RISE Working Paper 15-007

"Identification and Estimation of Preference Distributions When Voters Are Ideological"

by

Antonio Merlo and Áureo de Paula

RICE Department of Economics
Baker Hall, MS22
6100 Main Street, Houston, Texas 77005
https://economics.rice.edu

Identification and Estimation of Preference Distributions When Voters Are Ideological *

Antonio Merlo †

Rice University

Áureo de Paula ‡

UCL, São Paulo School of Economics-FGV, CeMMAP and IFS

This version: September, 2015

Abstract

This paper studies the nonparametric identification and estimation of voters' preferences when voters are ideological. We establish that voter preference distributions and other parameters of interest can be identified from aggregate electoral data. We also show that these objects can be consistently estimated and illustrate our analysis by performing an actual estimation using data from the 1999 European Parliament elections.
JEL: D72, C14; Keywords: Voting, Voronoi tessellation, identification, nonparametric.

*We would like to thank Stéphane Bonhomme, three anonymous referees, Micael Castanheira, Andrew Chesher, Eric Gautier, Ken Hendricks, Stefan Hoderlein, Bo Honoré, Frank Kleibergen, Dennis Kristensen, Ariel Pakes, Jim Powell, Bernard Salanié, Kevin Song and Dale Stahl for helpful suggestions. The paper also benefited from comments by seminar and conference participants at several institutions. Chen Han, Chamna Yoon and Nicolas Motz provided very able research assistance. de Paula gratefully acknowledges financial support from the European Research Council through Starting Grant 338187 and the Economic and Social Research Council through the ESRC Centre for Microdata Methods and Practice grant RES-589-28-0001.

†Department of Economics, Rice University, Houston, TX 77251. E-mail: `amerlo@rice.edu`

‡University College London, London, UK, São Paulo School of Economics-FGV, São Paulo, Brazil, C*e*MMAP, London, UK, and IFS, London, UK. E-mail: `apaula@ucl.ac.uk`

1 Introduction

Elections are the cornerstone of democracy and voters' decisions are essential inputs in the political process shaping the policies adopted by democratic societies. Understanding observed voting patterns and how they relate to voters' preferences is a crucial step in our understanding of democratic institutions and is of great relevance, both theoretically and practically. These considerations raise the following fundamental question: Is it possible to nonparametrically identify and estimate voters' preferences from aggregate data on electoral outcomes?

To address this question, one must first specify a theoretical framework that links voters' decisions to their preferences. The *spatial theory of voting*, formulated originally by Downs (1957) and Black (1958), building on Hotelling (1929)'s seminal work in industrial organization, and later extended by Davis, Hinich, and Ordeshook (1970), Enelow and Hinich (1984) and Hinich and Munger (1994), among others, is a staple of political economy.[1] This theory postulates that each individual has a most preferred policy or "bliss point" and evaluates alternative policies or candidates in an election according to how "close" they are to her ideal. More precisely, consider a situation where a group of voters is facing a contested election with any number of candidates. Suppose that each voter has preferences (i.e., their bliss point) that can be represented by a position in some common, multi-dimensional ideological (metric) space, and each candidate can also be represented by a position in the same ideological space. According to the spatial framework, each voter will cast her vote in favor of the candidate whose position is closest to her bliss point (given the positions of all the candidates in the election).[2] In this case, we say that voters *vote ideologically*.[3]

In this paper, we study the issue of nonparametric identification and estimation of voters' preferences using aggregate data on elections with arbitrary number of candidates,

[1]See, e.g., Hinich and Munger (1997).

[2]Data sets containing measures of the ideological positions of politicians based on their observed behavior in office are widely available (see, e.g., Poole and Rosenthal (1997) and Heckman and Snyder (1997) for the United States Congress or Hix, Noury, and Roland (2006) for the European Parliament).

[3]For a survey of alternative theories of voting, see, e.g., Merlo (2006).

under the maintained assumption that voters vote ideologically. Following Degan and Merlo (2009), we represent multi-candidate elections as Voronoi tessellations of the ideological space.[4] Using this geometric structure, we establish that voter preference distributions and other parameters of interest can be retrieved from aggregate electoral data. We also show that these objects can be estimated using the methodology proposed by Ai and Chen (2003), and perform an actual estimation using data from the 1999 European Parliament elections.

Since our analysis focuses on retrieving individual level fundamentals from aggregate data, it is related to the ecological inference problem.[5] It is also related to the vast literature on identification and estimation of discrete choice models.[6] In particular, our paper is most closely related to the industrial organization literature on discrete choice models with random coefficients and macro-level data (e.g., Berry, Levinsohn, and Pakes (1995) and, more recently, Berry and Haile (2009)), and pure characteristics models (see Berry and Pakes (2007) and references therein).[7]

In the language of the pure characteristics model, in our environment, the "consumer" (i.e., the voter) obtains utility $U^{\mathbf{t}}(C_i) = -(C_i - \mathbf{t})^\top W(C_i - \mathbf{t})$ from "product" (i.e., candidate) i, where $\mathbf{t}$ is a vector of individual "tastes" (i.e., the voter's bliss point), C_i is a vector of "product characteristics" (i.e., the candidate's position) and W is a matrix of weights. Also, the distribution of tastes depends on "market" (i.e., electoral precinct) level covariates, both observed and unobserved.[8] Whereas the distribution of tastes is typically taken to be parametric in pure characteristics models, we show that it can be *nonparametrically*

[4]Degan and Merlo (2009) characterize the conditions under which the hypothesis that voters vote ideologically is falsifiable using individual-level survey data on how the same individuals vote in multiple simultaneous elections (Henry and Mourifié (2013) extend their analysis and develop a formal test of the hypothesis). In this paper, we restrict attention to identification and inference based on aggregate data on electoral outcomes in environments where the hypothesis is non-falsifiable.

[5]Ecological inference refers to the use of aggregate data to draw conclusions about individual-level relationships when individual data are not available. See, e.g., King (1997) for a survey.

[6]Starting with McFadden (1974)'s seminal work, other important papers investigating the identification of discrete choice models include Manski (1988) and Matzkin (1992). See also Chesher and Silva (2002).

[7]Our work is also related to the spatial approach to individual discrete choice as a foundation for aggregate demand pioneered by Hotelling (1929). Spatial demand models are closely related to random coefficient models as pointed out, for example, by Caplin and Nalebuff (1991), who provide a unified synthesis of random coefficients, characteristics and spatial models.

[8]Clearly, the analogy is only partial since in the environment we consider there are no prices.

identified and estimated together with the finite dimensional components of the model (W). Our identification strategy relies on the geometric structure induced by the functional form of the utility function implied by the spatial theory of voting.

Part of the identification strategy we develop in this paper is related to previous work by Ichimura and Thompson (1998) and Gautier and Kitamura (2013) on binary choice models with random coefficients. In fact, in the special case where W is known and elections only have two candidates, the spatial model of voting is equivalent to a binary choice model with random coefficients. However, in the general setting where W is not known and elections have arbitrary numbers of candidates—the environment considered here—the identification strategy in Ichimura and Thompson (1998) and Gautier and Kitamura (2013) does not apply.

The remainder of the paper is organized as follows. In Section 2, we describe the model and in Section 3 discuss its identification. Nonparametric estimation is presented in Section 4. In Section 5, we illustrate our approach with an empirical application. Concluding remarks are presented in Section 6. All proofs are contained in the Appendix.

2 The model

Consider a situation where a population of voters has to elect representatives to public office (e.g., a legislature). Consistent with the spatial theory of voting, there is a common ideological space, Y, which is taken to be the k-dimensional Euclidean space (i.e., $Y = \mathbb{R}^k$ and the reference measurable space is this set equipped with the Borel sigma algebra: $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$). We observe a cross-section of elections $e \in \{1, ..., E\}$. An election e is a contest among $n_e \geq 2$ candidates. The number of candidates n_e may vary across elections, and we allow for this possibility in estimation. However, to simplify exposition, we refer to the number of candidates in a generic election by n, unless it is not clear from the context. Let $\mathcal{C} \equiv (C_1, \ldots, C_n) \in \mathbb{R}^{nk}$ denote a profile of candidates represented by the nk-dimensional vector concatenating all the candidate positions characterizing an election. Each candidate

$i \in \{1, ..., n\}$ is characterized by a distinct position in the ideological space, $C_i \in Y$, which is known to the voters and observed by the econometrician.

Each voter has an ideological position (or bliss point) $\mathbf{t}$, and her preferences are characterized by indifference sets that are ellipsoids in the k-dimensional Euclidean space, centered around her bliss point.[9] It follows that voter $\mathbf{t}$'s preferences over candidates in an election can be summarized by the utility function

$$U^{\mathbf{t}}(C_i) = u^{\mathbf{t}}\left(d^W(\mathbf{t}, C_i)\right), \tag{1}$$

where $u^{\mathbf{t}}(\cdot)$ is a decreasing function which may differ across voters and $d^W(\cdot,\cdot) \geq 0$ denotes the Euclidean distance with (positive definite, symmetric) weighting matrix W (i.e., for any two points $x, y \in \mathbb{R}^k$, $d^W(x,y) = \sqrt{(x-y)^\top W(x-y)}$). Other than monotonicity, we impose no additional restrictions on the $u^{\mathbf{t}}(\cdot)$ functions, which are therefore left unspecified. Given these preferences, a voter $\mathbf{t}$ (strictly) prefers candidate i to candidate j in an election if $d^W(\mathbf{t}, C_i) < d^W(\mathbf{t}, C_j)$. According to the spatial theory of voting (see, e.g., Hinich and Munger (1997)), the main diagonal elements in the matrix W subsume the relative importance to voters of the different dimensions of the ideological space. The off-diagonal elements, on the other hand, describe the way in which voters make trade-offs among these different dimensions.

As in Degan and Merlo (2009), for each position $C_i \in Y$ of a generic candidate i in an election, let $V_i^W(\mathcal{C}) \equiv \{\mathbf{t} \in Y : d^W(\mathbf{t}, C_i) \leq d^W(\mathbf{t}, C_j), j \neq i\}$ be the set of points in the ideological space Y that are closer to C_i than to the position of any other candidate in the election. Since $d^W(\cdot,\cdot)$ is the weighted Euclidean distance, it follows that for each pair of candidates in an election, C_i, C_j, the set of points in the ideological space Y that are equidistant from C_i and C_j is a hyperplane $H^W(C_i, C_j)$, which tessellates the ideological space Y into two regions (or half spaces), $Y_{C_i}^{C_j}$ and $Y_{C_j}^{C_i}$, where $Y_{C_j}^{C_i}$ is the set of ideological

[9] In one dimension, the restriction implies that each voter's utility function is single-peaked and symmetric.

positions that are closer to the position of candidate i than to the position of candidate j and vice versa for the set $Y_{C_i}^{C_j}$. Hence, for each candidate i, $V_i^W(\mathcal{C})$ is the intersection of the half spaces determined by the $n-1$ hyperplanes $(H^W(C_i, C_j))_{j\neq i}$ (i.e., $V_i^W(\mathcal{C}) = \cap_{j\neq i} Y_{C_j}^{C_i}$).

Note that, for all candidates $i \in \{1, ..., n\}$, $V_i^W(\mathcal{C})$ is non empty and convex. Hence, an election implies a tessellation of the ideological space Y into n convex regions, $\{V_i^W(\mathcal{C})\}_{i\in\{1,\ldots,n\}}$, where each region $V_i^W(\mathcal{C})$ is the set of voters voting for candidate i in the election.[10] The set $V^W(\mathcal{C}) \equiv \{V_i^W(\mathcal{C})\}_{i\in\{1,\ldots,n\}}$ defines what in computational and combinatorial geometry is called a *Voronoi tessellation* of $\mathbb{R}^k$ and each region $V_i^W(\mathcal{C})$ is a k-dimensional *Voronoi polyhedron* (or *Voronoi cell*).[11] Because the Voronoi cells $\{V_i^W(\mathcal{C})\}_{i\in\{1,\ldots,n\}}$ are the same for all weighting matrices αW with $\alpha > 0$, we impose the normalization that $||W||_{k\times k} = \sqrt{k}$, where $||W||_{k\times k} = \sqrt{Tr\,(W^\top W)}$ is the Frobenius norm. This, in particular, includes the k-order identity matrix as a possible weighting matrix W.

Figure 1 illustrates an example of the Voronoi tessellation that corresponds to an election with five candidates, $\{a, b, c, d, e\}$, with positions $\{C_a, C_b, C_c, C_d, C_e\}$ in the two-dimensional ideological space $Y = \mathbb{R}^2$ and weighting matrix equal to the identity matrix.

Voters are characterized by the random vector $\mathbf{T}$, representing their preference types in the ideological space $Y \subset \mathbb{R}^k$. The distribution of preference types (or bliss points) $\mathbf{T}$ in the population of voters is given by the conditional probability distribution $\mathbb{P}_{T|X,\epsilon}$, which is assumed to be absolutely continuous with respect to the Lebesgue measure on $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ given $\mathbf{X}$ and ϵ and the weighting matrix W. Here, $\mathbf{X}$ represents observable characteristics at the electoral precinct level, such as average demographic and economic features, and ϵ stands for unobservable electoral precinct characteristics. For example, in our empirical application, the French constituency of Paris is one such electoral precinct, for which we have data on observable characteristics such as age, gender, employment status and per-capita GDP of the precinct population at the time of the election. Together with the weighting matrix W, the

[10]Note that $V_i^W(\mathcal{C}) \cap V_j^W(\mathcal{C}) \subset H^W(C_i, C_j)$ for all $i \neq j$, and $\cup_{i\in\{1,\ldots,n\}} V_i^W(\mathcal{C}) = Y$.

[11]For a comprehensive treatment of Voronoi tessellations and their properties, see, e.g., Okabe, Boots, Sugihara, and Chiu (2000).

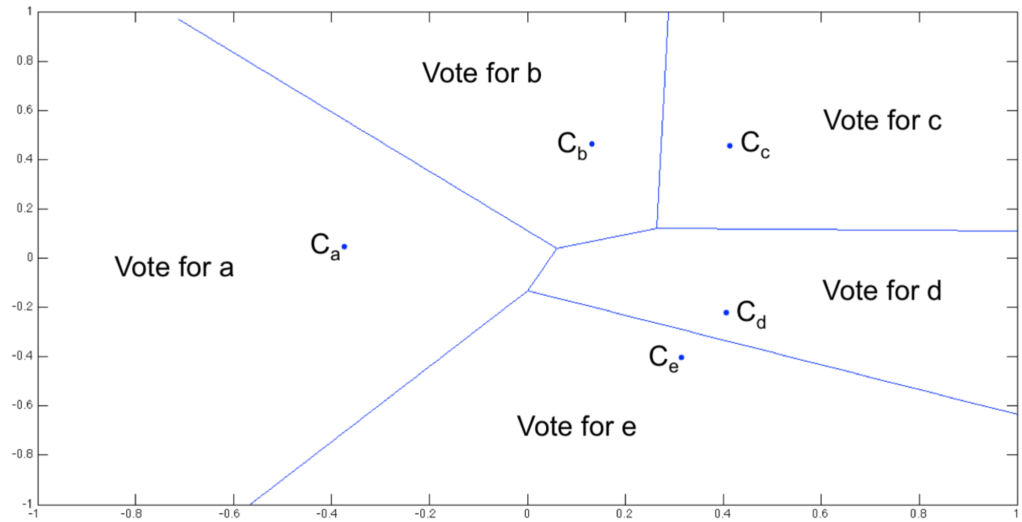

Figure 1: The Voronoi Tessellation for a 5-candidate election in $\mathbb{R}^2$ and $W = I$.

main object of interest is $\mathbb{P}_{T|X} \equiv \int \mathbb{P}_{T|X,\epsilon}\mathbb{P}_{\epsilon|X}(d\epsilon|X)$, the conditional probability distribution of preference types $\mathbf{T}$ in the population of voters given $\mathbf{X}$ only.

Conditional on $\mathbf{X}$, candidates are drawn from a distribution characterized by the measure $\mathbb{P}_{C|X}$, again absolutely continuous with respect to the Lebesgue measure on $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$. The proportion of votes obtained by each candidate is the probability of the Voronoi cell that contains the candidate's ideological position. For notational convenience, we omit the conditioning variable for most of this and the next section and refer to the distribution of voter locations simply as $\mathbb{P}_T$ and to the distribution of candidates as $\mathbb{P}_C$. Since the identification arguments can be repeated for strata defined by regressors, this is without loss of generality.

For each election, the observed data contain the number of candidates, the ideological position of each candidate and the electoral results (i.e., the proportion of votes obtained by each candidate). For any given profile of candidates $\mathcal{C}$, preference type distribution $\mathbb{P}_T$ and weighting matrix W, we can define the following object:

$$(\mathcal{C}, (\mathbb{P}_T, W)) \mapsto p(\mathcal{C}, (\mathbb{P}_T, W))$$

where $p(\mathcal{C}, (\mathbb{P}_T, W))$ takes values on the $n-$dimensional simplex and denotes the vector of the proportions of votes obtained by all the candidates in the profile $\mathcal{C}$ according to the preference type distribution $\mathbb{P}_T$ and weighting matrix W. The expected proportion of votes obtained by candidate i in an election with n candidates $\mathcal{C} = \{C_1, \ldots, C_n\}$ and Voronoi cell $V_i^W(\mathcal{C}) = \{\mathbf{t} \in \mathbb{R}^k : d^W(\mathbf{t}, C_i) \leq d^W(\mathbf{t}, C_j), j \neq i\}$ is given by:

$$\begin{aligned}
\mathbb{E}(\mathbf{1}_{\mathbf{t}\in V_i^W(\mathcal{C})}|\mathbf{X}, \mathcal{C}) &= \int \mathbf{1}_{\mathbf{t}\in V_i^W(\mathcal{C})} \mathbb{P}_{T|X,\mathcal{C},\epsilon}(d\mathbf{t}|\mathbf{X}, \mathcal{C}, \epsilon)\mathbb{P}_{\epsilon|X,\mathcal{C}}(d\epsilon|\mathbf{X}, \mathcal{C}) \\
&= \int \mathbf{1}_{\mathbf{t}\in V_i^W(\mathcal{C})} f_{T|X,\mathcal{C},\epsilon}(\mathbf{t}|\mathbf{X}, \mathcal{C}, \epsilon)\mathbb{P}_{\epsilon|X,\mathcal{C}}(d\epsilon|\mathbf{X}, \mathcal{C})d\mathbf{t} \\
&= \int \mathbf{1}_{\mathbf{t}\in V_i^W(\mathcal{C})} f_{T|X,\mathcal{C}}(\mathbf{t}|\mathbf{X}, \mathcal{C})d\mathbf{t}
\end{aligned}$$

where $f_{T|X,\mathcal{C},\epsilon}$ is the density of $\mathbb{P}_{T|X,\mathcal{C},\epsilon}$ and analogously for $f_{T|X,\mathcal{C}}$. It is important for identification that we require that $\mathbf{T}$ and $\mathcal{C}$ be conditionally independent (given $\mathbf{X}$) such that: $f_{T|X,\mathcal{C}} = f_{T|X}$. In this case,

$$\mathbb{E}(\mathbf{1}_{\mathbf{t}\in V_i^W(\mathcal{C})}|\mathbf{X}, \mathcal{C}) = \int \mathbf{1}_{\mathbf{t}\in V_i^W(\mathcal{C})} f_{T|X}(\mathbf{t}|\mathbf{X})d\mathbf{t}.$$

Notice that $\mathbf{T}$ and $\mathcal{C}$ are not (unconditionally) independent, but we assume that, upon conditioning on the demographic covariates $\mathbf{X}$, $\mathcal{C}$ carries no further information about the distribution of $\mathbf{T}$ (i.e., the random vectors $\mathcal{C}$ and $\mathbf{T}$ are conditionally independent given $\mathbf{X}$). This assumption is reasonable insofar as $\mathbf{X}$ lists all the guiding variables for the determination of a candidate's position and accommodates some partial strategic behavior.[12] It is similar to the independence assumption between regressors and coefficients typically required in the literature on discrete choice models with random coefficients (e.g., Ichimura and Thompson (1998) and Gautier and Kitamura (2013)). In our case, the variables $\mathcal{C}$ allow us to identify the

[12] As it is common in the political economy literature on the spatial model of voting, we treat the distribution of candidate positions as given. The assumption that, upon conditioning on the vector of observable characteristics $\mathbf{X}$, this distribution does not convey additional information on the distribution of voters' preferences is consistent, for example, with the "partisan" model of Hibbs (1977) and Alesina (1988). A full characterization of the distribution of candidates' positions as an equilibrium object in a general environment with more than two candidates and a multidimensional space is not feasible given the current status of the theoretical literature (e.g., Merlo (2006)). It is therefore outside of the scope of our analysis.

structure. Independent variation in characteristics is also used to identify the distributions of interest in Ichimura and Thompson (1998) and Gautier and Kitamura (2013). We also note that, except for prices, product characteristics are usually assumed to be exogenous in the differentiated products demand literature (e.g., Anderson, De Palma, and Thisse (1989) or Feenstra and Levinsohn (1995)). The assumption is made explicit below:

Assumption 1. *The random vectors* $\mathcal{C}$ *and* $\mathbf{T}$ *are conditionally independent given* $\mathbf{X}$.

