PIER Working Paper 13-055

"Identification and Estimation of Preference Distributions
When Voters Are Ideological**",** Second Version

by

Antonio Merlo and Aureo de Paula

http://ssrn.com/abstract=2337642

Identification and Estimation of Preference

Distributions When Voters Are Ideological [*]

Antonio Merlo [†]

University of Pennsylvania

Áureo de Paula [‡]

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

This version: October, 2013

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.

[†]Department of Economics, University of Pennsylvania, Philadelphia, PA 19104. E-mail: merloa@econ.upenn.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 some 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 $\mathcal{C} \equiv (C_1, \ldots, C_n) \in \mathbb{R}^k \times \cdots \times \mathbb{R}^k$ denote a profile of candidates in the n-fold Cartesian product of $\mathbb{R}^k$ which characterizes 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 $V_i^{W}(\mathcal{C}) \equiv \{\mathbf{t} \in Y : d^{W}(\mathbf{t}, C_i) < d^{W}(\mathbf{t}, C_j), j \neq i\}$ 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 partitions 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, $\{V_i^W(\mathcal{C})\}_{i \in \{1, ..., n\}}$, where each region $V_i^W(\mathcal{C})$ is the set of voters voting for candidate i in the election.[10] The set $\{V_i^W(\mathcal{C})\}_{i \in \{1, ..., n\}}$ 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 $\{V_i^W(\mathcal{C})\}_{i \in \{1, ..., n\}}$ 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 $\cup_{i \in \{1, ..., n\}} V_i^W(\mathcal{C}) = Y$.

[11]For a comprehensive treatment of Voronoi tessellations and their properties, see, e.g., Okabe, Boots, Sugihara, and Chiu (2000).
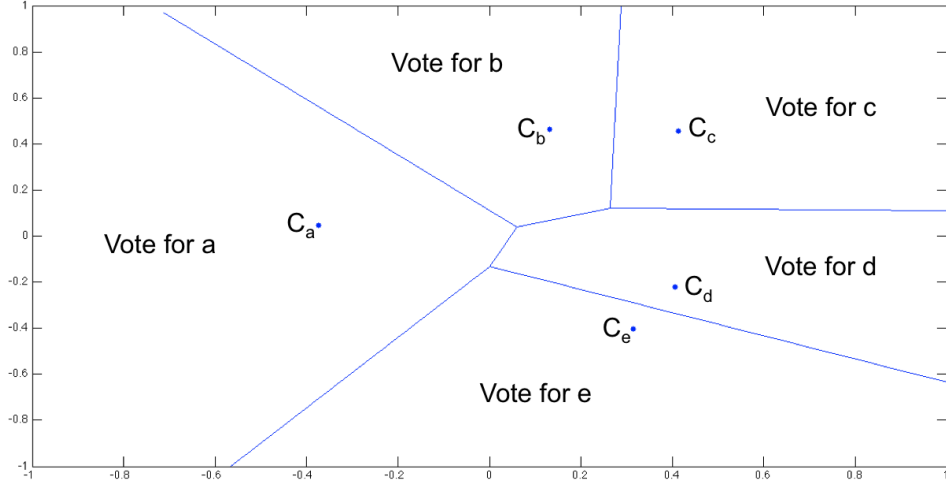
Figure 1: The Voronoi Tessellation for a 5-candidate election in $\mathbb{R}^2$ and $W = I$.

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:

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

7

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 $V_i^W(\mathcal{C}) = \{\mathbf{t} \in \mathbb{R}^k : d^W(\mathbf{t}, C^i) < d^W(\mathbf{t}, C^j), j \neq i\}$ 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 $\mathbf{X}$.*

Before presenting our identification results, 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.[13]

---

[13]In our identification strategy, we only use the expected "market shares" given the "market observable

3 Identification

The following definition qualifies our characterization of identifiability. We remind the reader that the analysis is conditional on **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.s. $\Rightarrow$ $(\mathbb{P}_{T_1}, W_1) = (\mathbb{P}_{T_2}, W_2)$.[14] $(\mathbb{P}_T, W)$ is (globally) identified if it is identified relative to any other probability measure and weighting matrix pair.*

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.

---

characteristics"as objects identified in the data and used to recover the underlying structural parameters. One important difference relative to BLP is that BLP impose that each product has a scalar unobservable product characteristic that varies by market and, in fact, recover this unobservable. No such structure is imposed on the unobservable in our framework and ϵ can have arbitrary dimension, so that the object of interest is the average distribution of tastes $\mathbb{P}_{T|X} \equiv \int \mathbb{P}_{T|X,\epsilon} \mathbb{P}_{\epsilon|X}(d\epsilon|X)$, averaged over the distribution of the unobservable.

[14]*Leb.-a.s.* refers to the fact that the underlying measure is the Lebesgue measure on $\mathbb{R}^k \times \cdots \times \mathbb{R}^k$, the n-fold Cartesian product of $\mathbb{R}^k$.

Lemma 1 *Suppose that W is known, Assumption 1 holds, and all probability measures are absolutely continuous with respect to the Lesbegue measure on $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ and defined on a common support. Then $\mathbb{P}_T$ is identified.*

The proof of Lemma 1 is given in the Appendix for elections with any number of candidates. 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. 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.[15]

The Cramér-Wold device is also used in Ichimura and Thompson (1998) (and later in Gautier and Kitamura (2013)) 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 $d^W(C_1, \mathbf{t}) - d^W(C_2, \mathbf{t}) \geq \mathbf{0}$. 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 $\mathbf{T}$. Even in the case where the weighting matrix is known, however, their arguments do not extend to the case where elections have more than two candidates, which

---

[15]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.

is the general environment considered here.[16]

We now turn attention to the general environment where the weighting matrix W is not known. The next step in our analysis considers the special case where there is an upper bound on the number of candidates related to the number of dimensions of the ideological space. The analysis of this case allows us to illustrate an important component of our identification argument, and provides another useful step for the proof of the main result of the paper contained in Theorem 1 below.

Lemma 2 establishes joint identification of the distribution of preference types in the population of voters and of the weighting matrix, when elections have no more than $k + 1$ candidates.

Lemma 2 *Suppose Assumption 1 holds, $||W||_{k \times k} = \sqrt{k}$, all probability measures are absolutely continuous with respect to the Lesbegue measure on $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ and defined on a common support and there are at most $k + 1$ candidates. Then $(\mathbb{P}_T, W)$ is identified.*

The proof of Lemma 2 is presented in the Appendix for arbitrary ideological space dimension k and elections with any number of candidates (up to $k + 1$). Here, we illustrate the identification argument for the case where the ideological space is two-dimensional and elections only have two candidates.[17]

To illustrate the argument, start with an arbitrary bounded set in $\mathbb{R}^2$, as indicated in the upper-left panel in Figure 2. Then, consider an election with candidates $\mathcal{C} = \{C_1, C_2\}$ such that this set belongs to $V_i^{\overline{W}}(\mathcal{C})$, but not to $V^W(\mathcal{C})$ 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

---

[16]Because Voronoi tessellations can also be defined on hyperspheres, and in Ichimura and Thompson (1998) and Gautier and Kitamura (2013) covariates and coefficients are both supported on a hypersphere, it may be possible to extend our methodology to study identification in discrete choice models with more than two alternatives. This generalization, however, is outside of the scope of this paper and is left for future work.

[17]As explained below, the generalization to more than two candidates is not straightforward (see, e.g., footnote 18). Also, the two-candidate case does not suffice to establish identification. In the theoretical model, the event that two or more candidates in an election share the same ideological position has probability zero. Hence, environments with more than two candidates cannot be reduced to the two-candidate scenario by "piling up" candidates in two locations of the ideological space. The possibility that three or more candidates in an election occupy at most two ideological positions also never occurs in the data.

cells when there are two candidates are separated by the line

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

and analogously for the weighted distance $d^{\overline{W}}$. 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 two systems $(\mathbb{P}_T, W)$ and $(\mathbb{P}_{\overline{T}}, \overline{W})$ are observationally equivalent, the two candidates 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 $\mathbb{P}_{\overline{W}}$.
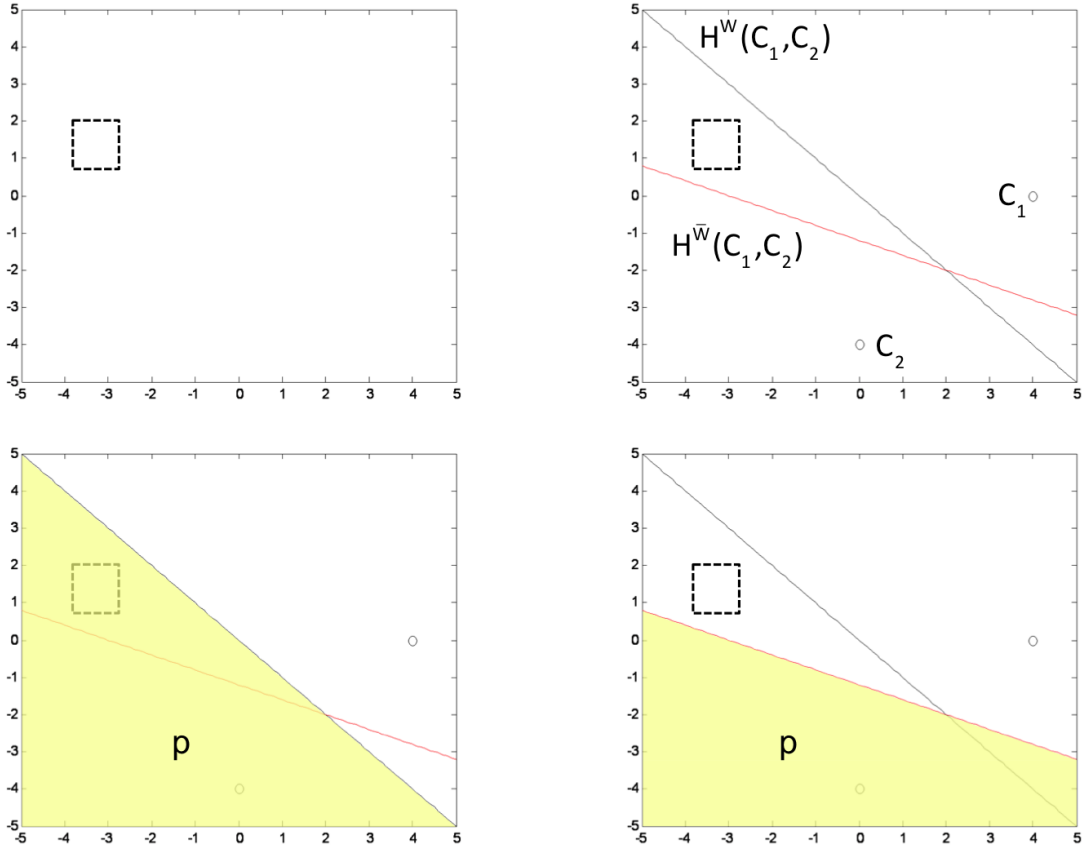

Figure 2: Voronoi Tessellations for Candidates C_1, C_2

13

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). It is also crucial for the ensuing argument that the initial set be contained in $V_i^W(\mathcal{C})$ but not in $V_i^{\overline{W}}(\mathcal{C})$ for one of the candidates in the second election. In the upper-right panel in Figure 3, this is achieved for candidate C_1'.

