Penn Institute for Economic Research
Department of Economics
University of Pennsylvania
3718 Locust Walk
Philadelphia, PA 19104-6297
pier@econ.upenn.edu
http://economics.sas.upenn.edu/pier

PIER Working Paper 11-001

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

by

Antonio Merlo and Aureo de Paula

http://ssrn.com/abstract=1734318

Identification and Estimation of Preference Distributions When Voters Are Ideological *

Antonio Merlo †
University of Pennsylvania

Áureo de Paula ‡
University of Pennsylvania

December, 2010

Abstract

This paper studies the nonparametric identification and estimation of voters' preferences when voters are ideological. We build on the methods introduced by Degan and Merlo (2009) representing elections as Voronoi tessellations of the ideological space. We exploit the properties of this geometric structure to 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 using the methodology proposed by Ai and Chen (2003) and we illustrate our analysis by performing an actual estimation using data from the 1999 European Parliament elections. *JEL*: D72, C14; Keywords: Voting, Voronoi tessellation, identification, nonparametric.

*We would like to thank Andrew Chesher, Eric Gautier, Ken Hendricks, Stefan Hoderlein, Bo Honoré, Frank Kleibergen, Dennis Kristensen, Ariel Pakes, Jim Powell, Bernard Salanié, Kevin Song and Dale Stahl for helpful suggestions. The paper also benefited from comments by seminar and conference participants at several institutions. Chen Han and Chamna Yoon provided very able research assistance.

†Department of Economics, University of Pennsylvania, Philadelphia, PA 19104. E-mail: merloa@sas.upenn.edu

‡Department of Economics, University of Pennsylvania, Philadelphia, PA 19104. E-mail: aureo@sas.upenn.edu

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) 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. 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 electoral data under the maintained assumption that voters vote ideologically. We build on the methods introduced by Degan and Merlo (2009)

[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).

representing elections as Voronoi tessellations of the ideological space.[4] We exploit the properties of this geometric structure to 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). We provide large sample results and Monte Carlo experiments for the estimator in the context of our application. We then illustrate our analysis by performing 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. Starting with McFadden (1974)'s seminal work, other important papers investigating the identification of discrete choice models include Manski (1988) and Matzkin (1992).[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). 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.[7] Whereas the distribution of tastes is typically taken to be parametric in pure characteristics models, we show that it can be *nonparametri-*

[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 (2010) extend their analysis and develop a formal test of the hypothesis). In this paper, we restrict attention to inference based on aggregate data on electoral outcomes in environments where the hypothesis is non-falsifiable.

[5]See, e.g., King (1997) for a survey.

[6]See also Chesher and Silva (2002).

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

cally identified and estimated together with the finite dimensional components of the model (W). Our identification strategy relies on the geometric structure (i.e., Voronoi tessellation) implied by the functional form of the utility function, but the main ideas also apply to more general utility functions.[8]

Part of the identification strategy we develop in this paper is related to previous work by Ichimura and Thompson (1998) and Gautier and Kitamura (2008) on binary choice models with random coefficients. However, these papers require a "dilatation invariance" property and only admit environments with two alternatives.[9] Since the property is not satisfied by our model and the environments we consider typically entail more than two alternatives, their identification strategy does not apply in our case. In fact, because Voronoi tessellations can also be defined on hyperspheres, and in Ichimura and Thompson (1998) and Gautier and Kitamura (2008) covariates and coefficients are both supported on a hypersphere, we believe our methodology may also be used to generalize their ideas.

The remainder of the paper is organized as follows. In Section 2, we describe the model and discuss its identification. Nonparametric estimation is presented in Section 3. Sections 4 and 5 contain Monte Carlo experiments and an empirical illustration, respectively. Concluding remarks are presented in Section 6. All proofs are contained in the Appendix.

2 Identification

Consider a situation where a population of voters has to elect some representatives to public office. Consistent with the spatial theory of voting, there is a common ideological space, Y, which is taken to be the d-dimensional Euclidean space (i.e., $Y = \mathbb{R}^d$ and the reference measurable space is this set equipped with the Borel sigma algebra: $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$). We observe a cross-section of elections $e \in \{1, ..., E\}$. An election is a contest among $n \geq 2$ candidates. Let $\mathcal{C} \equiv (C_1, \ldots, C_n) \in \mathbb{R}^d \times \cdots \times \mathbb{R}^d$ denote a profile of candidates in the n-fold Cartesian product of $\mathbb{R}^d$ which characterizes an election. Each candidate i in this set of candidates is

[8] We discuss this point in detail in Section 2.1.

[9] We provide a formal definition of this property in Section 2.

characterized by a distinct position in the ideological space, $C_i \in Y$, which is known to the voters.

Each voter has an ideological position (or bliss point) $\mathbf{t}$, and her preferences are characterized by indifference sets that are spheres in the d-dimensional Euclidean space (or d-spheres), centered around her bliss point.[10] It follows that voter $\mathbf{t}$'s preferences over candidates in election e can be summarized by the utility function

$$U_e^{\mathbf{t}}(C_i) = u_e^{\mathbf{t}}(d(\mathbf{t}, C_i)), \tag{1}$$

where $u_e^{\mathbf{t}}(\cdot)$ is a decreasing function which may differ across voters and elections and $d(\cdot,\cdot) \geq 0$ denotes the Euclidean distance (i.e., for any two points $x, y \in \mathbb{R}^d$, $d(x,y) = \sqrt{(x-y)^\top(x-y)}$). Other than monotonicity, we impose no additional restrictions on the $u_e^{\mathbf{t}}(\cdot)$ functions, which are therefore left unspecified. Given these preferences, a voter $\mathbf{t}$ (strictly) prefers candidate i to candidate j in election e if $d(\mathbf{t}, C_i) < d(\mathbf{t}, C_j)$.

As in Degan and Merlo (2009), for each election $e \in \{1, \ldots, E\}$, and position $C_i \in Y$ of a generic candidate i in the election, let $V_i(\mathcal{C}) \equiv \{\mathbf{t} \in Y : d(\mathbf{t}, C_i) < d(\mathbf{t}, C_j), j \neq i\}$ 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 $d(\cdot,\cdot)$ is the Euclidean distance, it follows that for each pair of candidates in election e, 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 $H(C_i, C_j)$, which partitions the ideological space Y into two regions (or half spaces), $Y_{C_i}^{C_j}$ and $Y_{C_j}^{C_i} = Y \backslash [Y_{C_i}^{C_j} \cup H(C_i, C_j)]$, where $Y_{C_j}^{C_i}$ is the set of ideological 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, $V_i(\mathcal{C})$ is the intersection of the half spaces determined by the $n-1$ hyperplanes $(H(C_i, C_j))_{j \neq i}$ (i.e., $V_i(\mathcal{C}) = \cap_{j \neq i} Y_{C_j}^{C_i}$). Note that, for all candidates i, $V_i(\mathcal{C})$ is non empty and convex. Hence, each election $e \in \{1, \ldots, E\}$ implies a partition of the ideological space Y into n convex regions,

[10]In one dimension, the restriction implies that each voter's utility function is single-peaked and symmetric. In the following subsection, we consider more general specifications of preferences where the voters' indifference sets are ellipsoids in the d-dimensional Euclidean space. When $d \geq 2$, such preferences allow for the possibility that voters may evaluate different ideological dimensions using different weights.

$\{V_i(\mathcal{C})\}_{i\in\{1,\ldots,n\}}$, where each region $V_i(\mathcal{C})$ is the set of voters voting for candidate i in election e.[11] For each election $e \in \{1, \ldots, E\}$, the set $\{V_i(\mathcal{C})\}_{i\in\{1,\ldots,n\}}$ defines what in computational and combinatorial geometry is called a *Voronoi tessellation* of R^d and each region $V_i(\mathcal{C})$ is a d-dimensional *Voronoi polyhedron* (or *Voronoi cell*).[12] Figure 1 illustrates an example of the Voronoi tessellation that corresponds to an election with 5 candidates, $\{a, b, c, d, e\}$, with positions $\{C_a, C_b, C_c, C_d, C_e\}$ in the two-dimensional ideological space $Y = R^2$.

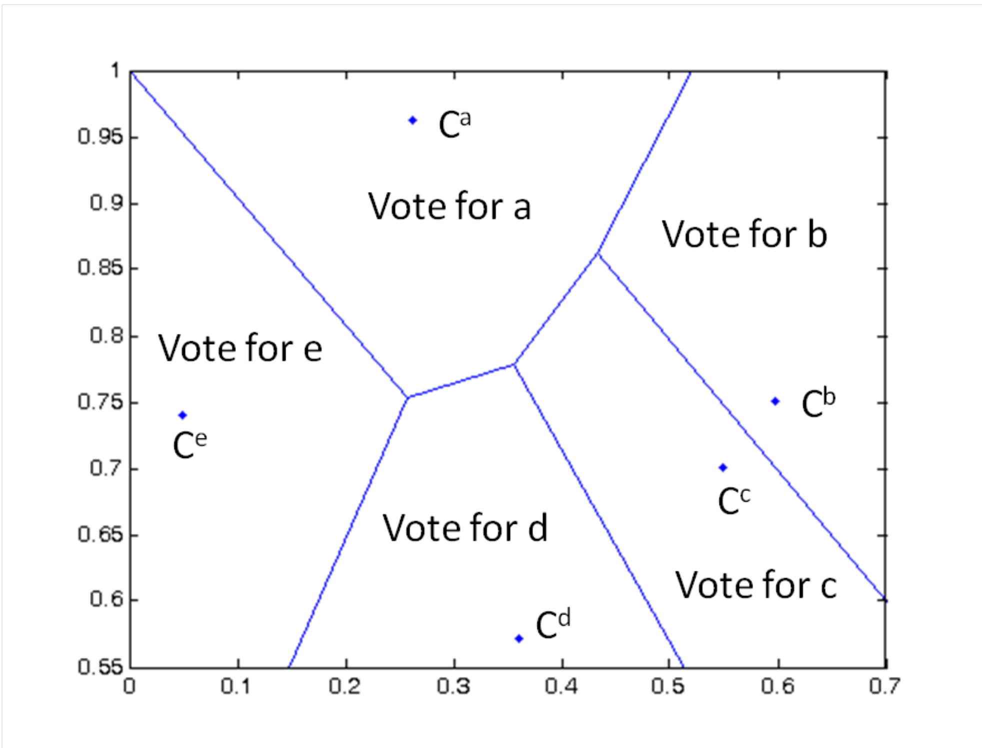

Figure 1: The Voronoi Tessellation for a 5-candidate election in $\mathbb{R}^2$.

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 $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ given $\mathbf{X}$ and ϵ.[13] Here, $\mathbf{X}$ represents observable characteristics at the electoral precinct level, such as average

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

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

[13]For a detailed discussion of conditional probability measures see Chapter 5 in Pollard (2002).

demographic and economic features, and ϵ stands for unobservable electoral precinct characteristics. For example, in our empirical illustration, the French constituency of Paris is one such electoral precinct, for which we have data on observable characteristics such as age, gender and education distribution of the precinct population at the time of the election. The 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 given $\mathbf{X}$ only. For notational convenience, we omit the conditioning variable for most of this section and refer to the distribution of voter locations simply as $\mathbb{P}_T$. Since the identification arguments can be repeated for strata defined by regressors, this is without loss of generality. Candidates are drawn from a distribution characterized by the measure $\mathbb{P}_C$, again absolutely continuous with respect to the Lebesgue measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. The proportion of votes obtained by each candidate is the probability of the Voronoi cell that contains the candidate's ideological position.

