Identification and Estimation of Preference Distributions When Voters Are Ideological *

Antonio Merlo [†]   Áureo de Paula [‡]

University of Pennsylvania   University of Pennsylvania

Please do not circulate (PRELIMINARY)

ABSTRACT

This paper studies the identification and estimation of voters' preferences under ideological voting. It builds on previous work by Degan and Merlo (2008), which explored the geometric structure of the *spatial theory of voting* to study its falsifiability. We use the structure delineated in that paper (Voronoi tessellations) to establish that voter preference distributions and other parameters can be retrieved from aggregate electoral data and suggest estimating these objects using Ai and Chen (2003)'s estimator. We provide large sample results for that estimator in our particular application and use data from the European Parliament to illustrate our analysis.

1   Introduction

Voting is a fundamental aspect of democracy and voters' decisions are essential factors of the political process shaping the policies adopted by democratic societies. Understanding observed voting patterns is a crucial step in the understanding of democratic institutions. In particular, identifying and estimating voters preferences has both practical and theoretical implications. From a theoretical standpoint, voters are an important primitive of political economy models. Different assumptions about their behavior have important consequences on the implications of these models and, more generally, on the interpretation of the induced behavior of politicians, parties and governments.

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 at some date a group of voters is facing some contested elections (i.e., there is at least one election and two or more candidates in each election). Suppose that each voter has political views (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, in each election, 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). In this case, we say that voters *vote ideologically*.

Hence, whether in reality voters vote ideologically (or whether other factors like for example instrumental considerations, or their assessment of candidates' personal characteristics determine their voting behavior) is clearly an important question. Because the positions of voters and candidates in a single ideological space are not immediately observable, a preliminary question is whether the hypothesis of ideological voting is even testable or falsifiable. In other words, which kind of data on candidates' positions and voting behavior would allow a researcher to potentially falsify and hence possibly reject the hypothesis that voters vote ideologically? Degan and Merlo (2008) address this question and find that in a variety of settings (e.g. single election data, few elections compared to ideological space dimension) ideological voting is not falsifiable.

Under the assumption that voters vote ideologically, a related natural question is whether the distribution of voter preferences is itself identifiable and estimable. There is a vast literature in political science about the estimation of the distribution of candidates in an ideological space and data sets containing mea-

---

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

sures of the positions of politicians in the ideological space based on their observed behavior in a variety of public offices are widely available (see Poole and Rosenthal (1997) as well as Heckman and Snyder (1997) for the United States or Hix, Noury, and Roland (2006) for the European Parliament[2]). Whereas Degan and Merlo parametrically estimate the distribution of voters, an important consideration is whether the distribution of voter preferences is nonparametrically identified and estimable using data on voting behavior.[3] Restrictions on the distribution of preferences may lead to meaningful theoretical results. Caplin and Nalebuff (1988), for instance, analyze electoral rules under the assumption that voters vote ideologically and the distribution of bliss points is concave. The empirical verification of such restrictions (e.g. concavity in the Caplin and Nalebuff (1988) case) would necessarily require the investigation of identification and estimation of the distribution of voter preferences. Even if the distribution can only be estimated parametrically, it is still important to know whether or not it is non-parametrically identified to gauge dependence of the results on particular parametric assumptions. The identification question in our context is also related to the standard revealed-preference argument which is prevalent in many fields of economics whereby one is interested in finding out what can be said about underlying preferences from observed behavior.