Candidates C_1' and C_1 obtain the same vote share, equal to $(1-p)$, in their respective elections under $(\mathbb{P}_W, W)$, 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* $H^{\overline{W}}(C_1', C_2')$ has probability $1-p$ under $\mathbb{P}_{\overline{W}}$ (see the lower-left panel in Figure 3).

Since, under $\mathbb{P}_{\overline{W}}$, 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). This would hold under either $\mathbb{P}_T$ or $\mathbb{P}_{\overline{T}}$ if they are observationally equivalent. 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 strategy illustrated above relies on the availability of multiple candidate profiles generating the same Voronoi tessellation (for weighting matrix W). The proof of Lemma 2 extends this argument for at most $k+1$ candidates and general ideological space dimension k. Note that even though this construction is clearly achieved with two candidates (as indicated in the upper-right panel of Figure 3), it is not immediate that with more than two candidates and general ideological space dimension k one can generate identical Voronoi tessellations under W for two different elections and have the initial bounded set be contained
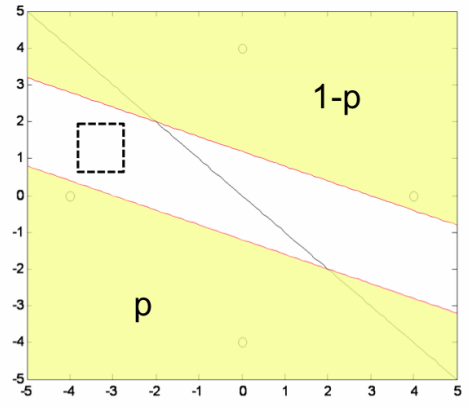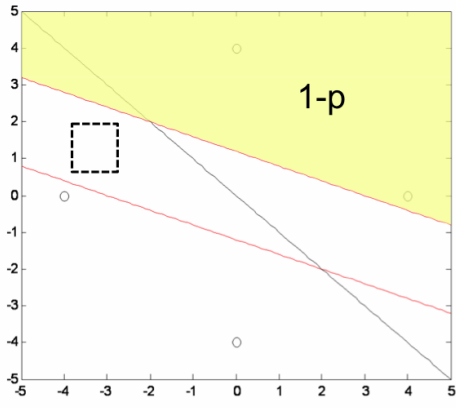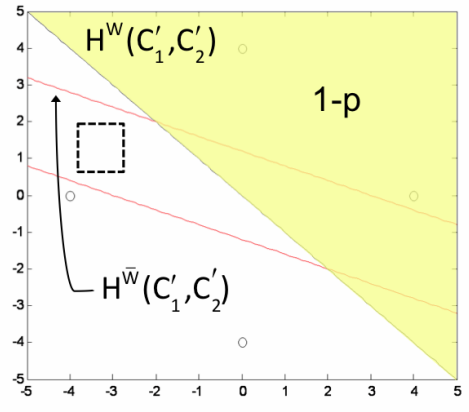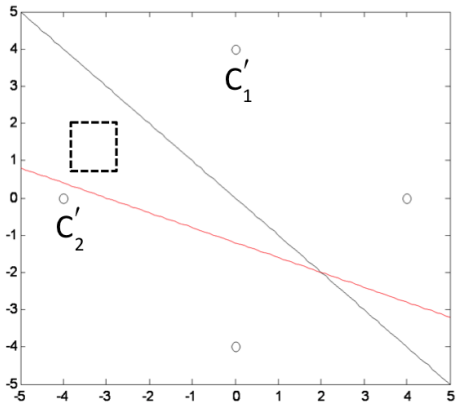
Figure 3: Voronoi Tessellations for Candidates C_1', C_2'

in the difference between Voronoi cells under W and $\overline{W}$ for one of the candidates in the second election. The proof of Lemma 2 exploits the geometric properties of Voronoi tessellations to show that this is indeed the case for elections with at most $k+1$ candidates.

When there are more than $k+1$ candidates, the same argument cannot be applied since the existence of multiple profiles generating the same Voronoi tessellation is no longer guaranteed. 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. If two environments are identified, for a set of candidate profiles with positive measure one can single out one candidate with different voting shares in the two environments. When there are $k+1$ candidates or more, a new candidate can be introduced without perturbing the W- or $\overline{W}$-Voronoi cells for the singled out candidate and identification is established.[18] This is the main result of the paper and is summarized in the following theorem:

Theorem 1 *Suppose Assumption 1 holds, $||W||_{k \times k} = \sqrt{k}$ and all probability measures are absolutely continuous with respect to the Lesbegue measure on $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ and defined on a common support. 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.

---

[18] One may be tempted at this juncture to establish the identification argument with only two candidates and then add candidates incrementally. This simpler strategy, however, would not work. The identification argument requires that the addition of a candidate does not perturb the Voronoi cells for the "identifying" candidates. When we start with only two candidates, the introduction of additional candidates would necessarily perturb the initial Voronoi cells for both candidates.

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.[19] 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

---

[19] 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.[20] It is nevertheless true that for a given election:

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

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

---

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

numbers of bidders.[21]

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 problem, 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).[22] 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 $k_0 > d/2$, $\delta_0 > d/2$, $\mathcal{B}_0 > 0$, a small $\varepsilon_0 > 0$ and let $\phi(\mathbf{t})$ denote the multivariate standard normal density. The class $\mathcal{H}$ admits densities f such that:

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

with

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

---

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

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

where $\int f(\mathbf{t})d\mathbf{t} = 1$, $\varepsilon > \varepsilon_0$,

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

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 $1 < \eta < \delta_0 - 1$. In practice, the term involving ε is either ignored (see Gallant and Nychka (1987), p. 370) or set to a very small number ($\varepsilon = 10^{-5}$ in Coppejans and Gallant (2002), for example).

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 $||f||_{\text{cons}} = \max_{|\lambda| \le k_0} \sup_{\mathbf{t}}$ $|D^\lambda f(\mathbf{t})|(1 + \mathbf{t}^\top \mathbf{t})^{\delta_0}$, which is the consistency norm we use in Proposition 1 below):

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

where H_i are Hermite polynomials, ϕ is the standard multivariate normal density and ε is a small positive number.[23] 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 $k \times \dim(\mathbf{X})$

---

[23]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:

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

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 monomials of $\tilde{\mathbf{X}}$.[24]

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 (with zero mean and identity variance-covariance)

---

[24]More precisely, in the application, $\hat{m}_i$ is the linear projection of ρ_i on $\tilde{\mathbf{X}}$ so that $J = 2$.

$\mathbf{z}_1, \ldots, \mathbf{z}_S$ and estimate ρ_i as

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

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 $1\left[d^W(\mathbf{t}_s, C_i) \leq d^W(\mathbf{t}_s, C_j), j \neq i\right]$ allows us to obtain the proportion of simulated types $\mathbf{t}_s$ that would choose candidate i and are positioned in $V_i^W(\mathcal{C})$. 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. These steps allow us to evaluate the objective function in (7) at given (f, W) once $\widehat{\Sigma}$ is computed.[25]

Because of the simulations, our objective function is not smooth. Hence, to minimize this function we use Nelder-Meade's non-gradient algorithm (though other non-gradient based methods can be employed alternatively). Using various 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 2 *(i) Elections are iid; (ii) supp($\tilde{\mathbf{X}}$) is compact with nonempty interior; (iii) the density of $\tilde{\mathbf{X}}$ is bounded and bounded away from 0.*

Assumption 3 *(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 $\mathbb{E}[\{g(\tilde{\mathbf{X}}) - b^J(\tilde{\mathbf{X}})^\top \pi\}^2] = o(1)$.*

---

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

Assumption 4 *(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 5 *(i)* $(n-1)J \geq J_t, (n-1)J \geq J_x, J_t \to \infty, J_x \to \infty$ *and* $J/E \to 0$ *as* $E \to \infty$.

The following proposition establishes consistency:

Proposition 1 *Under Assumptions 1-5 and* Θ *compact (with respect to the Frobenius norm),*

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

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]

---

[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. Even for this special case, however, their estimator would require data on individual voter choices whereas we focus on aggregate data.

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 the electoral precinct level for Austria, Belgium, Denmark, Ireland 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, Portugal, Spain, Sweden, and the United Kingdom in 1999.

The ideological positions of the parties were obtained from Hix, Noury, and Roland

---

[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.).

(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.[29] 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 policy stances.[30] 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

---

[29]The data are publicly available at `http://personal.lse.ac.uk/hix/HixNouryRolandEPdata.htm`.