For each election, the observed data contain 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 $\mathcal{C}$ and preference type distribution $\mathbb{P}_T$, we can define the following object:

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

where $p(\mathcal{C}, \mathbb{P}_T)$ 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 $\mathbb{P}_T$. 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 $V_i(\mathcal{C}) = \{\mathbf{t} \in \mathbb{R}^d : d(\mathbf{t}, C^i) < d(\mathbf{t}, C^j), j \neq i\}$ is given by:

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

where $f_{T|X,\epsilon}$ is the density of $\mathbb{P}_{T|X,\epsilon}$ and analogously for $f_{T|X}$.

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., $\mathcal{C}$ 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.[14] It is also weaker than (unconditional) independence between regressors and coefficients typically required in the literature on discrete choice models with random coefficients (e.g., Ichimura and Thompson (1998) or Gautier and Kitamura (2008)).[15] In our case, the variables $\mathcal{C}$ act as a "special regressor", allowing us to identify the structure. The assumption is made explicit below:

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

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

Definition 1 (Identification) *Let $\mathbb{P}_{T_1}$ and $\mathbb{P}_{T_2}$ be two measures on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, both absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^d$. $\mathbb{P}_{T_1}$ is identified relative to $\mathbb{P}_{T_2}$ if and only if $p(\cdot, \mathbb{P}_{T_1}) = p(\cdot, \mathbb{P}_{T_2})$, Leb-a.s. $\Rightarrow \mathbb{P}_{T_1} = \mathbb{P}_{T_2}$.*[16]

In words, two preference type distributions 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 measure. We can now state the identification result:

[14] 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.

[15] We also assume that ϵ and $\mathcal{C}$ are conditionally independent given $\mathbf{X}$.

[16] *Leb.-a.s.* refers to the fact that the underlying measure is the Lebesgue measure on $\mathbb{R}^d \times \cdots \times \mathbb{R}^d$, the n-fold Cartesian product of $\mathbb{R}^d$. The factors relate to the number of candidates in the elections.

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

The proof is given in the Appendix for elections with any number of candidates. It basically 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 identification on the intersection of the two supports. Since the Cramér-Wold device does not require absolute continuity, in principle the result could be extended to discrete types. Because in our application the relevant variables are continuous we did not pursue this extension further. Similar arguments are used in Ichimura and Thompson (1998) to show identification of the unknown distribution for the random coefficients in a binary choice model. In that paper, the distribution of random coefficients has to be restricted to a subset of their space (i.e., a hemisphere of the normalized hypersphere where random coefficients realizations take their values). This "dilatation invariance" property is due to the particular structure of the binary choice model analyzed by Ichimura and Thompson (1998) and also by Gautier and Kitamura (2008), which is not shared by our model.[17]

2.1 Extensions

The canonical spatial model of voting analyzed above can easily be extended to accommodate more general voters' preferences (e.g., Hinich and Munger (1997)). In particular,

[17] In particular, in the papers by Ichimura and Thompson (1998) and Gautier and Kitamura (2008), choices (y) follow a linear index threshold crossing condition which is essentially an inner product between covariates (x) and random coefficients. This implies the "dilatation invariance" property that $\mathbb{P}(y|x) = \mathbb{P}(y|cx)$, where c is any positive scalar. Our problem deals with multinomial choices and relies on a nonlinear index comparing alternative choices of candidates. Hence, the property does not apply.

consider the case in which individual utility functions are decreasing functions of a weighted Euclidean distance $d^W(x,y) = \sqrt{(x-y)^\top W(x-y)}$ with weighting matrix W, assumed to be symmetric and positive definite.[18] According to the spatial theory of voting, the main diagonal elements in the matrix W subsume the relative importance to a voter of the different dimensions of the ideological space in a given election. The off-diagonal elements, on the other hand, describe the way in which voters make trade-offs among these different dimensions.

Voters preferences are now described by the pair $(\mathbb{P}_T, W)$: the distribution of voter bliss points in the population $\mathbb{P}_T$ and the weighting matrix W. Our definition of identification is extended to this setting by ascertaining that two relatively identified pairs $(\mathbb{P}_{T_i}, W_i),\ i = 1, 2$ cannot give rise to the same voting proportions as a function of candidate positions across a certain number of elections.

As before, let the individual bliss points be represented by the variable $\mathbf{T}$ (distributed according to $\mathbb{P}_T$). Furthermore, consider preferences based on the weighted distance with weighting matrix W. For a given set of candidates $C_1, \ldots, C_n$, let $V_i^W((C_j)_{j=1,\ldots,k})$ represent the Voronoi cell for candidate i. In other words,

$$V_i^W((C_j)_{j=1,\ldots,k}) = \{\mathbf{t} \in \mathbb{R}^d : d^W(\mathbf{t}, C_i) < d^W(\mathbf{t}, C_j), j \neq i\}, \qquad i \in \{1, \ldots, n\}.$$

Accordingly, let $V^W((C_j)_{j=1,\ldots,k}) = \left(V_i^W((C_j)_{j=1,\ldots,k})\right)_{i=1,\ldots,k}$. Since these Voronoi cells are the same for weighting matrices $\alpha W,\ \alpha > 0$, we impose the normalization that $||W||_{d\times d} = \sqrt{d}$ where $||W||_{d\times d} = \sqrt{Tr(W^\top W)}$ is the Frobenius norm. This in particular includes the d-order identity matrix as a possible choice for W. Once this normalization is imposed, we obtain the following result:

Lemma 1 *Suppose Assumption 1 holds,* $||W||_{d\times d} = \sqrt{d}$ *and there are at most* $d+1$ *candidates. Then* $(\mathbb{P}_T, W)$ *is identified.*

[18] Degan and Merlo (2009) also consider this extension which Okabe, Boots, Sugihara, and Chiu (2000), p.197, refer to as the elliptic distance with weighting matrix W.

The proof for this result is presented in the Appendix for elections with any number of candidates. Intuitively, under the elliptic distance d^W, the Voronoi cells when there are two candidates are separated by the affine hyperplane

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

and analogously for the elliptic distance $d^{\overline{W}}$. The two affine hyperplanes ($H^W(C_1, C_2)$ 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 share of votes under $(\mathbb{P}_T, W)$ as they would under $(\mathbb{P}_{\overline{T}}, \overline{W})$ (see Figure 2).

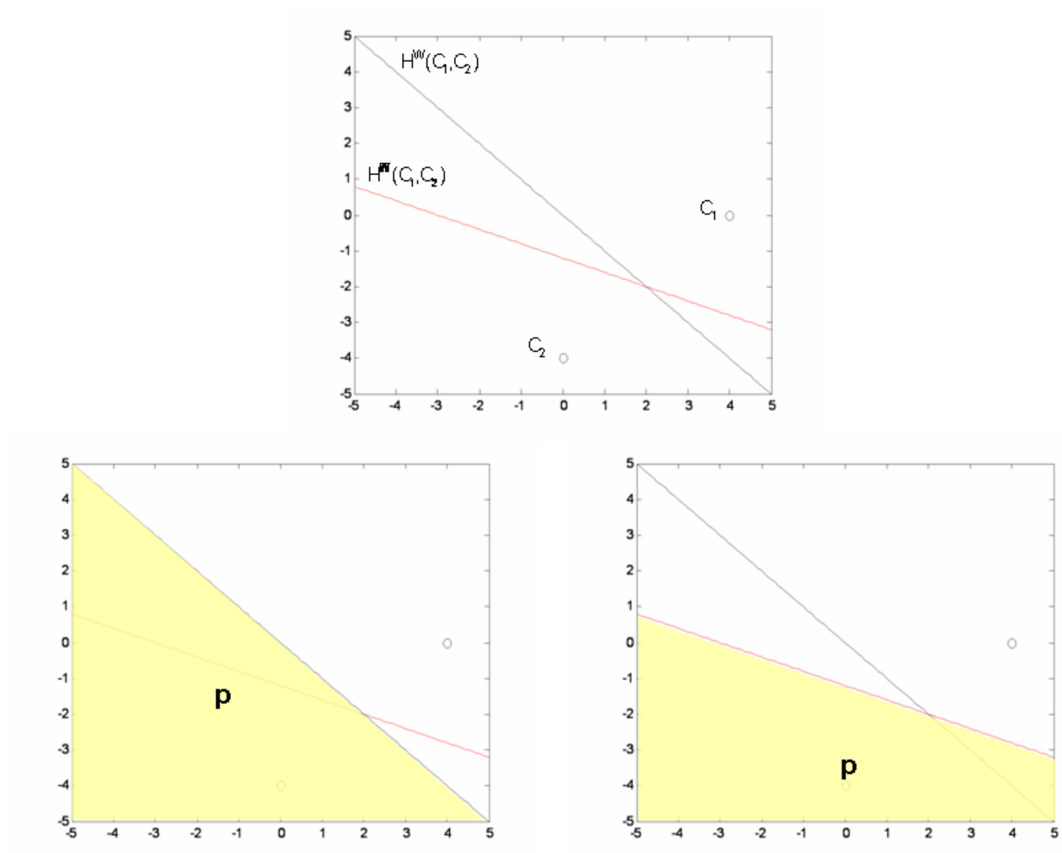

Figure 2: Voronoi Tessellations for Candidates C_1, C_2

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 under W is generated. The affine hyperplane characterizing the $\overline{W}$-Voronoi cells for the new pair (C_1', C_2') is parallel to the $\overline{W}$-Voronoi hyperplane for (C_1, C_2). Again, under the assumption of observational equivalence, these two cells under the $\overline{W}$ elliptic distance would have the same proportion of votes as with the unchanged Voronoi tessellation under W (see Figure 3).