Since we use variation in candidate locations to identify the distribution of voters' preference types $\mathbf{T}$, we also assume that there is sufficient variation in those variables.[13] We collect this and other assumptions on the underlying distributions in the following assumption:

Assumption 2. *The distribution of candidate profiles* $\mathcal{C}$ *is absolutely continuous with respect to the Lebesgue measure on* $(\mathbb{R}^{nk}, \mathcal{B}(\mathbb{R}^{nk}))$ *and the distribution of preference types* $\mathbf{T}$ *is absolutely continuous with respect to the Lesbegue measure on* $(\mathbb{R}^{k}, \mathcal{B}(\mathbb{R}^{k}))$.

3 Identification

The following definition qualifies our characterization of identifiability. We remind the reader that the analysis is conditional on $\mathbf{X}$ and notation is omitted for simplicity.

Definition 1 (Identification)**.** *Let* $(\mathbb{P}_{T_1}, W_1)$ *and* $(\mathbb{P}_{T_2}, W_2)$ *be two pairs where* $\mathbb{P}_i, i = 1, 2,$ *are probability measures on* $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$*, both absolutely continuous with respect to the Lebesgue measure on* $\mathbb{R}^k$ *and* $W_i, i = 1, 2,$ *are positive definite, symmetric weighting matrices.* $(\mathbb{P}_{T_1}, W_1)$ *is identified relative to* $(\mathbb{P}_{T_2}, W_2)$ *if and only if* $p(\cdot, (\mathbb{P}_{T_1}, W_1)) = p(\cdot, (\mathbb{P}_{T_2}, W_2))$*, Leb-a.e.* $\Rightarrow$ $(\mathbb{P}_{T_1}, W_1) = (\mathbb{P}_{T_2}, W_2)$.[14] $(\mathbb{P}_T, W)$ *is (globally) identified if it is identified relative to any other probability measure and weighting matrix pair.*

[13] For example, if the profile of candidate positions is the same in every election, it would not be possible to identify the distribution of preferences in the population of voters.

[14] *Leb.-a.e.* refers to the fact that the underlying measure is the Lebesgue measure on $(\mathbb{R}^{nk}, \mathcal{B}(\mathbb{R}^{nk}))$.

In words, two preference structures that for every possible configuration of candidates in an election (except for cases in a zero measure set) generate the same proportions of votes should correspond to the same (probability measure and weighting matrix) pair.

We begin our analysis by considering the case where the weighting matrix W is known. While our main goal is to show that the distribution of bliss points and the weighting matrix that characterize voters' preferences are jointly identified, the analysis of the simpler case allows us to clarify the relationship with the work by Ichimura and Thompson (1998) and Gautier and Kitamura (2013). Moreover, it provides a useful step for the proof of the main result of the paper contained in Theorem 1 below.

Lemma 1 establishes identification of the distribution of preference types in the population of voters when the weighting matrix is known.

Lemma 1. *Suppose that W is known and Assumptions 1 and 2 hold. Then $\mathbb{P}_T$ is identified.*

The proof of Lemma 1 is given in the Appendix for elections with any number of candidates and it is a straightforward extension of the argument in Ichimura and Thompson (1998), which can be directly applied to our setting to establish identification for the two-candidate case. The argument in the proof generalizes the simple insight that for two-candidate elections the Voronoi tessellation is given by an affine hyperplane. One can then sweep the space looking for an affine hyperplane that delivers different election outcomes for two distinct preference type distributions. That such an affine hyperplane exists is guaranteed by the Cramér-Wold device.[15] Consequently, even if candidate and voter types do not share the same support, the argument would deliver relative identification of voter type distributions that differ on the intersection of the supports.[16] In fact, even for the case where

[15]The Cramér-Wold device refers to the result that the distribution of a random vector is uniquely characterized by the family of distributions of all its linear combinations. This is related to the fact that the characteristic function for a multivariate distribution is also the characteristic function for the distribution of a linear combination of the random vector of interest (see Pollard (2002), p.202). Hence, one can also employ Fourier methods directly to obtain identification.

[16]As we require the probability measures to be absolutely continuous with respect to the Lebesgue measure, we rule out discrete voter type or candidate distributions. Since the Cramér-Wold device does not require absolute continuity, we conjecture that in principle the result could also be extended to discrete types. Because in our application the relevant variables are continuous we did not pursue this extension further.

there are more than two candidates, as long as one can sample elections where candidates are arbitrarily clustered around two positions (which is guaranteed by our assumptions), identification follows by continuity.

As pointed out above, the Cramér-Wold device is also used in Ichimura and Thompson (1998) to show identification of the unknown distribution for the random coefficients in a *binary* outcome model. When $n = 2$ and $\mathcal{C} = (C_1, C_2)$, the spatial model of voting postulates that a voter at $\mathbf{t}$ chooses C_2 when $d^W(\mathbf{t}, C_1) - d^W(\mathbf{t}, C_2) \geq 0$. This can be written as $Z(W)^\top \beta \geq 0$ where

$$Z(W) \equiv \frac{(-2(C_1 - C_2)^\top W, (C_1 - C_2)^\top W(C_1 + C_2))^\top}{||(-2(C_1 - C_2)^\top W, (C_1 - C_2)^\top W(C_1 + C_2))||} \in \mathbb{R}^{k+1} \text{ and } \beta \equiv \frac{(\mathbf{t}^\top, 1)^\top}{||(\mathbf{t}^\top, 1)||} \in \mathbb{R}^{k+1}.$$

Hence, when elections only have two candidates, the spatial model of voting reduces to a binary choice model with random coefficients as in Ichimura and Thompson (1998) and Gautier and Kitamura (2013). If W (and consequently Z) is known, one can then use their arguments to identify the distribution of β which can then be used to obtain the distribution of preference types $\mathbf{T}$.

We now turn attention to the general environment where the weighting matrix W is not known. We initially consider the special case where there are only two candidates and the weighting matrix is unknown. As in the case where the weighting matrix is known, the result can then be extended to elections with a general number of candidates by continuity arguments. We elaborate on this point in more detail later in this section.

Lemma 2 establishes joint identification of the distribution of preference types in the population of voters and of the weighting matrix, when elections have two candidates.

Lemma 2. *Suppose Assumptions 1 and 2 hold, $||W||_{k\times k} = \sqrt{k}$, and there are two candidates. Then $(\mathbb{P}_T, W)$ is identified.*

The proof of Lemma 2 is presented in the Appendix for arbitrary ideological space dimension k. Here, we provide the intuition for the identification in the case where the ideological space is two-dimensional. The result is established by showing that if there are two tuples, $(\mathbb{P}_{\overline{T}}, \overline{W})$ and $(\mathbb{P}_T, W)$, that are observationally equivalent, they would have to place zero probability on any arbitrary set in the ideological space.

Consider then two environments $(\mathbb{P}_{\overline{T}}, \overline{W})$ and $(\mathbb{P}_T, W)$ such that $\overline{W} \neq W$ and assume they are observationally equivalent. Start with an arbitrary bounded set in $\mathbb{R}^2$, as indicated by the square in the upper-left panel in Figure 2. Then, consider an election with candidates $\mathcal{C} = \{C_1, C_2\}$ such that this set is contained in $V_i^{\overline{W}}(\mathcal{C})$, but not in $V_i^W(\mathcal{C})$ for some i. In the upper-right panel in Figure 2, this is achieved for candidate C_1. Under the weighted distance d^W, the Voronoi cells when there are two candidates are separated by the line

$$H^W(C_1, C_2) \equiv \{\mathbf{t} \in \mathbb{R}^2 : \underbrace{C_1^\top W C_1 - C_2^\top W C_2 + 2(C_2 - C_1)^\top W \mathbf{t}}_{\equiv d^W(\mathbf{t}, C_1)^2 - d^W(\mathbf{t}, C_2)^2} = 0\}, \tag{2}$$

and analogously for the weighted distance $d^{\overline{W}}$. Hence, the area above $H^W(C_1, C_2)$ corresponds to $V_1^W(\mathcal{C})$ and the area below corresponds to $V_2^W(\mathcal{C})$. Similarly, the area above $H^{\overline{W}}(C_1, C_2)$ corresponds to $V_1^{\overline{W}}(\mathcal{C})$ and the area below corresponds to $V_2^{\overline{W}}(\mathcal{C})$. Note that the highlighted square is contained in $V_1^{\overline{W}}(\mathcal{C})$, but not in $V_1^W(\mathcal{C})$.

Note also that the two lines $H^W(C_1, C_2)$ and $H^{\overline{W}}(C_1, C_2)$ intersect at the midpoint $(C_1 + C_2)/2$. If the two tuples $(\mathbb{P}_T, W)$ and $(\mathbb{P}_{\overline{T}}, \overline{W})$ are observationally equivalent, the two candidates C_1 and C_2 should obtain the same shares of votes under $(\mathbb{P}_T, W)$ as they would under $(\mathbb{P}_{\overline{T}}, \overline{W})$. Denote by p the vote share of candidate C_2. As indicated in the two lower panels in Figure 2, this is the probability of the area *below* $H^W(C_1, C_2)$ under $\mathbb{P}_W$ and the area *below* $H^{\overline{W}}(C_1, C_2)$ under $\mathbb{P}_{\overline{W}}$.

One can then obtain a translation of the candidates, say (C_1', C_2'), such that $C_1 - C_2 = C_1' - C_2'$, and the same original Voronoi diagram is generated under W, as illustrated in the upper-left panel in Figure 3. The line characterizing the $\overline{W}$-Voronoi cells for the new pair (C_1', C_2') is parallel to the $\overline{W}$-Voronoi line for (C_1, C_2). From the upper-right panel in Figure 3 note that the original square is contained in $V_1^{\overline{W}}(\mathcal{C}')$, but not in $V_1^W(\mathcal{C}')$.

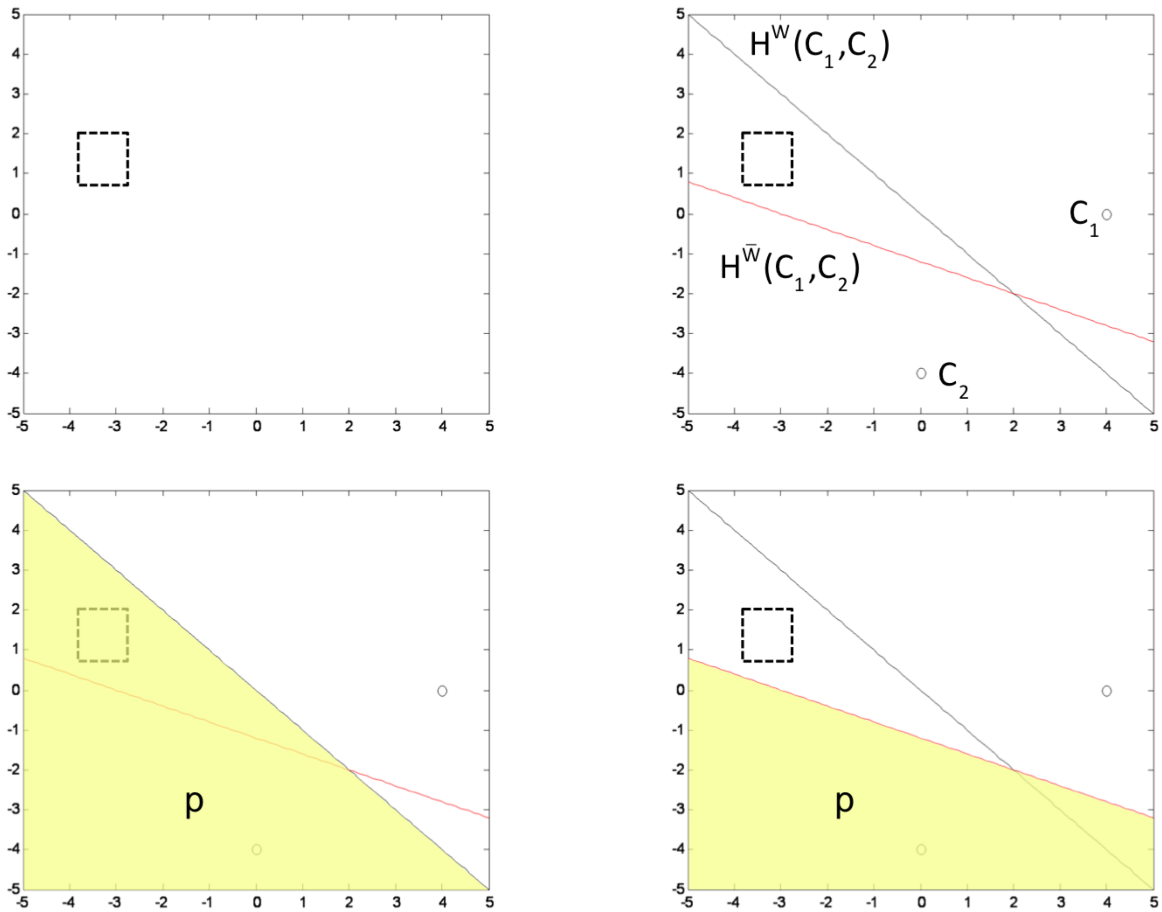

Figure 2: Voronoi Tessellations for Candidates C_1, C_2

Candidates C_1' and C_1 obtain the same vote share, equal to $(1-p)$, in their respective elections under $(\mathbb{P}_T, W)$, since they generate the same Voronoi cells (under W). In particular, this is the probability of the area *above* $H^W(C_1', C_2')$ (which is the same as $H^W(C_1, C_2)$), as indicated in the upper-right panel in Figure 3. Under observational equivalence, the share of candidate C_1' should also be $1-p$ under $(\mathbb{P}_{\overline{T}}, \overline{W})$. This means that the area *above* $H^{\overline{W}}(C_1', C_2')$ equals $1-p$ under $\mathbb{P}_{\overline{T}}$ (see the lower-left panel in Figure 3).

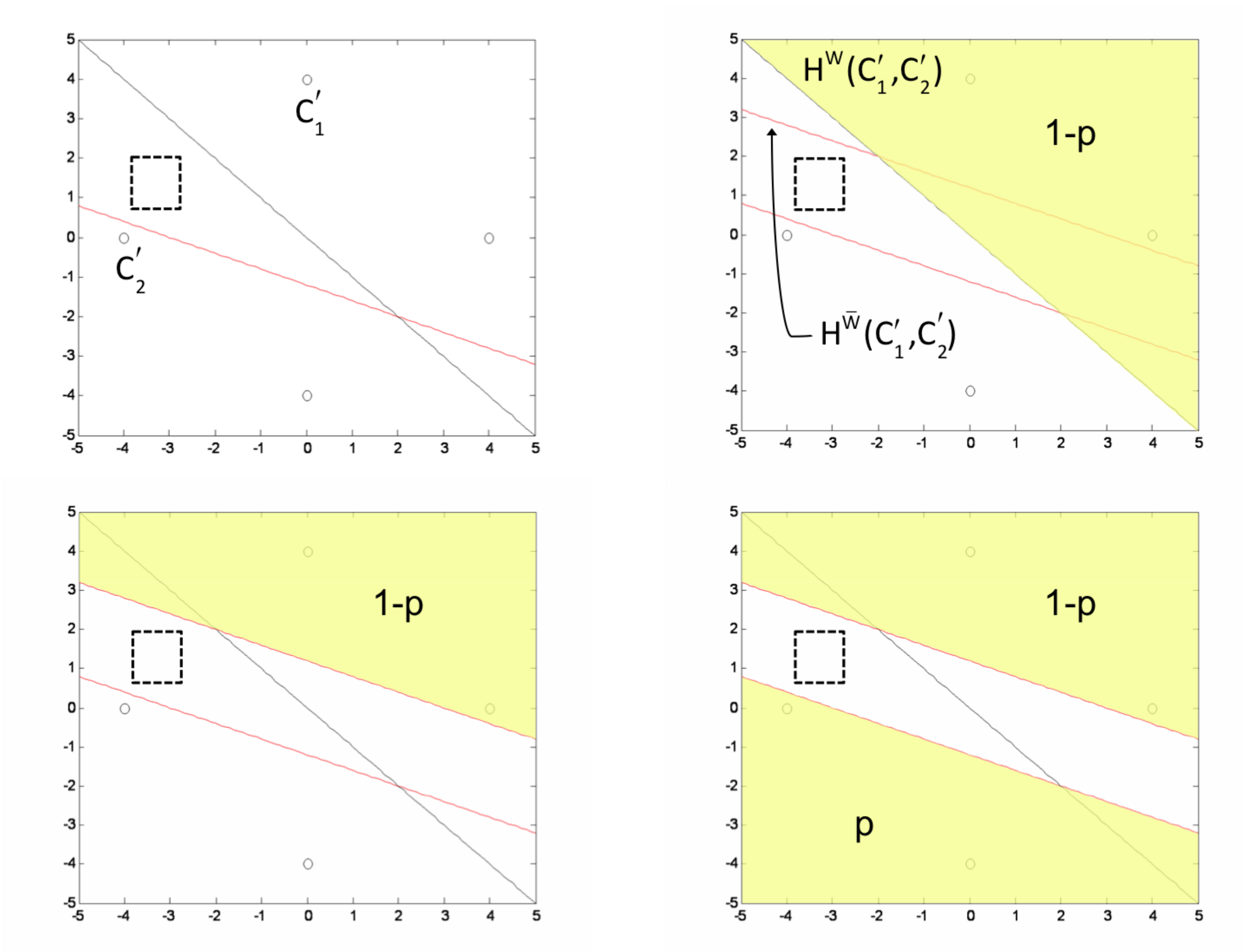

Figure 3: Voronoi Tessellations for Candidates C_1', C_2'

Since, under $\mathbb{P}_{\overline{T}}$, the area above $H^{\overline{W}}(C_1', C_2')$ equals $1-p$ and the probability of the area below $H^{\overline{W}}(C_1, C_2)$ equals p, the area between $H^{\overline{W}}(C_1, C_2)$ and $H^{\overline{W}}(C_1', C_2')$ would have zero probability (see the lower-right panel in Figure 3). Given that the rectangle is between $H^{\overline{W}}(C_1, C_2)$ and $H^{\overline{W}}(C_1', C_2')$, it also has zero probability. Since the argument can be repeated for any bounded set, any such set would have probability zero. We then reach a contradiction as this would lead to the conclusion that the probability of the entire ideological space ($\mathbb{R}^2$) is zero. The proof of Lemma 2 simply formalizes and extends this argument for a general ideological space dimension k.

When there are more than two candidates, the same argument cannot necessarily be applied since the existence of multiple profiles generating the same Voronoi tessellation is no longer guaranteed when the number of candidates is greater than $k+1$. It is nevertheless intuitive that the addition of more information with a larger number of candidates would still allow for identification. This is indeed so. As in Lemma 1, this is established if one can sample candidates arbitrarily close to two positions (as guaranteed by our assumptions) and appealing to continuity arguments. This is the main result of the paper, which is stated in the following theorem:

Theorem 1. *Suppose Assumptions 1 and 2 hold and $||W||_{k\times k} = \sqrt{k}$. Then $(\mathbb{P}_T, W)$ is identified.*

An important implication of Theorem 1 is that the distribution of voters' preferences in the ideological space can be recovered together with the relative weights voters ascribe to the various dimensions of the ideological space from cross-sectional, aggregate electoral data for any election. Using electoral data for different types of office (e.g., local vs. national legislatures), it is therefore possible, for example, to assess whether the recovered preference distributions are the same across elections, or whether voters care differently about specific ideological dimensions depending on the type of the political office. Similarly, using electoral data for the same office through time, it is possible to quantify the way voters' tastes evolve through time and how they correlate with economic conditions or other aggregate outcomes. We also note that W can potentially be made dependent on $\mathbf{X}$, since the identification results above are established for a given stratum of the covariates.

4 Estimation

In the simple case of a one-dimensional ideological space ($k = 1$), an election provides direct estimates of the cumulative distribution function $F_T(t|\mathbf{X}) = \int_{-\infty}^{t} f_{T|\mathbf{X}}(u|\mathbf{X})du$ at each of the midpoints separating any two contiguous candidates.[17] Estimation of the distribution of voters' preferences is therefore straightforward. Consider a generic election with n candidates and assume, without loss of generality, that $C_1 < C_2 < \cdots < C_n$. The sum of the proportions of votes received by candidate C_i and by all the candidates positioned to the left of C_i gives an estimate of the cdf F_T at $\overline{C}_i \equiv \frac{C_i + C_{i+1}}{2}$ where, $i = 1, \ldots, n-1$. As more elections are sampled (possibly with different numbers of candidates in each election), we obtain an increasing number of points at which we can estimate the cdf. Let p_i, $i = 1, \ldots, n$, be the vote shares obtained by candidates $C_1, \ldots, C_n$ in an election with n candidates. Notice that

$$\mathbb{E}(\mathbf{1}(T \leq \overline{C}_i)|\mathcal{C}, \mathbf{X}) = \mathbb{E}(p_i|\mathcal{C}, \mathbf{X}) = F_T(\overline{C}_i|\mathbf{X}),$$

and a natural estimator for F_T given a sample of elections with any number of candidates would be a multivariate kernel or local linear polynomial regression. Under usual conditions (see, e.g., Li and Racine (2007)), the estimator is consistent and has an asymptotically normal distribution. Other nonparametric techniques (splines, series) may also be employed. To impose monotonicity, one could appeal to monotone splines (Ramsay (1988), He and Shi (1998)) or smoothed isotonic regressions (Wright (1982), Friedman and Tibshirani (1984), Mukerjee (1988), Mammen (1991)), possibly conditioning on regressor strata if necessary.