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

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.[31] Figure 4 plots the positions for the "candidates" across all electoral constituencies in our data and indicates their EP group affiliation.
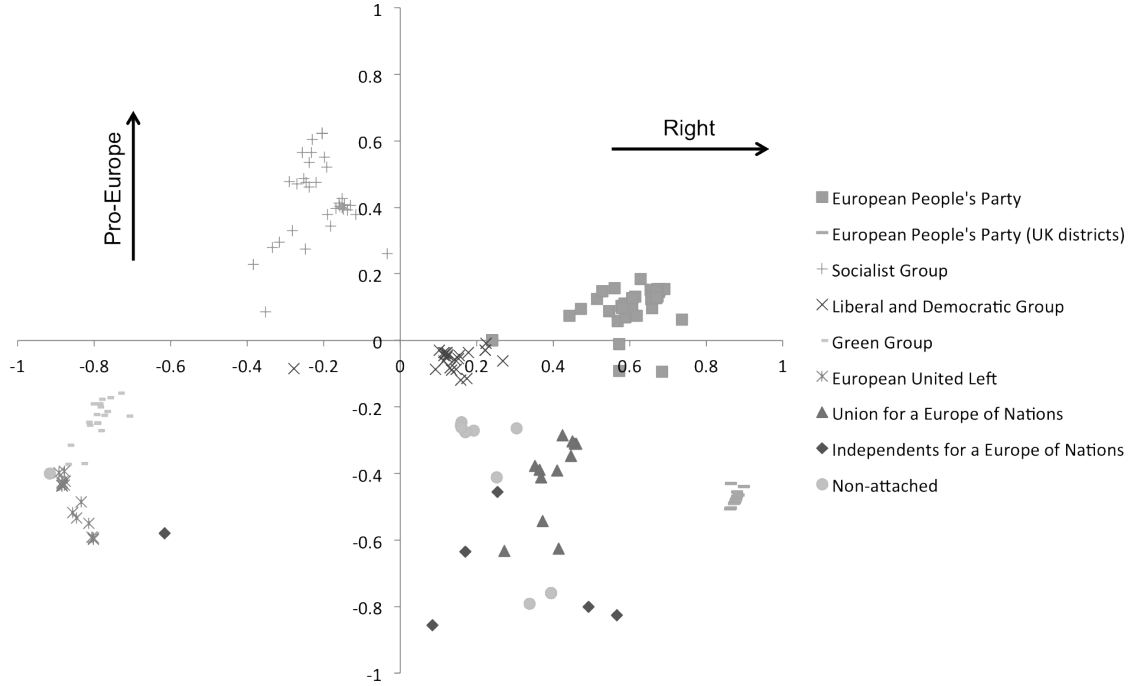

Figure 4: Candidate Positions, 1999

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

_____

[31]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.

ideological positions of a few notable politicians that 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 $(0.566, -0.825)$.[32]
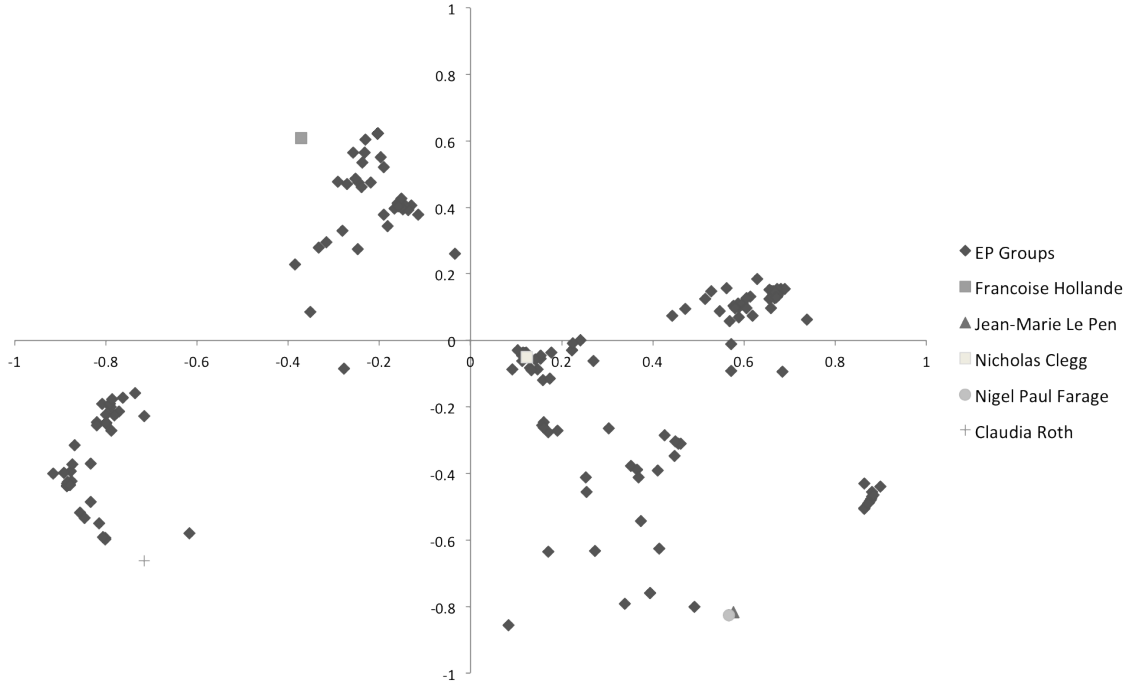
Figure 5: Individual Politician Positions, 1999

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,

---

[32]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. See https://twitter.com/MLP_officiel/status/330215806694744064.

France electoral precinct—with seven candidates, representing seven EP groups.

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

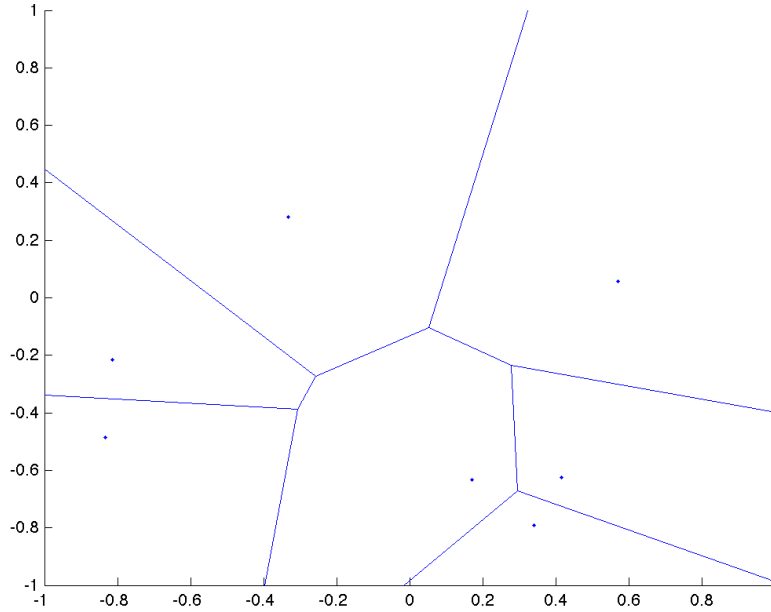Each electoral precinct corresponds to a different tessellation of the ideological space, and we measure the proportion of voters in each cell using the proportion of votes obtained by each of the candidates in that electoral unit. Figure 7 combines the Voronoi tessellations for all the elections in 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 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.

We combine the data on the ideological positions of electoral candidates with electoral
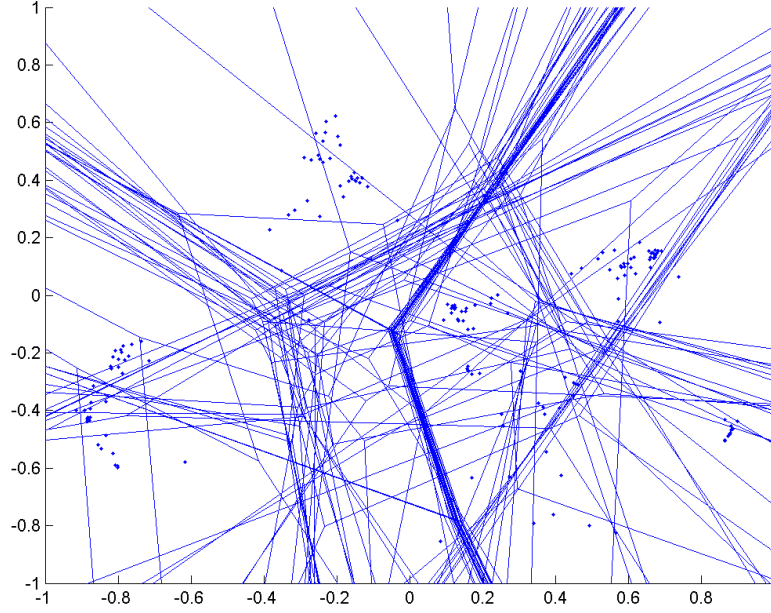
Figure 7: Superimposed Voronoi Diagrams, 1999

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
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.

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)
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.

outcomes in the 1999 elections and demographic and economic variables at the electoral precinct level from the 2001 European Census.[33] The election outcomes data were obtained from the CIVICACTIVE European Election Database.[34] The demographic and economic data were obtained from EUROSTAT and we extracted four variables at the electoral precinct level: the female-to-male ratio; the percentage of the population older than 35; GDP per capita; and the unemployment rate.[35] We present summary statistics for these variables in Table 2.

Using these data, which 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

---

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

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

[35]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.

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. Also, 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 $J_t = 3$ (types) and $J_x = 2$ (covariates), which provided the lowest values for the objective function among the specifications we experimented with. Finally, we use the identity matrix as our estimation weighting matrix ($\hat{\Sigma}$).