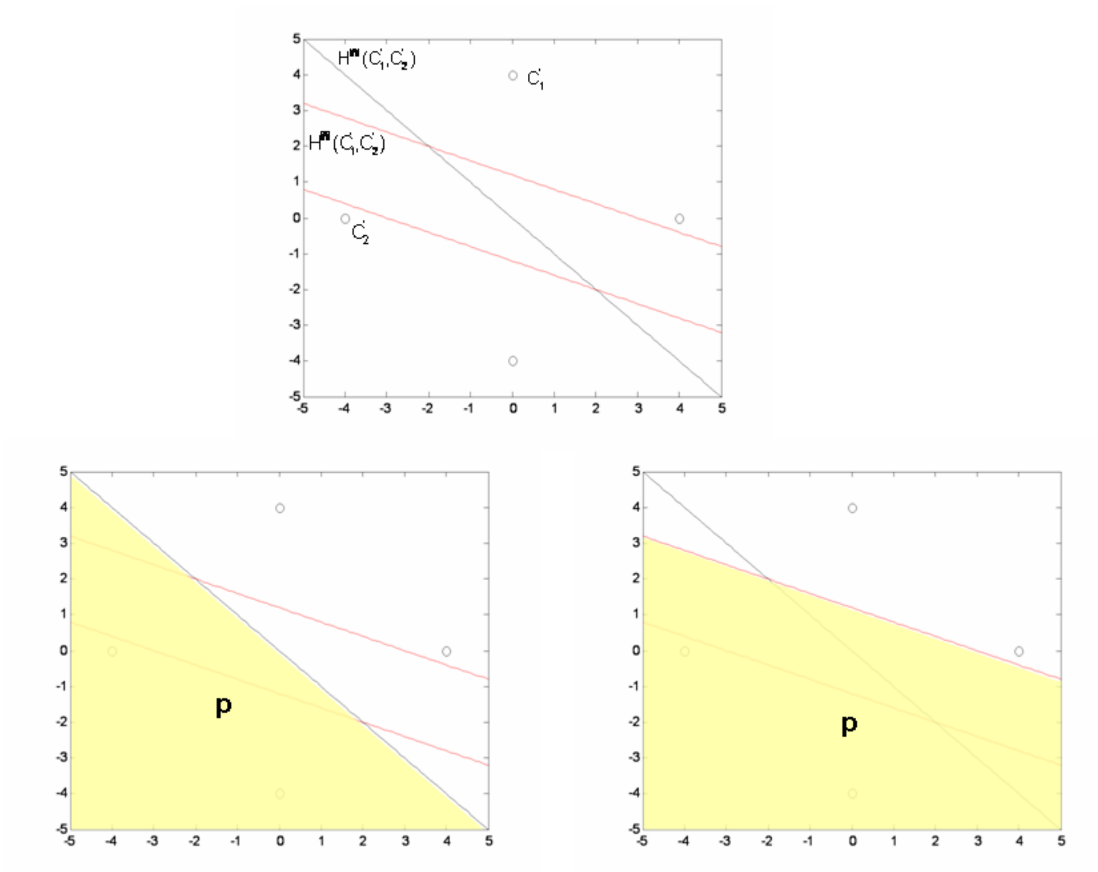

Figure 3: Voronoi Tessellations for Candidates C_1, C_2

This would imply the existence of a region with zero probability in the ideological space (under either $\mathbb{P}_T$ or $\mathbb{P}_{\overline{T}}$ as they are observationally equivalent). Since we can manipulate the argument to have any bounded set be contained in this region, any such set would have probability zero. We reach a contradiction as this would lead to the conclusion that the probability of the sample space ($\mathbb{R}^d$) is zero.

This proof strategy exploits the availability of multiple candidate profiles generating the same Voronoi tessellation (for weighting matrix W) and Lemma 1 extends the above argument for at most $d+1$ candidates. When there are more than $d+1$ candidates, the proof strategy cannot be applied since the existence of multiple profiles generating the same Voronoi tessellation is no longer guaranteed (see for instance the discussion in the proof for Theorem 14 in Ash and Bolker (1985) for $d = 2$). 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 $d+1$ 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. The following proposition summarizes the result:

Proposition 2 *Suppose Assumption 1 holds, $||W||_{d\times d} = \sqrt{d}$. Then $(\mathbb{P}_T, W)$ is identified.*

A natural corollary of Proposition 2 is that election specific (e.g., local, senatorial, gubernatorial, presidential) weights and bliss point distributions are identified (up to the normalization $||W^e||_{d\times d} = \sqrt{d}$). Consequently, voting records in multiple simultaneous elections are informative about different dimension weights ascribed by the voters in elections for different levels of government.

Another potential generalization would be to allow the weighting matrix W to be individual specific and to have the distribution of voter preferences range over bliss points and voting weights. We conjecture that in this case, identification would be lost as there would be too many degrees of freedom to fit the data. However, we were not able to find an appropriate argument to establish the result.

The ideas in Proposition 1 are also useful in more general settings using distance functions that are not Euclidean. The relative identifiability of two generic distance functions $d(\cdot,\cdot)$ and $\overline{d}(\cdot,\cdot)$ can be obtained in an analogous manner. We state this result below:

Proposition 3 *Suppose that Assumption 1 holds, that there are two candidates and that for two profiles* (C_1, C_2) *and* (C_1', C_2'),

$$\{\mathbf{t} \in \mathbb{R}^d : d(C_1, \mathbf{t}) = d(C_2, \mathbf{t})\} = \{\mathbf{t} \in \mathbb{R}^d : d(C_1', \mathbf{t}) = d(C_2', \mathbf{t})\}$$

and

$$\{\mathbf{t} \in \mathbb{R}^d : \overline{d}(C_1, \mathbf{t}) = \overline{d}(C_2, \mathbf{t})\} \cap \{\mathbf{t} \in \mathbb{R}^d : \overline{d}(C_1', \mathbf{t}) = \overline{d}(C_2', \mathbf{t})\} = \emptyset.$$

Then, $(\mathbb{P}_T, d(\cdot, \cdot))$ *and* $(\mathbb{P}_{\overline{T}}, \overline{d}(\cdot, \cdot))$ *are relatively identified.*

The proof follows along the lines of that for Lemma 1 and hence is omitted.

3 Estimation

Estimation in a one-dimensional ideological space is straightforward and is briefly discussed in the next subsection. In two or more dimensions, a different strategy is pursued and is discussed below.

3.1 A Simple Case

In the case of a one-dimensional ideological space, an election provides an estimate of the cumulative distribution function $F_T(t|\mathbf{X}) = \int_{-\infty}^{t} f_{u|\mathbf{X}}(T|\mathbf{X})du$ at the midpoints separating any two contiguous candidates. With n candidates in election e, assume without loss of generality that $C_{1,e} < C_{2,e} < \cdots < C_{n,e}$. The sum of the proportions of votes received by candidate $C_{i,e}$ and by all the candidates positioned to the left of $C_{i,e}$ gives an estimate of the cdf F_T at $\overline{C}_{i,e} \equiv \frac{C_{i,e}+C_{i+1,e}}{2}$ where , $i = 1, \ldots, n-1$. As more elections are sampled, we obtain an increasing number of points at which we can estimate the cdf. Let $Y_{i,e}$ be the proportion of votes obtained by candidates $C_{1,e}, \ldots, C_{i,e}$ in election e and assume there are v_e votes in this election. Notice that

$$\mathbb{E}(\mathbf{1}(\mathbf{T} \leq \overline{C}_{i,e})|\mathcal{C}_e, \mathbf{X}_e) = F_T(\overline{C}_{i,e}|\mathbf{X}_e).$$

Since $Y_{i,e} = \frac{\sum_{i=1}^{v_e} \mathbf{1}(\mathbf{T}_i \leq \overline{C}_{i,e})}{v_e}$,

$$\mathbb{E}(Y_{i,e}|\mathcal{C}_e, \mathbf{X}_e) = F_T(\overline{C}_{i,e}|\mathbf{X}_e)$$

and a natural estimator for F_T given m elections 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.

3.2 Multidimensional Ideological Space

When the number of dimensions of the ideological space is greater than one ($d > 1$), it is not possible to directly recover estimates for the cumulative distribution function as suggested above. It is nevertheless true that for a given election:

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

where $V(\mathcal{C})$ is the Voronoi cell for candidate i, $\tilde{\mathbf{X}} = (\mathbf{X}, \mathcal{C})$ and the expectation is taken with respect to the candidate positions, $\mathbf{X}$ and ϵ. The quantities $p_i, i \in \{1, \ldots, n\}$, are the electoral outcomes obtained from the data. This suggests estimating $f(\cdot)$ using a sieve minimum distance estimator as suggested in Ai and Chen (2003) (see also Newey and Powell (2003)). We follow here the notation in that paper. The estimator is the sample counterpart to the following minimization problem:

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

where $m(\tilde{\mathbf{X}}, f) = \mathbb{E}\left[\rho((p_i, C_i)_{i=1,\ldots,n}, \mathbf{X}, f)|\tilde{\mathbf{X}}\right]$ with

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

Notice that the n-th component of the above vector is omitted as the vector adds up to one. Here, we assume that elections have the same number of candidates. If this is not the case, 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 numbers of bidders.[19]

As pointed out by Ai and Chen (2003), two difficulties arise in constructing this estimator. First, the conditional expectation m is unknown. Second, the function space $\mathcal{H}$ may be too large. To address the first issue, a non-parametric estimator $\hat{m}$ is used in place of m. With regard to the second 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 $\{b_i(\tilde{\mathbf{X}}), i = 1, 2, \ldots\}$ denote a sequence of known basis functions (e.g., power series, splines, etc.) that approximate well square integrable real-valued functions of $\mathbf{X}$ and $\mathcal{C}$. With $b^J(\tilde{\mathbf{X}}) = \left(b_1(\tilde{\mathbf{X}}), \ldots, b_J(\tilde{\mathbf{X}})\right)^\top$, the sieve estimator for $m_i(\tilde{\mathbf{X}}, f)$, the i-th component in m, is given by

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

where $B = (b^J(\tilde{\mathbf{X}}), \ldots, b^J(\tilde{\mathbf{X}}))$ and, as before, e indexes the elections.

We consider the class $\mathcal{H}$ of densities studied by Gallant and Nychka (1987).[20] For simplicity, we omit the conditioning variable ($\mathbf{X}$) 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 $\epsilon_0 > 0$ and let $\phi(\mathbf{t})$ denote the multivariate standard normal density. The class $\mathcal{H}$ admits densities f such that:

$$f(\mathbf{t}, \xi) = h(\mathbf{t})^2 + \epsilon\phi(\mathbf{t})$$

with

$$\left(\sum_{|\lambda| \leq k_0} \int |D^\lambda h(\mathbf{t})|^2 (1 + \mathbf{t}^\top \mathbf{t})^{\delta_0} d\mathbf{t} \right)^{1/2} < \mathcal{B}_0 \tag{4}$$

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

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

where $\int f(\mathbf{t},\xi)d\mathbf{t}=1$, $\epsilon>\epsilon_0$,

$$D^{\lambda}f(\mathbf{t})=\frac{\partial_1^{\lambda}}{\partial x_1^{\lambda_1}}\frac{\partial_2^{\lambda}}{\partial x_2^{\lambda_2}}\cdots\frac{\partial_d^{\lambda}}{\partial x_d^{\lambda_d}}f(\mathbf{t}),\quad \lambda=(\lambda_1,\ldots,\lambda_d)^{\top}\in\mathbb{N}^d$$

and $|\lambda|=\sum_{i=1}^{d}\lambda_i$. Given a compact set on the ideological space, condition (4) 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 (4) 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 ($\epsilon=10^{-5}$ 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 consistency norm defined in the proof for Proposition 4):

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

where H_i are Hermite polynomials, ϕ is the standard multivariate normal density and ϵ is a small positive number.[21] As mentioned before, the set of densities on which $\cup_{E=1}^{\infty}\mathcal{H}_E$ is dense is fairly large. Because the parameter space is also compact with respect to the consistency

[21] In Gallant and Tauchen (1989) the functions are defined as follows. Let $z=R^{-1}(t-b-B\mathbf{x})$ where R and B are matrices of dimension $d\times d$ and $d\times\dim(\mathbf{x})$ respectively and b is a d-dimensional vector. Then,

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

where

$$h(\mathbf{z}|\mathbf{x})/\det(R)=\frac{\left[\sum_{|\alpha|=0}^{J_z}a_{\alpha}(\mathbf{x})\mathbf{z}^{\alpha}\right]^2\phi(\mathbf{z})}{\int\left[\sum_{|\alpha|=0}^{J_z}a_{\alpha}(\mathbf{x})\mathbf{U}^{\alpha}\right]^2\phi(\mathbf{U})d\mathbf{U}}$$

with $a(\mathbf{x})=\sum_{|\beta|=0}^{J_x}a_{\alpha\beta}\mathbf{x}^{\beta}$. The function $\mathbf{z}^{\alpha}$ maps the multi-index $\alpha=(\alpha_1,\ldots,\alpha_d)$ into the monomial $\mathbf{z}^{\alpha}=\Pi_{i=1}^{d}z_i^{\alpha_i}$ and analogously for $\mathbf{x}^{\beta}$ with respect to $\beta=(\beta_1,\ldots,\beta_{\dim(\mathbf{x})})$. Kim (2007) examines truncated versions of the GN sieve space on a compact support.

norm (see the proof for Proposition 4), ill-posedness of this inverse problem is not an issue (see Newey and Powell (2003)). The estimator is also very attractive computationally as integrals can be obtained analytically. We nevertheless approximate the integral of one such density over a particular Voronoi cell by simulation. 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. For the optimization, we use Nelder-Meade's non-gradient algorithm. As the number of draws increases, the approximation converges to the desired integral.

