911 million yen).

$\beta^{15}_{i,m}$: coefficient on a constant only (equivalently, the mean of $v_{il}(\{T^m_{ij}\}, \theta, \xi_l, \alpha_l)$).

$\beta^{16}_{i,m}$: the variance of $v_{il}(\{T^m_{ij}\}, \theta, \xi_l, \alpha_l)$. This is not obtained by a regression, but we include it here for notational convenience. This term is inluded in our estimation in order to reflect our identification discussion in footnote 22.

We construct $\gamma_{i,m}$ as follows:

$\gamma^1_{i,m}$ through $\gamma^4_{i,m}$ are the conditional mean of $v_{il}(\{T^m_{ij}\}, \theta, \xi_l, \alpha_l)$, as described below.

$\gamma^1_{i,m}$: the mean of $v_{il}(\{T^m_{ij}\}, \theta, \xi_l, \alpha_l)$ when the hometown of the candidate is outside the prefecture.

$\gamma^2_{i,m}$: the mean of $v_{il}(\{T^m_{ij}\}, \theta, \xi_l, \alpha_l)$ when the hometown of the candidate is outside the district but in the same prefecture.

$\gamma^3_{i,m}$: the mean of $v_{il}(\{T^m_{ij}\}, \theta, \xi_l, \alpha_l)$ when the hometown of the candidate is in the same district but not from the same municipality

$\gamma^4_{i,m}$: the value of $v_{il}(\{T^m_{ij}\}, \theta, \xi_l, \alpha_l)$ when the hometown of the candidate is the same as the municipality.

As the vote share of candidates add up to one, we calculate $\beta^1_{i,m}, ..., \beta^{16}_{i,m}$only for $N-1$ candidates, while we calculate $\gamma^1_{i,m}, ..., \gamma^4_{i,m}$ for each of the N candidates.

7.2 Appendix B

The amount of misaligned voting is given by the fraction of voters who did not vote for the most preferred candidate. As we discussed in the main text, the vote share is only observed at the municipality level, so we need to identify the extent of misaligned voting from aggregate data. We discuss this issue first (Step 1), assuming that we can precisely recover the outcome when everyone votes sincerely, at the municipality level. Then, we will discuss issues that arise in recovering this outcome (Step 2 to Step 4).

Step 1

Let $v^{i,data}$ denote the actual vote share for candidate i and $v^{i,sin}$ denote the vote share of candidate i when everyone votes sincerely (subscripts m, l are suppressed from now on). Also, let D_{ij} denote the total votes cast for candidate i by strategic voters who prefers candidate j most. Then the object of interest, the amount of misaligned voting, is expressed as $\sum_{i,j} D_{ij}$. The available data can be summarized as $v^{i,data} - v^{i,sin} = \sum_j D_{ij} - \sum_j D_{ji}$, where $\sum_j D_{ij}$ is the inflow of strategic votes into candidate i and $\sum_j D_{ji}$ is the outflow of strategic votes from candidate i. (Note that if $D_{ij} > 0$, then $D_{ji} = 0$. See footnote 23.). The question we are concerned with is the following: What can we learn about $\sum_j D_{ji}$ given that we only know $v^{i,data} - v^{i,sin}(\equiv \Delta^i) = \sum_j D_{ij} - \sum_j D_{ji}$?

We can show that for $N = 3$, $\sum_j D_{ji}$ can be bounded below by

$$lb(\{\Delta^i\}) = \max_i \{|\Delta^i|\}$$

and above by

$$ub(\{\Delta^i\}) = \max_i \{\Delta^i\} - \min_i \{\Delta^i\}.$$

We provide an analogous expression for $N = 4$ 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.[28] The proofs are provided in the Supplementary material.

Step 2 to Step 4

Now we discuss issues related to recovering the sincere voting outcome. Given preference parameters of the model, for any realization of ξ, we can compute what the outcome would be if all voters vote sincerely. We denote this predicted sincere-voting outcome as $v^{sin}(\widehat{\theta}, \xi)$. Ideally, we would know the actual realization of $\xi = \xi_0$, and compute the sincere voting outcome $v^{sin}(\widehat{\theta}, \xi_0)$ using this actual realization of ξ_0. Then the difference between the observed vote share, $v^{data} = v(\xi_0; \alpha_0, \{T_{ij}\}, \theta_0)$ and $v^{sin}(\widehat{\theta}, \xi_0)$, $(\Delta^i = v^{data} - v^{sin}(\widehat{\theta}, \xi_0))$ would allow us to compute the lower and upper bounds, $lb(\{\Delta^i\})$ and $ub(\{\Delta^i\})$. However, the actual realization of ξ, $\xi = \xi_0$, cannot be recovered uniquely. Also, the difference between $v^{data} = v(\xi_0)$ and $v^{sin}(\widehat{\theta}, \xi)$

[28] We do not know whether the bounds are sharp with regard to the class of DGP 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 $\{T_{ij}\}$, it is difficult to obtain bounds that are, at the same time, computable and sharp with regard to the DGP we considered in the estimation.