Degan and Merlo focus on falsifiability of the ideological voters hypothesis using individual-level data on how the same individuals vote in multiple simultaneous elections. Here, we focus on the issue of nonparametric identification and estimation of voters preferences using aggregate data under the maintained assumption that voters vote ideologically. In other words, we restrict attention to environments where the hypothesis is non-falsifiable. Since it focusses on retrieving individual level structure from aggregate data, our approach relates to the ecological inference problem as surveyed for instance in King (1997). In that sense, it also relates to the vast literature on identification and estimation of discrete choice models in industrial organization. Starting with McFadden (1974)'s seminal work, other important papers investigating the identification of discrete choice models include Manski (1988) and Matzkin (1992). Our paper is closer to the literature on discrete choice models with macro-level data (e.g. Berry, Levinsohn, and Pakes (2004) and more recently Berry and Haile (2009)). Close references in the econometrics literature are Ichimura and

---

[2]Note that in order to directly assess whether the behavior of voters is consistent with ideological voting one would need a consistent set of observations on the ideological positions of all voters and candidates in the same metric space. Hence, measures of citizens' self-reported ideological placements that are contained in some surveys (like, for example, the variable contained in the American National Election Studies, where voters are asked to place themselves on a 7-point liberal-conservative scale), cannot be used for this purpose, since, for instance, different people may interpret the scale differently.

[3]In this paper, as in Degan and Merlo (2008), we ignore the issue of abstention. For recent surveys of alternative theories of voter turnout see, e.g., Davis, Hinich, and Ordeshook (2002) and Merlo (2006).

Thompson (1998) and the recent analysis by Gautier and Kitamura (2008) on binary choice models with random coefficients.

One significant difference between our work and the previously cited literature is that we build on the method introduced in Degan and Merlo (2008), representing elections as Voronoi tessellations. Such objects 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)). Since this is relatively new, our methods can also be applied to other environments and inform for example literature above in industrial organization. Once we characterize our analysis in terms of Voronoi tessellations, we establish identification results for our basic model and an extension that accommodates more general preferences. We then suggest estimating the distribution of voter preferences and any parametric subcomponents using the methods proposed by Ai and Chen (2003) (see also Newey and Powell (2003)).[4]

To illustrate our results, we analyze data from the 1999 European Parliament election. More specifically, we obtain ideological positions for candidates generated by Hix, Noury, and Roland (2006), the proportion of votes obtained by each party in the election and demographics for different regions comprising age and gender distribution, education attainment and unemployment rates obtained from the 2001 European census.

2 Identification

2.1 Basic Model

In what follows, the ideological type space is a d-dimensional Euclidean space, $\mathbb{R}^d$ and the reference measurable space is this set equipped with the Borel sigma algebra: $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. The distribution of types 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.[5] Here $\mathbf{X}$ represents (electoral precinct) observable characteristics such as average 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 and gender distribution, education and unemployment 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

---

[4]For the large sample properties of the estimator we nevertheless rely on slightly different assumptions than Ai and Chen (2003).

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

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 $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$.

An election is a contest among n candidates. It is assumed that individuals vote ideologically and choose the candidate closest to them in the ideological space. As illustrated in Degan and Merlo (2008) an election defines a Voronoi tessellation on the Euclidean space. For a two-candidate election, the Voronoi tessellation is composed of two-half spaces separated by a hyperplane. The proportion of votes obtained by each candidate is the probability of the Voronoi cell that contains the candidate's ideological type. 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 $\mathbb{R}^d$. This characterizes an election.

We assume observed data compiles the proportion of votes obtained by each candidate in an election. For an election $\mathcal{C}$ and a pre-specified ideological type distribution $\mathbb{P}_T$, we can define the following object:

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

where $p(\mathcal{C}, \mathbb{P}_T)$ assembles the proportion of votes obtained by all the candidates in the profile $\mathcal{C}$ and takes values on the $n-$dimensional simplex. 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:

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

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

The following definition qualifies our characterization of identifiability:

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.[6] $\Rightarrow \mathbb{P}_{T_1} = \mathbb{P}_{T_2}$.*

In words, two type distributions that for every possible election configuration (except for cases in a zero measure set) give the same proportion of votes should correspond to the same measure. Now we are in shape to state the identification result:

---

[6]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 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. It basically generalizes the simple insight that for two candidates 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 ideological 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. 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 is due to the particular structure of the binary choice model analyzed by Ichimura and Thompson (1998) which is not shared by our model[7].