The estimator is formally defined as:

$$\widehat{f} = \text{argmin}_{f \in \mathcal{H}_{\mathcal{E}}} \frac{1}{E} \sum_{e=1}^{E} \widehat{m}(\tilde{\mathbf{X}}, f)^{\top} \left[\widehat{\Sigma}(\tilde{\mathbf{X}}) \right]^{-1} \widehat{m}(\tilde{\mathbf{X}}, f) \tag{5}$$

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

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) $(n-1)J \geq J_E$, $J_E \to \infty$ and $J/E \to 0$.*

The following proposition establishes consistency:

Proposition 4 *Under Assumptions 1-5,*

$$\widehat{f} \to_p f_T$$

with respect to the consistency norm defined by Gallant and Nychka (1987).

The estimator above can be easily extended for the weighted distance discussed in subsection 2.1. The parameters to be estimated are now given by $(W, f(\mathbf{t}|\mathbf{x}))$ where $W \in \Theta$, a (suitably normalized) space of matrices of dimension d. In this case, the estimator becomes:

$$(\widehat{W}, \widehat{f}) = \text{argmin}_{(W,f)\in\Theta\times\mathcal{H}_{\mathcal{E}}} \frac{1}{E}\sum_{e=1}^{E} \widehat{m}(\tilde{\mathbf{X}}, (W, f))^{\top} \left[\widehat{\Sigma}(\tilde{\mathbf{X}})\right]^{-1} \widehat{m}(\tilde{\mathbf{X}}, (W, f)) \tag{6}$$

where now

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

Consistency with respect to the product norm follows along the same lines as before.

Proposition 5 *Under Assumptions 2-5 and* Θ *compact (with respect to the Frobenius norm),*

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

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 changed 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.

4 Monte Carlo Experiments

In this section, we examine the small sample performance of the suggested estimation strategy in a few Monte Carlo experiments. We investigate models without covariates with three potential distribution of voter types. We use the distributions suggested in the Monte Carlo study by Ichimura and Thompson (1998) which are summarized in Table 1 and Figure 4.

For each of these, we postulate two different weighting matrices W for the elliptic distance function. The first one has $W_{1,2} = W_{2,1} = 0$ and $W_{2,2} = 2$, and the second $W_{1,2} = W_{2,1} = 0.5$ and $W_{2,2} = 2$. Both matrices are normalized to have $W_{1,1} = 1$. We assume that the analysis has 100 observations in each set of Monte Carlo experiments. Each observation contains the position and vote proportions for 2 candidates that are sampled uniformly over $[-1, 1]^2$. The proportions are estimated using (1000) draws from the voter type distribution in the data generating process. This introduces sampling error in the observed proportion of votes (i.e., an electoral precinct level ϵ) which differ in general from the numerical integration of the proposed type distribution over the candidate's Voronoi cell. We use 50 Monte Carlo repetitions for each one of the three models.

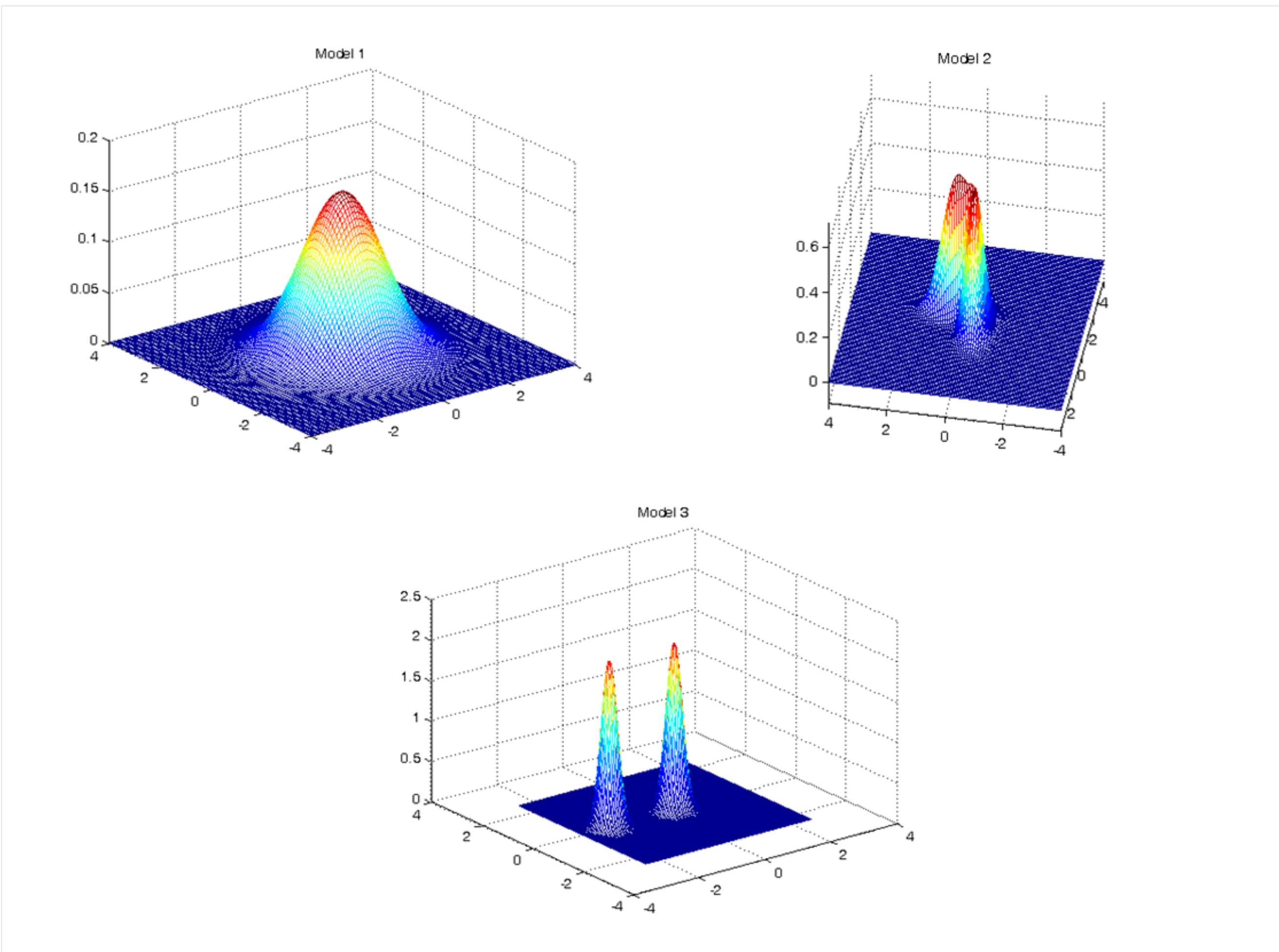

Figure 4: DGP Densities

Table 1: Data Generating Processes

Model 1:	$\mathcal{N}\left([0,0]',\mathbf{I}_2\right)$
Model 2:	Equiprobable mixture of
	$\mathbf{Y}_1 \sim \mathcal{N}\left(\begin{bmatrix} \mu \\ -\mu \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}\right)$
	and
	$\mathbf{Y}_2 \sim \mathcal{N}\left(\begin{bmatrix} -\mu \\ \mu \end{bmatrix}, \begin{bmatrix} \sigma_2^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_1^2 \end{bmatrix}\right)$
	$\mu = 0.3587, \sigma_1^2 = 0.2627, \sigma_2^2 = 0.06568, \rho = -0.1$
Model 3:	Y_1 and Y_2 independently distributed
	$Y_1 \sim \mathcal{N}\left(0,\sigma^2\right)$
	Y_2 an equally weighted mixture of X_1 and X_2
	$X_1 \sim \mathcal{N}\left(0.2806,\sigma^2\right), X_1 \sim \mathcal{N}\left(-1.6806,\sigma^2\right)$
	$\sigma^2 = 0.038462$

The estimation follows the guidelines prescribed in the previous section. For the estimation of $m(\cdot)$ we use linear splines (with cross-products) for Models 1 and 2 and simple linear projections for Model 3. The estimation weighting matrix $(\tilde{\Sigma})$ is the identity. In Tables 2, 3 and 4, we report squared bias, variance and MSE for the two parameters in the W matrix for each of the three models. We follow Blundell, Chen, and Kristensen (2007) in reporting similar quantities for the density estimates. Letting $\hat{f}_i$ be the estimate of f from the ith Monte Carlo simulation and letting $\overline{f}(\mathbf{t}) = \sum_{i=1}^{MC} \hat{f}_i(\mathbf{t})^2/MC$. The pointwise squared bias is then defined as $\left(\overline{f}(\mathbf{t}) - f(\mathbf{t})\right)^2$ and the pointwise variance is $\sum_{i=1}^{MC} \left(\hat{f}_i(\mathbf{t}) - \overline{f}_i(\mathbf{t})\right)/MC$. We report squared bias, variance and MSE integrated over a grid of 100×100 points.