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 $\widehat{\theta} \in \Theta_{CI}$. For any given draw of ξ from $\hat{F}_\xi$, we compute $\widehat{\Delta}^i(\xi)$,

$$\widehat{\Delta}^i(\xi) = v^i(\xi_0) - v^{i,sin}(\widehat{\theta}, \xi)$$

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

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

for each municipality, where $\hat{F}_\xi$ is the estimated distribution of ξ. Note that Lb_0 and Ub_0 do not necessarily coincide with $lb(\{\widehat{\Delta}^i(\xi_0)\})$ and $ub(\{\widehat{\Delta}^i(\xi_0)\})$, which are the lower and upper bounds of the extent of misaligned voting we would obtain if we had full knowledge of the realization of ξ, $\xi = \xi_0$. Therefore, we need to account for the parts of Lb_0 and Ub_0 that are induced by the randomness in ξ. 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 ξ. To do so, we compute the mean effects of the randomness components by considering

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

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

$$\widetilde{\Delta}^i(\widetilde{\xi}, \widetilde{\widetilde{\xi}}) = v^{i,sin}(\widehat{\theta}, \widetilde{\xi}) - v^{i,sin}(\widehat{\theta}, \widetilde{\widetilde{\xi}}).$$

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

$$LB = Lb_0 - Lb_\xi, \text{ and}$$
$$UB = Ub_0 - Ub_\xi.$$

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.

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

7.3 Appendix C

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 $\xi, \xi = \xi_0$. Denote the counterfactual vote share as $v_{m,l}^{sin}(\widehat{\theta}, \xi_0)$. The problem is that we cannot compute this because ξ_0 is unobserved. But we can obtain bounds for $v_{m,l}^{sin}(\widehat{\theta}, \xi_0)$ by following the same procedure as in Appendix B. We can also compute the number of seats in the same way. Note that we do not need to do Step 1.

7.4 Supplementary Material

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

7.4.1 Case of $N = 3$

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

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

and $\#\{\Delta^i > 0\}$ indicates the number of Δ^is that are positive, and $1\{\cdot\}$ is an indicator function.Let D_{ij} denote the votes cast for candidate i by strategic voters who prefers candidate j most. Then the amount of misaligned voting is $\sum_{ij} D_{ij}$ (Note that if $D_{ij} > 0$, then $D_{ji} = 0$. See footnote 23.).

First, we prove that the extent of strategic voting is bound by $lb(\{\Delta^i\})$ and $ub(\{\Delta^i\})$. 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

$$\Delta^1 = D_{12} + D_{13}, \tag{A1}$$

$$\Delta^2 = D_{23} - D_{12}, \tag{A2}$$

$$\Delta^3 = -D_{13} - D_{23}. \tag{A3}$$

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) $\{\#\{\Delta^i > 0\} = 1\}$, and ii) $\{\#\{\Delta^i > 0\} = 2\}$. 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),

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

In case (ii),

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

We can also see that $\max_i\{|\Delta^i|\}$ 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 $ub(\{\Delta^i\})$ is sharp. Let $h(\Delta^1, \Delta^2, \Delta^3) \leq ub(\{\Delta^i\})$ for all $\{\Delta^i_{m,l}\}$, and moreover $h(\Delta^{1*}, \Delta^{2*}, \Delta^{3*}) < ub(\{\Delta^i\})$. 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 max operator. 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, $D_{12}+D_{13}+D_{23} =\Delta^{1*}-\min\{\Delta^{2*}, \Delta^{3*}\} =ub(\{\Delta^{i*}\})$, 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, $D_{12} + D_{13} + D_{23} =\max\{\Delta^{1*}, \Delta^{2*}\} - \Delta^{3*} =ub(\{\Delta^{i*}\})$.

Third, we prove by contradiction that the lower bound $lb(\{\Delta^i\})$ is sharp. Let $h(\Delta^1, \Delta^2, \Delta^3) \geq lb(\{\Delta^i\})$ for all $\{\Delta^i_{m,l}\}$, and moreover $h(\Delta^{1*}, \Delta^{2*}, \Delta^{3*}) > lb(\{\Delta^i\})$. 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, $D_{12} + D_{13} + D_{23} = -\Delta^{2*} - \Delta^{3*} = \Delta^{1*} = lb(\{\Delta^{i*}\})$. 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 $D_{12} + D_{13} + D_{23} = -\Delta^{3*} = lb(\{\Delta^{i*}\})$. Thus, h cannot be a lower bound.

7.4.2 Case of $N = 4$

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

$$\begin{aligned} lb(\{\Delta^i\}) = 1\{\#\{\Delta^i > 0\} = 3\} \max \left\{ \min_{i,j \neq i}\{\Delta^i + \Delta^j | \Delta^i, \Delta^j > 0\}, -\min_i\{\Delta^i | \Delta^i < 0\} \right\} \\ + 1\{\#\{\Delta^i > 0\} = 2\} \max \left\{ \min_i\{\Delta^i | \Delta^i > 0\}, -\min_i\{\Delta^i | \Delta^i < 0\} \right\} \\ + 1\{\#\{\Delta^i > 0\} = 1\} \max \left\{ \max_i\{\Delta^i\}, -\max_i\{\Delta^i | \Delta^i < 0\} \right\}, \text{ and} \\ ub(\{\Delta^i\}) = \max_{i,j \neq i}\{2\Delta^i + \Delta^j\} - \max_i\{\Delta^i | \Delta^i < 0\} \end{aligned}$$

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