In the general case where the number of dimensions of the ideological space is greater than one ($k > 1$), however, it is not possible to directly recover estimates for the cumulative

[17] If the ideological space has only one dimension, $F_T(t|\mathbf{X})$ is the only object of interest, since W is a scalar that plays no role.

distribution function as in the previous case.[18] It is nevertheless true that for a given election:

$$\mathbb{E}\left[\int \mathbf{1}_{\mathbf{t}\in V_i^W(\mathcal{C})} f_{T|X}(\mathbf{t}|\mathbf{X})d\mathbf{t} - p_i \middle| \tilde{\mathbf{X}}\right] = 0, \quad i \in \{1, \ldots, n\}$$

where $V_i^W(\mathcal{C})$ is the Voronoi cell for candidate i, $\tilde{\mathbf{X}} = (\mathbf{X}, \mathcal{C})$, and the expectation is taken with respect to ϵ given candidate positions and $\mathbf{X}$. As before, the quantities $p_i, i \in \{1, \ldots, n\}$, are the electoral outcomes obtained from the data (i.e., the vote shares obtained by each candidate in the election).

In a parametric context, this structure suggests searching for parameters characterizing W and f that minimize the distance between sample analogs of the moments above and zero. Because $f(\cdot)$ is non-parametric, we use a sieve minimum distance estimator as suggested in Ai and Chen (2003) (see also Newey and Powell (2003) and Ai and Chen (2007)). We follow here the notation in that paper. Letting $W \in \Theta$ and $f \in \mathcal{H}$, the estimator is the sample counterpart to the following minimization problem:

$$\inf_{(W,f)\in\Theta\times\mathcal{H}} \mathbb{E}\left[m(\tilde{\mathbf{X}}, f)^\top \left[\Sigma(\tilde{\mathbf{X}})\right]^{-1} m(\tilde{\mathbf{X}}, f)\right] \tag{3}$$

where $\Sigma(\tilde{\mathbf{X}})$ is a positive definite matrix for every $\tilde{\mathbf{X}}$ and $m(\tilde{\mathbf{X}}, (W, f)) = \mathbb{E}\left[\rho(\mathbf{p}, \tilde{\mathbf{X}}, (W, f))|\tilde{\mathbf{X}}\right]$ with

$$\rho(\mathbf{p}, \tilde{\mathbf{X}}, (W, f)) = \left(\int \mathbf{1}_{\mathbf{t}\in V_i^W(\mathcal{C})} f_{T|X}(\mathbf{t}|\mathbf{X})d\mathbf{t} - p_i\right)_{i=1,\ldots,n-1} \tag{4}$$

where $\mathbf{p} = (p_i)_{i=1,\ldots,n}$ denotes the vector of vote shares in the data. Notice that the n-th component of the above vector is omitted as the vector adds up to one. For ease of exposition, here we consider the case where elections have the same number of candidates. If the number of candidates differs across elections (as is the case in our empirical application), the objective function can be rewritten as the sum of similarly defined functions for different candidate numbers and treated, for example, as in the analysis of auctions with different

[18]If the ideological space is multi-dimensional, the weighting matrix W is also an object of interest.

numbers of bidders.[19]

As pointed out by Ai and Chen (2003), two difficulties arise in constructing this estimator. First, the conditional expectation m is unknown. Second, the function space $\mathcal{H}$ may be too large. To address the first issue, a non-parametric estimator $\hat{m}$ is used in place of m. With regard to the second issue, the domain $\mathcal{H}$ is replaced by a sieve space $\mathcal{H}_E$ which increases in complexity as the sample size grows.

For the estimation of the function m, let $\{b_j(\cdot), j = 1, 2, \ldots\}$ denote a sequence of known basis functions (e.g., power series, splines, etc.) that approximate well square integrable real-valued functions of $\tilde{\mathbf{X}} = (\mathbf{X}, \mathcal{C})$. With $b^J(\cdot) = (b_1(\cdot), \ldots, b_J(\cdot))^\top$ and given a particular parameter vector (W, f), the sieve estimator for the function $m_i(\cdot, (W, f))$, the i-th component in m, is given by

$$\widehat{m}_i(\cdot, (W, f)) = \sum_{e=1}^{E} \rho_i(\mathbf{p}_e, \tilde{\mathbf{X}}_e, (W, f)) b^J(\tilde{\mathbf{X}}_e)^\top (B^\top B)^{-1} b^J(\cdot) \qquad i = 1, \ldots, n-1 \qquad (5)$$

where $B_{E\times J} = (b^J(\tilde{\mathbf{X}}_1), \ldots, b^J(\tilde{\mathbf{X}}_E))^\top$ and $e = 1, \ldots, E$ indexes the elections in the data.

We consider the class $\mathcal{H}$ of densities studied by Gallant and Nychka (1987).[20] For simplicity, we initially omit the conditioning variables $(\mathbf{X})$, but notice that the approach can be extended to conditional densities as in Gallant and Tauchen (1989), for example. Fix $k_0 > d/2$, $\delta_0 > d/2$, $\mathcal{B}_0 > 0$, and let $\phi(\mathbf{t})$ denote the multivariate standard normal density. The class $\mathcal{H}$ admits densities f such that:[21]

$$f(\mathbf{t}) = h(\mathbf{t})^2$$

[19]See for instance the treatment in Donald and Paarsch (1993).

[20]See also Fenton and Gallant (1996a), Fenton and Gallant (1996b), Coppejans and Gallant (2002) and references therein.

[21]Since Gallant and Nychka (1987) study a likelihood-based estimator, they focus on $f(\mathbf{t}) = h(\mathbf{t})^2 + \varepsilon\phi(\mathbf{t})$. The additional term $\varepsilon\phi(\mathbf{t})$ is meant to steer the (log-)likelihood function away from $-\infty$. We use a moment-based objective function and Theorems 1 and 2 in Gallant and Nychka (1987), which do not require this additional term, so we drop this term from our presentation.

with

$$\left(\sum_{|\lambda| \leq k_0} \int |D^\lambda h(\mathbf{t})|^2 (1 + \mathbf{t}^\top \mathbf{t})^{\delta_0} d\mathbf{t} \right)^{1/2} < \mathcal{B}_0 \tag{6}$$

where $\int f(\mathbf{t}) d\mathbf{t} = 1$,

$$D^\lambda h(\mathbf{t}) = \frac{\partial^{|\lambda|}}{\partial t_1^{\lambda_1} \partial t_2^{\lambda_2} \ldots \partial t_k^{\lambda_k}} h(\mathbf{t}), \quad \lambda = (\lambda_1, \ldots, \lambda_k)^\top \in \mathbb{N}^k$$

and $|\lambda| = \sum_{i=1}^k \lambda_i$. Given a compact set on the ideological space, condition (6) essentially constrains the smoothness of the densities and prevents strongly oscillatory behaviors over this compact set. Out of this set, the condition imposes some reasonable restrictions on the tail behavior of the densities. Nevertheless, condition (6) allows for tails as fat as $f(\mathbf{t}) \propto (1 + \mathbf{t}^\top \mathbf{t})^{-\eta}$ for $\eta > \delta_0$ or as thin as $f(\mathbf{t}) \propto e^{-\mathbf{t}^\top \mathbf{t}^\eta}$ for $1 < \eta < \delta_0 - 1$.

Gallant and Nychka (1987) show that the following sequence of sieve spaces is dense on the (closure of the) above class of densities (with respect to the norm $||f||_{\text{cons}} = \max_{|\lambda| \leq k_0} \sup_{\mathbf{t}} |D^\lambda f(\mathbf{t})|(1 + \mathbf{t}^\top \mathbf{t})^{\delta_0}$, which is the consistency norm we use in Proposition 1 below):

$$\mathcal{H}_E = \left\{ f : f(\mathbf{t}) = \left[\sum_{i=0}^{J_t} H_i(\mathbf{t}) \right]^2 \exp\left(-\frac{\mathbf{t}^\top \mathbf{t}}{2} \right), \int f(\mathbf{t}) d\mathbf{t} = 1 \right\}$$

where H_i are Hermite polynomials, ϕ is the standard multivariate normal density and ε is a small positive number.[22] As mentioned before, the set of densities on which $\cup_{E=1}^\infty \mathcal{H}_E$ is dense is fairly large. Because the (closure of) the parameter space is also compact with respect to the consistency norm (see the proof for Proposition 1), the inverse operator is continuous (see p. 1569 in Newey and Powell (2003)). Hence, ill-posedness of this inverse problem is not an issue.

As in Gallant and Tauchen (1989), when the conditioning variables $\mathbf{X}$ are introduced, let $\mathbf{z} = R^{-1}(\mathbf{t} - b - A\mathbf{X})$ where R and A are matrices of dimension $k \times k$ and $k \times \dim(\mathbf{X})$

[22] Kim (2007) examines truncated versions of the Gallant-Nychka sieve space on a compact support.

respectively and b is a k-dimensional vector. Then,

$$f(\mathbf{t}|\mathbf{X}) = h(\mathbf{z}|\mathbf{X})/\det(R)$$

where

$$h(\mathbf{z}|\mathbf{X}) = \frac{\left[\sum_{|\alpha|=0}^{J_t} a_\alpha(\mathbf{X})\mathbf{z}^\alpha\right]^2 \phi(\mathbf{z})}{\int \left[\sum_{|\alpha|=0}^{J_t} a_\alpha(\mathbf{X})\mathbf{U}^\alpha\right]^2 \phi(\mathbf{U})d\mathbf{U}}$$

with $a_\alpha(\mathbf{X}) = \sum_{|\beta|=0}^{J_x} a_{\alpha\beta}\mathbf{X}^\beta$. The function $\mathbf{z}^\alpha$ maps the multi-index $\alpha = (\alpha_1, \ldots, \alpha_d)$ into the monomial $\mathbf{z}^\alpha = \Pi_{i=1}^k z_i^{\alpha_i}$ and analogously for $\mathbf{X}^\beta$ with respect to $\beta = (\beta_1, \ldots, \beta_{\dim(\mathbf{X})})$.

The estimator is formally defined as:

$$(\widehat{W}, \widehat{f}) = \text{argmin}_{(W,f)\in\Theta\times\mathcal{H}_E} \frac{1}{E}\sum_{e=1}^{E} \widehat{m}(\tilde{\mathbf{X}}, (W,f))^\top \left[\widehat{\Sigma}(\tilde{\mathbf{X}})\right]^{-1} \widehat{m}(\tilde{\mathbf{X}}, (W,f)) \tag{7}$$

For a given pair (W, f), the components of the vector $\widehat{m}(\cdot, (W, f))$ are calculated as in (5). In our empirical application $\{b_j(\cdot), j = 1, 2, \ldots\}$ is a polynomial sieve and the ith component of $\widehat{m}(\tilde{\mathbf{X}}, (W, f))$ is the linear projection of $\rho_i(\mathbf{p}, \tilde{\mathbf{X}}, (W, f))$ on $\tilde{\mathbf{X}}$.

To calculate $\rho_i(\mathbf{p}, \tilde{\mathbf{X}}, (W, f))$ for a given (W, f) one needs to compute the integral in $\int \mathbf{1}_{\mathbf{t}\in V_i^W(\mathcal{C})} f_{T|X}(\mathbf{t}|\mathbf{X})d\mathbf{t} - p_i$. The estimator is very attractive computationally as integrals for putative densities f over a particular Voronoi cell can be easily obtained by simulation. Practically, we sample many draws from a bivariate normal density and take the average of the Hermite factors of the density evaluated at each draw times an indicator for whether the draw is closer to the candidate corresponding to the Voronoi cell of interest than to any other candidate. More precisely, for given parameter values and $\mathbf{X}$, we simulate S independent multivariate normal random variables[23] (with zero mean and identity variance-covariance)

[23] In our empirical application, we use $S = 1000$.

$\mathbf{z}_1, \ldots, \mathbf{z}_S$ and estimate $\rho_i(W, f)$ as

$$\rho_{iS}(W,f) \equiv \frac{1}{det(R)} S^{-1} \sum_{s=1}^{S} \frac{\left[\sum_{|\alpha|=0}^{J_t} a_\alpha(\mathbf{X}) \mathbf{z}_s^\alpha\right]^2}{\int \left[\sum_{|\alpha|=0}^{J_x} a_\alpha(\mathbf{X}) \mathbf{U}^\alpha\right]^2 \phi(\mathbf{U}) d\mathbf{U}} \times 1\left[d^W(\mathbf{t}_s, C_i) \le d^W(\mathbf{t}_s, C_j), j \neq i\right] \tag{8}$$

where $\mathbf{t}_s = b + A\mathbf{X} + R\mathbf{z}_s$. Given the parameters, the integral in the denominator can be analytically computed as it corresponds to the sum of even moments of normal variables. We use `Mathematica` to compute these integrals. The indicator $1\left[d^W(\mathbf{t}_s, C_i) \le d^W(\mathbf{t}_s, C_j), j \neq i\right]$ allows us to obtain the proportion of simulated types $\mathbf{t}_s$ that would choose candidate i and are positioned in $V_i^W(\mathcal{C})$. The construction of $\rho_{iS}(\cdot,\cdot)$ allows us to evaluate the objective function in (7) at given (f, W) once $\widehat{\Sigma}$ is computed.[24] Denote the estimator based on $\rho_{iS}(\cdot,\cdot)$ using S simulations by $(\hat{W}_S, \hat{f}_S)$. As the number of draws S increases, the approximation converges to the desired integral of $f(\mathbf{t}|\mathbf{X})$ over the Voronoi cell for candidate C_i by the Law of Large Numbers. In fact, because the convergence is uniform in (W, f), $plim_S(\hat{W}_S, \hat{f}_S) = (\hat{W}, \hat{f})$ as defined in (7).

Because of the simulations, our implementation of ρ_i and hence our objective function are not smooth. Hence, to minimize this function we use Nelder-Meade's non-gradient algorithm (though other non-gradient based methods could also be employed).[25] Using ten randomly drawn initial parameter proposals we proceed incrementally, first minimizing the objective function for values of J_t and J_x and using the optimal values as starting parameters for higher orders. The program is executed in `Fortran` using a High Performance Computing cluster. In our estimation, we follow Gallant and Tauchen (1989) and rescale the covariates (see Section 5 for further details).

To establish consistency we rely on the following assumptions:

Assumption 3. *(i) Elections are iid; (ii) supp*$(\tilde{\mathbf{X}})$ *is compact with nonempty interior; (iii)*

[24] In the empirical application we use $\widehat{\Sigma} = I$.

[25] If changes of variables are used to make the domain of integration (i.e., $d^W(\mathbf{t}_s, C_i) \le d^W(\mathbf{t}_s, C_j), j \neq i$) rectangular, the objective function may be made smooth (see, Genz and Bretz (2009)).

the density of $\tilde{\mathbf{X}}$ is bounded and bounded away from 0.

Assumption 4. *(i) The smallest and largest eigenvalues of $\mathbb{E}\{b^J(\tilde{\mathbf{X}})b^J(\tilde{\mathbf{X}})^\top\}$ are bounded and bounded away from zero for all J; (ii) for any $g(\cdot)$ with $\mathbb{E}[g(\tilde{\mathbf{X}})^2] < \infty$, there exist $b^J(\tilde{\mathbf{X}})^\top \pi$ such that $\mathbb{E}[\{g(\tilde{\mathbf{X}}) - b^J(\tilde{\mathbf{X}})^\top \pi\}^2] = o(1)$.*

Assumption 5. *(i) $\widehat{\Sigma}(\tilde{\mathbf{X}}) = \Sigma(\tilde{\mathbf{X}}) + o_p(1)$ uniformly over $supp(\tilde{\mathbf{X}})$; (ii) $\Sigma(\tilde{\mathbf{X}})$ is finite positive definite over $supp(\tilde{\mathbf{X}})$.*

Assumption 6. *Let $dim(J)$ be the number of parameters in the sieve approximation for $m(\cdot)$ and let $dim(J_t)$ and $dim(J_x)$ be the number of parameters for the sieve approximation of the distribution of types defined in equation (8). Analogously, let $dim(W)$ be the dimension of the parametric component W. Then, $(n-1)dim(J) \geq dim(J_t) + dim(J_x) + dim(W)$, $J_t, J_x \to \infty$ and $J/E \to 0$ as $E \to \infty$.*

The following proposition establishes consistency:

Proposition 1. *Under Assumptions 1-6 and Θ compact (with respect to the Frobenius norm),*

$$plim_S(\widehat{W}_S, \widehat{f}_S) = (\widehat{W}, \widehat{f}) \to_p (W, f_T)$$

with respect to the norm

$$||(W, f)|| = \max_{|\lambda| \leq k_0} \sup_{\mathbf{t}} |D^\lambda f(\mathbf{t})|(1 + \mathbf{t}^\top \mathbf{t})^{\delta_0} + \sqrt{tr(W^\top W)}.$$

The proof for the above result is a slightly modified version of Lemma 3.1 in Ai and Chen (2003), where instead of appealing to Holder continuity in demonstrating stochastic equicontinuity of the objective function we adapt Lemma 3 in Andrews (1992) using dominance conditions. Because we do not rely on Holder continuity, however, the results on rates of convergence in Ai and Chen (2003) do not directly apply here. Hence, we do not provide asymptotic standard errors for the parametric components and functionals of the non-parametric components as in Ai and Chen (2003).[26] In the empirical application, we do, however, provide bootstrap standard errors.

5 Empirical Application

In this section, we illustrate the methodology described above with an empirical analysis of the 1999 elections of the European Parliament.[27] Elections for the European Parliament take place under the proportional representation system and typically with closed party lists. This means that voters in each electoral precinct do not vote for specific candidates, but for parties, and the total fraction of votes received by a party across all electoral precincts determines its proportion of seats in the Parliament. The identity of the politicians elected to Parliament is then determined by the parties' lists (e.g., if a party obtains three seats, the first three candidates on its list are elected).[28] Hence, in this context, the electoral candidates in an election are the parties competing in the election. As pointed out by Spenkuch (2013), among others, under proportional representation "it is in practically every voter's best interest to reveal his true preferences over which party he wishes to gain the marginal seat by voting for said party" (p. 1). In other words, in elections with proportional representations, voters have no incentives to behave strategically, and the maintained assumption that voters vote ideologically is particulary well suited for the European Parliament elections.

Our data consist of ideological positions of the candidates/parties competing in the election, electoral outcomes, and demographic and economic characteristics, for each electoral precinct. Since data on all demographic and economic variables are not available at

[26] Deriving rates of convergence in the context of our model is not straightforward and we leave it for future research. When W is known and elections have no more than two candidates, Gautier and Kitamura (2013) suggest an alternative estimator and provide rates of convergence.

[27] A description of the rules and composition of the European Parliament since its inception in 1979 can be found at `http://www.elections-europeennes.org/en/`.

[28] More precisely, Germany, Spain, France, Greece, Portugal and the United Kingdom have closed party lists; Austria, Belgium, Denmark, Finland, Italy, Sweden and the Netherlands have a preferential vote system (where voters can express a preference for the candidates on the list, but votes that do not express a preference are counted as votes for the party list); and Ireland has a single transferable vote system (where the voter indicates his/her first choice, then his/her secondary choice, etc.).

the electoral precinct level for Austria, Belgium, Denmark and Italy, we exclude these countries from the empirical analysis. Hence, our data set is a cross-section of elections for the European Parliament in the 693 electoral precincts of Finland, France, Germany, Greece, the Netherlands, Ireland, Portugal, Spain, Sweden, and the United Kingdom in 1999.[29]

The ideological positions of the parties were obtained from Hix, Noury, and Roland (2006), who used roll-call data for the 1999-2004 Legislature of the European Parliament to generate two-dimensional ideological positions for each member of parliament along the lines of the NOMINATE scores of Poole and Rosenthal (1997) for the US Congress.[30] As indicated in Heckman and Snyder (1997), ideological positions are obtained essentially through a (nonlinear) factor model with a large number of roll-call votes and parliament members. Given the magnitude of these dimensions, we follow the empirical literature on "large N and large T" factor models and take these scores as data (see, e.g., Stock and Watson (2002), Bai and Ng (2006a) or Bai and Ng (2006b)).

Hix, Noury, and Roland (2006) provide an interpretation of the two dimensions of the ideological space based on an extensive statistical analysis which combines parties' manifestos and expert judgements by political analysts. They relate the first dimension to a general left-right scale on socio-economic issues, and the second dimension to positions regarding European integration policies.