Table 2: Monte Carlo Results: Model 1

$(W_{1,2}, W_{2,2}) = (0, 2)$			
Bias^2	Variance	MSE	J_E
$(0.0001, 0.0141, 4.4566 \times 10^{-5})$	$(0.0011, 0.0408, 2.2451 \times 10^{-5})$	$(0.0012, 0.0549, 6.7017 \times 10^{-5})$	0
$(0.2781, 0.3837, 4.8133) \times 10^{-5}$	$(0.0005, 0.0093, 3.8467 \times 10^{-5})$	$(0.0005, 0.0093, 5.1980 \times 10^{-5})$	1

$(W_{1,2}, W_{2,2}) = (0.5, 2)$			
Bias^2	Variance	MSE	J_E
$(0.0008, 0.0230, 6.4572 \times 10^{-5})$	$(0.0034, 0.0477, 4.6148 \times 10^{-5})$	$(0.0042, 0.0707, 1.1072 \times 10^{-4})$	0
$(0.0000, 0.0011, 4.0107 \times 10^{-5})$	$(0.0008, 0.0089, 5.6757 \times 10^{-4})$	$(0.0008, 0.0100, 4.5782 \times 10^{-5})$	1

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

Table 3: Monte Carlo Results: Model 2

$(W_{1,2}, W_{2,2}) = (0, 2)$			
Bias^2	Variance	MSE	J_E
$(0.0023, 0.4658, 0.0025)$	$(0.2206, 0.2212, 5.5267 \times 10^{-4})$	$(0.2228, 0.6871, 0.0031)$	0
$(0.0010, 0.0853, 0.0016)$	$(0.1215, 0.2108, 4.6720 \times 10^{-4})$	$(0.1224, 0.2961, 0.0020)$	1
$(0.0001, 0.0201, 0.0012)$	$(0.0912, 0.1316, 4.3440 \times 10^{-4})$	$(0.0913, 0.1517, 0.0016)$	2
$(0.0006, 0.0120, 9.3694 \times 10^{-4})$	$(0.0693, 0.0952, 3.9928 \times 10^{-4})$	$(0.0699, 0.1072, 0.0013)$	3
$(0.0001, 0.0088, 8.4013 \times 10^{-4})$	$(0.0556, 0.0900, 3.6408 \times 10^{-4})$	$(0.0557, 0.0988, 0.0012)$	4

$(W_{1,2}, W_{2,2}) = (0.5, 2)$			
Bias^2	Variance	MSE	J_E
$(0.2346, 2.5381, 0.0042)$	$(0.2391, 0.8908, 0.0042)$	$(0.4737, 3.4289, 0.0042)$	0
$(0.2435, 2.0193, 0.0038)$	$(0.2473, 1.0363, 0.0007)$	$(0.4908, 3.0556, 0.0046)$	1
$(0.2005, 1.9740, 0.0037)$	$(0.2497, 1.0579, 8.1940 \times 10^{-4})$	$(0.2005, 3.0319, 0.0045)$	2
$(0.1958, 1.0874, 0.0036)$	$(0.2458, 1.0874, 8.2810 \times 10^{-4})$	$(0.4416, 3.0167, 0.0044)$	3
$(0.1937, 1.9216, 0.0036)$	$(0.2439, 1.0867, 8.7007 \times 10^{-4})$	$(0.0180, 0.5403, 0.0045)$	4

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

Table 4: Monte Carlo Results: Model 3

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

Bias^2	Variance	MSE	J_E
$(0.0015, 0.0002, 0.0274)$	$(0.0442, 0.1275, 0.0014)$	$(0.0457, 0.1277, 0.0287)$	0
$(0.0008, 0.0111, 0.0274)$	$(0.0221, 0.0633, 0.0015)$	$(0.0229, 0.0633, 0.0289)$	1
$(0.0007, 0.0164, 0.0136)$	$(0.0938, 0.0317, 0.0125)$	$(0.0946, 0.0481, 0.0260)$	2
$(0.0036, 0.0064, 0.0149)$	$(0.0775, 0.0258, 0.0136)$	$(0.0812, 0.0323, 0.0285)$	3
$(0.0008, 0.0389, 0.0073)$	$(0.0244, 0.2360, 0.0131)$	$(0.0252, 0.2749, 0.0204)$	4

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

Bias^2	Variance	MSE	J_E
$(0.0021, 0.0208, 0.0279)$	$(0.0752, 0.4363, 0.0019)$	$(0.0773, 0.4571, 0.0019)$	0
$(0.0002, 0.0056, 0.0274)$	$(0.0186, 0.0226, 0.0016)$	$(0.0187, 0.0282, 0.0289)$	1
$(0.0004, 0.0561, 0.0133)$	$(0.1189, 0.1552, 0.0138)$	$(0.1193, 0.2113, 0.0271)$	2
$(0.0010, 0.0099, 0.0140)$	$(0.0880, 0.0226, 0.0139)$	$(0.0890, 0.0326, 0.0279)$	3
$(0.0001, 0.0301, 0.0071)$	$(0.0097, 0.1467, 0.0115)$	$(0.0098, 0.1768, 0.0186)$	4

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

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

5 Empirical Illustration

In this section, we illustrate the methodology described above with an empirical analysis of the 1999 election of the European Parliament.[22] Elections for the European Parliament take place under the proportional representation system and typically with closed party lists. This means that voters do not vote for specific candidates, but for parties. The fraction of votes received by a party 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 in its list are elected). Our data consist of ideological positions of the parties competing in the election, electoral outcomes and demographic characteristics, for each electoral precinct.

The ideological positions of the parties were obtained from Hix, Noury, and Roland (2006), who used roll-call data for the 1999-2004 Legislature of the European Parliament to generate two-dimensional ideological positions for each MP along the lines of the NOMINATE scores of Poole and Rosenthal (1997) for the US Congress.[23] 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)). Since the closed-list proportional representation system induces strong party cohesion, where elected representatives systematically vote along party lines, to obtain the ideological positions of the parties competing in the election, for each dimension, we use the average coordinate of individual candidates from a given party as the coordinate for the party's position (see, e.g., Degan and Merlo (2009)). Figure 5 depicts a typical datapoint (the Milano, Italy electoral precinct) with seven parties and reports the proportion of votes obtained by each party.

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

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

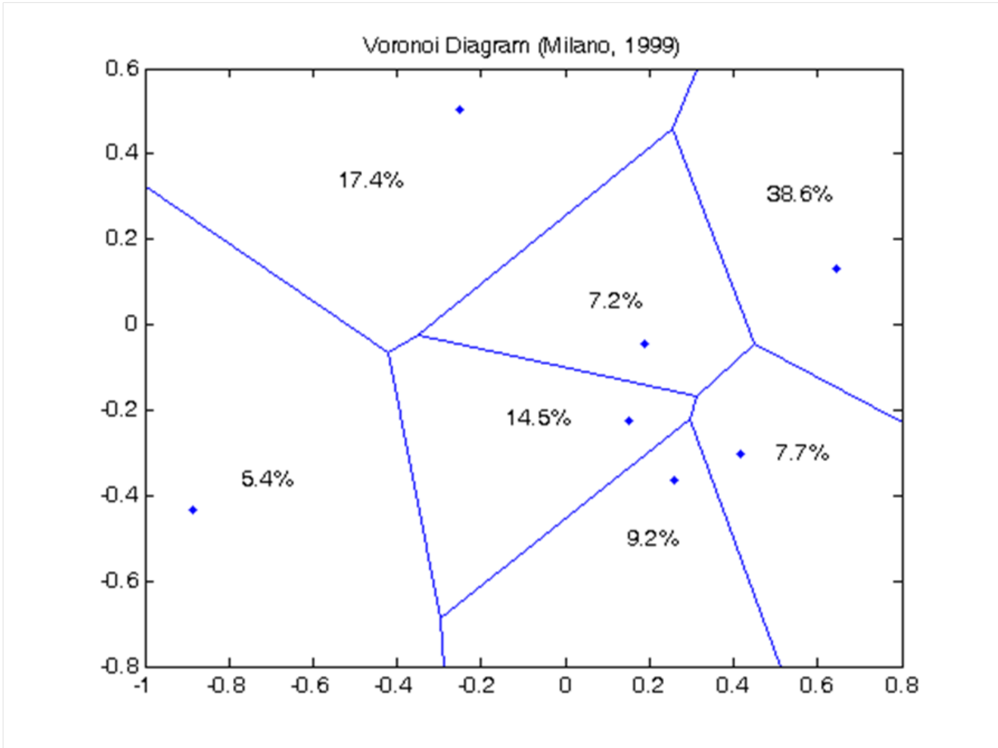

Figure 5: Voronoi Diagram for Milano (Italy), 1999

We combine the data on the ideological positions of parties with electoral outcomes in the 1999 elections and demographic information at the electoral precinct level (age, gender and education distribution of the precinct population) from the 2001 European Census. The election outcomes data were obtained from the CIVICACTIVE European Election Database.[24] The demographic data were obtained from EUROSTAT and we extracted three electoral precinct variables: the female-to-male ratio, the percentage of the population with secondary education and the percentage of the population older than 35.[25]

We estimate two versions of the model. In the first version, we include only one

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

[25]Our variables were obtained from `http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/search_database` under "Data Navigation Tree > Database by themes > Population and social conditions > Population (populat) > Census (cens) > Regional level census 2001 round (cens_r2001) > Educational level (cens_redu) > Population by sex, age group, highest educational attainment and occupation (cens_reisco)". The female-to-male ratio in this tabulation is lower than typically publicized figures and one may want to compute this variable using another tabulation.

covariate: the female-to-male ratio in the electoral precinct. There are 846 electoral precincts in our dataset with observations on this covariate. The results are summarized in Figure 6. The Hermite polynomials used were of order 2 (types) and 1 (demographic covariate).

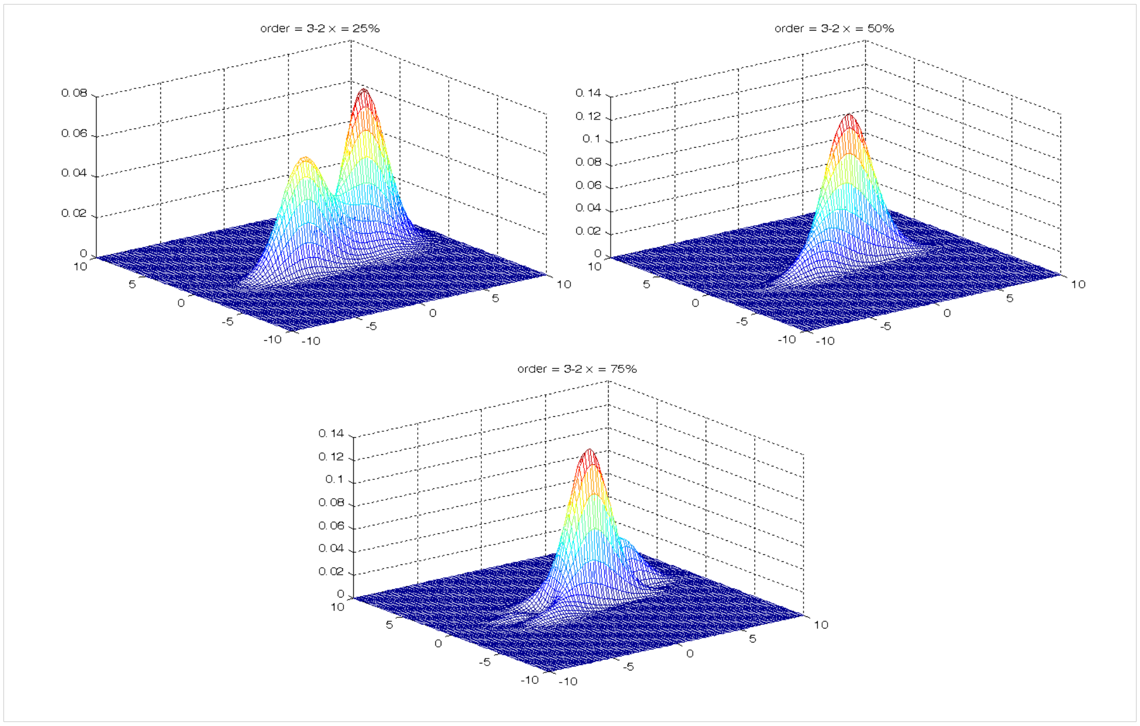

Figure 6: Estimation Results at Percentiles of Conditioning Variable

In our second specification we use all three covariates: the female-to-male ratio (X_1), the percentage of the population with secondary education (X_2) and the percentage of the population older than 35 (X_3). There are 437 electoral precincts in the data with observations on all three covariates. We had to exclude Germany as we could not find information on the two additional covariates at the electoral precinct level. Following Gallant and Tauchen (1989), we re-scale the data to avoid situations in which extremely large or small values of the polynomial part of the conditional density are required to compensate for extremely small or large values of the exponential part. Following these authors, we transform the data so that $\tilde{\mathbf{x}}_e = \mathbf{S}^{-1/2}(\mathbf{x}_e - \overline{\mathbf{x}})$ 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 results are summarized in Figures 7 and 8.