The estimates of the weighting matrix W we obtain are $W_{2,2} = 0.273$ and $W_{1,2} = W_{2,1} = -0.339$. 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 25th percentile of the female-to-male ratio (approximately 1.02 in our data) and the 75th percentile of the proportion of residents above 35 years-old (approximately 0.65 in the data) and various percentile combinations for the other two variables (per-capita GDP and the unemployment rate).[36] Similarly, Figure 10 plots level curves for the voter type

---

[36]Electoral precincts with about 1.02 female/male ratio and 65% of the population above 35 years-old in the data correspond to localities such as Noord-Dreenthe (NL), Haute Saone (FR) or Hautes Alpes (FR), for example.
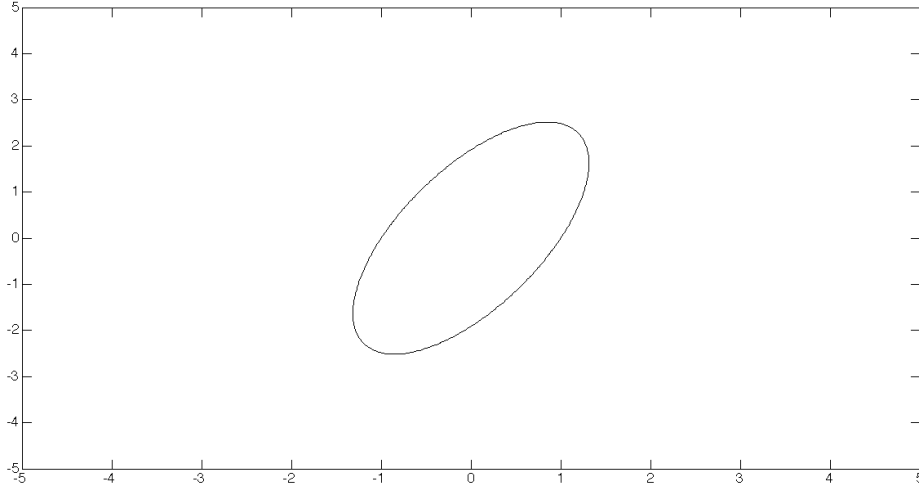
Figure 8: Indifference Curve for $d^W(\mathbf{0}, C) = 1$.

distribution for electoral precincts at the 25th percentile of the GDP per capita and the 75th percentile of the unemployment rate and various percentile combinations for the other two variables (female-to-male ratio and the proportion of residents above 35 years-old). As we can see from the figures, multi-modality and non-concavities are pronounced features of the recovered distribution of voter preferences. Similar patterns emerge for any combination of the conditioning variables. These findings represent a 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)).

To investigate the relationships between demographic and economic variables and the distribution of voters' preferences, Tables 3 and 4 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.[37] As we can see from the tables, electoral precincts with a relatively larger female-to-male ratio, precincts with a relatively larger share

---

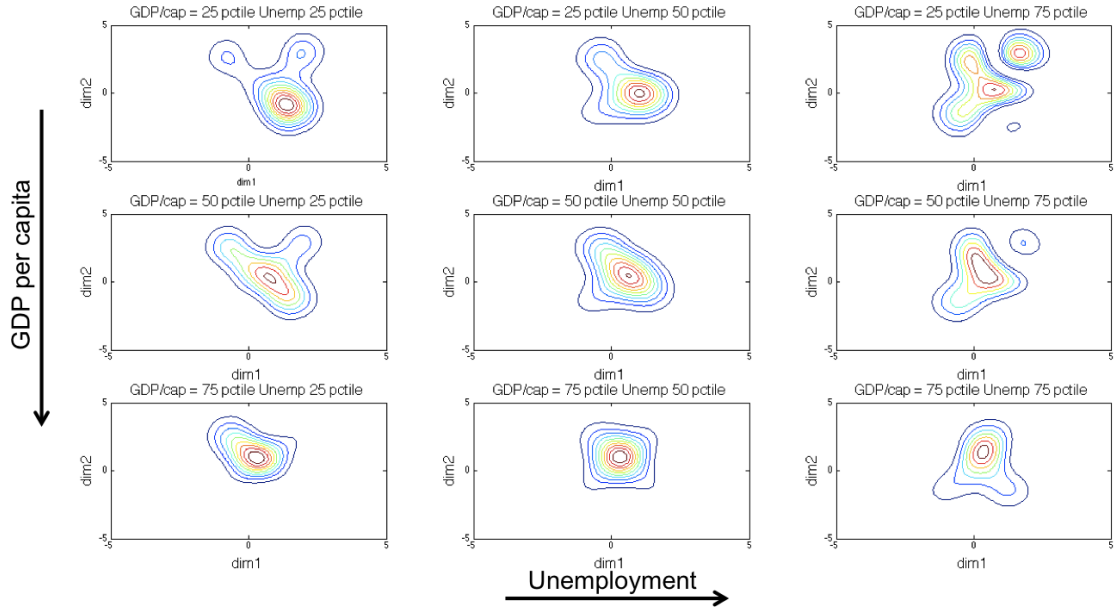[37]Loosely speaking, the table reports the "marginal effects" of each covariate.

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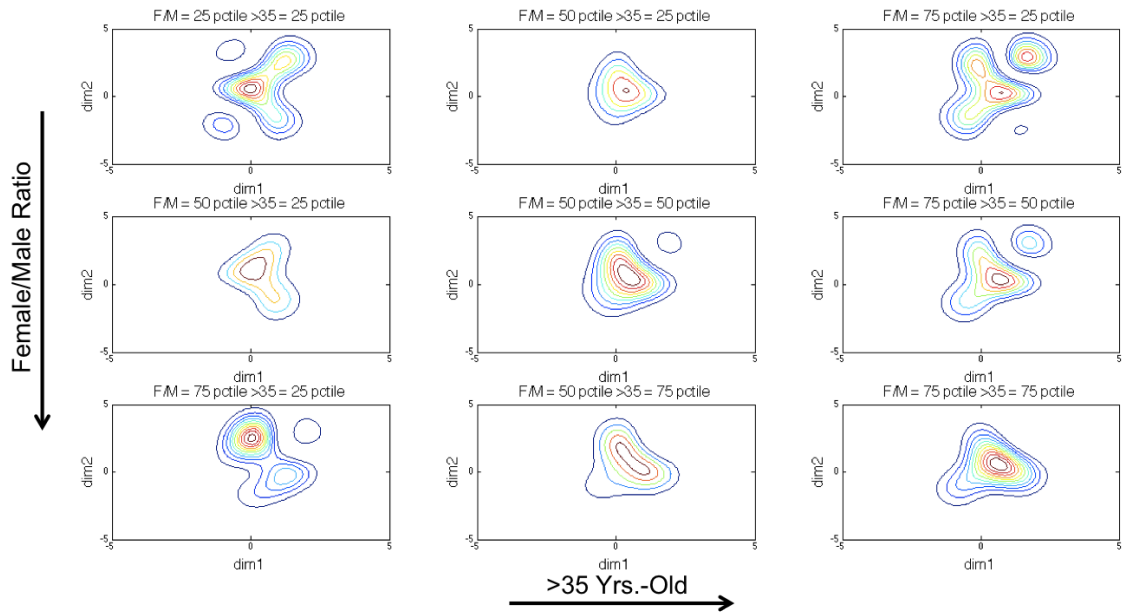

Figure 10: Results at Percentiles of Conditioning Variables (GDP per capita: 25 pctile and Unemployment: 75 pctile)

of the population above the age of 35 and precincts with a relatively higher level of GDP per-capita are relatively less conservative (or right-leaning) and more pro-Europe. On the other hand, electoral precincts with a relatively higher unemployment rate are relatively more liberal but less pro-Europe.

By and large, our findings are consistent with those of the EUROBAROMETER surveys, which document similar correlations between the age, gender, and employment status of European citizens and their sentiments toward European policies.[38] In particular, in the 1999 survey, 62% of the women say that they consider themselves to be left-leaning toward European socio-economic policies compared to 59% of the men, 57% of the citizens between the ages of 18 and 35 compared to 60% of those between the ages of 36 and 65, and 68% of the unemployed as opposed to 60% of those who have a job. Also, in the 1995 EUROBAROMETER survey, 26% of the citizens between the ages of 18 and 35 say that they are very much in favor of European integration compared to 28% of those between the ages of 36 and 65.[39] Among the unemployed, 23% say that are very much in favor, in comparison to 27% of those with a job. On the other hand, relatively more men (30%) than women (24%) say that they are very much in favor of European integration, which is somewhat at odd with our finding.[40]

As a measure of within-sample fit, we calculate the Pearson correlation between realized and predicted vote shares which is equal to 0.83. 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

---

[38]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`.

[39]The 1999 survey did not ask this question.

[40]All the statistics reported here are calculated using the Mannheim Eurobarometer Trend File, 1970-2002 (ICPSR 4357), which is available on-line at `http://www.icpsr.umich.edu`.

Table 3: Fraction of right-leaning voters

Percentiles	Female/ Male	% > 35 Yrs.-old	GDP per capita	Unempl.
25th	0.647	0.611	0.654	0.701
50th	0.620	0.601	0.622	0.653
75th	0.560	0.568	0.599	0.566

Table 4: Fraction of pro-Europe voters

Percentiles	Female/ Male	% > 35 Yrs.-old	GDP per capita	Unempl.
25th	0.615	0.624	0.596	0.668
50th	0.632	0.638	0.626	0.651
75th	0.666	0.638	0.653	0.618

0.61. Overall, these results indicate that the model fits the data relatively well.

6 Conclusion

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*, 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.

Voronoi tessellations are extensively studied in computational geometry and have found wide applicability in computer science, statistics and many other applied mathematics areas (see Okabe, Boots, Sugihara, and Chiu (2000)). The results developed in that literature, however, are relatively new in the social sciences. We believe the methods developed in this paper can also be applied to other economic environments and in particular

to applications in industrial organization (see, e.g., the parametric models in Anderson, De Palma, and Thisse (1989) or Feenstra and Levinsohn (1995)).

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.

ASH, P., AND E. BOLKER (1985): "Recoginizing Dirichlet Tessellations," *Geometriae Dedicata*, 19, pp.175–206.

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.

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.

HARTVIGSEN, D. (1992): "Recognizing Voronoi Diagrams with Linear Programming," *ORSA Journal on Computing*, 4, pp.369–374.

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.

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.

MUNKRES, J. R. (2000): *Topology*. Prentice Hall, 2nd edn.

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.

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 $W = I$. 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 $d((A_1, b_1), (A_2, b_2)) = \sqrt{||A_1 - A_2||_{r \times l}^2 + ||b_1 - b_2||_m^2}$. We use "$\leq$" and "$<$" to denote componentwise weak and strong inequalities and "$\nleq$" and "$\nless$" when the inequality does not hold for at least one component.