The members of the European Parliament (MEPs) organize themselves into ideological party groups (EP groups) as in traditional national legislatures. Each EP group contains all the MEPs representing the parties that belong to that group. Within each country, it is typically the case that parties that belong to the same EP groups form electoral coalitions, where all the parties in the same EP group run a common electoral campaign based on a unified message representing the ideological positions of their group. Often, these positions vary across electoral constituencies within a country, representing regional differences in pol-

[29]We only have complete data on one electoral precinct in Ireland, Dublin, which is included in our analysis.
[30]The data are publicly available at `http://personal.lse.ac.uk/hix/HixNouryRolandEPdata.htm`.

icy stances.[31] Since the closed-list proportional representation system induces strong party cohesion (see, e.g., Diermeier and Feddersen (1998)), where elected representatives systematically (though not always) vote along party lines, we identify the ideological position of each "candidate" running in an electoral constituency by the ideological position of his/her EP group in that constituency. In particular, for each dimension of the ideological space, we use the average coordinate of individual MEPs from each EP group in a constituency as the coordinate for the position of the "candidate" representing that EP group in that constituency.[32] Figure 4 plots the positions for the "candidates" across all electoral constituencies in our data and indicates their EP group affiliation. All elections had more than two candidates: 68 elections had three, 396 elections had four, 43 elections had five, 40 elections had six, and 146 elections had 7 candidates.

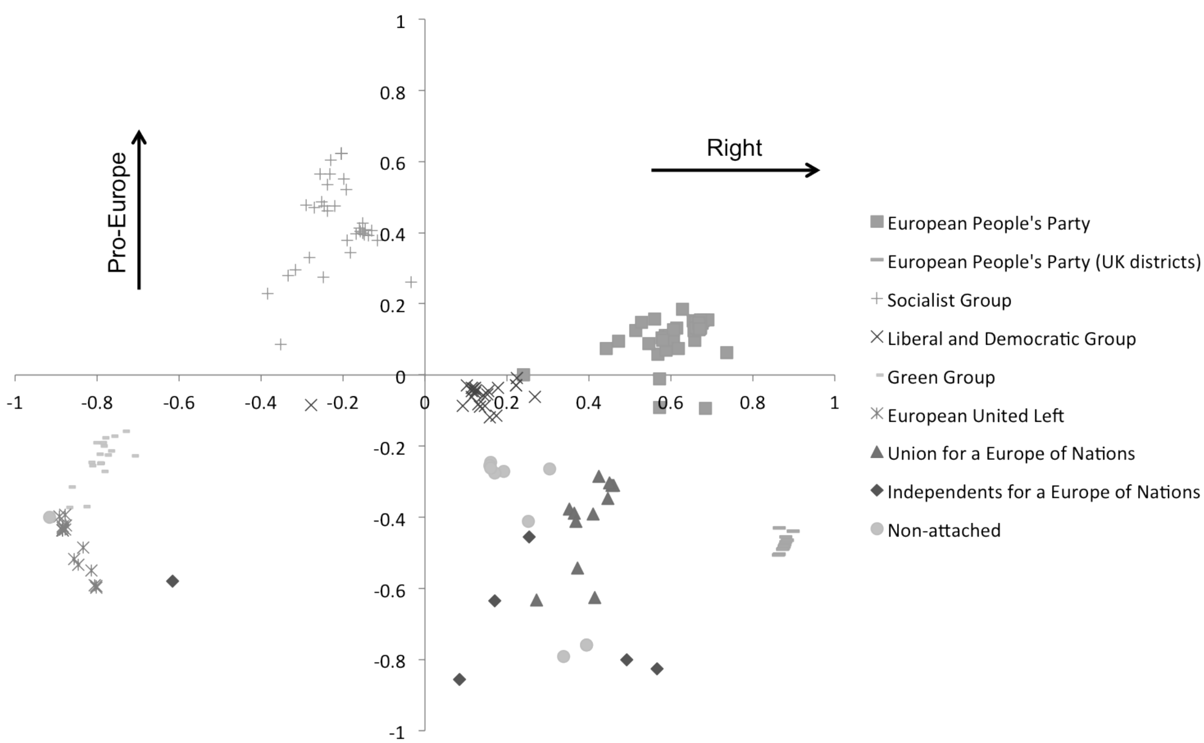

Figure 4: Candidate Positions, 1999

[31]Note that some countries have a single electoral constituency (Finland, France, Greece, Netherlands, Portugal, Spain, and Sweden), while others (Germany, Ireland and UK) have many sub-national constituencies. Each constituency contains many electoral precincts.

[32]Degan and Merlo (2009) use a similar procedure for U.S. congressional elections. Note that very similar positions are obtained if instead of the average we use the median coordinate.

In accordance with the interpretation of Hix, Noury, and Roland (2006): "On the first dimension (. . .) the Radical Left and Greens [are] on the furthest left, then the Socialists on the center-left, the Liberals in the center, the European People's Party on the center-right, the British Conservatives and allies and French Gaullists and allies to the right," whereas on the second dimension "the main pro-European parties (the Socialists, Liberals, and European People's Party) [are] at the top (. . .) and the main anti-Europeans (the Radical Left, Greens, Gaullists, Extreme Right and Anti-Europeans) at the bottom" (p. 499).

To further illustrate the data on ideological positions, in Figure 5 we also plot the ideological positions of a few notable politicians who ran in the 1999 European Parliament elections as front runners on their parties' lists. On the left-wing/pro-Europe quadrant, for example, we can locate François Hollande, current president of France, at coordinates $(-0.372, 0.609)$, whereas in the Southwest quadrant (left, anti-Europe integration), we find Claudia Roth, leader of the German Green Party, at coordinates (-0.715,-0.663). In the right-wing/anti-Europe quadrant, we find Nicholas Clegg, leader of the UK Liberal Democrat Party, at $(0.123, -0.049)$; Jean-Marie Le Pen, founder and former leader of the French National Front Party, at $(0.576, -0.816)$; and Nigel Paul Farage, leader of the UK Independence Party, at $(0.566, -0.825)$.[33]

An observation unit in the data comprises information on candidate positions and vote shares at the electoral precinct level. Figure 6 depicts a typical data point—the Paris, France electoral precinct—with seven candidates, representing seven EP groups.

Each electoral precinct corresponds to a different tessellation of the ideological space, and we measure the proportion of voters in each cell using the proportion of votes obtained by each of the candidates in that electoral unit. Figure 7 combines the Voronoi tessellations for all the elections in our data. It is apparent from the figure that these tessellations cover the ideological space and provide sufficient variation that allows us to identify and estimate the

[33]Note that Le Pen and Farage are remarkably aligned in the ideological space. This may not come as a surprise after Marine Le Pen, daughter of Jean-Marie Le Pen, tweeted "congratulations" to the UK Independence Party after their recent success in local elections.

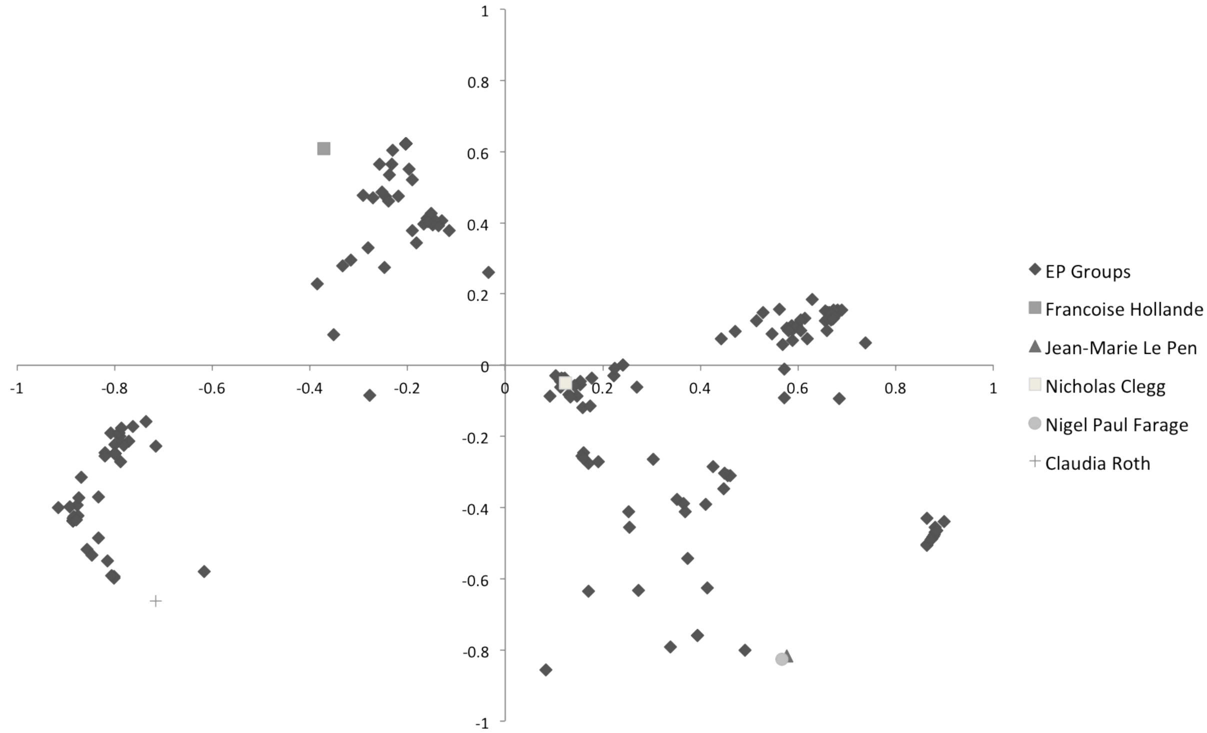

Figure 5: Individual Politician Positions, 1999

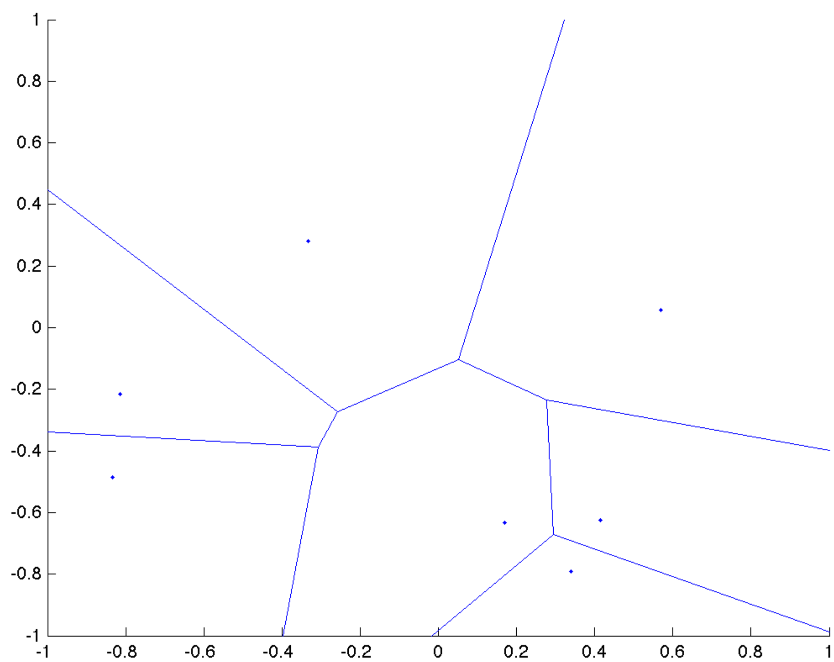

Figure 6: Voronoi Diagram for Paris (France), 1999

distribution of voter types (see our discussion of the conditions for identification in Section 3 above).

Table 1 contains minima and maxima for candidate coordinates. As we can see from the table, there is wide variability of candidate positions within each country, while the support of candidate distributions does not vary much across countries. Hence, there is no evidence of ideological segregation (or clustering) of electoral candidates by country.

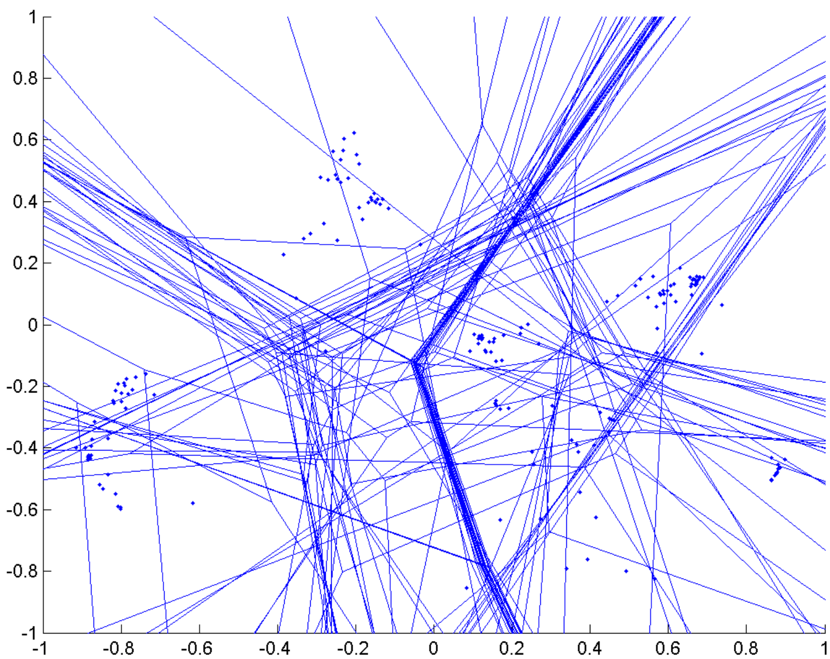

Figure 7: Superimposed Voronoi Diagrams, 1999

We combine the data on the ideological positions of electoral candidates with electoral outcomes in the 1999 elections and demographic and economic variables at the electoral precinct level from the 2001 European Census.[34] The election outcomes data were obtained from the CIVICACTIVE European Election Database.[35] The demographic and economic data were obtained from EUROSTAT and we extracted four variables at the electoral precinct

[34]Since the European Census is conducted every ten years, we use data from the 2001 census, which is the closest to 1999.

[35]The data is available at `http://extweb3.nsd.uib.no/civicactivecms/opencms/civicactive/en/`.

Table 1: Candidate Position Coordinates (Min and Max)

	Dimension 1		Dimension 2	
	Min	**Max**	**Min**	**Max**
Finland	-0.802	0.572	-0.597	0.474
France	-0.834	0.569	-0.792	0.280
Germany	-0.885	0.690	-0.438	0.622
Greece	-0.815	0.587	-0.550	0.551
Ireland	-0.874	0.547	-0.376	0.564
Netherlands	-0.856	0.577	-0.518	0.461
Portugal	-0.846	0.580	-0.632	0.475
Spain	-0.916	0.629	-0.400	0.603
Sweden	-0.833	0.571	-0.591	0.274
UK	-0.868	0.899	-0.855	0.521

Source: Hix, Noury and Roland. We define candidate positions as the (average) position for MEPs from a given EP group within each available constituency.

level: the female-to-male ratio; the percentage of the population older than 35; GDP per capita; and the unemployment rate.[36] We present summary statistics for these variables in Table 2.

Using these data, which as noted above contain a cross-section of 693 elections, we estimate our model. Following Gallant and Tauchen (1989), we re-scale the data to avoid situations in which extremely large (or small) values of the polynomial part of the conditional density are required to compensate for extremely small (or large) values of the exponential part. We transform the data so that $\check{\mathbf{X}}_e = \mathbf{S}^{-1/2}(\mathbf{X}_e - \overline{\mathbf{X}})$ where $\mathbf{S} = (1/E)\sum_{e=1}^{E}(\mathbf{X}_e - \overline{\mathbf{X}})(\mathbf{X}_e - \overline{\mathbf{X}})^\top, \overline{\mathbf{X}} = (1/E)\sum_{e=1}^{E}\mathbf{X}_e$ and $\mathbf{S}^{-1/2}$ is the Cholesky factorization of the inverse of $\mathbf{S}$. The estimates for $m(\cdot)$ as defined in (5) are linear projections on covariates. We use Hermite polynomials of order $J_t = 2$ (types) and $J_x = 2$ (covariates).[37] Finally, we use the identity matrix as our estimation weighting matrix ($\hat{\Sigma}$).

[36] Female-to-male ratio is obtained from a combination of the variable `cens_01rsctz` (where available) and `demo_r_d3avg` (otherwise), where `cens_01rsctz` is based on census data, while `demo_r_d3avg` contains yearly estimates. The number of individuals above 35 years-old comes from `cens_01rapop`. GDP per capita comes from `nama_r_e3gdp`. Unemployment figures are obtained from `lfst_r_lfu3rt`.

[37] In total, we have 78 parameters, including the parametric component ($= dim(J_t) + dim(J_x) + dim(W)$). Since we have up to 7 candidates per election ($n - 1 = 6$) and a constant plus four covariates and candidate positions in our linear estimation of $m(\cdot)$, $dim(J) = 1 + 4 + 2 \times 7 = 19$, the bounds in Assumption 6 are comfortably satisfied.

Table 2: Summary Statistics

Mean St Dev.	Female/ Male	% > 35 Yrs.-old	GDP per capita	Unempl.
Overall	1.040	0.616	21,989.10	0.074
	(0.034)	(0.051)	(9,165.46)	(0.047)
Finland	1.035	0.579	23,990.00	0.102
	(0.026)	(0.026)	(5,336.84)	(0.036)
France	1.052	0.679	21,820.83	0.083
	(0.023)	(0.037)	(7,140.55)	(0.024)
Germany	1.046	0.632	23,899.88	0.074
	(0.031)	(0.029)	(9,696.70)	(0.051)
Greece	0.985	0.563	12,058.82	0.108
	(0.038)	(0.034)	(2,947.06)	(0.039)
Ireland	1.065	0.446	40,600.00	0.030
	.	.	.	.
Netherlands	1.018	0.549	25,502.50	0.022
	(0.023)	(0.026)	(5,057.15)	(0.011)
Portugal	1.067	0.572	10,876.67	0.039
	(0.032)	(0.050)	(3,122.79)	(0.019)
Spain	1.027	0.561	15,516.00	0.099
	(0.030)	(0.048)	(3,467.58)	(0.046)
Sweden	1.014	0.574	25,742.86	0.054
	(0.015)	(0.019)	(3,349.14)	(0.012)
UK	1.050	0.562	25,672.73	0.049
	(0.017)	(0.038)	(9,083.06)	(0.015)

Source: EUROSTAT. GDP per capita is in Euros. We only have complete data on one precinct in Ireland: Dublin. Hence, no standard deviations are provided for Ireland.

The estimates of the weighting matrix W we obtain are $W_{2,2} = 0.287$ and $W_{1,2} = W_{2,1} = -0.316$. Bootstrap standard errors for $W_{2,2}$ and $W_{1,2}$ are equal to 0.052 and 0.049, respectively. Given J_t and J_x, the estimator is essentially an (overidentified) GMM estimator. We compute the standard errors from estimates obtained from 200 bootstrap samples (after recentering the targeted moments as recommended by standard practice, see Horowitz (2001)).[38] Bootstrap standard errors are also presented for functionals of the estimated distributions of voter types.

These estimates quantify the relative importance of the European integration dimension (dimension 2) versus the socio-economic policy dimension (dimension 1), $W_{2,2}$ (with $W_{1,1}$ normalized to one), and the extent to which voters are willing to trade-off the two dimensions, $W_{1,2}$. Figure 8 plots an indifference curve for a voter with ideological position $(0, 0)$ implied by these estimates. In particular, the figure depicts the *loci* of candidates at distance 1 from a voter with ideological position $(0, 0)$. Our results indicate that when a candidate adopts a more right-leaning position on the left-right socio-economic policy scale, voters need to be "compensated" by a more pro-European integration posture to attain the same utility level. At the same time, voters attribute more importance to candidates' ideological positions on socio-economic issues than to their stance on European integration.

Turning attention to the estimates of the distribution of the ideological positions of voters, $\mathbb{P}_{T|X}$, Figure 9 plots level curves for the voter type distribution for electoral precincts at the 75th percentile of the female-to-male ratio (approximately 1.06 in our data) and the 25th percentile of the proportion of residents above 35 years-old (approximately 0.58 in the data) and various percentile combinations for the other two variables (per-capita GDP and the unemployment rate).[39] As we can see from the figure, multi-modality and non-concavity are pronounced features of the recovered distribution of voter preferences.

[38] Chen and Pouzo (2009) suggest a weighted bootstrap when the generalized residual $\rho(\cdot)$ is nonsmooth (as in our case), but require that $m(\cdot)$ be smooth (which is not our case).

[39] Electoral precincts with about 1.06 female/male ratio and 58% of the population above 35 years-old in the data correspond approximately to localities such as Leziria do Tejo (PT) or North Yorkshire (UK), for example.