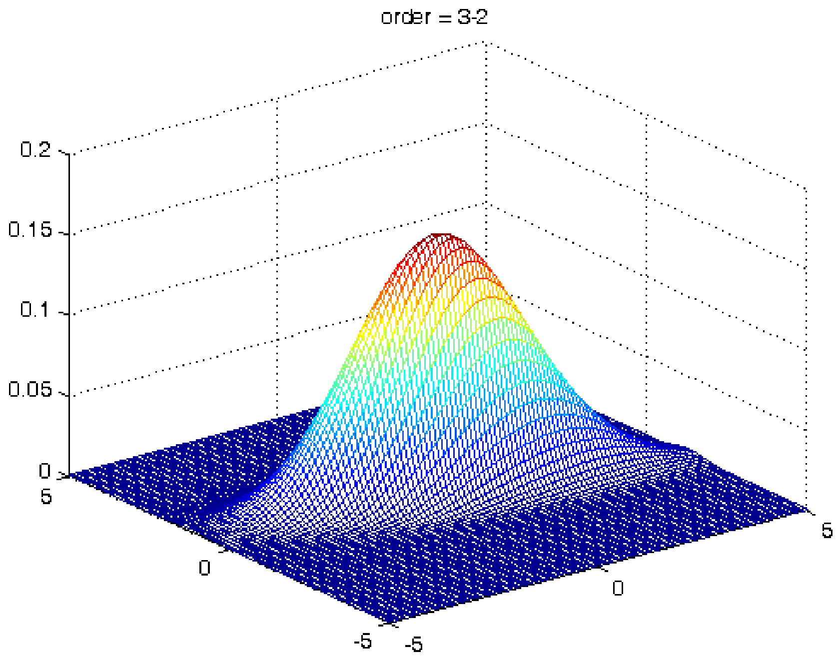

Figure 7: Estimation Results at Mean of Conditioning Variables

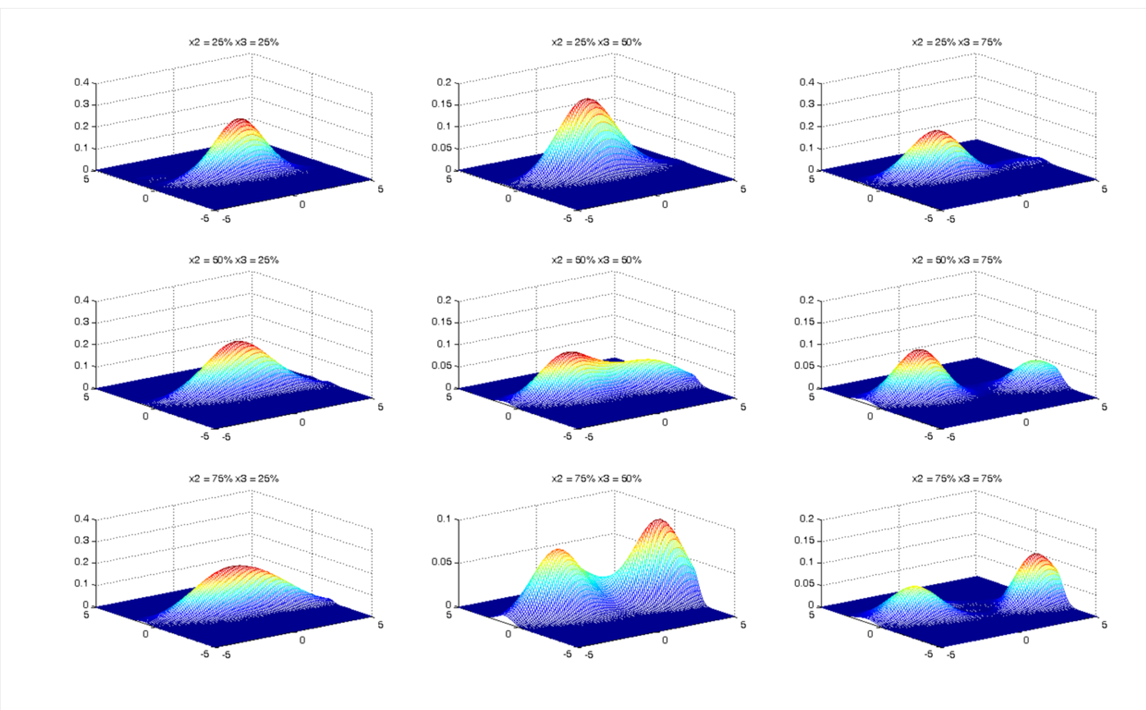

Figure 8: Estimation Results at Percentiles of Conditioning Variable ($X_1 = 50\%$)

The estimates for $m(\cdot)$ were linear projections on covariates. In both the one-covariate and the three covariate cases, we use Hermite polynomials of order 2 (types) and 1 (demographic covariates), which provided the lowest values for the objective function among the specifications we experimented with.

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 domestic socio-economic issues, and the second dimension to positions regarding European integration policies. For both our specifications, we observe a negative association between the two dimensions of the preference type distribution. Following Hix, Noury and Roland's interpretation, this implies that voters who are on the right of the left-right scale on domestic policies also tend to be less supportive of European integration policies. This relation is by no means a simple one though, as multiple local modes seem to be present in the distribution of voter types for different values of the demographic covariates. Another finding that is robust to the covariates specification is that the first ideological dimension appears to be substantially more important to voters than the second one. The parameter W_{22} for the weighting matrix is estimated at $\hat{W}_{22} = 0.0939$ for the three covariate specification and $\hat{W}_{22} = 0.0974$ for the one covariate specification. On the other hand, the estimates of the voters' trade-off between the two ideological dimensions we obtain vary with the model specification. The off-diagonal element W_{12} is estimated at $\hat{W}_{12} = 0.2451$ for the three covariate case and at $\hat{W}_{12} = -0.2195$ for the one covariate specification.

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)). They are, however, 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.

References

Ai, C., and X. Chen (2003): "Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions," *Econometrica*, 71(6), 1795–1843.

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

Andrews, D. W. K. (1992): "Generic Uniform Convergence," *Econometric Theory*, 8(2), 241–257.

Ash, P., and E. Bolker (1985): "Recoginizing Dirichlet Tessellations," *Geometriae Dedicata*, 19, 175–206.

Bai, J., and S. Ng (2006a): "Confidence Intervals for Diffusion Index Forecasts and Inference for Factor-Augmented Regressions," *Econometrica*, 74(4), 1133–1150.

——— (2006b): "Evaluating latent and observed factors in macroeconomics and finance," *Journal of Econometrics*, 131(1-2), 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(4), pp. 841–890.

Berry, S., and A. Pakes (2007): "The Pure Characteristics Demand Model," *International Economic Review*, 48, 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(6), 1613–1669.

Chesher, A., and J. M. C. S. Silva (2002): "Taste Variation in Discrete Choice Models," *The Review of Economic Studies*, 69(1), pp. 147–168.

Coppejans, M., and A. R. Gallant (2002): "Cross-Validated SNP Density Estimates," *Journal of Econometrics*, 110, 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, 426–448.

Degan, A., and A. Merlo (2009): "Do Voters Vote Ideologically?," *Journal of Economic Theory*, 144, 1869–1894.

Donald, S. G., and H. J. Paarsch (1993): "Piecewise Pseudo-Maximum Likelihood Estimation in Empirical Models of Auctions," *International Economic Review*, 34(1), 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.

Fenton, V. M., and A. R. Gallant (1996a): "Convergence Rates of SNP Density Estimators," *Econometrica*, 64(3), 719–727.

——— (1996b): "Qualitative and Asymptotic Performance of SNP Density Estimators," *Journal of Econometrics*, 74(1), 77–118.

Friedman, J., and R. Tibshirani (1984): "The Monotone Smoothing of Scatterplots," *Technometrics*, 26(3), 243–250.

Gallant, A. R., and D. W. Nychka (1987): "Semi-Nonparametric Maximum Likelihood Estimation," *Econometrica*, 55(2), 363–390.

Gallant, A. R., and G. Tauchen (1989): "Seminonparametric Estimation of Conditionally Constrained Heterogeneous Processes: Asset Pricing Applications," *Econometrica*, 57(5), 1091–1120.

Gautier, E., and Y. Kitamura (2008): "Nonparametric Estimation in Random Coefficients Binary Choice Models," Working Paper, Yale University.

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

He, X., and P. Shi (1998): "Monotone B-Spline Smoothing," *Journal of the American Statistical Association*, 93(442), 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, S142–S189.

Henry, M., and I. Mourifié (2010): "Revealed Euclidean Preferences," Université de Montreal Working Paper.

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, 494–511.

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, 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, 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(2), 724–740.

Manski, C. (1988): “Identification of Binary Response Models,” *Journal of the American Statistical Association*, 83(403).

Matzkin, R. (1992): “Nonparametric Identification and Estimation of Polychotomous Choice Models,” *Journal of Econometrics*, 58.

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(2), 741–750.

Newey, W. K. (1991): “Uniform Convergence in Probability and Stochastic Equicontinuity,” *Econometrica*, 59(4), 1161–1167.

Newey, W. K., and J. L. Powell (2003): “Instrumental Variable Estimation of Nonparametric Models,” *Econometrica*, 71(5), 1565–1578.

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

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(4), 425–441.

Scheike, T. H. (1994): "Anisotropic Growth of Voronoi Cells," *Advances in Applied Probability*, 26(1), 43–53.

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(460), 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(1), 278–286.

Appendix: Proofs

Proof of Proposition 1

It is enough to consider a single election with n candidates. In what follows, $\mathcal{M}_{k\times l}$ is the space of $k \times l$ real matrices which is endowed with the typical Frobenius matrix norm $||A||_{k\times l} = \sqrt{\mathrm{Tr}(A^\top A)}$ for $A \in \mathcal{M}_{k\times l}$. Accordingly, $||A||_k$ is the typical Euclidean norm in $\mathbb{R}_k$. The product metric space $\mathcal{M}_{k\times l} \times \mathbb{R}_m$ is endowed with the normed product metric $d((A_1, b_1), (A_2, b_2)) = \sqrt{||A_1 - A_2||^2_{k\times l} + ||b_1 - b_2||^2_m}$.