Step 1: $(\exists (A^*, b^*) \in \mathcal{M}_{n-1 \times k} \times \mathbb{R}^{n-1} : \mathbb{P}_{T_1}(\{\mathbf{T} \in \mathbb{R}^k : A^*\mathbf{T} \leq b^*\}) \neq \mathbb{P}_{T_2}(\{\mathbf{T} \in \mathbb{R}^k : A^*\mathbf{T} \leq b^*\}).)$ Suppose that $\mathbb{P}_{T_1}(\{\mathbf{T} \in \mathbb{R}^k : A\mathbf{T} \leq b\}) = \mathbb{P}_{T_2}(\{\mathbf{T} \in \mathbb{R}^k : A\mathbf{T} \leq b\}), \forall A, b$. For a given A, let $\mathbf{Z} \equiv A\mathbf{T}$ and define the joint cdfs of $\mathbf{Z}$ under $\mathbb{P}_{T_1}$ and $\mathbb{P}_{T_2}$ as

$$F_{T_1, A}(b) \equiv \mathbb{P}_{T_1}(\{\mathbf{T} \in \mathbb{R}^k : A\mathbf{T} \leq b\})$$

and

$$F_{T_2, A}(b) \equiv \mathbb{P}_{T_2}(\{\mathbf{T} \in \mathbb{R}^k : A\mathbf{T} \leq b\}).$$

Since the probabilities of $\{\mathbf{T} \in \mathbb{R}^k : A\mathbf{T} \leq b\}$ coincide for any A and b,

$$F_{T_1, A} = F_{T_2, A}, \quad \forall A.$$

By the Cramér-Wold device (see (Pollard 2002), p.202), this implies that the cdfs for any linear combination $c^\top \mathbf{Z}$ of $\mathbf{Z}$ will coincide under $\mathbb{P}_{T_1}$ and $\mathbb{P}_{T_2}$. Since a linear combination of $\mathbf{Z}$ is a linear combination of $\mathbf{T}$, the cdf for an arbitrary linear combination of $\mathbf{T}$ under $\mathbb{P}_{T_1}$ coincides with the cdf for that combination under $\mathbb{P}_{T_2}$. Again, by the Cramér-Wold device,

this implies that $\mathbb{P}_{T_1} = \mathbb{P}_{T_2}$. Consequently,

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

In what follows, it is necessary that A^* does not possess any zero rows, but we show in Step 2 that we can always select such an A^*.

Step 2: $(\exists \eta > 0 : \mathbb{P}_{T_1}(\{\mathbf{T} \in \mathbb{R}^k : A\mathbf{T} \leq b\}) \neq \mathbb{P}_{T_2}(\{\mathbf{T} \in \mathbb{R}^k : A\mathbf{T} \leq b\}), \forall (A, b) \in \mathcal{N}((A^*, b^*), \eta)$, **an** η**-Euclidean ball around** (A^*, b^*)**.**) We claim that

$$(A, b) \mapsto h(A, b) \equiv \mathbb{P}_{T_1}(\{\mathbf{T} \in \mathbb{R}^k : A\mathbf{T} \leq b\})$$

is continuous at $(A^*, b^*) \in \mathcal{M}_{n-1 \times k} \times \mathbb{R}^{n-1}$. To see that, first note that

$$\{\mathbf{t} \in \mathbb{R}^k : A\mathbf{t} \leq b\} = \{\mathbf{t} \in \mathbb{R}^k : A\mathbf{t} < b\} \cup \{\mathbf{t} \in \mathbb{R}^k : A\mathbf{t} \leq b \wedge A\mathbf{t} \not< b\} \tag{8}$$

and

$$\{\mathbf{t} \in \mathbb{R}^k : A\mathbf{t} < b\} \cap \{\mathbf{t} \in \mathbb{R}^k : A\mathbf{t} \leq b \wedge A\mathbf{t} \not< b\} = \emptyset,$$

where "$\wedge$" symbolizes logical conjunction (and should be read as "and").

Before proceeding, we argue that either $\{\mathbf{t} \in \mathbb{R}^k : A\mathbf{t} \leq b \wedge A\mathbf{t} \not< b\}$ has Lebesgue measure zero, or there is $(\tilde{A}, \tilde{b})$ such that $\{\mathbf{t} \in \mathbb{R}^k : A\mathbf{t} \leq b\} = \{\mathbf{t} \in \mathbb{R}^k : \tilde{A}\mathbf{t} \leq \tilde{b}\}$ and $\{\mathbf{t} \in \mathbb{R}^k : \tilde{A}\mathbf{t} \leq \tilde{b} \wedge \tilde{A}\mathbf{t} \not< \tilde{b}\}$ has zero Lebesgue measure. This will allow us to focus on the probability of events defined by strict inequalities later on.

To see this, first notice that the set $\{\mathbf{t} \in \mathbb{R}^k : A\mathbf{t} \leq b \wedge A\mathbf{t} \not< b\}$ contains $\mathbf{t}$ for which $A\mathbf{t} \leq b$ and equality holds for at least one equation. This set can then be further partitioned as the union of sets where equality is required to hold for all $n-1$ rows, the $n-1$ sets where strict inequality is required to hold for the ith row (where $i = 1, \ldots, n-1$) and equality is

required to hold for the remaining rows, the $\binom{n-1}{2}$ sets where strict inequality is required to hold for two of the rows and equality is required to hold for the remaining rows and so on, up to the $n - 1$ sets where equality is required to hold for only one of the rows and strict inequality is required to hold for the remaining rows. If any of these sets has non-zero Lebesgue measure, it must be that the rows in A corresponding to the equalities are all equal to zero. Consider, for instance, the first set in this partition: $\{\mathbf{t} \in \mathbb{R}^k : A\mathbf{t} = b\}$. Since the solution set for the linear system of equations $A\mathbf{t} = b$ has dimension $k - \text{rank}(A)$, this set has zero Lebesgue measure except when $\text{rank}(A) = 0$, which can only happen if A is the zero matrix. More generally, for a given set in the partition where l rows are required to hold with equality and the remaining are required to hold with strict inequality, let $\overline{A}_{l \times k}$ and $\overline{b}_{l \times 1}$ be the submatrix of A and subvector of b containing those rows where equality is required. Likewise, let $\underline{A}_{(n-1-l) \times k}$ and $\underline{b}_{(n-1-l) \times 1}$ be the submatrix of A and subvector of b containing those rows where strict inequality is required. The dimension of the solution set for the whole system of equalities and inequalities is at most the dimension of the solution set for the (sub-)system of l equations: $k - \text{rank}(\overline{A})$. Hence, it has non-zero Lebesgue measure only if $\overline{A} = \mathbf{0}_{l \times k}$, where $\mathbf{0}_{l \times k}$ denotes an l by k matrix of zeroes. In that case,

$$\mathbf{0}_{l \times k} \not\leq \overline{b} \Rightarrow \{\mathbf{t} \in \mathbb{R}^k : A\mathbf{t} \leq b\} = \emptyset \tag{9}$$

and

$$\mathbf{0}_{l \times k} \leq \overline{b} \Rightarrow \{\mathbf{t} \in \mathbb{R}^k : A\mathbf{t} \leq b\} = \{\mathbf{t} \in \mathbb{R}^k : \underline{A}_{(n-1-l) \times k}\mathbf{t} \leq \underline{b}_{(n-1-l) \times 1}\}. \tag{10}$$

If $\mathbf{0}_{l \times k} \not\leq \overline{b}$, expression (9) implies that $\{\mathbf{t} \in \mathbb{R}^k : A\mathbf{t} \leq b \wedge A\mathbf{t} \not\leq b\}$ is empty and has zero Lebesgue measure. When $\mathbf{0}_{l \times k} \leq \overline{b}$, expression (10) above points out that the rows corresponding to the zero rows in A are redundant. In this case, $\{\mathbf{t} \in \mathbb{R}^k : A\mathbf{t} \leq b\} = \{\mathbf{t} \in \mathbb{R}^k : \tilde{A}_{(n-1) \times k}\mathbf{t} \leq \tilde{b}_{(n-1) \times 1}\}$ where the $\tilde{A}$ is obtained from A by replacing each zero row in A with any of its non-zero rows and $\tilde{b}$ is obtained by replacing the corresponding rows in b. Notice that this is possible as long as $A \neq \mathbf{0}_{n-1 \times k}$. In this case, the Lebesgue measure of

$\{\mathbf{t} \in \mathbb{R}^k : \tilde{A}\mathbf{t} \leq \tilde{b} \wedge \tilde{A}\mathbf{t} \not< \tilde{b}\}$ is zero since $\tilde{A}$ does not have any zero rows.

To show that $h(\cdot, \cdot)$ is continuous at (A^*, b^*), take a sequence $(A_s, b_s)_{s=1}^{\infty}$ such that $d((A_s, b_s), (A^*, b^*)) \xrightarrow{s \to \infty} 0$. Then, notice that

$$
\begin{aligned}
|h(A^*, b^*) - h(A_s, b_s)| \leq \quad & \mathbb{P}_{T_1}(\{\mathbf{T} \in \mathbb{R}^k : A^*\mathbf{T} \leq b^* \wedge A_s\mathbf{T} \not\leq b_s\}) + \\
& \mathbb{P}_{T_1}(\{\mathbf{T} \in \mathbb{R}^k : A_s\mathbf{T} \leq b_s \wedge A^*\mathbf{T} \not\leq b^*\})
\end{aligned}
$$

We then note that $A^* \neq \mathbf{0}_{n-1 \times k}$, since in this case $\{\mathbf{t} : \mathbf{0}_{n-1 \times k}\mathbf{t} \leq b^*\}$ is either $\mathbb{R}^k$ (if $\mathbf{0}_{n-1 \times 1} \leq b^*$) or empty (if $\mathbf{0}_{n-1 \times 1} \not\leq b^*$). In both cases $\mathbb{P}_{T_1}$ and $\mathbb{P}_{T_2}$ yield the same probability (i.e., 1 or 0, respectively), and the zero matrix would not be selected in Step 1. For the same reason, we also note that if A^* has l zero rows and $\mathbf{0}_{l \times k} \not\leq \bar{b}^*$ (using the previous notation for the b^* subvector corresponding to A^* zero rows), both $\mathbb{P}_{T_1}$ and $\mathbb{P}_{T_2}$ would yield the same probability (i.e., zero) at A^* and b^* would not be picked in Step 1. If, on the other hand, A^* contains zero rows and $\mathbf{0}_{l \times k} \leq \bar{b}^*$, one can always select A^* such that all its rows are non-zero. If this were not case, obtain $\tilde{A}^*$ and $\tilde{b}^*$ as indicated above and notice that

$$
\mathbb{P}_{T_i}(\{\mathbf{T} \in \mathbb{R}^k : A^*\mathbf{T} \leq b^*\}) = \mathbb{P}_{T_i}(\{\mathbf{T} \in \mathbb{R}^k : \tilde{A}^*\mathbf{T} \leq \tilde{b}^*\})
$$

for $i = 1, 2$ and we could reach the same conclusion as in Step 1 using $\tilde{A}^*$ and $\tilde{b}^*$.