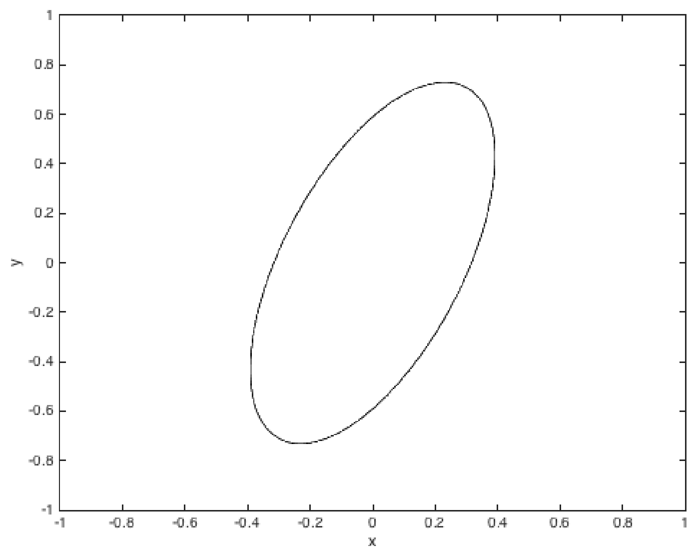

Figure 8: Indifference Curve for $d^W(\mathbf{0}, C) = 1$.

These findings represent a potential challenge for theoretical research in political economy, which systematically assumes that the distribution of voters' preferences is uni-modal and/or log-concave (see, e.g., Persson and Tabellini (2000), Austen-Smith and Banks (2000) and Austen-Smith and Banks (2005)).

Another summary of our estimates is presented in Table 3, where we present the average coordinates of the estimated distribution of voter preferences and the correlation between coordinates for each country in our sample (see the table's notes for the exact construction). For purposes of comparison, Table 3 also reports the average coordinates of the distribution of candidate positions in the data and the correlation between coordinates for each country. According to our estimates there is a positive correlation between candidates' (average) coordinates and voters' (average) coordinates, which is equal to 0.76 for dimension 1 and 0.23 for dimension 2.[40] With respect to the correlation between (average) coordinates, the signs of the correlation for voters and for candidates are the same for six out of the ten countries.

[40]Recall that Assumption 1 postulates that, after conditioning on observable characteristics, the distributions of voter preferences and candidate positions are independent. The correlations reported in Table 3 are not conditional on covariates.

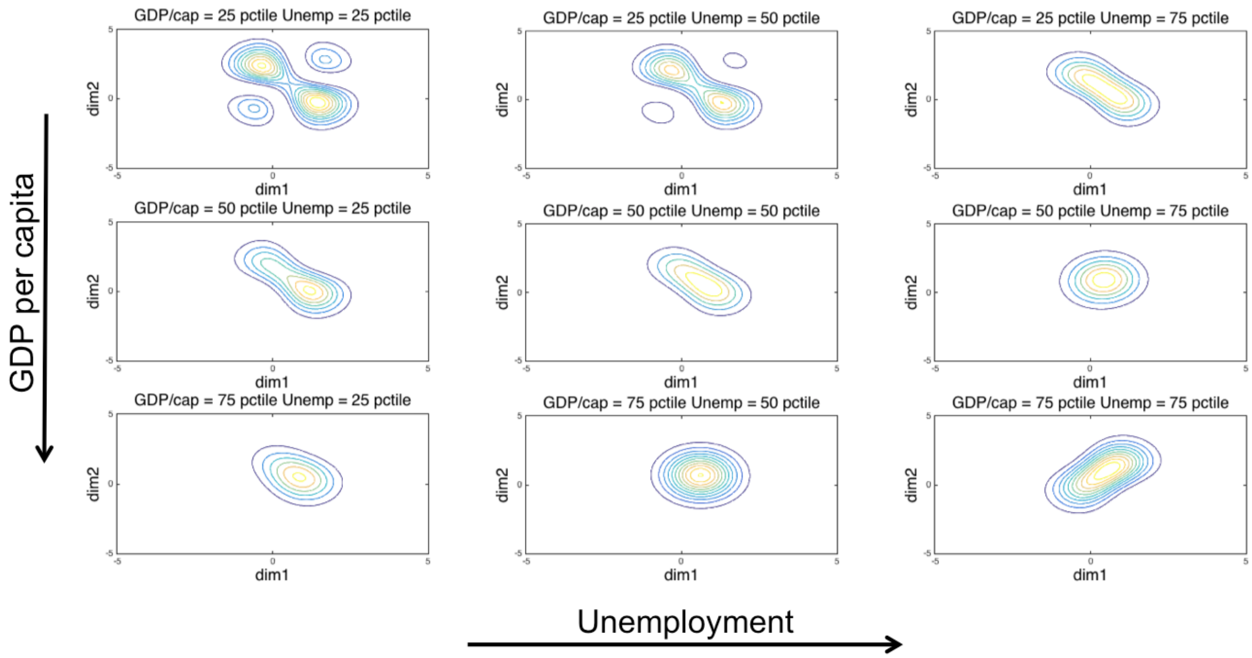

Figure 9: Results at Percentiles of Conditioning Variables (Female/Male: 75 pctile and % > 35 Years-old: 25 pctile)

Table 3: Distribution of Voters and Candidates Coordinates

	Voters			Candidates		
Country	Dim. 1 (mean)	Dim. 2 (mean)	Correl.	Dim. 1 (mean)	Dim. 2 (mean)	Correl.
Finland	0.366	0.543	0.673	-0.017	-0.029	0.222
France	0.371	0.396	-0.692	-0.002	-0.161	-0.027
Germany	0.465	0.488	-0.410	0.197	0.246	-0.079
Greece	0.093	0.401	-0.579	-0.094	0.079	0.469
Ireland	0.652	0.967	0.322	0.243	-0.175	0.018
Netherlands	0.728	0.597	-0.227	0.074	0.019	0.112
Portugal	0.656	1.543	0.587	0.091	0.148	-0.091
Spain	0.505	0.888	0.127	0.068	0.232	0.158
Sweden	0.265	0.416	0.629	-0.015	-0.057	0.267
UK	0.666	0.699	0.492	0.283	-0.117	-0.689

Note: Average voter coordinates for each country are a (population weighted) average of precinct distributional means given that precinct covariates. The voter type correlation is conditional on national average values for the covariates. Candidate averages and correlation are constructed using the ideological positions for each of the MEPs available in the data.

To investigate the relationships between demographic and economic variables and the distribution of voters' preferences, Tables 4 and 5 report the fraction of voters who are on the right of the left-right socio-economic policy dimension and the fraction of voters who are pro-Europe, respectively, for electoral precincts at the 25th, 50th and 75th percentiles of each covariate and average levels for all other covariates.[41] As we can see from the tables, electoral precincts with a relatively larger female-to-male ratio, precincts with a relatively larger share of the population above the age of 35, and precincts with a relatively higher level of GDP per-capita are relatively less conservative (or less right-leaning on socio-economic policies), and more pro-Europe. On the other hand, electoral precincts with a relatively higher unemployment rate are relatively less conservative, but also less pro-Europe.

Although not directly comparable, many of our findings are consistent with those of the EUROBAROMETER surveys, which document similar correlations between the gender and employment status of European citizens and their sentiments toward European policies.[42] In particular, according to the 1999 survey, relatively fewer women (37.04%) and relatively fewer people who are unemployed (29.71%) locate themselves on the "right" of the political spectrum than men (39.48%) and people who are employed (38.09%), respectively.[43] Moreover, according to the 1995 EUROBAROMETER survey, relatively fewer women (14.92%) and relatively more people who are unemployed (18.17%) consider their country's membership of the European Union "a bad thing" than men (15.87%) and people who are employed

[41]Loosely speaking, the table reports the "marginal effects" of each covariate.

[42]The EUROBAROMETER surveys are public opinion surveys conducted annually by the European Commission. They interview a representative sample of European citizens in all European Union member nations asking a variety of questions, that may differ from year to year, about the citizens' attitude toward Europe and European policies. Detailed descriptions of the surveys can be found at `http://ec.europa.eu/public_opinion/index_en.htm`. The statistics we report here are for the ten countries in our estimation sample only and are calculated using the Mannheim Eurobarometer Trend File, 1970-2002 (ICPSR 4357), which is available on-line at `http://www.icpsr.umich.edu`.

[43]These statistics are based on the answer to the following question: "In political matters people talk of the 'left' and the 'right'. How would you place your views on this [10-point] scale?" where "right" corresponds to an answer of 6 and above. Note that the relative comparisons between men and women and between employed and unemployed hold for any value of the cutoff used to classify answers as "right." Also, note that the EUROBAROMETER 10-point scale does not necessarily map into the spatial representation of the ideological space we consider.

Table 4: Fraction of right-leaning voters

Percentiles	Female/ Male	% > 35 Yrs.-old	GDP per capita	Unempl.
25th	0.616	0.655	0.665	0.694
	(0.096)	(0.083)	(0.102)	(0.093)
50th	0.642	0.644	0.650	0.670
	(0.067)	(0.065)	(0.071)	(0.074)
75th	0.665	0.636	0.640	0.614
	(0.093)	(0.066)	(0.072)	(0.070)

Note: Proportion of voters with first component (left-right) > 0. Standard errors in parentheses computed by bootstrap from 200 samples.

(15.43%), respectively.[44] On the other hand, according to the same EUROBAROMETER surveys, relatively more people older than 35 locate themselves on the "right" of the political spectrum (39.39%) and consider their country's membership of the European Union "a bad thing" (16.57%) than their younger counterparts (36.35% and 12.92%, respectively), which is somewhat at odds with our findings.

As a measure of within-sample fit, we calculate the Pearson correlation between realized and predicted vote shares which is equal to 0.84. In order to assess the out-of-sample performance of the model, we also perform an additional estimation. We exclude Portugal and its 108 electoral precincts from the estimation sample, and use the estimated model to predict the voting shares in the excluded Portuguese electoral precincts. The Pearson correlation between realized and predicted vote shares we obtain for Portugal is equal to 0.81. Overall, these results indicate that the model fits the data relatively well.

6 Discussion

In this paper, we have addressed the issue of nonparametric identification and estimation of voters' preferences using aggregate data on electoral outcomes. Starting from the basic tenets of one of the fundamental models of political economy, the *spatial theory of voting*,

[44]These statistics are based on the answer to the following question: "Generally speaking, do you think that [your country's] membership of the European Community (common market) is ...?" The 1999 survey did not ask this question.

Table 5: Fraction of pro-Europe voters

Percentiles	Female/ Male	% > 35 Yrs.-old	GDP per capita	Unempl.
25th	0.624	0.691	0.675	0.679
	(0.093)	(0.094)	(0.110)	(0.097)
50th	0.677	0.683	0.683	0.688
	(0.081)	(0.082)	(0.087)	(0.082)
75th	0.730	0.677	0.685	0.670
	(0.098)	(0.083)	(0.078)	(0.087)

Note: Proportion of voters with second component (against-pro Europe) > 0. Standard errors in parentheses computed by bootstrap from 200 samples.

and building on the work of Degan and Merlo (2009), which represents elections as Voronoi tessellations of the ideological space, we have established that voter preference distributions and other parameters of interest can be retrieved from aggregate electoral data. We have also shown that these objects can be consistently estimated using the methods by Ai and Chen (2003), and have provided an empirical illustration of our analysis using data from the 1999 European Parliament elections.

One potential extension of interest allows for electoral candidates to differ not only with respect to their locations in the ideological space, but also with respect to (non-spatial) individual characteristics related to their quality. These quality characteristics, which are commonly referred to as "valence" in the literature (see, e.g., Enelow and Hinich (1984) and the discussion in Degan and Merlo (2009)), are typically assumed to be known to the voters, but not the econometrician. In the context of our model, this extension can be accommodated within our framework by assuming that voter $\mathbf{t}$'s preferences over candidates in an election can be summarized by the utility function

$$U^{\mathbf{t}}\left(C_i, \delta_i\right) = u^{\mathbf{t}}\left(d^W\left(\mathbf{t}, C_i\right)^2 + \delta_i\right),$$

where δ_i is a candidate-specific valence term and $u^{\mathbf{t}}(\cdot)$ is a decreasing function as in Section 2 above. For this "linear-quadratic" specification of voter preferences, which is widely used in the political economy literature (see, e.g., Enelow and Hinich (1984)), Degan and Merlo

(2009) have shown that the set of nearest neighbors to a given candidate is still given by an intersection of halfspaces as when utility functions are given by expression (1). Specifically, the Voronoi cells for each candidate are now given by:

$$V_i^W(\mathcal{C},\delta) \equiv \{\mathbf{t} \in Y : d^W(\mathbf{t}, C_i)^2 + \delta_i \leq d^W(\mathbf{t}, C_j)^2 + \delta_j, j \neq i\},$$

where $\delta \equiv (\delta_1, \ldots, \delta_n)$.

In order to establish identification for this alternative specification of the model that incorporates valence terms, we need to modify our previous assumptions accordingly.

Assumption 1'. *The random vectors $(\mathcal{C},\delta)$ and $\mathbf{T}$ are conditionally independent given $\mathbf{X}$.*

Assumption 2'. *The distribution of $(\mathcal{C},\delta)$ is absolutely continuous with respect to the Lebesgue measure on $(\mathbb{R}^{n(k+1)}, \mathcal{B}(\mathbb{R}^{n(k+1)}))$ and the distribution of preference types $\mathbf{T}$ is absolutely continuous with respect to the Lesbegue measure on $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$.*

Under these assumptions, we can then establish that voter preference distributions and other parameters of interest (which include the valence parameter vector $\delta \equiv (\delta_1, \ldots, \delta_n)$) can be identified from aggregate electoral data.

Theorem 2. *Suppose Assumptions 1' and 2' hold and $||W||_{k\times k} = \sqrt{k}$. Then $(\mathbb{P}_T, W)$ is identified.*

The result is demonstrated along the same lines of our previous results and we provide further details on the necessary modifications to our arguments in the Appendix. We note that while in the proof of Theorem 2 the profile of valence terms $(\delta_1, \ldots, \delta_n)$ is assumed to be fixed across elections (which is consistent, for example, with our empirical application, where it would correspond to the valence of parties or party groups), in principle, it could also be allowed to vary across elections.

To conclude, it may be useful to cast our model into the broader context of a general spatial model of preferences with generic products, where the "consumer"obtains utility

$U^{\mathbf{t}}(C_i) = -(C_i - \mathbf{t})^\top W(C_i - \mathbf{t})$ from "product" i, $\mathbf{t}$ is a vector of individual "tastes", C_i is a vector of "product characteristics", W is a matrix of weights, and the distribution of tastes in the population of consumers $\mathbb{P}_{T|X,\epsilon}$ depends on "market" level covariates, both observed ($\mathbf{X}$), and unobserved (ϵ). Since this framework abstracts away from price endogeneity, the results of our analysis do not immediately translate to demand estimation, and generalizing our framework to address this broader class of problems is outside of the scope of this paper.

Nevertheless, (parametric) identification of individual taste heterogeneity is also important in demand estimation with aggregate data à la Berry, Levinsohn, and Pakes (1995) (BLP). Hence, our results have some relevance for this broader class of problems. Like those demand models, our framework allows for unobservable covariates to impact the distribution of tastes, and could in principle explain why there is not a perfect fit between market shares as predicted by the model and the observed market shares in a particular market. This is important since, as BLP also note, without an unobservable one would typically reject the model using a standard chi-squared goodness-of-fit test. This is because the number of sampled consumers that enter the measured market shares is typically quite large, so observed shares should equal predicted shares in each market. However, the unobservable breaks this equality: for a given market (i.e., product locations) the model will still predict a distribution of market shares.

References

AI, C., AND X. CHEN (2003): "Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions," *Econometrica*, 71, pp.1795–1843.

——— (2007): "Estimation of Possibly Misspecified Semiparametric Conditional Moment Restriction Models with Different Conditioning Variables," *Journal of Econometrics*, 141, pp.5–43.

ALESINA, A. (1988): "Credibility and Policy Convergence in a Two-Party System with Rational Voters," *American Economic Review*, 78, pp. 796–806.

ANDERSON, S. P., A. DE PALMA, AND J.-F. THISSE (1989): "Demand for Differentiated Products, Discrete Choice Models and the Characteristics Approach," *The Review of Economic Studies*, 56, pp. 21–35.

ANDREWS, D. W. K. (1992): "Generic Uniform Convergence," *Econometric Theory*, 8, pp.241–257.

AUSTEN-SMITH, D., AND J. S. BANKS (2000): *Positive Political Theory I: Collective Preferences*. Ann Arbor: University of Michigan Press.

——— (2005): *Positive Political Theory II: Strategy and Structure*. Ann Arbor: University of Michigan Press.

BAI, J., AND S. NG (2006a): "Confidence Intervals for Diffusion Index Forecasts and Inference for Factor-Augmented Regressions," *Econometrica*, 74, pp.1133–1150.

——— (2006b): "Evaluating latent and observed factors in macroeconomics and finance," *Journal of Econometrics*, 131, pp.507–537.

BERRY, S., AND P. HAILE (2009): "Identification of Discrete Choice Demand from Market Level Data," Working Paper, Yale University.

BERRY, S., J. LEVINSOHN, AND A. PAKES (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 63, pp.841–890.

BERRY, S., AND A. PAKES (2007): "The Pure Characteristics Demand Model," *International Economic Review*, 48, pp.1193–1226.

BLACK, D. (1958): *The Theory of Committees and Elections*. New York: Cambridge University Pressy.

BLUNDELL, R., X. CHEN, AND D. KRISTENSEN (2007): "Semi-Nonparametric IV Estimation of Shape-Invariant Engel Curves," *Econometrica*, 75, pp.1613–1669.

CAPLIN, A., AND B. NALEBUFF (1991): "Aggregation and Imperfect Competition: On the Existence of Equilibrium," *Econometrica*, 59, pp. 25–59.

CHEN, X., AND D. POUZO (2009): "Efficient Estimation of Semiparametric Conditional Moment Models with Possibly Nonsmooth Residuals," *Journal of Econometrics*, 152, 46–60.

CHESHER, A., AND J. M. C. S. SILVA (2002): "Taste Variation in Discrete Choice Models," *The Review of Economic Studies*, 69, pp. 147–168.

COPPEJANS, M., AND A. R. GALLANT (2002): "Cross-Validated SNP Density Estimates," *Journal of Econometrics*, 110, pp.27–65.

DAVIS, O., M. HINICH, AND P. ORDESHOOK (1970): "An Expository Development of a Mathematical Model of the Electoral Process," *American Political Science Review*, 64, pp.426–448.

DEGAN, A., AND A. MERLO (2009): "Do Voters Vote Ideologically?," *Journal of Economic Theory*, 144, pp.1869–1894.

DIERMEIER, D., AND T. J. FEDDERSEN (1998): "Cohesion in Legislatures and the Vote of Confidence Procedure," *The American Political Science Review*, 92, pp. 611–621.

Donald, S. G., and H. J. Paarsch (1993): "Piecewise Pseudo-Maximum Likelihood Estimation in Empirical Models of Auctions," *International Economic Review*, 34, pp.121–148.

Downs, A. (1957): *An Economic Theory of Democracy.* New York: Harper and Row.

Enelow, J., and M. Hinich (1984): *Economic Theories of Voter Turnout.* New York: Cambridge University Press.

Feenstra, R. C., and J. A. Levinsohn (1995): "Estimating Markups and Market Conduct with Multidimensional Product Attributes," *The Review of Economic Studies*, 62, pp. 19–52.

Fenton, V. M., and A. R. Gallant (1996a): "Convergence Rates of SNP Density Estimators," *Econometrica*, 64, pp.719–727.

——— (1996b): "Qualitative and Asymptotic Performance of SNP Density Estimators," *Journal of Econometrics*, 74, pp.77–118.

Friedman, J., and R. Tibshirani (1984): "The Monotone Smoothing of Scatterplots," *Technometrics*, 26, pp.243–250.

Gallant, A. R., and D. W. Nychka (1987): "Semi-Nonparametric Maximum Likelihood Estimation," *Econometrica*, 55, pp.363–390.

Gallant, A. R., and G. Tauchen (1989): "Seminonparametric Estimation of Conditionally Constrained Heterogeneous Processes: Asset Pricing Applications," *Econometrica*, 57, pp.1091–1120.

Gautier, E., and Y. Kitamura (2013): "Nonparametric Estimation in Random Coefficients Binary Choice Models," *Econometrica*, 81, pp.581–607.

Genz, A., and F. Bretz (2009): "Computation of Multivariate Normal and t Probabilities," in *Lecture Notes in Statistics*, vol. 195. Springer-Verlag, New York.

He, X., and P. Shi (1998): "Monotone B-Spline Smoothing," *Journal of the American Statistical Association*, 93, pp.643–650.

Heckman, J., and J. Snyder (1997): "Linear Probability Models of the Demand for Attributes with an Empirical Application to Estimating the Preferences of Legislators," *The RAND Journal of Economics*, 28, pp.S142–S189.

Henry, M., and I. Mourifié (2013): "Revealed Euclidean Preferences," *Journal of Applied Econometrics*, 28, pp.650–666.