Step 1: $(\exists (A^*, b^*) \in \mathcal{M}_{n-1\times d} \times \mathbb{R}_{n-1} : \mathbb{P}_{T_1}(\{\mathbf{T} \in \mathbb{R}_d : A^*\mathbf{T} \leq b^*\}) \neq \mathbb{P}_{T_2}(\{\mathbf{T} \in \mathbb{R} : A^*\mathbf{T} \leq b^*\}))$ Suppose that $\mathbb{P}_{T_1}(\{\mathbf{T} \in \mathbb{R}_d : A\mathbf{T} \leq b\}) = \mathbb{P}_{T_2}(\{\mathbf{T} \in \mathbb{R}_d : A\mathbf{T} \leq b\}), \forall A, b$. 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

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

and

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

Since the probabilities of $\{\mathbf{T} \in \mathbb{R}_d : A\mathbf{T} \leq b\}$ 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,

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

Step 2: $(\exists \eta > 0 : \mathbb{P}_{T_1}(\{\mathbf{T} \in \mathbb{R}_d : A\mathbf{T} \leq b\}) \neq \mathbb{P}_{T_2}(\{\mathbf{T} \in \mathbb{R}_d : A\mathbf{T} \leq b\}), \forall (A, b) \in \mathcal{N}((A^*, b^*), \eta))$ We claim that

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

is continuous at $(A,b) \in \mathcal{M}_{n-1\times d}\times\mathbb{R}_{n-1}$. Take a sequence $(A_k,b_k)_{k=1}^{\infty}$ such that $d((A_k,b_k),(A,b)) \stackrel{k\to\infty}{\longrightarrow} 0$. Then

$$\begin{aligned}|h(A,b)-h(A_k,b_k)| \leq \quad & |\mathbb{P}_{T_1}(\{\mathbf{T}\in\mathbb{R}_d : A\mathbf{T}\leq b \wedge A_k\mathbf{T} > b_k\}| + \\ & |\mathbb{P}_{T_1}(\{\mathbf{T}\in\mathbb{R}_d : A_k\mathbf{T}\leq b_k \wedge A\mathbf{T} > b\}|\end{aligned}$$

Note that

$$\limsup_k \{\mathbf{t}\in\mathbb{R}_d : A\mathbf{t}\leq b\wedge A_k\mathbf{t} > b_k\} = \cap_m \cup_{k\geq m} \{\mathbf{t}\in\mathbb{R}_d : A\mathbf{t}\leq b\wedge A_k\mathbf{t} > b_k\}$$

and $\mathbf{t}$ belongs to this set if it belongs to $\{\mathbf{t}\in\mathbb{R}_d : A\mathbf{t}\leq b\wedge A_k\mathbf{t} > b_k\}$ for infinitely many k. Since $d((A_k,b_k),(A,b)) \stackrel{k\to\infty}{\longrightarrow} 0 \Rightarrow ||A_k - A||_{n-1\times d} \stackrel{k\to\infty}{\longrightarrow} 0$ and $||b_k - b||_{n-1}\to 0$, for any fixed $\mathbf{t}\in\mathbb{R}_d$, $||(A_k\mathbf{T}-b_k)-(A\mathbf{t}-b)||_{n-1} \stackrel{k\to\infty}{\longrightarrow} 0$. Hence, $A\mathbf{t}-b\geq 0$ if and only if there is K such that $k > K$ implies that $A_k\mathbf{t}-b_k\geq 0$. This means that

$$\limsup_k \{\mathbf{t}\in\mathbb{R}_d : A\mathbf{t}\leq b\wedge A_k\mathbf{t} > b_k\} = \emptyset.$$

Likewise,

$$\liminf_k \{\mathbf{t}\in\mathbb{R}_d : A\mathbf{t}\leq b\wedge A_k\mathbf{t} > b_k\} = \cup_m \cap_{k\geq m} \{\mathbf{t}\in\mathbb{R}_d : A\mathbf{t}\leq b\wedge A_k\mathbf{t} > b_k\}$$

and $\mathbf{t}$ belongs to this set if there is m such that $A_k\mathbf{T}\leq b_k$ for every $k\geq m$. Again because $||(A_k\mathbf{t}-b_k)-(A\mathbf{t}-b)||_{n-1} \stackrel{k\to\infty}{\longrightarrow} 0$, $A\mathbf{t}\geq b$ if and only if there is m such that $A_k\mathbf{t}\geq b_k$ for every $k\geq m$. Hence,

$$\liminf_k \{\mathbf{t}\in\mathbb{R}_d : A_k\mathbf{t}\leq b_k\} = \emptyset.$$

Finally, this means that

$$\lim_k \{\mathbf{t}\in\mathbb{R}_d : A\mathbf{t}\leq b\wedge A_k\mathbf{t} > b_k\} = \emptyset.$$

Countable additivity then implies that

$$\lim_k \mathbb{P}_{T_1}(\{\mathbf{T}\in\mathbb{R}_d : A\mathbf{T}\leq b\wedge A_k\mathbf{T} > b_k\}) = \mathbb{P}_{T_1}(\lim_k \{\mathbf{T}\in\mathbb{R}_d : A\mathbf{T}\leq b\wedge A_k\mathbf{T} > b_k\}) = 0.$$

A similar argument holds for $\{\mathbf{t}\in\mathbb{R}_d : A_k\mathbf{t}\leq b_k\wedge A\mathbf{t} > b\}$. Consequently,

$$|h(A,b)-h(A_k,b_k)| \stackrel{k\to\infty}{\longrightarrow} 0$$

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

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

is a continuous function of (A, b).

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

<u>Step 3</u>: **($\{\mathbf{t} \in \mathbb{R}_d : A\mathbf{t} < b\}$ 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

$$R_1 \equiv \{\mathbf{t} \in \mathbb{R}_d : A\mathbf{t} < b\}$$

represents a Voronoi cell for some set of candidates, we use the fact that a tessellation of $\mathbb{R}_d$ into polyhedra $R_1, R_2, \ldots, R_n$ is a Voronoi tesselation if and only if there are points $\mathcal{C} = \{C_1, \ldots, C_n\} \subset \mathbb{R}_p$ 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 $\{\mathbf{t} \in \mathbb{R}_d : A\mathbf{t} \leq b\} \neq \emptyset$ (otherwise $\mathbb{P}_{T_1}(\emptyset) = \mathbb{P}_{T_2}(\emptyset) = 0$ contradicting Step 1). Furthermore, $\{\mathbf{t} \in \mathbb{R}_d : A\mathbf{t} < b\} \neq \emptyset$ as well. Otherwise, since both $\mathbb{P}_{T_1}$ and $\mathbb{P}_{T_2}$ are absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}_d$, $\mathbb{P}_{T_1}(\{\mathbf{T} \in \mathbb{R}_d : A\mathbf{T} \leq b\}) = \mathbb{P}_{T_2}(\{\mathbf{T} \in \mathbb{R}_d : A\mathbf{T} \leq b\}) = 0$, again contradicting Step 1. 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. 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 from each of the $\mathcal{N}(C_i^*, \epsilon)$ balls and let $\mathcal{S}(\epsilon) \equiv \{(A, b) \in \mathcal{M}_{n-1,d} \times \mathbb{R}_{n-1} : \{\mathbf{t} \in \mathbb{R}_d : A\mathbf{t} > b\}$ 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 ($\epsilon_1 \geq \epsilon_2 \Rightarrow \mathcal{S}(\epsilon_1) \supset \mathcal{S}(\epsilon_2)$) and, given the order topology, 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 $\times_{i=1}^n \mathcal{N}(C_i^*, \epsilon)$ has positive Lebesgue measure on the n-fold Cartesian product of $\mathbb{R}_d$ 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 $\mathbb{P}_{T_2}$. ■

Proof of Lemma 1

Consider two different spatial voting models characterized by $(\mathbb{P}_T, W)$ and $(\mathbb{P}_{\overline{T}}, \overline{W})$. If $W = \overline{W}$, identification follows along the lines of the first proposition. 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 $\mathbb{R}^d$ are shared by (the closure of) $d+1$ cells (see Theorem 9 and subsequent remark on Ash and Bolker (1985), p.185). Consequently, if there are at most $d+1$ 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 k candidates.

If $k = d + 1$ 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

$$\begin{aligned}
& d^W(C_i', \mathbf{t}) - d^W(C_j', \mathbf{t}) = 0 & \Leftrightarrow \\
\Leftrightarrow \quad & (C_i' - \mathbf{t})^\top W (C_i' - \mathbf{t}) - (C_j' - \mathbf{t})^\top W (C_j' - \mathbf{t}) = 0 & \Leftrightarrow \\
\Leftrightarrow \quad & C_i'^\top W C_i' - C_j'^\top W C_j' - 2(C_i' - C_j')^\top W \mathbf{t} = 0 & \Leftrightarrow \\
\Leftrightarrow \quad & (2C_i - P)^\top W (2C_i - P) - (2C_j - P)^\top W (2C_j - P) - 4(C_i - C_j)^\top W \mathbf{t} = 0 & \Leftrightarrow \\
\Leftrightarrow \quad & C_i^\top W C_i - C_j^\top W C_j - (C_i - C_j)^\top W P - (C_i - C_j)^\top W \mathbf{t} = 0 &
\end{aligned}$$

Since P is 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

$$\begin{aligned}
& C_i^\top W C_i - C_j^\top W C_j - 2(C_i - C_j)^\top W \mathbf{t} = 0 \quad \Leftrightarrow \\
\Leftrightarrow \quad & d^W(C_i, \mathbf{t}) - d^W(C_j, \mathbf{t}) = 0.
\end{aligned}$$

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

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

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

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

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

These define $k-1$ linear equations on $\mathbf{t} \in \mathbb{R}^d$. Since $k-1 < d$, the solution set for this equation contains at least one element. 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 $\mathcal{C}'$, sets of size $n \leq d+1$ 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

$$\begin{aligned} \mathbf{t} \in H^{\overline{W}}(C_i', C_j') \quad &\Rightarrow C_i'^{\top}\overline{W}C_i' - C_j'^{\top}\overline{W}C_j' - 2(C_j' - C_i')^{\top}\overline{W}\mathbf{t} = 0 \\ &\Rightarrow \tfrac{1}{2}\left(C_i'^{\top}\overline{W}C_i' - C_j'^{\top}\overline{W}C_j'\right) - (C_j - C_i)^{\top}\overline{W}\mathbf{t} = 0. \end{aligned} \tag{7}$$

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

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

By definition $H^{\overline{W}}(C_i, C_j)$ is also a translation of this hyperplane.

Step 3: ($\exists i$ 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 $\mathbb{R}^d$ (see Hartvigsen (1992)). It can then be represented as:

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

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

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

Similarly,

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

Because $V^{\overline{W}}(\mathcal{C})$ and $V^{\overline{W}}(\mathcal{C}')$ have parallel faces (see (7)), we have

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

Furthermore, expression (7) gives that the $d-1$ 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, $V_i^{\overline{W}}(\mathcal{C}) = V_i^{\overline{W}}(\mathcal{C}')$. Since $||W|| = ||\overline{W}||$ and $V_i^W(\mathcal{C}) = V_i^W(\mathcal{C}')$ this can only happen in a set of candidates of zero measure. This follows because a given set of candidates defines its Voronoi cells by "growing ellipsoids", all at the same rate and with axes determined by the weighting matrix (W or $\overline{W}$) (see Scheike (1994), p.45). The axes of the ellipsoid associated with W are vectors proportional to its eigenvectors (see Strang (1988), pp.334-336). Hence, if W and $\overline{W}$ have different sets of eigenvectors, their Voronoi cells for a given set of candidates grow with different orientations and there cannot be i such that $V_i^W(\mathcal{C}) = V_i^W(\mathcal{C}')$ and $V_i^{\overline{W}}(\mathcal{C}) = V_i^{\overline{W}}(\mathcal{C}')$. If W and $\overline{W}$ have the same eigenspaces, $V_i^W(\mathcal{C}) = V_i^W(\mathcal{C}') = V_i^{\overline{W}}(\mathcal{C}) = V_i^{\overline{W}}(\mathcal{C}')$ as long as $C_i - C_j$ belongs to an eigenspace of W (or equivalently in this case, $\overline{W}$). This configuration has measure zero (by application of Fubini's theorem for null sets for example).