Since A^* can be chosen such that no rows are zero, the dimension of $\{\mathbf{t} \in \mathbb{R}^k : A^*\mathbf{t} \leq b^* \wedge A^*\mathbf{t} \not< b^*\}$ is less than k and, consequently, it has Lebesgue measure zero as shown above. Then, because $\mathbb{P}_{T_i}$ is assumed to be absolutely continuous with respect to the Lebesgue measure, we can then use the decomposition in (8) and replace weak inequalities with strict inequalities in the events above:

$$
\mathbb{P}_{T_1}(\{\mathbf{T} \in \mathbb{R}^k : A^*\mathbf{T} \leq b^* \wedge A_s\mathbf{T} \not\leq b_s\}) = \mathbb{P}_{T_1}(\{\mathbf{T} \in \mathbb{R}^k : A^*\mathbf{T} < b^* \wedge A_s\mathbf{T} \not\leq b_s\}).
$$

Note then that

$$\limsup_s \{\mathbf{t} \in \mathbb{R}^k : A^*\mathbf{t} < b^* \wedge A_s\mathbf{t} \not\leq b_s\} = \cap_m \cup_{s \geq m} \{\mathbf{t} \in \mathbb{R}^k : A^*\mathbf{t} < b^* \wedge A_s\mathbf{t} \not\leq b_s\}$$

and $\mathbf{t}$ belongs to this set if it belongs to $\{\mathbf{t} \in \mathbb{R}^k : A^*\mathbf{t} < b^* \wedge A_s\mathbf{t} \not\leq b_s\}$ for infinitely many s. Since $d((A_s, b_s), (A^*, b^*)) \overset{s\to\infty}{\longrightarrow} 0 \Rightarrow ||A_s - A^*||_{n-1 \times k} \overset{s\to\infty}{\longrightarrow} 0$ and $||b_s - b^*||_{n-1} \overset{s\to\infty}{\longrightarrow} 0$, for any fixed $\mathbf{t} \in \mathbb{R}^k$, $||(A_s\mathbf{t} - b_s) - (A^*\mathbf{t} - b^*)||_{n-1} \overset{s\to\infty}{\longrightarrow} 0$. Hence, $A^*\mathbf{t} - b^* < 0$ if and only if there is m such that $s > m$ implies that $A_s\mathbf{t} - b_s < 0$. This means that

$$\limsup_s \{\mathbf{t} \in \mathbb{R}^k : A^*\mathbf{t} < b^* \wedge A_s\mathbf{t} \not\leq b_s\} = \emptyset.$$

Likewise,

$$\liminf_s \{\mathbf{t} \in \mathbb{R}^k : A^*\mathbf{t} < b^* \wedge A_s\mathbf{t} \not\leq b_s\} = \cup_m \cap_{s \geq m} \{\mathbf{t} \in \mathbb{R}^k : A^*\mathbf{t} < b^* \wedge A_k\mathbf{t} \not\leq b_s\}$$

and $\mathbf{t}$ belongs to this set if there is m such that $A^*\mathbf{t} < b^*$ and $A_s\mathbf{t} \not\leq b_s$ for every $s > m$. Again because $||(A_s\mathbf{t} - b_s) - (A^*\mathbf{t} - b^*)||_{n-1} \overset{s\to\infty}{\longrightarrow} 0$, $A^*\mathbf{t} < b^*$ if and only if there is m such that $A_s\mathbf{t} < b_s$ for every $s > m$. Hence,

$$\liminf_s \{\mathbf{t} \in \mathbb{R}^k : A_s\mathbf{t} < b_s\} = \emptyset.$$

Finally, this means that

$$\lim_s \{\mathbf{t} \in \mathbb{R}^k : A^*\mathbf{t} < b^* \wedge A_s\mathbf{t} \not\leq b_s\} = \emptyset.$$

Countable additivity then implies that

$$\lim_s \mathbb{P}_{T_1}(\{\mathbf{T} \in \mathbb{R}^k : A^*\mathbf{T} < b^* \wedge A_s\mathbf{T} \not\leq b_s\}) = \mathbb{P}_{T_1}(\lim_s \{\mathbf{T} \in \mathbb{R}^k : A^*\mathbf{T} < b^* \wedge A_s\mathbf{T} \not\leq b_s\}) = 0.$$

Once again, because $||(A_s \mathbf{t} - b_s) - (A^* \mathbf{t} - b^*)||_{n-1} \overset{s \to \infty}{\longrightarrow} 0$ and A^* can be chosen so that no rows are zero, eventually no rows in A_s are zero and a similar argument holds for $\{\mathbf{t} \in \mathbb{R}^k : A_s \mathbf{t} \leq b_s \wedge A^* \mathbf{t} \not\leq b^*\}$. Consequently,

$$|h(A^*, b^*) - h(A_s, b_s)| \overset{s \to \infty}{\longrightarrow} 0$$

and $h(\cdot, \cdot)$ is continuous. Finally, if $\mathbb{P}_{T_2}$ is substituted for $\mathbb{P}_{T_1}$ the same conclusion is obtained and this shows that

$$\mathbb{P}_{T_1}(\{\mathbf{T} \in \mathbb{R}^k : A\mathbf{T} \leq b\}) - \mathbb{P}_{T_2}(\{\mathbf{T} \in \mathbb{R} : A\mathbf{T} \leq b\})$$

is a continuous function at (A^*, b^*).

By Step 1, $\mathbb{P}_{T_1}(\{\mathbf{T} \in \mathbb{R}^k : A^* \mathbf{T} \leq b^*\}) - \mathbb{P}_{T_2}(\{\mathbf{T} \in \mathbb{R} : A^* \mathbf{T} \leq b^*\}) \neq 0$. Since this function is continuous at (A^*, b^*), this inequality should hold for any (A, b) in some η-ball around (A^*, b^*): $\mathcal{N}((A^*, b^*), \eta)$.

Step 3: ($\{\mathbf{t} \in \mathbb{R}_k : A\mathbf{t} \leq b\}$ **is a Voronoi cell for any** $(A, b) \in \mathcal{N}((A^*, b^*), \eta)$.) With n candidates, a Voronoi cell is characterized by the intersection of $n - 1$ half-spaces (see Okabe, Boots, Sugihara, and Chiu (2000), p.49). To see that

$$R_1 \equiv \{\mathbf{t} \in \mathbb{R}^k : A\mathbf{t} \leq b\}$$

represents a Voronoi cell for some set of candidates, we use the fact that a tessellation of $\mathbb{R}^k$ into polyhedra $R_1, R_2, \ldots, R_n$ is a Voronoi tesselation if and only if there are points $\mathcal{C} = \{C_1, \ldots, C_n\} \subset \mathbb{R}^k$ such that

   i. C_i belongs to the interior of R_i, for $i = 1, \ldots, n$

   ii. If R_i and R_j are neighboring polyhedra, then C_i is the reflection of C_j in the hyperplane

containing $R_i \cap R_j$

(see Theorem 1.1 in Hartvigsen (1992)). Now, note that $\{\mathbf{t} \in \mathbb{R}^k : A\mathbf{t} \leq b\} \neq \emptyset$ (otherwise $\mathbb{P}_{T_1}(\{\mathbf{T} \in \mathbb{R}^k : A\mathbf{T} \leq b\}) = \mathbb{P}_{T_2}(\{\mathbf{T} \in \mathbb{R}^k : A\mathbf{T} \leq b\}) = 0$, contradicting Step 2). Furthermore, $\{\mathbf{t} \in \mathbb{R}^k : A\mathbf{t} < b\} \neq \emptyset$ as well. Otherwise, since both $\mathbb{P}_{T_1}$ and $\mathbb{P}_{T_2}$ are absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^k$ and A^* has no zero rows (and consequently η can be chosen so that no A has zero rows), we would have $\mathbb{P}_{T_1}(\{\mathbf{T} \in \mathbb{R}^k : A\mathbf{T} \leq b\}) = \mathbb{P}_{T_2}(\{\mathbf{T} \in \mathbb{R}^k : A\mathbf{T} \leq b\}) = 0$, again contradicting Step 2. Consequently, R_1 has non-empty interior and any point C_1 in the interior of R_1 satisfies (i) above. We can also find $C_2, \ldots, C_n$ such that the segment $\overline{C_1 C_j}$ is perpendicularly bisected by one of the hyperplanes defined by the system $A\mathbf{t} = b$ and condition (ii) above is satisfied. (Note that this is facilitated as we only rely on one Voronoi cell.)

Step 4: $(p(\cdot, \mathbb{P}_{T_1}) \neq p(\cdot, \mathbb{P}_{T_2})$ **with positive Lebesgue measure.)** Consider a set of candidate positions $\mathcal{C}^* = \{C_1^*, \ldots, C_n^*\}$ that generates $A^* \mathbf{t} < b^*$ as a Voronoi cell. Without loss of generality, assume that this cell corresponds to candidate C_1^*. For each of these points C_i^*, $i = 1, \ldots, n$, define an ϵ-ball $\mathcal{N}(C_i^*, \epsilon)$, $\epsilon > 0$. Consider the Voronoi tessellation generated by the selection of n points $C_i \in \mathcal{N}(C_i^*, \epsilon), i = 1, \ldots, n$. Collect all A's and b's representing the Voronoi cell for the first candidate from such selections into the set $\mathcal{S}(\epsilon) \equiv \{(A, b) \in \mathcal{M}_{n-1,k} \times \mathbb{R}^{n-1} : \{\mathbf{t} \in \mathbb{R}^k : A\mathbf{t} > b\}$ is the Voronoi cell containing C_1 and $C_i \in \mathcal{N}(C_i^*, \epsilon)$, $i = 1, \ldots, n\}$. Notice that $\mathcal{S}(\epsilon) \xrightarrow{\epsilon \to 0} \{(A^*, b^*)\} \subset \mathcal{N}((A^*, b^*), \eta)$. Furthermore, $\{\mathcal{S}(\epsilon)\}_{\epsilon > 0}$ is totally ordered by set inclusion ($\epsilon_1 \geq \epsilon_2 \Rightarrow \mathcal{S}(\epsilon_1) \supseteq \mathcal{S}(\epsilon_2)$) and, given the order topology (see Munkres (2000), p.84), the mapping $\epsilon \mapsto \mathcal{S}(\epsilon)$ is continuous. Hence, $\exists \epsilon > 0$ so that $\mathcal{S}(\epsilon) \subset \mathcal{N}((A^*, b^*), \eta)$. Since $\epsilon > 0$, the set $\times_{i=1}^n \mathcal{N}(C_i^*, \epsilon)$ (i.e., the Cartesian product of $\mathcal{N}(C_i^*, \epsilon)$) has positive Lebesgue measure on the n-fold Cartesian product of $\mathbb{R}^k$ and identification follows as candidate points obtained in this set generate a Voronoi cell that attains a different proportion of votes under $\mathbb{P}_{T_1}$ and $\mathbb{P}_{T_2}$. $\blacksquare$

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 leads to the same tessellation across the two environments and, since Voronoi cells are obtained as intersections of half-spaces, identification follows along the lines of Lemma 1. Assume then that $W \neq \overline{W}$ and $(\mathbb{P}_T, W)$ and $(\mathbb{P}_{\overline{T}}, \overline{W})$ are observationally equivalent: for almost every candidate-election profile $\mathcal{C} = (C_1, \ldots, C_n)$, the proportion of votes obtained under the two different systems is identical.