2.2   Extensions

Degan and Merlo (2008) also consider extensions of the canonical model examined in the previous sections. In particular, 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. Okabe, Boots, Sugihara, and Chiu (2000) refer to this as the elliptic distance with weighting matrix W (see page 197). According to the spatial theory of voting the main diagonal elements in the matrix W subsume the relative importance to a particular 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 individuals make trade-offs among these different dimensions (see for example Hinich and Munger (1997)).

We would like to analyze the identifiability of voter preferences which 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.

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

---

[7]In particular, choices follow linear index threshold crossing condition, essentially an inner product between covariates and random coefficients. Our problem deals with multinomial choices and relies on a nonlinear index comparing alternative choices of candidates.

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 αW, $\alpha > 0$, 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 includes the d-order identity matrix as a particular choice for W. It is perhaps remarkable that, once this normalization is imposed, we obtain the following result:

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

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

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

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 1).

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 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 a 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

7

of more information with an 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* $||W||_{d \times d} = \sqrt{d}$. *Then* $(\mathbb{P}_T, W)$ *is identified.*

A natural corollary of Proposition 2 is that election specific (senatorial, gubernatorial, presidential) weights and bliss point distributions are identified (up to the normalization $||W^e||_{d \times d} = \sqrt{d}$). Consequently, repeated voting records 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 but were not able to find an appropriate argument.

The ideas in Proposition 1 are useful in more general settings. The relative identifiability of two 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 there are two candidates and 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 Strategy

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.

3.1 A Simple Case

In only one dimension, an election provides an estimate of the cumulative distribution function $F_T(t|\mathbf{X}) = \int_{-\infty}^t f_{T|\mathbf{X}}(T|\mathbf{X})dT$ at the midpoints separating the candidates. With two candidates in election e, $C_{1e} < C_{2e}$, the proportion of voters for C_{1e} gives an estimate for the cdf at $\overline{C}_e = \frac{C_{1e}+C_{2e}}{2}$. As more elections are sampled, we obtain an increasing number of points at which we can estimate the cdf. Let Y_e be the proportion of votes obtained by the candidate with smaller position in election e and assume there are n_e votes in this election. Notice that

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

where $i = 1, \ldots, n_e$. Since $Y_e = \frac{\sum_{i=1}^{n_e} \mathbf{1}(T_i \leq \overline{C}_e)}{n_e}$,

$$\mathbb{E}(Y_e|\overline{C}_e, \mathbf{X}_e) = F_T(\overline{C}_e|\mathbf{X}_e)$$

and a natural estimator for F_T given m elections would be a multivariate kernel or local linear polynomial regression. Under usual conditions (see for instance Li and Racine (2007)), the estimator is consistent and has an asymptotically normal distribution and can presumably be extended to more than two candidates using the theory amenable to weakly dependent data processes (if elections are iid, dependence would only exist within a given election). 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 Many Dimensions

When the dimension of the ideological space is larger 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:

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

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 quantity $p_i, i \in \{1, \ldots, n\}$ 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:

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

9

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

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

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 as in the analysis of auctions with different numbers of bidders for example.[8]

As pointed out by Ai and Chen (2003), two difficulties arise in constructing this estimator. First of all, 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 $\{b_i(\tilde{\mathbf{X}}), i = 1, 2, \dots\}$ 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 $b^J(\tilde{\mathbf{X}}) = \left(b_1(\tilde{\mathbf{X}}), \dots, b_J(\tilde{\mathbf{X}})\right)'$, the sieve estimator for $m_i(\tilde{\mathbf{X}}, f)$, the i-th component in m, is given by

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

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

We consider the class $\mathcal{H}$ of densities studied by Gallant and Nychka (1987)[9] 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

$$\left(\sum_{|\lambda| \leq k_0} \int |D^\lambda h(\mathbf{t})|^2 (1 + \mathbf{t}'\mathbf{t})^{\delta_0} d\mathbf{t}\right)^{1/2} < \mathcal{B}_0 \tag{2}$$