Hibbs, D. (1977): "Political Parties and Macroeconomic Policy," *American Political Science Review*, 71, pp. 1467–1487.

Hinich, M., and M. C. Munger (1997): *Analytical Politics*. Cambridge University Press, Cambridge, UK.

Hinich, M. J., and M. C. Munger (1994): *Ideology and the Theory of Political Choice*. Ann Arbor: University of Michigan Press.

Hix, S., A. Noury, and G. Roland (2006): "Dimensions of Politics in the European Parliament," *American Journal of Political Science*, 50, pp.494–511.

Horowitz, J. (2001): "The Bootstrap," in *The Handbook of Econometrics*, ed. by J. Heckman, and E. Leamer, vol. 5. Elsevier.

Hotelling, H. (1929): "Stability in Competition," *Economic Journal*, 39, pp. 41–57.

Ichimura, H., and T. S. Thompson (1998): "Maximum Likelihood Estimation of a Binary Choice Model with Random Coefficients of Unknown Distribution," *Journal of Econometrics*, 86, pp.269–295.

Kim, K. i. (2007): "Uniform Convergence Rate of the SNP Density Estimator and Testing for Similarity of Two Unknown Densities," *Econometrics Journal*, 10, pp.1–34.

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

Li, Q., and J. Racine (2007): *Nonparametric Econometrics*. Princeton University Press, 1st edn.

Mammen, E. (1991): "Estimating a Smooth Monotone Regression Function," *The Annals of Statistics*, 19, pp.724–740.

Manski, C. (1988): "Identification of Binary Response Models," *Journal of the American Statistical Association*, 83, pp. 729–738.

Matzkin, R. (1992): "Nonparametric Identification and Estimation of Polychotomous Choice Models," *Journal of Econometrics*, 58, pp.137–168.

McFadden, D. (1974): "Conditional Logit Analysis of Qualitative Choice Behavior," in *Frontiers of Econometrics*, ed. by P. Zarembka. New York: Academic Press.

Merlo, A. (2006): "Whither Political Economy? Theories, Facts and Issues," in *Advances in Economics and Econometrics, Theory and Applications: Ninth World Congress of the Econometric Society*, ed. by R. Blundell, W. Newey, and T. Persson. Cambridge: Cambridge University Press.

Mukerjee, H. (1988): "Monotone Nonparametric Regression," *The Annals of Statistics*, 16, pp.741–750.

Newey, W. K. (1991): "Uniform Convergence in Probability and Stochastic Equicontinuity," *Econometrica*, 59, pp.1161–1167.

Newey, W. K., and J. L. Powell (2003): "Instrumental Variable Estimation of Nonparametric Models," *Econometrica*, 71, pp.1565–1578.

Okabe, A., B. Boots, K. Sugihara, and S. N. Chiu (2000): *Spatial Tessellations*. Wiley, Chichester, UK.

Pakes, A., and D. Pollard (1989): "Simulations and the Asymptotics of Optimization Estimators," *Econometrica*, 57(5).

Persson, T., and G. Tabellini (2000): *Political Economics*. Cambridge: MIT Press.

Pollard, D. (2002): *A User's Guide to Measure Theoretic Probability*. Cambridge University Press, Cambridge, UK.

Poole, K. T., and H. Rosenthal (1997): *Congress: A Political Economic History of Roll Call Voting*. New York: Oxford University Press.

Ramsay, J. O. (1988): "Monotone Regression Splines in Action," *Statistical Science*, 3, pp.425–441.

Spenkuch, J. (2013): "On the Extent of Strategic Voting," Northwestern University Working Paper.

Stock, J. H., and M. W. Watson (2002): "Forecasting Using Principal Components From a Large Number of Predictors," *Journal of the American Statistical Association*, 97, pp.1167–1179.

Strang, G. (1988): *Linear Algebra and Its Applications*. Harcourt Brace Jovanovich, 3rd edn.

Wright, F. T. (1982): "Monotone Regression Estimates for Grouped Observations," *The Annals of Statistics*, 10, pp.278–286.

Appendix: Proofs

Proof of Lemma 1.

Since W is known, without loss of generality we assume that $W = I$.[45] It is enough to consider a single election with n candidates. In what follows, for any integers l, m and r: $\mathcal{M}_{r\times l}$ is the space of $r \times l$ real matrices which is endowed with the typical Frobenius matrix norm $||A||_{r\times l} = \sqrt{\text{Tr}(A^\top A)}$ for $A \in \mathcal{M}_{r\times l}$; $||b||_l$ is the typical Euclidean norm in $\mathbb{R}^l$; and the product metric space $\mathcal{M}_{r\times l} \times \mathbb{R}^m$ is endowed with the normed product metric $d((A_1, b_1), (A_2, b_2)) = \sqrt{||A_1 - A_2||^2_{r\times l} + ||b_1 - b_2||^2_m}$.

Step 1: **(If $n = 2, \mathbb{P}_T$ is identified)** It suffices to show that any two distinct distributions $\mathbb{P}_{T_1}$ and $\mathbb{P}_{T_2}$ are relatively identified. When there are only two candidates, say C_1 and C_2, voters for whom $d(\mathbf{t}, C_1) - d(\mathbf{t}, C_2) < 0$ will vote for candidate C_1. Those for whom $d(\mathbf{t}, C_1) - d(\mathbf{t}, C_2) > 0$ will vote for candidate C_2. Equidistant voters determine the boundary of these two sets (which are the Voronoi cells for each candidate), which is defined by an affine hyperplane $\{\mathbf{t} \in \mathbb{R}^k : d(\mathbf{t}, C_1) = d(\mathbf{t}, C_2)\} = \{\mathbf{t} \in \mathbb{R}^k : A\mathbf{t} = b\}$ where it can be seen that $b = C_2^\top C_2 - C_1^\top C_1$ and $A_{1\times k} = 2(C_2 - C_1)^\top$.

Suppose that $\mathbb{P}_{T_1}$ and $\mathbb{P}_{T_2}$ are observationally equivalent. For any two candidates C_1 and C_2, vote shares will be identifical under either $\mathbb{P}_{T_1}$ or $\mathbb{P}_{T_2}$. This means that $\mathbb{P}_{T_1}(\{\mathbf{T} \in \mathbb{R}^k : A\mathbf{T} \leq b\}) = \mathbb{P}_{T_2}(\{\mathbf{T} \in \mathbb{R}^k : A\mathbf{T} \leq b\}), \forall A, b$. Hence, the cummulative distribution function for any linear combination of $\mathbf{T}$ will coincide under $\mathbb{P}_{T_1}$ and $\mathbb{P}_{T_2}$. By the Cramér-Wold device (see Pollard (2002), p.202), this implies that $\mathbb{P}_{T_1} = \mathbb{P}_{T_2}$. Consequently,

$$\begin{aligned}
&\mathbb{P}_{T_1} \neq \mathbb{P}_{T_2} \Rightarrow \\
&\exists (A^*, b^*) \in \mathcal{M}_{n-1\times k} \times \mathbb{R}^{n-1} : \mathbb{P}_{T_1}(\{\mathbf{T} \in \mathbb{R}^k : A^*\mathbf{T} \leq b^*\}) \neq \mathbb{P}_{T_2}(\{\mathbf{T} \in \mathbb{R}^k : A^*\mathbf{T} \leq b^*\}).
\end{aligned}$$

Given A^* and b^*, one can then find two candidates C_1^* and C_2^* such that $b^* = C_2^{*\top} C_2^* - C_1^{*\top} C_1^*$

[45] This is because one can then focus on the identification of the distribution of $W^{-1/2}\mathbf{T}$, which would yield identification of the distribution of $\mathbf{T}$ if W is known.

and $A^*_{1\times k} = 2(C_2^* - C_1^*)^\top$ for whom vote shares under $\mathbb{P}_{T_1}$ would be different from vote shares under $\mathbb{P}_{T_2}$.

For any two candidates, $\mathbb{P}_T(\{\mathbf{T} \in \mathbb{R}^k : d(\mathbf{T}, C_1) - d(\mathbf{T}, C_2) \leq 0\}) = \mathbb{P}_T(\{\mathbf{T} \in \mathbb{R}^k : C_1^\top C_1 - C_2^\top C_2 + 2(C_2 - C_1)^\top \mathbf{T} \leq 0\}) = \int_{C_1^\top C_1 - C_2^\top C_2 + 2(C_2 - C_1)^\top \mathbf{t} \leq 0} d\mathbb{P}(\mathbf{t})$. Since this is a continuous function of C_1 and C_2, any pair of candidates in a neighborhood of the candidate pair (C_1^*, C_2^*) obtained above would also present vote shares that are different under $\mathbb{P}_{T_1}$ than under $\mathbb{P}_{T_2}$. If the candidate and voter type distributions have a common support, elections allowing the distinction of any two distinct type distributions occur with positive probability. Otherwise, elections can only discriminate type distributions that differ on the candidate distribution support.

Step 2: (If $n > 2, \mathbb{P}_T$ is identified) If $\mathbb{P}_{T_1} \neq \mathbb{P}_{T_2}$, Step 1 demonstrates that there is a pair of locations C_1^* and C_2^* such that the proportion of votes closest to each one of the two locations differs under $\mathbb{P}_{T_1}$ and $\mathbb{P}_{T_2}$. Given our assumptions, even if candidates are not exactly positioned on these two locations, there is positive probability that an election occurs with candidates situated in small open balls around C_1^* and C_2^* with diameter $\eta > 0$.

For a particular voter, let $\overline{d}(\mathbf{t}, C_j^*; \eta)$ be the largest distance between that voter and any vector in the neighborhood of $C_j^*, j = 1, 2$. Likewise, let $\underline{d}(\mathbf{t}, C_j^*; \eta)$ be the smallest distance between that voter and any vector in the neighborhood of $C_j^*, j = 1, 2$. It should be clear that $\underline{d}(\mathbf{t}, C_j^*; \eta) \leq d(\mathbf{t}, C_j^*) \leq \overline{d}(\mathbf{t}, C_j^*; \eta), j = 1, 2$. The proportion of votes going to the candidates in the neighbohood of C_1^* is bounded above by $\mathbb{P}_T(\{\mathbf{T} \in \mathbb{R}^k : \underline{d}(\mathbf{T}, C_1^*; \eta) - \overline{d}(\mathbf{T}, C_2^*; \eta) \leq 0\})$ and bounded below by $\mathbb{P}_T(\{\mathbf{T} \in \mathbb{R}^k : \overline{d}(\mathbf{T}, C_1^*; \eta) - \underline{d}(\mathbf{T}, C_2^*; \eta) \leq 0\})$.

As the diameter η of these neighborhoods shrinks to zero, $\overline{d}(\mathbf{t}, C_j^*; \eta) \to d(\mathbf{t}, C_j^*)$ and $\underline{d}(\mathbf{t}, C_j^*; \eta) \to d(\mathbf{t}, C_j^*)$. Then the proportion of votes going to candidates in the neighborhood around C_1^* and the proportion of votes for the candidates in the neighborhood around C_2^* converge to the proportion of votes obtained by candidates situated exactly at C_1^* and C_2^*, respectively. Since these two positions distinguish the two distributions $\mathbb{P}_{T_1}$ and $\mathbb{P}_{T_2}$, continuity guarantees that elections where candidates are situated in a small enough neighborhood of these two positions also distinguish the two distributions. ■

Proof of Lemma 2.

Consider two different spatial voting models characterized by $(\mathbb{P}_T, W)$ and $(\mathbb{P}_{\overline{T}}, \overline{W})$. If $W = \overline{W}$, any election between two candidates will lead to the same partition of voters across the two environments and identification follows along the lines of Lemma 1. Furthermore, if $k = 1$, the weighting matrix W is a scalar and the normalization sets it equal to one. The identification again follows along the lines of Lemma 1. Assume then that $W \neq \overline{W}$ and $k > 1$. Suppose that $(\mathbb{P}_T, W)$ and $(\mathbb{P}_{\overline{T}}, \overline{W})$ are observationally equivalent: for almost every candidate-election profile $\mathcal{C} = (C_1, C_2)$, the proportion of votes obtained under the two different systems is identical.

Step 1: **(There is more than one set of candidates that generates the same partition of voters for a given weighting matrix W.)** Draw an election $\mathcal{C} = (C_1, C_2)$ and note that, with probability one, $C_1 \neq C_2$. Consider the set of vectors $\mathbf{t}$ such that

$$d^W(\mathbf{t}, C_1) = d^W(\mathbf{t}, C_2).$$

The above equation can be explicitly written as

$$(C_1 - C_2)^\top W\mathbf{t} = (C_1^\top W C_1 - C_2^\top W C_2)/2.$$

The solution set for this equation contains at least one element as long as $(C_1 - C_2)^\top W$ is different from the zero vector. Since W is positive definite and, consequently, has full rank, its nullspace is a singleton (comprised of the vector zero). Hence, $(C_1 - C_2)^\top W \neq \mathbf{0}$ (except when $C_1 = C_2$, which happens with zero probability given our assumptions). In this case, let P denote one solution to the equation above and let $\mathcal{C}'$ be such that

$$C_i' = 2C_i - P, \ \forall i.$$

Notice that $d^W(\mathbf{t}, C_1') - d^W(\mathbf{t}, C_2')$ is equal to

$$\begin{aligned}
& (C_1' - \mathbf{t})^\top W(C_1' - \mathbf{t}) - (C_2' - \mathbf{t})^\top W(C_2' - \mathbf{t}) \\
= \ & C_1'^\top W C_1' - C_2'^\top W C_2' - 2(C_1' - C_2')^\top W\mathbf{t} \\
= \ & (2C_1 - P)^\top W(2C_1 - P) - (2C_2 - P)^\top W(2C_2 - P) - 4(C_1 - C_2)^\top W\mathbf{t} \\
= \ & C_1^\top W C_1 - C_2^\top W C_2 - (C_1 - C_2)^\top W P - (C_1 - C_2)^\top W\mathbf{t}.
\end{aligned}$$

P is such that $d^W(P, C_1) - d^W(P, C_2) = 0$ and consequently $\frac{1}{2}\left(C_1^\top W C_1 - C_2^\top W C_2\right) = (C_1 - C_2)^\top W P$. This, in turn, implies that

$$\begin{aligned}
d^W(\mathbf{t}, C_1') - d^W(\mathbf{t}, C_2') = 0 \quad & \Leftrightarrow \\
C_1^\top W C_1 - C_2^\top W C_2 - 2(C_1 - C_2)^\top W\mathbf{t} = 0 \quad & \Leftrightarrow \\
d^W(\mathbf{t}, C_1) - d^W(\mathbf{t}, C_2) = 0. &
\end{aligned}$$

This establishes that the partition of voters under W (i.e., the W-Voronoi diagram) is the same across the two elections.

Step 2: (For $\mathcal{C} \neq \mathcal{C}'$ such that $V^W(\mathcal{C}) = V^W(\mathcal{C}')$, $H^W(C_1, C_2)$ and $H^{\overline{W}}(C_1, C_2)$ are different hyperplanes and $H^{\overline{W}}(C_1, C_2)$ and $H^{\overline{W}}(C_1', C_2')$ are parallel.) With only two candidates, Voronoi cells are simply half-spaces of $\mathbb{R}^k$ defining the nearest-neighbor sets for each candidate. Consider $\mathcal{C}$ and $\mathcal{C}'$ such that their Voronoi tessellations under W coincide, i.e., $V^W(\mathcal{C}) = V^W(\mathcal{C}')$ where $\mathcal{C}$ and $\mathcal{C}'$ are obtained as in Step 1.

To see that $H^{W}(C_1, C_2) \neq H^{\overline{W}}(C_1, C_2)$, first note that

$$H^{W}(C_1, C_2) \equiv \{\mathbf{t} \in \mathbb{R}^k : C_1^\top W C_1 - C_2^\top W C_2 + 2(C_2 - C_1)^\top W \mathbf{t} = 0\}, \tag{9}$$

is the solution set to a linear equation. Since W is positive definite and $C_1 \neq C_2$ with probability one, $2(C_2 - C_1)^\top W$ is nonzero with probability one. Then, with probability one the solution set to the equation defining $H^{W}(C_1, C_2)$ above has dimension $k - 1$, which is the dimension of the nullspace of $2(C_2 - C_1)^\top W$ (see, e.g., Strang (1988)). The same holds for $H^{\overline{W}}(C_1, C_2)$, which is defined as in (9) using $\overline{W}$ instead of W.

On the other hand, the intersection of $H^{W}(C_1, C_2)$ and $H^{\overline{W}}(C_1, C_2)$ is the solution set (in $\mathbb{R}^k$) to the system of equations given by:

$$\begin{cases} C_1^\top W C_1 - C_2^\top W C_2 + 2(C_2 - C_1)^\top W \mathbf{t} = 0 \\ C_1^\top \overline{W} C_1 - C_2^\top \overline{W} C_2 + 2(C_2 - C_1)^\top \overline{W} \mathbf{t} = 0. \end{cases}$$

This solution set has dimension $k - 2$ as long as $2(C_2 - C_1)^\top W$ and $2(C_2 - C_1)^\top \overline{W}$ are linearly independent. This is because in this case the nullspace for the matrix of coefficients (which stacks these two row vectors) would have dimension $k-2$. Again, since $C_1 \neq C_2$ with probability one, this happens (with probability one) if W and $\overline{W}$ are linearly independent. This is implied by $W, \overline{W}$ positive definite, $W \neq \overline{W}$ and $\|W\| = \|\overline{W}\|$. To see this, suppose instead that W and $\overline{W}$ are linearly dependent. In this case, $\overline{W} = cW$ for $c \in \mathbb{R}$. Then $\|W\| = |c|\|\overline{W}\|$, which implies that $|c| = 1$ (since $\|W\| = \|\overline{W}\|$). Given that $W \neq \overline{W}$, this means that $W = -\overline{W}$ and hence $\det(W) = -\det(\overline{W})$. Then, either W or $\overline{W}$ cannot be positive definite. Consequently, because the dimension of their intersection is smaller than the dimension of either $H^{W}(C_1, C_2)$ or $H^{\overline{W}}(C_1, C_2)$, these two sets are different.

We now show that $H^{\overline{W}}(C_1, C_2)$ and $H^{\overline{W}}(C_1', C_2')$ are parallel. Given our definition of $\mathcal{C}$ and $\mathcal{C}'$, note that

$$C_1' - C_2' = 2(C_1 - C_2).$$

Then see that

$$\begin{aligned}\mathbf{t} \in H^{\overline{W}}(C_1', C_2') &\Rightarrow C_1'^\top \overline{W} C_1' - C_2'^\top \overline{W} C_2' - 2(C_2' - C_1')^\top \overline{W}\mathbf{t} = 0 \\ &\Rightarrow \tfrac{1}{2}\left(C_1'^\top \overline{W} C_1' - C_2'^\top \overline{W} C_2'\right) - 2(C_2 - C_1)^\top \overline{W}\mathbf{t} = 0.\end{aligned} \tag{10}$$

where $H^{\overline{W}}(C_1', C_2')$ is defined as in (2). This shows that $H^{\overline{W}}(C_1', C_2')$ is a translation of the hyperplane

$$H^{\overline{W}}(C_1, C_2) = \{\mathbf{t} \in \mathbb{R}^d : \left(C_1^\top \overline{W} C_1 - C_2^\top \overline{W} C_2\right) - 2(C_2 - C_1)^\top \overline{W}\mathbf{t} = 0\}.$$

In other words, the linear equations defining the two hyperplanes differ only by a constant.

Step 3: (For $\mathcal{C} \neq \mathcal{C}'$ such that $V^W(\mathcal{C}) = V^W(\mathcal{C}')$, $\exists i$ such that $V_i^{\overline{W}}(\mathcal{C})$ is strictly contained in $V_i^{\overline{W}}(\mathcal{C}')$.) The Voronoi cell $V_i^{\overline{W}}(\mathcal{C})$ is a half-space in $\mathbb{R}^k$. It is defined by:

$$\begin{aligned}V_i^{\overline{W}}(\mathcal{C}) &= \{\mathbf{t} \in \mathbb{R}^k : d^{\overline{W}}(\mathbf{t}, C_i) \leq d^{\overline{W}}(\mathbf{t}, C_j)\} \\ &= \{\mathbf{t} \in \mathbb{R}^k : 2(C_j - C_i)^\top \overline{W}\mathbf{t} \leq C_j^\top \overline{W} C_j - C_i^\top \overline{W} C_i\}.\end{aligned}$$

Similarly, using the result (from Step 2) that $H^{\overline{W}}(C_1, C_2)$ and $H^{\overline{W}}(C_1', C_2')$ are parallel,

$$V_i^{\overline{W}}(\mathcal{C}') = \{\mathbf{t} \in \mathbb{R}^k : 2(C_j - C_i)^\top \overline{W}\mathbf{t} \leq \frac{1}{2}(C_j'^\top \overline{W} C_j' - C_i'^\top \overline{W} C_i')\}.$$

Let

$$\Delta_{ij} \equiv C_j^\top \overline{W} C_j - C_i^\top \overline{W} C_i - \frac{1}{2}\left(C_j'^\top \overline{W} C_j' - C_i'^\top \overline{W} C_i'\right)$$

for $j \neq i$. It can be seen that $\Delta_{ij} \neq 0$, except perhaps in a set of candidate profiles that has zero probability, since we assume that the distribution of candidate profiles is absolutely continuous.