Step 1: (There is more than one set of candidates that generates a Voronoi tessellation.) Generically (i.e., except for a set of measure zero), all the vertices of a Voronoi tessellation in $\mathbb{R}^k$ are shared by $k + 1$ cells (see Theorem 9 and subsequent remark on Ash and Bolker (1985), p.185). Consequently, if there are at most $k + 1$ candidates, there is at most one vertex (a point on the boundary of three or more regions). A generalization of case (3) in Theorem 14 of Ash and Bolker (1985) (p.191) then implies that a given Voronoi tessellation can be generated by more than one set of candidates. We will rely on a particular set of alternative candidates generating the same W-Voronoi tessellation. The argument relies on the existence of a point which is equidistant from all the n candidates.

If $n = k + 1$ and no three candidates are collinear, there will be a vertex. Since collinearity of three candidates is an event of measure zero, generically there will be a vertex. Let the vertex be denoted by P and let $\mathcal{C}'$ be such that

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

Notice that

$$d^W(C_i', \mathbf{t}) - d^W(C_j', \mathbf{t}) = 0 \qquad \Leftrightarrow$$

$$(C_i' - \mathbf{t})^\top W(C_i' - \mathbf{t}) - (C_j' - \mathbf{t})^\top W(C_j' - \mathbf{t}) = 0 \qquad \Leftrightarrow$$

$$C_i'^\top W C_i' - C_j'^\top W C_j' - 2(C_i' - C_j')^\top W\mathbf{t} = 0 \qquad \Leftrightarrow$$

$$(2C_i - P)^\top W(2C_i - P) - (2C_j - P)^\top W(2C_j - P) - 4(C_i - C_j)^\top W\mathbf{t} = 0 \quad \Leftrightarrow$$

$$C_i^\top W C_i - C_j^\top W C_j - (C_i - C_j)^\top W P - (C_i - C_j)^\top W\mathbf{t} = 0$$

Since P is the vertex shared by all the regions, $d^W(C_i, P) - d^W(C_j, P) = 0$ for any i and j and consequently $\frac{1}{2}\left(C_i^\top W C_i - C_j^\top W C_j\right) = (C_i - C_j)^\top W P$. This in turn implies that

$$C_i^\top W C_i - C_j^\top W C_j - 2(C_i - C_j)^\top W\mathbf{t} = 0 \quad \Leftrightarrow$$

$$d^W(C_i, \mathbf{t}) - d^W(C_j, \mathbf{t}) = 0.$$

Since this holds for any choice of i and j, the W-Voronoi diagram is the same.

If $n < k + 1$, the set of vectors $\mathbf{t}$ such that

$$d^W(C_1, \mathbf{t}) = \cdots = d^W(C_n, \mathbf{t})$$

will have dimension at least one. To see this, note that the above is equivalent to

$$d^W(C_1, \mathbf{t}) - d^W(C_n, \mathbf{t}) = \cdots = d^W(C_{n-1}, \mathbf{t}) - d^W(C_n, \mathbf{t}) = 0.$$

These define $n - 1$ linear equations on $\mathbf{t} \in \mathbb{R}^k$: $(C_1 - C_s)^\top W\mathbf{t} = (C_1^\top W C_1 - C_s^\top W C_s)/2, s = 2, \ldots, n$. The solution set for this system contains at least one element as long as $\mathrm{rank}([C_1 - C_2 \quad C_1 - C_3 \ldots C_1 - C_n]^\top W) = n - 1$ (see Strang (1988), p.96). Since $n - 1 < k$ and W is positive definite (and, consequently, has full rank), this is the case except in a set of candidate profiles with zero Lebesgue measure. In this case, let P denote one such solution

and proceed as before in defining $\mathcal{C}'$.

Step 2: (**For $\mathcal{C} \neq \mathcal{C}'$ such that $V^W(\mathcal{C}) = V^W(\mathcal{C}')$, $V^{\overline{W}}(\mathcal{C})$ and $V^{\overline{W}}(\mathcal{C}')$ have parallel faces.**) Consider $\mathcal{C}$ and $\mathcal{C}'$, profiles of size $n \leq k+1$ such that their Voronoi tessellations under W coincide, i.e., $V^W(\mathcal{C}) = V^W(\mathcal{C}')$. As before, let $V_i^W, i = 1, \ldots, n$ denote the n cells in this Voronoi tessellation. Accordingly, denote by C_i and C_i' the corresponding candidates in $\mathcal{C}$ and $\mathcal{C}'$.

Given our definition of $\mathcal{C}$ and $\mathcal{C}'$, note that

$$C_i' - C_j' = 2(C_i - C_j)$$

for every i and j. Then see that

$$
\begin{aligned}
\mathbf{t} \in H^{\overline{W}}(C_i', C_j') \quad &\Rightarrow C_i'^{\top} \overline{W} C_i' - C_j'^{\top} \overline{W} C_j' - 2(C_j' - C_i')^{\top} \overline{W} \mathbf{t} = 0 \\
&\Rightarrow \tfrac{1}{2} \left(C_i'^{\top} \overline{W} C_i' - C_j'^{\top} \overline{W} C_j' \right) - 2(C_j - C_i)^{\top} \overline{W} \mathbf{t} = 0.
\end{aligned}
\tag{11}
$$

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

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

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 convex polyhedron in $\mathbb{R}^k$ (see Hartvigsen (1992)). It can then be represented as:

$$V_i^{\overline{W}}(\mathcal{C}) = \{\mathbf{t} \in \mathbb{R}^k : \mathbf{A}_i \mathbf{t} \leq \mathbf{b}_i\}$$

where the rows of the vector $\mathbf{A}_i \mathbf{t} - \mathbf{b}_i$ are the "defining hyperplanes" (see Hartvigsen (1992)):

$$\underbrace{-2(C_i - C_j)^\top \overline{W} \, \mathbf{t}}_{\equiv \mathbf{a}_{ij}^\top} + \underbrace{C_i^\top \overline{W} C_i - C_j^\top \overline{W} C_j}_{\equiv b_{ij}} = 0, \ j \neq i.$$

Similarly,

$$V_i^{\overline{W}}(\mathcal{C}') = \{\mathbf{t} \in \mathbb{R}^k : \mathbf{A}_i' \mathbf{t} \leq \mathbf{b}_i'\}.$$

Because $V_i^{\overline{W}}(\mathcal{C})$ and $V_i^{\overline{W}}(\mathcal{C}')$ have parallel faces (see (11)), we have

$$\mathbf{A}_i' = \mathbf{A}_i.$$

Furthermore, expression (11) gives that the $n-1$ rows of $\mathbf{b}_i - \mathbf{b}_i'$ are given by

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

for $j \neq i$.

For every i, there exists j such that $\Delta_{ij} \neq 0$. Otherwise, $V_i^{\overline{W}}(\mathcal{C}) = V_i^{\overline{W}}(\mathcal{C}')$, which can only happen in a set of candidates of zero Lebesgue measure. To see this, note that

$$d^{\overline{W}}(C_i, \mathbf{t}) - d^{\overline{W}}(C_j, \mathbf{t}) = 0 \ \Leftrightarrow \ C_i^\top \overline{W} C_i - C_j^\top \overline{W} C_j - (C_i - C_j)^\top \overline{W} \mathbf{t} = 0.$$

Given that $C_i' = C + i + P$ where P is defined in Step 1 to attain $V_i^W(\mathcal{C}) = V_i^W(\mathcal{C}')$ (and hence depends on $\mathcal{C}$)

$$d^{\overline{W}}(C_i', \mathbf{t}) - d^{\overline{W}}(C_j', \mathbf{t}) = 0 \qquad\qquad \Leftrightarrow$$

$$C_i^\top \overline{W} C_i - C_j^\top \overline{W} C_j - (C_i - C_j)^\top \overline{W} \mathbf{t} - (C_i - C_j)^\top \overline{W} P = 0.$$

Since $V_i^{\overline{W}}(\mathcal{C}) = V_i^{\overline{W}}(\mathcal{C}')$ means that

$$d^{\overline{W}}(C_i, \mathbf{t}) - d^{\overline{W}}(C_j, \mathbf{t}) = 0 \quad \Leftrightarrow \quad d^{\overline{W}}(C_i', \mathbf{t}) - d^{\overline{W}}(C_j', \mathbf{t}) = 0,$$

we have that $(C_i - C_j)^\top \overline{W} P = 0$. Because $\overline{W}$ is positive definite, this can only happen in a set of candidates of zero Lebesgue measure.