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

$$D^\lambda f(\mathbf{t}) = \frac{\partial_1^\lambda}{\partial x_1^{\lambda_1}} \frac{\partial_2^\lambda}{\partial x_2^{\lambda_2}} \cdots \frac{\partial_d^\lambda}{\partial x_d^{\lambda_d}} f(\mathbf{t}), \quad \lambda = (\lambda_1, \dots, \lambda_d)' \in \mathbb{N}^d$$

---

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

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

and $|\lambda| = \sum_{i=1}^{d} \lambda_i$. Given a compact set on the ideological space, condition (2) 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 (2) allows for tails as fat as $f(\mathbf{t}) \propto (1+\mathbf{t}'\mathbf{t})^{-\eta}$ for $\eta > \delta_0$ or as thin as $f(\mathbf{t}) \propto e^{-\mathbf{t}'\mathbf{t}^{\eta}}$ 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:

$$\mathcal{H}_E = \left\{ f : f(\mathbf{t}, \xi) = \left[\sum_{i=0}^{J_E} H_i(\mathbf{t}) \right]^2 \exp\left(-\frac{\mathbf{t}'\mathbf{t}}{2} \right) + \epsilon\phi(\mathbf{t}), \int f(\mathbf{t},\xi)d\mathbf{t} = 1 \right\}$$

where H_i are Hermite polynomials, ϕ is the standard multivariate normal density and ϵ is a small positive number.[10] As mentioned before, the set of densities on which $\cup_{E=1}^{\infty}\mathcal{H}_E$ is dense is fairly large. The estimator is also very attractive computationally as integrals can be obtained analytically.

The estimator is formally defined as:

$$\widehat{f} = \operatorname{argmin}_{f \in \mathcal{H}_{\mathcal{E}}} \frac{1}{E} \sum_{e=1}^{E} \widehat{m}(\tilde{\mathbf{X}}, f)' \left[\widehat{\Sigma}(\tilde{\mathbf{X}}) \right]^{-1} \widehat{m}(\tilde{\mathbf{X}}, f) \tag{3}$$

To establish consistency we rely on the following assumptions:

Assumption 1 *(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 2 *(i) The smallest and largest eigenvalues of $\mathbb{E}\{b^J(\tilde{\mathbf{X}})b^J(\tilde{\mathbf{X}})'\}$ 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 $b^J(\tilde{\mathbf{X}})'\pi$ such that $\mathbb{E}[\{g(\tilde{\mathbf{X}}) - b^J(\tilde{\mathbf{X}})'\pi\}^2] = o(1)$.*

---

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

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

where

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

with $a(\tilde{x}) = \sum_{|\beta|=0}^{J_x} a_{\alpha\beta}\tilde{x}^\beta$. 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 $\tilde{x}^\beta$ with respect to $\beta = (\beta_1, \ldots, \beta_{\sharp\tilde{x}})$.

Assumption 3 *(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 4 *(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 1-4,*

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

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 2.2. The parameter to be estimated is 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:

$$(\widehat{W}, \widehat{f}) = \text{argmin}_{(W,f) \in \Theta \times \mathcal{H}_{\varepsilon}} \frac{1}{E} \sum_{e=1}^{E} \widehat{m}(\tilde{\mathbf{X}}, (W, f))' \left[\widehat{\Sigma}(\tilde{\mathbf{X}})\right]^{-1} \widehat{m}(\tilde{\mathbf{X}}, (W, f)) \tag{4}$$

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 1-4 and Θ compact (with respect to the Frobenius norm),*

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

with respect to the norm

$$||(W, f)|| = \max_{|\lambda| \leq k_0} \sup_{\mathbf{t}} |D^{\lambda} f(\mathbf{t})| (1 + \mathbf{t}'\mathbf{t})^{\delta_0} + \sqrt{tr(W'W)}$$

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   Empirical Illustration

In this section, we illustrate the methodology suggested previously with an analysis of the election for the July 1999-May 2004 European Parliament.[11]

---

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

4.1 Data Description

Our data consist of ideological positions for the candidates, electoral outcomes within each country and demographic data for each of these.