If we have that $\Delta_{ij} > 0$,

$$2(C_j - C_i)^\top \overline{W}\mathbf{t} < \frac{1}{2}(C_j'^\top \overline{W} C_j' - C_i'^\top \overline{W} C_i') \Rightarrow 2(C_j - C_i)^\top \overline{W}\mathbf{t} < C_j^\top \overline{W} C_j - C_i^\top \overline{W} C_i$$

and $V_i^{\overline{W}}(\mathcal{C}') \subset V_i^{\overline{W}}(\mathcal{C})$. Furthermore, because the inequality is strict, $\text{int}(V_i^{\overline{W}}(\mathcal{C}) \backslash V_i^{\overline{W}}(\mathcal{C}')) \neq \emptyset$ (where for any set B, $\text{int}(B)$ denotes the interior of that set). If $\Delta_{ij} < 0$, the inclusion is reversed.

Step 4: ($\mathbb{P}_{\overline{T}}(\mathbb{R}^k) = 0$, leading to a contradiction.) From the previous steps, given $\mathcal{C}$, we can generate $\mathcal{C}' \neq \mathcal{C}$ such that $V^W(\mathcal{C}) = V^W(\mathcal{C}')$, and $V_i^{\overline{W}}(\mathcal{C})$ is strictly contained in $V_i^{\overline{W}}(\mathcal{C}')$ for some i. (Notice that this can be done for any $\mathcal{C}$, except perhaps on a set of Lebesgue measure zero.) Take an arbitrary vector $\mathbf{t}^\triangledown \in \text{int}(V_i^{\overline{W}}(\mathcal{C}') \backslash V_i^{\overline{W}}(\mathcal{C}))$. Then, for any $\mathbf{t} \in \mathbb{R}^k$, let $\mathcal{C}_{\mathbf{t}-\mathbf{t}^\triangledown}$ denote the candidate profile where each candidate position in the original candidate profile is translated by $\mathbf{t} - \mathbf{t}^\triangledown$, i.e. $\mathcal{C}_{\mathbf{t}-\mathbf{t}^\triangledown} = (C_i + \mathbf{t} - \mathbf{t}^\triangledown)_{i=1,\ldots,n}$. Because $C_i' = 2C_i - P$ (see Step 1), each component in the candidate profile $\mathcal{C}'$ will also be translated by the same vector $\mathbf{t} - \mathbf{t}^\triangledown$. Accordingly, denote the translated profile by $\mathcal{C}'_{\mathbf{t}-\mathbf{t}^\triangledown}$. It can then be established that $\mathbf{t} \in \text{int}(V_i^{\overline{W}}(\mathcal{C}'_{\mathbf{t}-\mathbf{t}^\triangledown}) \backslash V_i^{\overline{W}}(\mathcal{C}_{\mathbf{t}-\mathbf{t}^\triangledown}))$.

Now, note that because $\mathbb{Q}^k$, the k-Cartesian product of the set of rational numbers $\mathbb{Q}$, is dense in $\mathbb{R}^k$, we have that $\cup_{\mathbf{t}\in\mathbb{Q}^k} \text{int}(V_i^{\overline{W}}(\mathcal{C}'_{\mathbf{t}-\mathbf{t}^\triangledown}) \backslash V_i^{\overline{W}}(\mathcal{C}_{\mathbf{t}-\mathbf{t}^\triangledown})) = \mathbb{R}^k$ (i.e., this is a countable cover of $\mathbb{R}^k$). (Because $\mathbb{R}^k$ is a separable metric space and consequently second-countable, it can be covered by a countable family of bounded, open sets.)

Since $(\mathbb{P}_T, W)$ and $(\mathbb{P}_{\overline{T}}, \overline{W})$ are observationally equivalent, for (almost) every candidate profile

$$p(\mathcal{C}; \mathbb{P}_T, W) = p(\mathcal{C}; \mathbb{P}_{\overline{T}}, \overline{W}),$$

where $p(\cdot; \mathbb{P}_T, W)$ is the vector of shares that each candidate gets under $(\mathbb{P}_T, W)$. Consider one of the translated profiles $\mathcal{C}_{\mathbf{t}-\mathbf{t}^\triangledown}$. For this profile, let $p_{\mathbf{t}-\mathbf{t}^\triangledown}$ denote the proportion of votes obtained by candidate $C_i + \mathbf{t} - \mathbf{t}^\triangledown$:

$$p_{\mathbf{t}-\mathbf{t}^\triangledown} = \mathbb{P}_T(V_i^W(\mathcal{C}_{\mathbf{t}-\mathbf{t}^\triangledown})) = \mathbb{P}_{\overline{T}}(V_i^{\overline{W}}(\mathcal{C}_{\mathbf{t}-\mathbf{t}^\triangledown})),$$

where the second equality follows from the assumption of observational equivalence.

Then consider $\mathcal{C}'_{\mathbf{t}-\mathbf{t}^\triangledown}$. Notice that the Voronoi tessellations generated by $\mathcal{C}_{\mathbf{t}-\mathbf{t}^\triangledown}$ and $\mathcal{C}'_{\mathbf{t}-\mathbf{t}^\triangledown}$ are translations of the Voronoi tessellations generated by $\mathcal{C}$ and $\mathcal{C}'$, respectively. Because $V^W(\mathcal{C}) = V^W(\mathcal{C}')$, we then have that $V^W(\mathcal{C}_{\mathbf{t}-\mathbf{t}^\triangledown}) = V^W(\mathcal{C}'_{\mathbf{t}-\mathbf{t}^\triangledown})$ and the proportion of votes obtained by candidate $C'_i + \mathbf{t} - \mathbf{t}^\triangledown$ under W is also $p_{\mathbf{t}-\mathbf{t}^\triangledown}$:

$$p_{\mathbf{t}-\mathbf{t}^\triangledown} = \mathbb{P}_T(V_i^W(\mathcal{C}'_{\mathbf{t}-\mathbf{t}^\triangledown})).$$

Since (almost) every candidate profile generates observationally equivalent outcomes under $(\mathbb{P}_T, W)$ and $(\mathbb{P}_{\overline{T}}, \overline{W})$, we can assume that this is also the case for almost every profile $\mathcal{C}'$ generated from a profile $\mathcal{C}$ according to Step 1. If that is not the case, there is a set of $\mathcal{C}$ with positive measure that leads to $\mathcal{C}'$ which are not observationally equivalent under $(\mathbb{P}_T, W)$ and $(\mathbb{P}_{\overline{T}}, \overline{W})$. Because this set of $\mathcal{C}'$ candidate profiles has positive measure and the outcomes under $(\mathbb{P}_T, W)$ and $(\mathbb{P}_{\overline{T}}, \overline{W})$ are distinct, we would attain identification.

Otherwise, if the outcomes for $\mathcal{C}'_{\mathbf{t}-\mathbf{t}^\triangledown}$ are observationally equivalent under $(\mathbb{P}_T, W)$ and $(\mathbb{P}_{\overline{T}}, \overline{W})$, it is then the case that

$$\mathbb{P}_{\overline{T}}(V_i^{\overline{W}}(\mathcal{C}'_{\mathbf{t}-\mathbf{t}^\triangledown})) = p_{\mathbf{t}-\mathbf{t}^\triangledown}.$$

Furthermore, note that

$$\begin{aligned} 0 &= \mathbb{P}_{\overline{T}}(V_i^{\overline{W}}(\mathcal{C}'_{\mathbf{t}-\mathbf{t}^\triangledown})) - \mathbb{P}_{\overline{T}}(V_i^{\overline{W}}(\mathcal{C}_{\mathbf{t}-\mathbf{t}^\triangledown})) = \\ &= \mathbb{P}_{\overline{T}}(V_i^{\overline{W}}(\mathcal{C}'_{\mathbf{t}-\mathbf{t}^\triangledown}) \backslash V_i^{\overline{W}}(\mathcal{C}_{\mathbf{t}-\mathbf{t}^\triangledown})). \end{aligned}$$

The second equality follows from the fact that $V_i^{\overline{W}}(\mathcal{C})$ is a strict subset of $V_i^{\overline{W}}(\mathcal{C}')$. But since

$$\cup_{\mathbf{t}\in\mathbb{Q}^k}\text{int}(V_i^{\overline{W}}(\mathcal{C}'_{\mathbf{t}-\mathbf{t}^\triangledown})\backslash V_i^{\overline{W}}(\mathcal{C}_{\mathbf{t}-\mathbf{t}^\triangledown})) = \mathbb{R}^k,$$

countable subadditivity implies that

$$\mathbb{P}_{\overline{T}}(\mathbb{R}^k) \leq \sum_{\mathbf{t}\in\mathbb{Q}^k} \mathbb{P}_{\overline{T}}(\text{int}(V_i^{\overline{W}}(\mathcal{C}'_{\mathbf{t}-\mathbf{t}^\triangledown})\backslash V_i^{\overline{W}}(\mathcal{C}_{\mathbf{t}-\mathbf{t}^\triangledown}))) = 0.$$

This implies that $\mathbb{P}_{\overline{T}}(\mathbb{R}^k) = 0$, a contradiction. ■

Proof of Theorem 1

The argument follows along the lines of Step 2 in Lemma 1. Lemma 2 demonstrates identification for two candidate profiles. Focussing on elections where candidates are concentrated in a vicinity of two positions in the ideological space delivers identification by continuity. ■

Proof of Proposition 1.

We first show that $plim_S(\hat{W}_S, \hat{f}_S) = (\hat{W}, \hat{f})$. This can be established by showing that $\rho_{iS}(\cdot,\cdot)$ converges in probability to $\rho_i(\cdot,\cdot)$ uniformly over the parameter space (i.e., Θ and the set of coefficient vectors characterising f). Note first that, given $\mathbf{X} = \mathbf{x}$, $C_1, \ldots, C_n$,

$$g(\mathbf{z}_s) = \frac{1}{det(R)} \frac{\left[\sum_{|\alpha|=0}^{J_t} a_\alpha(\mathbf{x})\mathbf{z}_s^\alpha\right]^2}{\int \left[\sum_{|\alpha|=0}^{J_x} a_\alpha(\mathbf{x})\mathbf{U}^\alpha\right]^2 \phi(\mathbf{U})d\mathbf{U}} \times 1\left[d^W(\mathbf{t}_s, C_i) \leq d^W(\mathbf{t}_s, C_j), j \neq i\right],$$

with $\mathbf{t}_s = b + A\mathbf{x} + R\mathbf{z}_s$ is Euclidean as defined in Pakes and Pollard (1989). This is because the first factor is essentially a polynomial in $\mathbf{z}_s$ with bounded coefficients and the second factor is an indicator for $\mathbf{z}_s$ belonging to a Voronoi cell, which is an intersection of halfspaces.

Both are Euclidean classes (see Example 2.9 in Pakes and Pollard (1989) for the first factor and Lemmas 2.4, 2.5 and the discussion before Lemma 2.8 also in Pakes and Pollard (1989) for the second factor). Finally, the product of two functions in Euclidean classes forms an Euclidean class (Lemma 2.14). Lemma 2.8 in Pakes and Pollard (1989) then shows that $|\rho_{iS}(W, f) = \rho_i(W, f)|$ converges almost surely to zero uniformly in the parameters.

We then show consistency of $(\hat{W}, \hat{f})(\equiv plim_S(\hat{W}_S, \hat{f}_S))$. The result follows from an adaptation of the consistency result in Lemma 3.1 of Ai and Chen (2003) (which in turn uses Theorem 4.1 and Lemma A1 from Newey and Powell (2003)).

Seven assumptions are employed by Ai and Chen (2003) in demonstrating consistency. Our Assumptions 3-6 directly reproduce assumptions 3.1, 3.2, 3.4 and 3.7 in Lemma 3.1 in Ai and Chen (2003). Assumption 3.3 in Ai and Chen (2003) is an identification assumption that is attained from the identification results in Theorem 1. Theorem 2 in Gallant and Nychka (1987) says that $\cup_{E=1}^{\infty}\mathcal{H}_E$ is dense in (the closure) of $\mathcal{H}$. This corresponds to Assumption 3.5(ii) in Ai and Chen (2003). The compactness of Θ with respect to the topology induced by the Frobenius norm and the compactness of (the closure of) $\mathcal{H}$ with respect to the topology induced by the consistency norm (which follows from Theorem 1 in Gallant and Nychka (1987)) imply that the product space is also compact (with respect to the product topology) by Tychonoff's Theorem. This delivers Assumption 3.5(i) in Lemma 3.1 from Ai and Chen (2003).

Given compactness, pointwise convergence can be established easily given Assumptions 3.1-3.5, 3.7 in Lemma 3.1 from Ai and Chen (2003). Assumption 3.6 in that paper is then used to establish the uniform convergence of the objective function, which corresponds to condition (ii) from Newey and Powell (2003). Once this is done, Ai and Chen apply Lemma A1 from Newey and Powell (2003) to obtain consistency. Instead of appealing to Holder continuity (as in Assumption 3.6 from Ai and Chen (2003)), here we use alternative results to show that the objective function is stochastically equicontinuous and hence converges uniformly (see Theorem 2.1 in Newey (1991)). This can be obtained once we show stochastic equicontinuity of

$$g_E(f,W) = \frac{1}{E}\sum_{e=1}^{E}\left(\rho_i(p_e,\mathcal{C}_e,\mathbf{X}_e,W,f)^2\right)_{i=1,\ldots,n-1} = \frac{1}{E}\sum_{e=1}^{E}\left(\rho_{i,e}(W,f)^2\right)_{i=1,\ldots,n-1}.$$

We let $\rho_{i,e}(W,f) \equiv \rho_i(p_e,\mathcal{C}_e,\mathbf{X}_e,W,f)$ and $\rho_i = [\rho_{i,1},\ldots,\rho_{i,E}]^\top$. To obtain stochastic equicontinuity, notice that the $E\times(n-1)$ matrix of estimates

$$\widehat{M} = B(B^\top B)^{-1}B^\top\rho(W,f) = P\rho(W,f),$$

where ρ is an $E\times(n-1)$ matrix stacking $\left(\int \mathbf{1}_{\mathbf{t}\in V_i^W(\mathcal{C})} f(\mathbf{t})d\mathbf{t} - p_i\right)^\top_{i=1,\ldots,n-1}$ for all observations and P is an $E\times E$ idempotent matrix with rank (= trace) at most J. Since we have Assumption 5, we can assume without loss of generality that $\widehat{\Sigma}(\mathbf{X}_e,\mathcal{C}) = I$. This in turn implies an objective function equal to

$$Q_n(W,f) \equiv \frac{1}{E}\sum_{e=1}^{E}||\widehat{m}(\mathbf{X}_e,\mathcal{C}_e,(W,f))||^2 = \frac{1}{E}\mathrm{tr}\left(\widehat{M}^\top\widehat{M}\right) = \frac{1}{E}\mathrm{tr}\left(\rho^\top P^\top P\rho\right),$$

which in turn delivers

$$\begin{aligned} |Q_n(W_1,f_1) - Q_n(W_2,f_2)| &= \left|\textstyle\sum_{i=1}^{n-1}\left(\frac{1}{E}||P\rho_i(W_1,f_1)||^2 - \frac{1}{E}||P\rho_i(W_2,f_2)||^2\right)\right| \\ &\le \textstyle\sum_{i=1}^{n-1}\left|\frac{1}{E}||P\rho_i(W_1,f_1)||^2 - \frac{1}{E}||P\rho_i(W_2,f_2)||^2\right|, \end{aligned} \tag{11}$$

where $\|\cdot\|$ is the usual Euclidean norm. Because, for any vectors A and B and positive scalar c,

$$\left|\frac{||A||}{\sqrt{c}} - \frac{||B||}{\sqrt{c}}\right| \le \frac{||A-B||}{\sqrt{c}} \Rightarrow \left|\frac{||A||^2}{c} - \frac{||B||^2}{c}\right| \le \frac{||A-B||(||A||+||B||)}{c},$$

each of the terms in the sum in expression (11) is bounded by

$$\left|\frac{1}{E}||P\left(\rho_i(W_1,f_1)-\rho_i(W_2,f_2)\right)||\left(||P\rho_i(W_1,f_1)||+||P\rho_i(W_2,f_2)||\right)\right| \leq \\ \left|\frac{1}{E}||\rho_i(W_1,f_1)-\rho_i(W_2,f_2)||\left(||\rho_i(W_1,f_1)||+||\rho_i(W_2,f_2)||\right)\right|,$$

where the inequality follows because P is idempotent and consequently $||Pa|| \leq ||a||$ for conformable a (see the proof for Corollary 4.2 in Newey (1991)). Now, since

$$||\rho_i(W,f)||^2 = \sum_{e=1}^{E} \rho_{i,e}(W,f)^2 \leq 4E,$$

we have

$$\left|\frac{1}{E}||\rho_i(W_1,f_1)-\rho_i(W_2,f_2)||\left(||\rho_i(W_1,f_1)||+||\rho_i(W_2,f_2)||\right)\right| \leq \\ \left|4\sqrt{\frac{||\rho_i(W_1,f_1)-\rho_i(W_2,f_2)||^2}{E}}\right|$$

This in turn gives

$$\sup_{(W_1,f_1)\in\Theta\times\mathcal{H}} \sup_{(W_2,f_2)\in\mathcal{N}((W_1,f_1),\delta)} |Q_n(W_1,f_1)-Q_n(W_2,f_2)| \\ \leq \sum_{i=1}^{n-1} \sup_{(W_1,f_1)\in\Theta\times\mathcal{H}} \sup_{(W_2,f_2)\in\mathcal{N}((W_1,f_1),\delta)} \left|4\sqrt{\frac{||\rho_i(W_1,f_1)-\rho_i(W_2,f_2)||^2}{E}}\right|,$$

where $\mathcal{N}((W_1,f_1),\delta)$ is a ball of radius δ centered at (W_1,f_1). These imply that

$$Prob\left(\sup_{(W_1,f_1)\in\Theta\times\mathcal{H}} \sup_{(W_2,f_2)\in\mathcal{N}((W_1,f_1),\delta)} |Q_n(W_1,f_1)-Q_n(W_2,f_2)| > \epsilon\right) \\ \leq \sum_{i=1}^{n-1} Prob\left(\sup_{(W_1,f_1)\in\Theta\times\mathcal{H}} \sup_{(W_2,f_2)\in\mathcal{N}((W_1,f_1),\delta)} \left|4\sqrt{\frac{||\rho_i(W_1,f_1)-\rho_i(W_2,f_2)||^2}{E}}\right| > \frac{\epsilon}{n-1}\right) \\ = \sum_{i=1}^{n-1} Prob\left(\sup_{(W_1,f_1)\in\Theta\times\mathcal{H}} \sup_{(W_2,f_2)\in\mathcal{N}((W_1,f_1),\delta)} \frac{||\rho_i(W_1,f_1)-\rho_i(W_2,f_2)||^2}{E} > \frac{\epsilon^2}{16(n-1)^2}\right).$$

Consequently, if we show for each $i = 1, \ldots, n-1$ that

$$\lim_{\delta \to 0} \lim_{n \to \infty} Prob \left(\sup_{(W_1, f_1) \in \Theta \times \mathcal{H}} \sup_{(W_2, f_2) \in \mathcal{N}((W_1, f_1), \delta)} \frac{||\rho_i(W_1, f_1) - \rho_i(W_2, f_2)||^2}{E} > \epsilon \right) = 0$$

for any $\epsilon > 0$, we obtain stochastic equicontinuity of the objective function:

$$\lim_{\delta \to 0} \lim_{n \to \infty} Prob \left(\sup_{(W_1, f_1) \in \Theta \times \mathcal{H}} \sup_{(W_2, f_2) \in \mathcal{N}((W_1, f_1), \delta)} |Q_n(W_1, f_1) - Q_n(W_2, f_2)| > \epsilon \right) = 0$$

for any $\epsilon > 0$.