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

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

and $V_i^{\overline{W}}(\mathcal{C}^\top) \subset V_i^{\overline{W}}(\mathcal{C})$. 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

$$\Delta_{il} \leq 0 \Leftrightarrow \Delta_{li} \geq 0.$$

Then,

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

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

To generalize the above argument by induction, assume the claim is true if i is such that $\Delta_{ij} \geq 0$ for all but $s-1$ indices. Above we showed that this holds for $s = 2$. We will now show that one can obtain l such that $\mathbf{b}_l - \mathbf{b}_l' \geq 0$ there is i such that for all but s indices, $\Delta_{ij} \geq 0$. For those indices l such that $\Delta_{il} \leq 0$, we have $\Delta_{li} \geq 0$. Take one of them and, as in (6), $\Delta_{lj} \geq 0$ for all those indices j such that $\Delta_{ij} \geq 0$. Since $\Delta_{li} \geq 0$, there are at most $s-1$ indices such that $\Delta_{lm} \leq 0$. By induction, $\mathbf{b}_l - \mathbf{b}_l' \geq 0$ and $V_i^{\overline{W}}(\mathcal{C}') \subset V_i^{\overline{W}}(\mathcal{C})$. If inequalities are reversed, $\mathbf{b}_l - \mathbf{b}_l' \leq 0$, the inclusion is itself reversed.

Step 4: ($\mathbb{P}_{\overline{T}}(\mathbb{R}^d) = 0$, leading to a contradiction) Select a neighborhood $\mathcal{N}$ in $\mathbb{R}^d \times \cdots \times \mathbb{R}^d$ such that for any candidate profile $\mathcal{C} \in \mathcal{N}$, the corresponding $\mathcal{C}'$ (generated as in Step 1) and assume without loss of generality that for i, $V_i^{\overline{W}}(\mathcal{C}) \subset V_i^{\overline{W}}(\mathcal{C}')$.

Because $\mathbb{R}^d$ is a separable metric space and consequently second-countable, it can be covered by a countable family of bounded, open sets (start with the cover $\{\mathcal{N}(x,\epsilon)\}_{x\in\mathbb{R}^d}$ where $\mathcal{N}(x,\epsilon)$ is an ϵ-ball around x for some $\epsilon > 0$ and use Lindelöf's Theorem to obtain a countable subcover). Let B be a set in this sub-cover such that for any candidate profile $\mathcal{C} \in \mathcal{N}$

$$B \subset V_i^{\overline{W}}(\mathcal{C}') \backslash V_i^{\overline{W}}(\mathcal{C}).$$

This can always be achieved by selecting a small enough neighborhood $\mathcal{N}$. Since the $(\mathbb{P}_T, W)$ and $(\mathbb{P}_{\overline{T}}, \overline{W})$ are observationally equivalent, for (almost) every profile in $\mathcal{N}$,

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

where $p(\cdot; \mathbb{P}_T, W)$ is the vector of shares that each candidate gets under $(\mathbb{P}_T, W)$. Consider one such profile $\mathcal{C}$.

For this profile, let p denote the proportion of votes obtained by candidate C_i:

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

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

Then consider $\mathcal{C}'$ generated as in Step 1. Because $V^W(\mathcal{C}) = V^W(\mathcal{C}')$, the proportion of votes obtained by candidate $C_i^\top$ under W is also p:

$$p = \mathbb{P}_T(V_i^W(\mathcal{C}'))$$

Since (almost) every candidate profile in $\mathcal{N}$ 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 $\mathcal{C}'$. Otherwise, focus on the set of generated candidates of $\mathcal{C}'$. Then this set of candidate profiles has positive measure and the outcomes under $(\mathbb{P}_T, W)$ and $(\mathbb{P}_{\overline{T}}, \overline{W})$ are distinct. This would establish identification.

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

$$p = \mathbb{P}_{\overline{T}}(V^{\overline{W}}(\mathcal{C}'))$$

Furthermore, note that

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

and consequently $\mathbb{P}_{\overline{T}}(B) = 0$. But since this is an arbitrary B in the subcover, this implies that $\mathbb{P}_{\overline{T}}(\mathbb{R}^d) = 0$, a contradiction. ■

Proof of Proposition 2

If there are at most $d+1$ candidates, Lemma 1 establishes the results. If $n > d+1$, consider first a Voronoi tessellation with $d+1$ candidates $C_1, \ldots, C_{d+1}$. In this case, apply Lemma

1 to obtain the existence of a set of candidate profiles with positive measure such that $p((C_1,\ldots,C_{d+1});\mathbb{P}_T,W) \neq p((C_1,\ldots,C_{d+1});\mathbb{P}_{\overline{T}},\overline{W})$.

Now, let i be a candidate for which voting shares are distinct under the two environments. Since there are at least $d+1$ 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 $Y \in \mathbb{C}, C_i + \alpha(Y - C_i) \in \mathbb{C},\ \forall \alpha > 0$). Since P is equidistant from all $d+1$ candidates by definition, there is a hypersphere $\mathbb{S}^W$ centered at P that contains every candidate. For any points $C_{d+2},\ldots,C_n$ in $\mathbb{C}\backslash\mathbb{S}^W$, $V_i^W((C_1,\ldots,C_{d+1})) = V_i^W((C_1,\ldots,C_n))$ as the hyperplane bisecting C_i and C_{d+2} so chosen does not intersect $V_i^W((C_1,\ldots,C_{d+1}))$. This is the case when $C_{d+2} = \alpha(C_j - C_i)$ where $\alpha > 1$ and C_j is any other candidate in the hypersphere. By locating C_{d+2} in $\mathbb{C}\backslash\mathbb{S}^W$ we assure that the same holds for any other point in a neighborhood containing C_{d+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 Proposition 1 we can show that the same holds for a neighborhood of $(C_1,\ldots,C_n)$ in $\mathbb{R}^d \times \ldots \mathbb{R}^d$. ■

Proof of Proposition 4

Let

$$||f||_{\text{cons}} = \max_{|\lambda|\le k_0} \sup_{\mathbf{t}} |D^\lambda f(\mathbf{t})|(1+\mathbf{t}^\top\mathbf{t})^{\delta_0}.$$

This is the consistency norm defined by Gallant and Nychka (1987) on $\mathcal{H}$.

The result follows from Lemma 3.1 in Ai and Chen (2003) (which in turn relies on Theorem 4.1 and Lemma A1 of Newey and Powell (2003)). Assumptions 2-5 correspond to assumptions 3.1, 3.2, 3.4 and 3.7 in that lemma. Assumption 3.3 is attained from the previous identification results (Propositions (1) and (2)). Assumption 3.5 follow from compactness and denseness results in Gallant and Nychka (1987). For Assumption 3.6, notice that

$$\mathbb{E}\left[|\rho_i(p_e,\mathcal{C}_e,\mathbf{X}_e,f)|^2\,\middle|\,\tilde{\mathbf{X}}\right] \le 4$$

for $i = 1, \ldots, n-1$. Then take $f_1, f_2 \in \mathcal{H}$ and note that

$$(1 + \mathbf{t}^\top \mathbf{t})^{\delta_0} |f_1(\mathbf{t}) - f_2(\mathbf{t})| \leq ||f_1 - f_2||_{\text{cons}}$$

for any $\mathbf{t}$. Consequently

$$\begin{aligned} \left| \int \mathbf{1}_{\mathbf{t} \in V_i(\mathcal{C}_e)} (f_1(\mathbf{t}) - f_2(\mathbf{t}))\, d\mathbf{t} - p_i(\mathcal{C}_e) + p_i(\mathcal{C}_e) \right| &\leq \int |f_1(\mathbf{t}) - f_2(\mathbf{t})|\, d\mathbf{t} \\ &\leq \int \frac{1}{(1 + \mathbf{t}^\top \mathbf{t})^{\delta_0}} d\mathbf{t}\, ||f_1 - f_2||_{\text{cons}} \end{aligned}$$

which establishes Holder continuity of ρ with respect to f. This completes the proof for the consistency of $\widehat{f_T}$. ■

Proof of Proposition 5

The proof again relies on the same results as the previous result. These amount to the verification of Lemma A1 in Newey and Powell (2003).

First notice that compactness of Θ with respect to the topology induced by the Frobenius norm and that of $\mathcal{H}$ with respect to topology induced by the consistency norm in Gallant and Nychka (1987) implies that the product space is also compact (with respect to the product topology) by Tychonoff's Theorem. This observation plus assumptions 2-5 and our identification results imply conditions (i) and (iii) in Lemma A1 of Newey and Powell (2003).

Because of compactness and since pointwise convergence can be established easily given the assumptions we impose, the uniform convergence condition (iii) in Newey and Powell (2003) is attained once we show that the objective function is stochastically equicontinuous (see Theorem 2.1 in Newey (1991)). This can be obtained once we show stochastic equicontinuity of

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

To see this, notice that the $E \times (n-1)$ matrix of estimates

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

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

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

which in turn delivers

$$\begin{aligned} |Q_n(W_1, f_1) - Q_n(W_2, f_2)| &= \left| \textstyle\sum_{i=1}^{n-1} \left(\frac{1}{E} ||P\rho_i(W_1, f_1)||^2_E - \frac{1}{E} ||P\rho_i(W_2, f_2)||^2_E \right) \right| \\ &\leq \textstyle\sum_{i=1}^{n-1} \left| \frac{1}{E} ||P\rho_i(W_1, f_1)||^2_E - \frac{1}{E} ||P\rho_i(W_2, f_2)||^2_E \right| \end{aligned} \tag{8}$$

Because

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

after multiplication of both terms in the inverse triangle inequality by $(||A|| + ||B||)/\sqrt{C}$, each of the terms in the sum (8) is bounded by

$$\begin{aligned} &\left| \frac{1}{E} ||P\left(\rho_i(W_1, f_1) - \rho_i(W_2, f_2)\right)||_E \left(||P\rho_i(W_1, f_1)||_E + ||P\rho_i(W_2, f_2)||_E\right) \right| \leq \\ &\qquad \left| \frac{1}{E} ||\rho_i(W_1, f_1) - \rho_i(W_2, f_2)||_E \left(||\rho_i(W_1, f_1)||_E + ||\rho_i(W_2, f_2)||_E\right) \right| \end{aligned}$$

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

$$||\rho_i(W, f)||^2_E = \sum_{e=1}^{E} \rho^2_{ie} \leq 4E$$

we have

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

This in turn gives

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

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

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

Consequently, once we show that

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

for any $\epsilon > 0$, we have stochastic equicontinuity of the objective function. Let then

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

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

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

To show stochastic equicontinuity we essentially follow the proof for Lemma 3 in Andrews (1992). Given $\epsilon > 0$, take $4 < M < \infty$ and $\delta > 0$ such that $\mathbb{P}\left(Y_{e\delta} > \epsilon^2/2\right) < \epsilon^2/(2M)$. 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 $\lim_{\delta\to 0} \mathbb{P}\left(Y_{e\delta} > \epsilon\right) = 0$ for any $\epsilon > 0$. Now, for such a δ,

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

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

■