The ideological positions for the parties were obtained from Hix, Noury, and Roland (2006), who used roll-call data for that legislature to generate two-dimensional ideological positions for each of the members of the parliament along the lines of the NOMINATE scores generated in Poole and Rosenthal (1997).[12] As indicated in Heckman and Snyder (1997), the generation of ideological positions is done 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, for example, Stock and Watson (2002), Bai and Ng (2006a) or Bai and Ng (2006b) for analyses on this practice). Since 1999, elections for the European Parliament have taken place under the proportional representation system and typically with closed lists. By focussing on the parties as the major political entities and since the system presumably leads to increased party loyalty, we are able to identify the political candidate by the party, using the aggregate positions of the individual candidates as the party's position in a given election.

We amend the data on ideological positions with electoral outcomes in 1999 obtained from the European Parliament and demographic information (age and gender distribution, education attainment, unemployment) from the 2001 European Census. The election outcomes data was obtained from the CIVI-CACTIVE European Election Database.[13] The demographic data was obtained from EUROSTAT.

4.2 Estimates

To be added.

5 Conclusion

To be added.

---

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

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

References

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

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 (2004): "Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Vehicle Market," *Journal of Political Economy*, 112(1), 68–105.

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

CAPLIN, A., AND B. NALEBUFF (1988): "On 64Rule," *Econometrica*, 56(4), 787–814.

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.

——— (2002): "Economic Theories of Voter Turnout," *Economic Journal*, 112, F332–F352.

DEGAN, A., AND A. MERLO (2008): "Do Voters Vote Ideologically?," Working Paper, University of Pennsylvania.

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.

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.

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 A: 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 $||A||_{k \times l} = \sqrt{\text{Tr}(A^\top A)}$ 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 $d((A_1, b_1), (A_2, b_2)) = \sqrt{||A_1 - A_2||_{k \times l}^2 + ||b_1 - b_2||_m^2}$.

<u>Step 1:</u> $(\exists (A^*, b^*) \in \mathcal{M}_{n-1 \times d} \times \mathbb{R}_{n-1} : \mathbb{P}_{T_1}(\{\mathbf{T} \in \mathbb{R}_d : A^* \mathbf{T} \leq b^*\}) \neq \mathbb{P}_{T_2}(\{\mathbf{T} \in \mathbb{R} : A^* \mathbf{T} \leq b^*\}))$
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 $c'\mathbf{Z}$ of $\mathbf{Z}$ will coincide under $\mathbb{P}_{T_1}$ and $\mathbb{P}_{T_2}$. Since a linear combination of $\mathbf{Z}$ is a linear combination of $\mathbf{T}$, the cdf for an arbitrary linear combination of $\mathbf{T}$ under $\mathbb{P}_{T_1}$ coincides with the cdf for that combination under $\mathbb{P}_{T_2}$. Again, by the Cramér-Wold device, this implies that $\mathbb{P}_{T_1} = \mathbb{P}_{T_2}$. Consequently,

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

<u>Step 2:</u> $(\exists \eta > 0 : \mathbb{P}_{T_1}(\{\mathbf{T} \in \mathbb{R}_d : A\mathbf{T} \leq b\}) \neq \mathbb{P}_{T_2}(\{\mathbf{T} \in \mathbb{R}_d : A\mathbf{T} \leq b\}), \forall (A, b) \in \mathcal{N}((A^*, b^*), \eta))$ 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 $d((A_k, b_k), (A, b)) \overset{k \to \infty}{\longrightarrow} 0$. Then