If there is i such that $\Delta_{ij} \geq 0$ for any $j \neq i$ (or $\Delta_{ij} \leq 0$ for any $j \neq i$) with at least one strict inequality, $\mathbf{b}_i - \mathbf{b}_i' \geq 0$ (or $\mathbf{b}_i - \mathbf{b}_i' \leq 0$) and $\mathbf{b}_i \neq \mathbf{b}_i'$. But then

$$\mathbf{A}_i \mathbf{t} \leq \mathbf{b}_i' \Rightarrow \mathbf{A}_i \mathbf{t} \leq \mathbf{b}_i$$

and $V_i^{\overline{W}}(\mathcal{C}') \subset V_i^{\overline{W}}(\mathcal{C})$. Furthermore, because the inequality is strict for at least one i, $\mathrm{int}(V_i^{\overline{W}}(\mathcal{C}) \backslash V_i^{\overline{W}}(\mathcal{C}')) \neq \emptyset$ (where for any set B, $\mathrm{int}(B)$ denotes the interior of that set). If $\mathbf{b}_i - \mathbf{b}_i' \leq 0$, the inclusion is reversed.

If this is not the case, but there exists i such that $\Delta_{ij} \geq 0$ for all $j \neq i$ except for $j = l$, note that

$$\Delta_{il} < 0 \Leftrightarrow \Delta_{li} > 0.$$

Then,

$$\begin{aligned}
\Delta_{li} + \Delta_{ij} &= C_l^\top \overline{W} C_l - C_i^\top \overline{W} C_i - \tfrac{1}{2}\left(C_l'^\top \overline{W} C_l' - C_i'^\top \overline{W} C_i'\right) + \\
&\quad + C_i^\top \overline{W} C_i - C_j^\top \overline{W} C_j - \tfrac{1}{2}\left(C_i'^\top \overline{W} C_i' - C_j'^\top \overline{W} C_j'\right) = \\
&= C_l^\top \overline{W} C_l - C_j^\top \overline{W} C_j - \tfrac{1}{2}\left(C_l'^\top \overline{W} C_l' - C_j'^\top \overline{W} C_j'\right) = \Delta_{lj} > 0
\end{aligned} \tag{12}$$

for any $j \neq i, l$. That means that $\Delta_{lj} > 0$ for any $j \neq l$ and consequently $\mathbf{b}_l - \mathbf{b}_l' > 0$.

To generalize the above argument by induction, assume the claim is true if i is such that $\Delta_{ij} \geq 0$ for all but $s - 1$ of the j indices. Above we showed that this holds for $s = 2$. We will now show that one can obtain l such that $\mathbf{b}_l - \mathbf{b}_l' \geq 0$ when there is i such that for all but s indices, $\Delta_{ij} \geq 0$. For those indices l such that $\Delta_{il} < 0$, we have $\Delta_{li} > 0$. Take one of them

and, as in (12), $\Delta_{lj} > 0$ for all those indices j such that $\Delta_{ij} \geq 0$. Since $\Delta_{li} > 0$, there are at most $s - 1$ indices such that $\Delta_{lm} \leq 0$. By induction, $\mathbf{b}_l - \mathbf{b}'_l > 0$ and $V_l^{\overline{W}}(\mathcal{C}') \subset V_l^{\overline{W}}(\mathcal{C})$. As before, because the inequality is strict, $\mathrm{int}(V_l^{\overline{W}}(\mathcal{C}) \backslash V_l^{\overline{W}}(\mathcal{C}')) \neq \emptyset$. If inequalities are reversed, $\mathbf{b}_l - \mathbf{b}'_l < 0$, the inclusion is reversed.

Step 4: $(\mathbb{P}_{\overline{T}}(\mathbb{R}^k) = 0,$ **leading to a contradiction.)** Given 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, so the argument that follows is robust to perturbations in $\mathcal{C}$.) Take an arbitrary vector $\mathbf{t}^{\triangledown} \in \mathrm{int}(V_i^{\overline{W}}(\mathcal{C}') \backslash V_i^{\overline{W}}(\mathcal{C}))$. 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 $\mathbf{t} \in \mathrm{int}(V_i^{\overline{W}}(\mathcal{C}'_{\mathbf{t}-\mathbf{t}^{\triangledown}}) \backslash V_i^{\overline{W}}(\mathcal{C}_{\mathbf{t}-\mathbf{t}^{\triangledown}}))$.

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

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

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

If there are at most $k+1$ candidates, Lemma 2 establishes the results. If $n > k+1$, consider first a Voronoi tessellation with $k+1$ candidates $C_1, \ldots, C_{k+1}$. In this case, apply Lemma 2 to obtain the existence of a set of candidate profiles with positive measure such that $p((C_1, \ldots, C_{k+1}); \mathbb{P}_T, W) \neq p((C_1, \ldots, C_{k+1}); \mathbb{P}_{\overline{T}}, \overline{W})$.

Now, let i be a candidate for which voting shares are distinct under the two environments. Since there are at least $k+1$ candidates, V_i^W has (generically) at least one vertex (see Theorem 9 in Ash and Bolker (1985)). Select one of these vertices P and let $\mathbb{C}$ be the *smallest* cone that has C_i as its vertex and contains all the candidates that share the vertex P with C_i (i.e., if Y is a generic element in $\mathbb{C}$, then $C_i + \alpha(Y - C_i) \in \mathbb{C}$, $\forall \alpha > 0$). Since P is equidistant from all $k+1$ candidates by definition, there is a hypersphere $\mathbb{S}^W$ centered at P that contains every candidate. For any points $C_{k+2}, \ldots, C_n$ in $\mathbb{C} \backslash \mathbb{S}^W$, $V_i^W((C_1, \ldots, C_{k+1})) = V_i^W((C_1, \ldots, C_n))$ as the hyperplane bisecting C_i and C_{k+2} so chosen does not intersect $V_i^W((C_1, \ldots, C_{k+1}))$. This is the case when $C_{k+2} = \alpha(C_j - C_i)$ where $\alpha > 1$ and C_j is any other candidate in the hypersphere. By locating C_{k+2} in $\mathbb{C} \backslash \mathbb{S}^W$ we assure that the same holds for any other point in a neighborhood containing C_{k+2}.

An analogous argument can be made for $(\mathbb{P}_{\overline{T}}, \overline{W})$. Consequently, for any point in the set $(\mathbb{C} \backslash \mathbb{S}^W) \cap (\mathbb{C} \backslash \mathbb{S}^{\overline{W}}) = \mathbb{C} \backslash (\mathbb{S}^W \cup \mathbb{S}^{\overline{W}})$ we have $\mathbb{P}_T(V_i^W((C_1, \ldots, C_n))) \neq \mathbb{P}_{\overline{T}}(V_i^{\overline{W}}((C_1, \ldots, C_n)))$.

Using arguments akin to those in Steps 2 and 4 of the proof of Lemma 2 we can show that the same holds for a neighborhood of $(C_1, \ldots, C_n)$ in $\mathbb{R}^k \times \ldots \mathbb{R}^k$. ∎

Proof of Proposition 1

The result follows from an adaptation of Lemma 3.1 in Ai and Chen (2003) (which in turn uses Theorem 4.1 and Lemma A1 from Newey and Powell (2003)).

Assumptions 2-5 correspond to assumptions 3.1, 3.2, 3.4 and 3.7 in that Lemma 3.1 in Ai and Chen (2003). Assumption 3.3 for that lemma is attained from the identification results in Theorem 1. Assumption 3.5(ii) follows from Theorem 2 in Gallant and Nychka (1987). The compactness of Θ with respect to the topology induced by the Frobenius norm and that of (the closure of) $\mathcal{H}$ with respect to topology induced by the consistency norm (see Theorem 1 in Gallant and Nychka (1987)) implies 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 in condition (ii) from Newey and Powell (2003). Once this is done, Ai and Chen can 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

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

58

where ρ is an $E \times (n-1)$ matrix stacking $\left(\int \mathbf{1}_{\mathbf{t} \in V_i^W(\mathcal{C})} f(\mathbf{t}) d\mathbf{t} - p_i \right)^\top_{i=1,\dots,n-1}$ for all observations and P is an $E \times E$ idempotent matrix with rank (= trace) at most J. Since we have Assumption 4, 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

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

which in turn delivers

$$
\begin{aligned}
|Q_n(W_1, f_1) - Q_n(W_2, f_2)| &= \left| \sum_{i=1}^{n-1} \left(\frac{1}{E} ||P\rho_i(W_1, f_1)||^2 - \frac{1}{E} ||P\rho_i(W_2, f_2)||^2 \right) \right| \\
&\leq \sum_{i=1}^{n-1} \left| \frac{1}{E} ||P\rho_i(W_1, f_1)||^2 - \frac{1}{E} ||P\rho_i(W_2, f_2)||^2 \right|
\end{aligned}
\tag{13}
$$

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

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

each of the terms in the sum in expression (13) is bounded by

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

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

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

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

59

we have

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

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, once we show 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 have stochastic equicontinuity of the objective function. Let then

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

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

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

To show stochastic equicontinuity we essentially follow the proof for Lemma 3 in Andrews (1992). Given $\epsilon > 0$, take $M > 4$ and $\delta > 0$ such that $Prob\left(Y_{e\delta} > \epsilon^2/2\right) < \epsilon^2/(2M)$. That such a δ can be chosen follows because Assumption TSE-1D from Andrews (1992) holds in our application. Given its Lemma 4 (replacing $\|\cdot\|$ by $(\cdot)^2$) we obtain Termwise Stochastic Equicontinuity (TSE), which essentially states that $\lim_{\delta\to 0} Prob\left(Y_{e\delta} > \epsilon\right) = 0$ for any $\epsilon > 0$. Now, for such a δ,

$$\lim_{E\to\infty} Prob\left(\frac{1}{E}\sum_{e=1}^{E} Y_{e\delta} > \epsilon\right) \leq \frac{1}{\epsilon}\mathbb{E}(Y_{e\delta})$$

$$= \frac{1}{\epsilon}\left[\mathbb{E}\left(Y_{e\delta}\mathbf{1}\left(Y_{e\delta} \leq \frac{\epsilon^2}{2}\right)\right) + \mathbb{E}\left(Y_{e\delta}\mathbf{1}\left(\frac{\epsilon^2}{2} < Y_{e\delta} \leq M\right)\right) + \mathbb{E}\left(Y_{e\delta}\mathbf{1}\left(Y_{e\delta} > M\right)\right)\right]$$

$$\leq \frac{1}{\epsilon}\left(\frac{\epsilon^2}{2} + MProb(Y_{e\delta} > \frac{\epsilon^2}{2})\right) \leq \epsilon$$

Since this argument can be repeated for $i = 1, \ldots, n-1$, we have stochastic equicontinuity.

∎