Let then

$$Y_{e\delta} = \sup_{(W_1, f_1) \in \Theta \times \mathcal{H}} \sup_{(W_2, f_2) \in \mathcal{N}((W_1, f_1), \delta)} \left(\rho_{i,e}(W_1, f_1) - \rho_{i,e}(W_2, f_2)\right)^2$$

(for $i \in \{1, \ldots, n-1\}$) and notice that

$$\sup_{(W_1, f_1) \in \Theta \times \mathcal{H}} \sup_{(W_2, f_2) \in \mathcal{N}((W_1, f_1), \delta)} \frac{||\rho_i(W_1, f_1) - \rho_i(W_2, f_2)||^2}{E} = \frac{1}{E} \sum_{e=1}^{E} Y_{e\delta}.$$

To show stochastic equicontinuity we adapt the proof of Lemma 3 in Andrews (1992). Consider $\epsilon > 0$ and take $M > 4$ and $\delta > 0$ such that $Prob\,(Y_{e\delta} > \epsilon^2/2) < \epsilon^2/(2M)$. That such a δ can be chosen follows because of compactness of $\Theta \times \mathcal{H}$ and continuity of $\rho_i(\cdot, \cdot)$. (This corresponds to Assumption TSE-1D in Andrews (1992).) For such a δ,

$$\begin{aligned} \lim_{E \to \infty} Prob \left(\frac{1}{E} \sum_{e=1}^{E} Y_{e\delta} > \epsilon \right) &\leq \lim_{E \to \infty} \frac{1}{\epsilon} \mathbb{E} \left(\frac{1}{E} \sum_{e=1}^{E} Y_{e\delta} \right) = \frac{1}{\epsilon} \mathbb{E}(Y_{e\delta}) \\ = \frac{1}{\epsilon} \left[\mathbb{E} \left(Y_{e\delta} \mathbf{1} \left(Y_{e\delta} \leq \frac{\epsilon^2}{2} \right) \right) + \mathbb{E} \left(Y_{e\delta} \mathbf{1} \left(\frac{\epsilon^2}{2} < Y_{e\delta} \leq M \right) \right) \right. &\left. + \mathbb{E} \left(Y_{e\delta} \mathbf{1} \left(Y_{e\delta} > M \right) \right) \right] \\ &\leq \frac{1}{\epsilon} \left(\frac{\epsilon^2}{2} + M Prob(Y_{e\delta} > \frac{\epsilon^2}{2}) \right) \leq \epsilon \end{aligned}$$

The first inequality follows from Markov's Inequality. The following equality holds since observations are i.i.d. The second inequality follows because $\mathbb{E} \left(Y_{e\delta} \mathbf{1} \left(Y_{e\delta} \leq \frac{\epsilon^2}{2} \right) \right) \leq \frac{\epsilon^2}{2}$,

$\mathbb{E}\left(Y_{e\delta}\mathbf{1}\left(\frac{\epsilon^2}{2} < Y_{e\delta} \leq M\right)\right) \leq MProb(Y_{e\delta} > \frac{\epsilon^2}{2})$ and, finally, $Prob(Y_{e\delta} > M) = 0$ since $M > 4$. The last inequality then stems from $Prob\left(Y_{e\delta} > \epsilon^2/2\right) < \epsilon^2/(2M)$. Since this argument can be repeated for $i = 1, \ldots, n-1$, we have stochastic equicontinuity. ■

Proof of Theorem 2

The result follows along the lines of Theorem 1 and here we elaborate on the necessary modifications to the intermediate steps in establishing the statement. The alterations take into consideration the fact that the set of voters for candidate i are now given by

$$V_i^W(\mathcal{C}, \delta) \equiv \{\mathbf{t} \in Y : d^W(\mathbf{t}, C_i)^2 + \delta_i \leq d^W(\mathbf{t}, C_j)^2 + \delta_j, j \neq i\}.$$

Lemma 1 ($W = \mathbf{I}$) follows with minor alterations. In Step 1, with $n = 2$, the scalar b should incorporate the valence terms and is now equal to $C_2^\top C_2 - C_1^\top C_1 + \delta_2 - \delta_1$. The vector $\mathbf{A}$ is unaltered. The same argument delivers an election profile $(C_i^*, \delta_i^*)_{i=1,2}$ for which the two voter distributions $\mathbb{P}_{T_1}$ and $\mathbb{P}_{T_2}$ produce different voting proportions. Step 2 then uses the fact that voting proportions are continuous in $\mathcal{C}^*$ to demonstrate that elections with $n > 2$ where candidates are situated in η-neighborhoods of C_1^* and C_2^* produce different voting proportions under each voter type distribution. The argument can be easily adapted using now (η-)neighborhoods around δ_1^* and δ_2^* as well.

Lemma 2 then assumes that $W \neq \overline{W}$ and $n = 2$ to show that $(\mathbb{P}_T, W)$ is identified. The first step in the proof shows that there is more than one set of candidates that generates the same partition of voters for a given weighting matrix W. Given two candidates and their valence terms, the set of voters $\mathbf{t}$ who are equidistant from both candidates is given by

$$H^W(C_1, C_2, \delta_1, \delta_2) \equiv \{\mathbf{t} \in \mathbb{R}^k : 2(C_1 - C_2)^\top W\mathbf{t} = (C_1^\top WC_1 - C_2^\top WC_2) + \delta_1 - \delta_2\}.$$

Consider P in this set such that $P \neq aC_1 + (1-a)C_2$ and such that $2(C_1 - C_2)^\top \overline{W}(P -$

$aC_1 - (1-a)C_2) \neq 0$, where

$$a = \frac{(C_2 - C_1)^\top W (C_2 - C_1) + \delta_1 - \delta_2}{2(C_2 - C_1)^\top W (C_2 - C_1)}.$$

(Note that $aC_1 + (1-a)C_2$ also pertains to $H^W(C_1, C_2, \delta_1, \delta_2)$.) The requirement that $2(C_1 - C_2)^\top \overline{W}(P - aC_1 - (1-a)C_2) \neq 0$ is important for Step 3. The set of vectors P satisfying such restrictions is nonempty. To see this, remember that $2(C_1 - C_2)^\top W$ and $2(C_1 - C_2)^\top \overline{W}$ are linearly independent (with probability one) (see Step 2 in Lemma 2). Hence, the set $H^W(C_1, C_2, \delta_1, \delta_2)$ has dimension $k-1$ and its intersection with $\{\mathbf{t} \in \mathbb{R}^k : 2(C_1 - C_2)^\top \overline{W}(\mathbf{t} - aC_1 - (1-a)C_2) = 0\}$ forms a system of 2 equations in k unknowns and has dimension $k-2$. Then, consider

$$C_i' = C_i + (P - aC_1 - (1-a)C_2)$$

and $\delta_i' = \delta_i, i = 1, 2$. It is immediate to obtain that $C_1' - C_2' = C_1 - C_2$. Furthermore, one gets

$$C_1'^\top W C_1' - C_2'^\top W C_2' = C_1^\top W C_1 - C_2^\top W C_2 + 2(C_1 - C_2)^\top W (P - aC_1 - (1-a)C_2).$$

Since both P and $aC_1 + (1-a)C_2$ belong to $H^W(C_1, C_2, \delta_1, \delta_2)$, the last term in the right-hand side is zero. Consequently, $C_1'^\top W C_1' - C_2'^\top W C_2' = C_1^\top W C_1 - C_2^\top W C_2$. This in turn implies that

$$2(C_1 - C_2)^\top W \mathbf{t} = (C_1^\top W C_1 - C_2^\top W C_2) + \delta_1 - \delta_2$$

$$\Leftrightarrow$$

$$2(C_1' - C_2')^\top W \mathbf{t} = (C_1'^\top W C_1' - C_2'^\top W C_2') + \delta_1' - \delta_2',$$

which establishes the first step in the Lemma. Upon redefining $H^W(C_1, C_2)$ and $H^{\overline{W}}(C_1, C_2)$

in Step 2 to accomodate the valence terms, this step is also straightforward. For Step 3, though, it is relevant to point out that, since $C_1' - C_2' = C_1 - C_2$, the last line in (10) now equals $C_1'^{\top}\overline{W}C_1' - C_2'^{\top}\overline{W}C_2' + \delta_1' - \delta_2' - 2(C_2 - C_1)^{\top}\overline{W}\mathbf{t} = 0$. Since $\delta_i = \delta_i', i = 1, 2$, this implies that Δ_{ij} is now given by $C_1^{\top}\overline{W}C_1 - C_2^{\top}\overline{W}C_2 - (C_1'^{\top}\overline{W}C_1' - C_2'^{\top}\overline{W}C_2')$, which using the definition of $C_i', i = 1, 2$ equals $2(C_1 - C_2)^{\top}\overline{W}(P - aC_1 - (1-a)C_2)$. Given the choice of P, this last quantity is nonzero and Step 3 follows to demonstrate that for these two elections, $\exists i$ such that $V_i^{\overline{W}}(\mathcal{C}, \delta)$ is strictly contained in $V_i^{\overline{W}}(\mathcal{C}', \delta')$. Step 4 can then be carried out with the obvious notational modifications to incorporate δ. Using a continuity argument as in Theorem 1, we then obtain the result for $n > 2$. ■

Online Appendix: Monte Carlo Experiments

In this Appendix, we examine the small sample performance of the suggested estimation strategy in a few Monte Carlo experiments. We investigate models without covariates with three potential distribution of voter types. We use the distributions suggested by Ichimura and Thompson (1998) and summarized in Table 6 and Figure 10. For each of these, we postulate two different weighting matrices W for the weighted distance function. The first one has $W_{1,2} = W_{2,1} = 0$ and $W_{2,2} = 2$, and the second $W_{1,2} = W_{2,1} = 0.5$ and $W_{2,2} = 2$. Both matrices are normalized to have $W_{1,1} = 1$. We assume that the analysis has 100 observations in each set of Monte Carlo experiments.[46] Each observation contains the position and vote proportions for 2 candidates that are sampled uniformly over $[-1, 1]^2$. The proportions are estimated using (1000) draws from the voter type distribution in the data generating process. This introduces sampling error in the observed proportion of votes (i.e., an electoral precinct level ϵ) which differ in general from the numerical integration of the proposed type distribution over the candidate's Voronoi cell. We use 50 Monte Carlo repetitions for each one of the three models.

Table 6: Data Generating Processes

Model 1:	$\mathbf{T} \sim \mathcal{N}\left([0,0]', \mathbf{I}_2\right)$
Model 2:	$\mathbf{T}$ is an equiprobable mixture of
	$\mathbf{T}_a \sim \mathcal{N}\left(\begin{bmatrix} \mu \\ -\mu \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}\right)$
	and
	$\mathbf{T}_b \sim \mathcal{N}\left(\begin{bmatrix} -\mu \\ \mu \end{bmatrix}, \begin{bmatrix} \sigma_2^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_1^2 \end{bmatrix}\right)$
	$\mu = 0.3587, \sigma_1^2 = 0.2627, \sigma_2^2 = 0.06568, \rho = -0.1$

[46] This sample size is much smaller than in actual datasets (e.g., in our empirical illustration we use between 270 and 693 elections) and should depict the usefulness of the methodology even in relatively data-scarce scenarios. Of course, performance will improve in larger datasets.

Table 6: Data Generating Processes (Continued)

Model 3:	$\mathbf{T} = (T_1, T_2)^\top$ with T_1 and T_2 independently distributed
	$T_1 \sim \mathcal{N}(0, \sigma^2)$
	T_2 an equally weighted mixture of T_a and T_b
	$T_a \sim \mathcal{N}(0.2806, \sigma^2), T_1 \sim \mathcal{N}(-1.6806, \sigma^2)$
	$\sigma^2 = 0.038462$

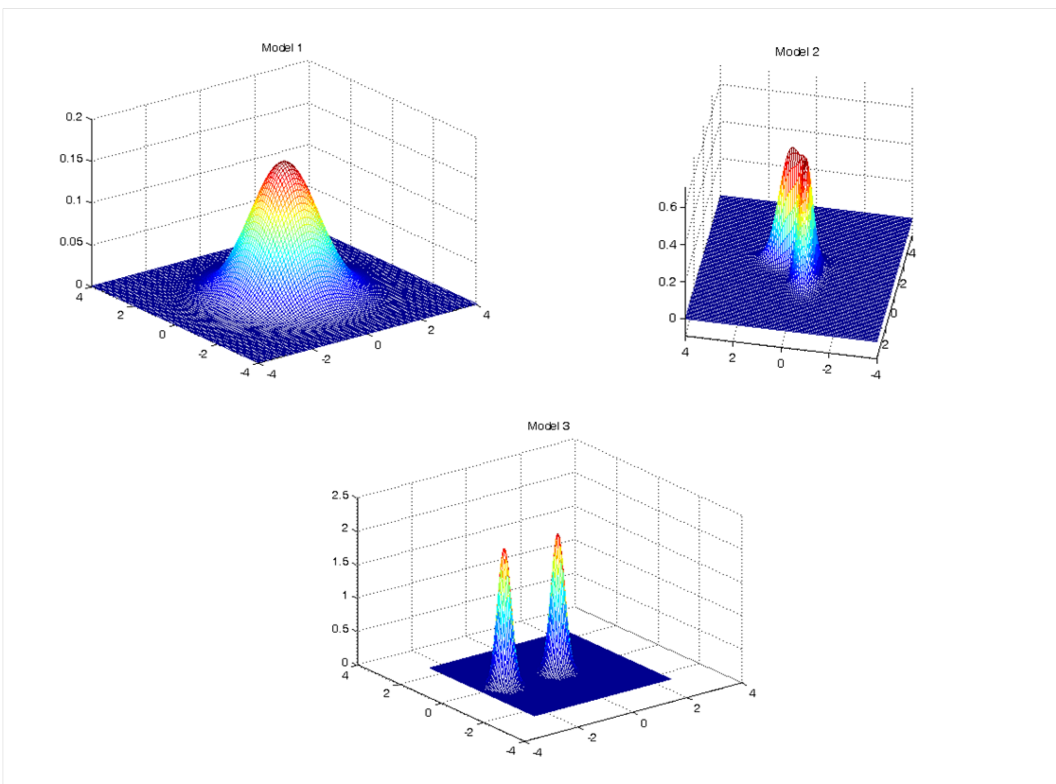

Figure 10: DGP Densities

The estimation follows the guidelines prescribed in the previous section. For the estimation of $m(\cdot)$ we use linear splines (with cross-products) for Models 1 and 2 and simple linear projections for Model 3. The estimation weighting matrix ($\tilde{\Sigma}$) is the identity. In Tables 7, 8 and 9, we report squared bias, variance and MSE for the two parameters in the W matrix for each of the three models. We follow Blundell, Chen, and Kristensen (2007) in reporting similar quantities for the density estimates. Letting $\hat{f}_i$ be the estimate of f from the ith Monte Carlo simulation and letting $\overline{f}(\mathbf{t}) = \sum_{i=1}^{MC} \hat{f}_i(\mathbf{t})/MC$. The pointwise squared bias is then defined as $\left(\overline{f}(\mathbf{t}) - f(\mathbf{t})\right)^2$ and the pointwise variance is $\sum_{i=1}^{MC} \left(\hat{f}_i(\mathbf{t}) - \overline{f}_i(\mathbf{t})\right)^2 /MC$. We report squared bias, variance and MSE integrated over a grid of 100×100 points.

Table 7: Monte Carlo Results: Model 1

Bias^2	Variance	MSE	J_t
$(W_{1,2}, W_{2,2}) = (0, 2)$			
$(0.0001, 0.0141, 4.4566 \times 10^{-5})$	$(0.0011, 0.0408, 2.2451 \times 10^{-5})$	$(0.0012, 0.0549, 6.7017 \times 10^{-5})$	1
$(0.2781, 0.3837, 4.8133) \times 10^{-5}$	$(0.0005, 0.0093, 3.8467 \times 10^{-5})$	$(0.0005, 0.0093, 5.1980 \times 10^{-5})$	2
$(W_{1,2}, W_{2,2}) = (0.5, 2)$			
$(0.0008, 0.0230, 6.4572 \times 10^{-5})$	$(0.0034, 0.0477, 4.6148 \times 10^{-5})$	$(0.0042, 0.0707, 1.1072 \times 10^{-4})$	1
$(0.0000, 0.0011, 4.0107 \times 10^{-5})$	$(0.0008, 0.0089, 5.6757 \times 10^{-4})$	$(0.0008, 0.0100, 4.5782 \times 10^{-5})$	2

The three arguments correspond to $W_{1,2}$, $W_{2,2}$ and the integrated quantities for the density as described in the text. The order refers to the Hermite polynomial order. $m(\cdot)$ is estimated using linear splines. 50 Monte Carlo repetitions for 100 elections with two candidates sampled uniformly on $[0, 1]^2$.

Table 8: Monte Carlo Results: Model 2

$(W_{1,2}, W_{2,2}) = (0, 2)$

Bias^2	Variance	MSE	J_t
$(0.0023, 0.4658, 0.0025)$	$(0.2206, 0.2212, 5.5267 \times 10^{-4})$	$(0.2228, 0.6871, 0.0031)$	1
$(0.0010, 0.0853, 0.0016)$	$(0.1215, 0.2108, 4.6720 \times 10^{-4})$	$(0.1224, 0.2961, 0.0020)$	2
$(0.0001, 0.0201, 0.0012)$	$(0.0912, 0.1316, 4.3440 \times 10^{-4})$	$(0.0913, 0.1517, 0.0016)$	3
$(0.0006, 0.0120, 9.3694 \times 10^{-4})$	$(0.0693, 0.0952, 3.9928 \times 10^{-4})$	$(0.0699, 0.1072, 0.0013)$	5
$(0.0001, 0.0088, 8.4013 \times 10^{-4})$	$(0.0556, 0.0900, 3.6408 \times 10^{-4})$	$(0.0557, 0.0988, 0.0012)$	5

$(W_{1,2}, W_{2,2}) = (0.5, 2)$

Bias^2	Variance	MSE	J_t
$(0.2346, 2.5381, 0.0042)$	$(0.2391, 0.8908, 0.0042)$	$(0.4737, 3.4289, 0.0042)$	1
$(0.2435, 2.0193, 0.0038)$	$(0.2473, 1.0363, 0.0007)$	$(0.4908, 3.0556, 0.0046)$	2
$(0.2005, 1.9740, 0.0037)$	$(0.2497, 1.0579, 8.1940 \times 10^{-4})$	$(0.2005, 3.0319, 0.0045)$	3
$(0.1958, 1.0874, 0.0036)$	$(0.2458, 1.0874, 8.2810 \times 10^{-4})$	$(0.4416, 3.0167, 0.0044)$	4
$(0.1937, 1.9216, 0.0036)$	$(0.2439, 1.0867, 8.7007 \times 10^{-4})$	$(0.0180, 0.5403, 0.0045)$	5

The three arguments correspond to $W_{1,2}$, $W_{2,2}$ and the integrated quantities for the density as described in the text. The order refers to the Hermite polynomial order. $m(\cdot)$ is estimated using linear splines. 50 Monte Carlo repetitions for 100 elections with two candidates sampled uniformly on $[0, 1]^2$.

Table 9: Monte Carlo Results: Model 3

$(W_{1,2}, W_{2,2}) = (0, 2)$

Bias^2	Variance	MSE	J_t
$(0.0015, 0.0002, 0.0274)$	$(0.0442, 0.1275, 0.0014)$	$(0.0457, 0.1277, 0.0287)$	1
$(0.0008, 0.0111, 0.0274)$	$(0.0221, 0.0633, 0.0015)$	$(0.0229, 0.0633, 0.0289)$	2
$(0.0007, 0.0164, 0.0136)$	$(0.0938, 0.0317, 0.0125)$	$(0.0946, 0.0481, 0.0260)$	3
$(0.0036, 0.0064, 0.0149)$	$(0.0775, 0.0258, 0.0136)$	$(0.0812, 0.0323, 0.0285)$	4
$(0.0008, 0.0389, 0.0073)$	$(0.0244, 0.2360, 0.0131)$	$(0.0252, 0.2749, 0.0204)$	5

$(W_{1,2}, W_{2,2}) = (0.5, 2)$

Bias^2	Variance	MSE	J_t
$(0.0021, 0.0208, 0.0279)$	$(0.0752, 0.4363, 0.0019)$	$(0.0773, 0.4571, 0.0019)$	1
$(0.0002, 0.0056, 0.0274)$	$(0.0186, 0.0226, 0.0016)$	$(0.0187, 0.0282, 0.0289)$	2
$(0.0004, 0.0561, 0.0133)$	$(0.1189, 0.1552, 0.0138)$	$(0.1193, 0.2113, 0.0271)$	3
$(0.0010, 0.0099, 0.0140)$	$(0.0880, 0.0226, 0.0139)$	$(0.0890, 0.0326, 0.0279)$	4
$(0.0001, 0.0301, 0.0071)$	$(0.0097, 0.1467, 0.0115)$	$(0.0098, 0.1768, 0.0186)$	5

The three arguments correspond to $W_{1,2}$, $W_{2,2}$ and the integrated quantities for the density as described in the text. The order refers to the Hermite polynomial order. $m(\cdot)$ is estimated using linear projections. 50 Monte Carlo repetitions for 100 elections with two candidates sampled uniformly on $[0, 1]^2$.

As expected, the estimator attains low bias and variance for relatively low orders of the Hermite polynomial in Model 1. An order 0 polynomial ($J_t = 1$) already offers good properties. Moving to an order 1 polynomial ($J_t = 2$) leads to improvements particularly for the weighting matrix parameters. For Model 2, with a diagonal weighting matrix, substantial gains are observed before one reaches an order 3 polynomial ($J_t = 4$) when incremental improvements are then minor. With a non-diagonal weighting matrix, the type distribution seems to be accurately estimated even at lower orders, but the parameters are less precisely estimated. For Model 3, even with a non-diagonal weighting matrix the estimator seems to behave well.