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

18

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 $d((A_k, b_k), (A, b)) \overset{k \to \infty}{\longrightarrow} 0 \Rightarrow ||A_k - A||_{n-1 \times d} \overset{k \to \infty}{\longrightarrow} 0$ and $||b_k - b||_{n-1} \to 0$, for any fixed $\mathbf{T} \in \mathbb{R}_d$, $||(A_k\mathbf{T} - b_k) - (A\mathbf{T} - b)||_{n-1} \overset{k \to \infty}{\longrightarrow} 0$. 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 $||(A_k\mathbf{T} - b_k) - (A\mathbf{T} - b)||_{n-1} \overset{k \to \infty}{\longrightarrow} 0$, $A\mathbf{T} \geq b$ 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,

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

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^*): $\mathcal{N}((A^*, b^*), \eta)$.

Step 3: ($\{\mathbf{T} \in \mathbb{R}_d : A\mathbf{T} < b\}$ **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 $\mathcal{S}(\epsilon) \overset{\epsilon \to 0}{\longrightarrow} \{(A^*, b^*)\} \subset \mathcal{N}((A^*, b^*), \eta)$. 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.

<u>Step 1:</u> (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

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

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

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

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

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

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

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

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

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

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

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

21

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

$$
\begin{aligned}
\mathbf{T} \in H^{\overline{W}}(C_i', C_j') \quad &\Rightarrow C_i'^{\top} \overline{W} C_i' - C_j'^{\top} \overline{W} C_j' - 2(C_j' - C_i')^{\top} \overline{W} \mathbf{T} = 0 \\
&\Rightarrow \tfrac{1}{2} \left(C_i'^{\top} \overline{W} C_i' - C_j'^{\top} \overline{W} C_j' \right) - (C_j - C_i)^{\top} \overline{W} \mathbf{T} = 0.
\end{aligned}
\tag{5}
$$

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 (5)), we have

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

Furthermore, expression (5) 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 $V_i^{\overline{W}}(\mathcal{C}') \subset V_i^{\overline{W}}(\mathcal{C})$. 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 (5), $\Delta_{lj} \geq 0$ 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' 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

$$||f||_{\text{cons}} = \max_{|\lambda| \leq k_0} \sup_{\mathbf{t}} |D^\lambda f(\mathbf{t})|(1 + \mathbf{t}'\mathbf{t})^{\delta_0}.$$

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 1-4 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

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

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

$$(1 + \mathbf{t}'\mathbf{t})^{\delta_0} |f_1(\mathbf{t}) - f_2(\mathbf{t})| \leq ||f_1 - f_2||_{\text{cons}}$$

for any $\mathbf{t}$. Consequently

$$\left| \int \mathbf{1}_{\mathbf{t} \in V_i(\mathcal{C}_e)} (f_1(\mathbf{t}) - f_2(\mathbf{t}))\, d\mathbf{t} - p_i(\mathcal{C}_e) + p_i(\mathcal{C}_e) \right| \leq \int |f_1(\mathbf{t}) - f_2(\mathbf{t})|\, d\mathbf{t}$$

$$\leq \int \frac{1}{(1 + \mathbf{t}'\mathbf{t})^{\delta_0}}\, d\mathbf{t}\, ||f_1 - f_2||_{\text{cons}}$$

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

26

the product space is also compact (with respect to the product topology) by Tychonoff's Theorem. This observation plus assumptions 1-4 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

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

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

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

where ρ is an $E \times (n-1)$ matrix stacking $\left(\int \mathbf{1}_{\mathbf{t} \in V_i(\mathcal{C}_e, W)} f(\mathbf{t}) d\mathbf{t} - p_{i,e} \right)'_{i=1,\dots,n-1}$ for all observations and P is an $E \times E$ idempotent matrix with rank (= trace) at most J. Since we have Assumption 3, 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

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

which in turn delivers

$$
\begin{aligned}
|Q_n(W_1, f_1) - Q_n(W_2, f_2)| &= \left| \sum_{i=1}^{n-1} \left(\frac{1}{E} ||P\rho_i(W_1, f_1)||_E^2 - \frac{1}{E} ||P\rho_i(W_2, f_2)||_E^2 \right) \right| \\
&\leq \sum_{i=1}^{n-1} \left| \frac{1}{E} ||P\rho_i(W_1, f_1)||_E^2 - \frac{1}{E} ||P\rho_i(W_2, f_2)||_E^2 \right|
\end{aligned}
\tag{6}
$$

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 (6) is bounded by

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

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

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

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

we have

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

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

This in turn gives

$$\sup_{(W_1, f_1) \in \Theta \times \mathcal{H}} \quad \sup_{(W_2, f_2) \in \mathcal{N}((W_1, f_1), \delta)} |Q_n(W_1, f_1) - Q_n(W_2, f_2)|$$

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

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

$$\mathbb{P} \left(\sup_{(W_1, f_1) \in \Theta \times \mathcal{H}} \quad \sup_{(W_2, f_2) \in \mathcal{N}((W_1, f_1), \delta)} |Q_n(W_1, f_1) - Q_n(W_2, f_2)| > \epsilon \right)$$

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

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

Consequently, once we show that

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

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

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

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

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

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,

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

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

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

Since this argument can be repeated for $i = 1, \ldots, n-1$, we have stochastic equicontinuity.   ∎
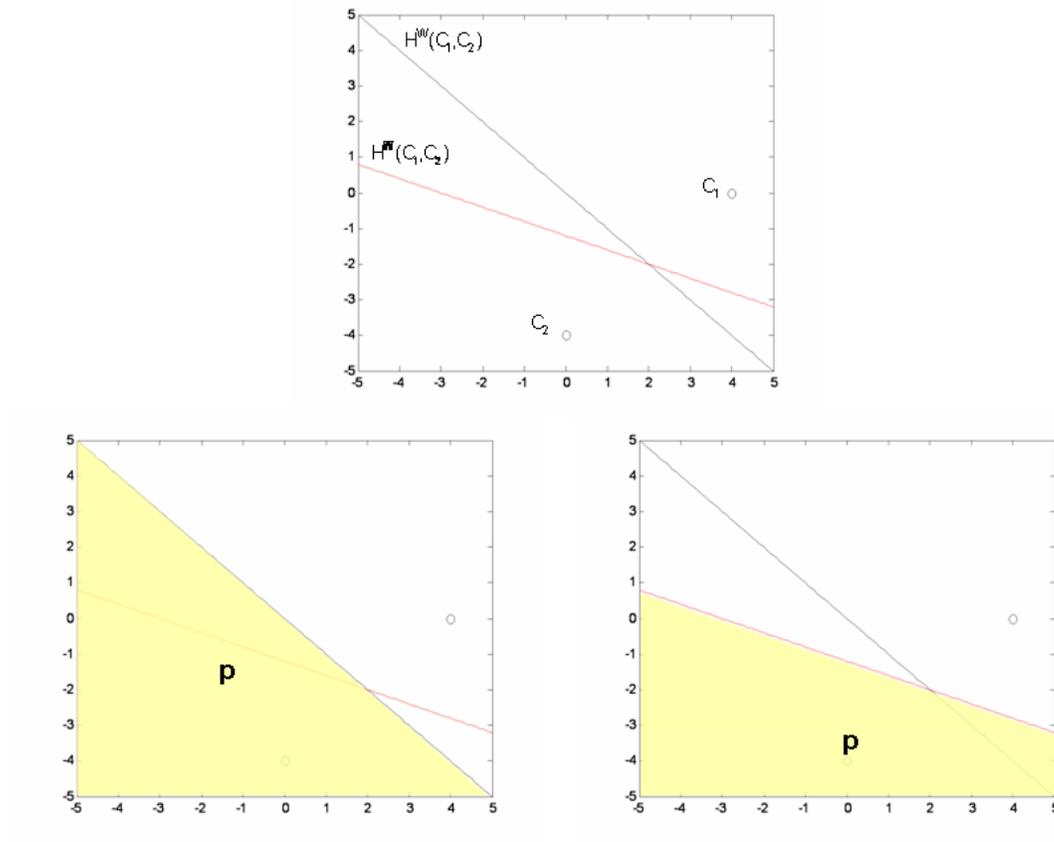
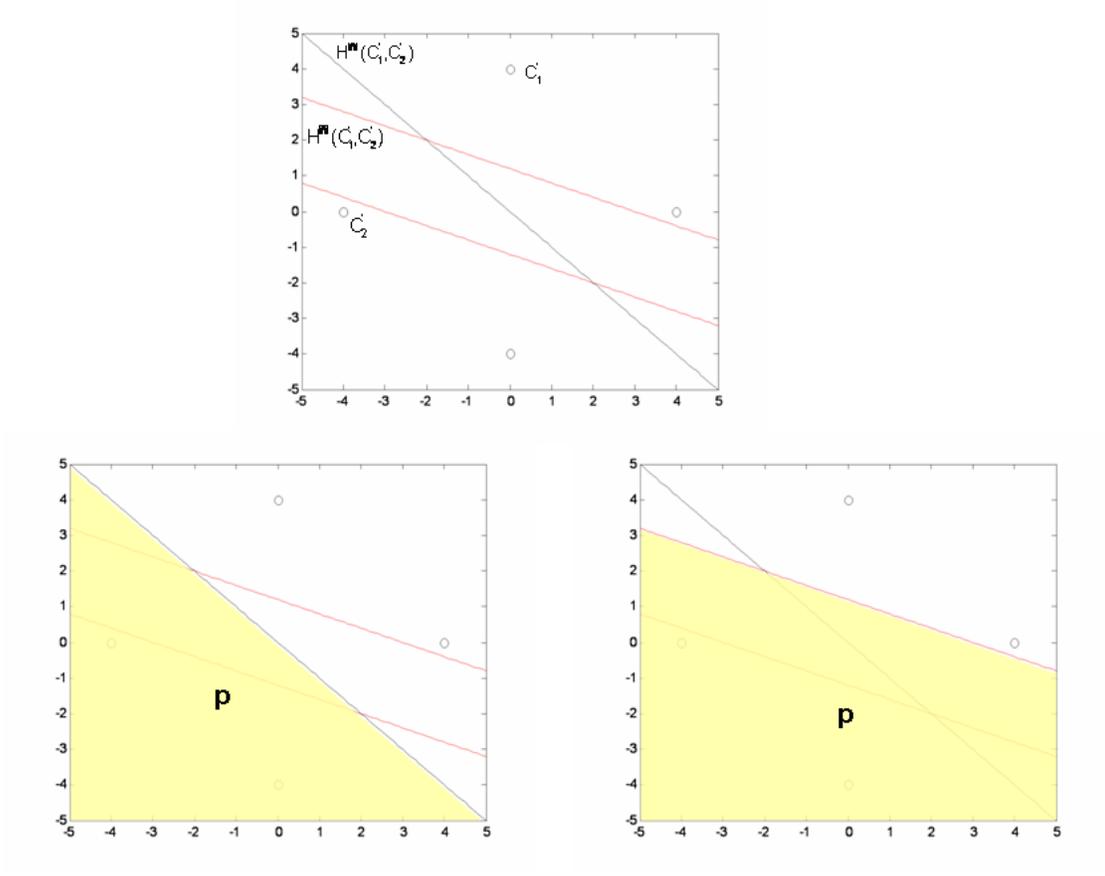# Appendix B: Figures

Figure 1: Voronoi Tessellations for Candidates C_1, C_2

Figure 2: Voronoi Tessellations for Candidates C_1, C_2