Online Appendix to "Identification and Estimation of Preference Distributions When Voters Are Ideological" *

Antonio Merlo †

Rice University

Áureo de Paula ‡

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

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

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

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

Online Appendix: Proofs

Proof of Lemma 1.

Since W is known, without loss of generality we assume that $W = I$.[1] It is enough to consider a single election with n candidates. In what follows, for any integers l, m and r: $\mathcal{M}_{r\times l}$ is the space of $r \times l$ real matrices which is endowed with the typical Frobenius matrix norm $||A||_{r\times l} = \sqrt{\mathrm{Tr}(A^\top A)}$ for $A \in \mathcal{M}_{r\times l}$; $||b||_l$ is the typical Euclidean norm in $\mathbb{R}^l$; and the product metric space $\mathcal{M}_{r\times l} \times \mathbb{R}^m$ is endowed with the normed product metric $d((A_1, b_1), (A_2, b_2)) = \sqrt{||A_1 - A_2||^2_{r\times l} + ||b_1 - b_2||^2_m}$.

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

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

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

Given A^* and b^*, one can then find two candidates C_1^* and C_2^* such that $b^* = C_2^{*\top} C_2^* - C_1^{*\top} C_1^*$ and $A^*_{1\times k} = 2(C_2^* - C_1^*)^\top$ for whom vote shares under $\mathbb{P}_{T_1}$ would be different from vote shares

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

under $\mathbb{P}_{T_2}$.

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

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

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

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

Proof of Lemma 2.

Consider two different spatial voting models characterized by $(\mathbb{P}_T, W)$ and $(\mathbb{P}_{\overline{T}}, \overline{W})$. If

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

Throughout, we focus on elections such that $C_{1i} \neq C_{2i}$ for every ith coordinate $(i = 1, \ldots, k)$ in C_1 and C_2 and $W(C_1 - C_2)$ and $\overline{W}(C_1 - C_2)$ are linearly independent. These elections happen with probability one given Assumption 2. First, note that the set $\{(C_{11}, \ldots, C_{1k}, C_{21}, \ldots, C_{2k}) \in \mathbb{R}^{2k} : C_{1i} = C_{2i}\}$ has dimension $2k - 1$ for each $i = 1, \ldots, k$ and hence, (Lebesgue-) measure zero. Hence, $\cup_{i=1,\ldots,k}\{(C_{11}, \ldots, C_{1k}, C_{21}, \ldots, C_{2k}) \in \mathbb{R}^{2k} : C_{1i} = C_{2i}\}$ has (Lebesgue-) measure zero.

To see that the event $\{(C_1, C_2) \in \mathbb{R}^{2k} : W(C_1 - C_2)$ and $\overline{W}(C_1 - C_2)$ are linearly *dependent*$\}$ also has (Lebesgue-) measure zero, suppose that this were not the case. Then, one can find k linearly independent vectors $\mathbf{t}_1, \ldots, \mathbf{t}_k$ such that $W\mathbf{t}_1 = \alpha_1 \overline{W}\mathbf{t}_1, \ldots, W\mathbf{t}_k = \alpha_k \overline{W}\mathbf{t}_k$ for some $\alpha_i \in \mathbb{R} - \{0\}, i = 1, \ldots, k$. Since $\mathbf{t}_1, \ldots, \mathbf{t}_k$ are linearly independent, they form a basis for $\mathbb{R}^k$ and any vector in $\mathbb{R}^k$ is uniquely expressed as a linear combination of those vectors. Let $\mathbf{t}$ then be any other vector in $\mathbb{R}^k$ expressed as $\mathbf{t} = a_1\mathbf{t}_1 + \cdots + a_k\mathbf{t}_k$ with non-zero coefficients $a_1, \ldots, a_k$. (The set of vectors in $\mathbb{R}^k$ for which this is not the case has zero Lebesgue-measure.) Then $W\mathbf{t} = W(a_1\mathbf{t}_1 + \cdots + a_k\mathbf{t}_k) = \overline{W}(a_1\alpha_1\mathbf{t}_1 + \cdots + a_k\alpha_k\mathbf{t}_k)$. If $W\mathbf{t}$ and $\overline{W}\mathbf{t}$ are linearly dependent, we have $W\mathbf{t} = \alpha\overline{W}\mathbf{t}$ for some $\alpha \in \mathbb{R} - \{0\}$. These two expressions deliver that

$$\overline{W}\alpha\mathbf{t} = \overline{W}(a_1\alpha_1\mathbf{t}_1 + \cdots + a_k\alpha_k\mathbf{t}_k) \Leftrightarrow \mathbf{t} = a_1\frac{\alpha_1}{\alpha}\mathbf{t}_1 + \cdots + a_k\frac{\alpha_k}{\alpha}\mathbf{t}_k,$$

where we use the fact that $\overline{W}$ is positive definite (and, hence, invertible) and that $\alpha \neq 0$. Since $\mathbf{t}$ is uniquely expressed as a linear combination of the basis vectors, we then have that $\alpha = \alpha_1 = \cdots = \alpha_k$. Since this holds for every $\mathbf{t}$ and $\alpha_1, \ldots, \alpha_k$ depend on the basis but do not depend on the specific vector $\mathbf{t}$ (with non-zero coefficients $a_1, \ldots, a_k$), α does not depend on $\mathbf{t}$. This then means that $(W - \alpha\overline{W})\mathbf{t} = 0$ holds for every $\mathbf{t}$, implying that the dimension of the nullspace of $W - \alpha\overline{W}$ is k and its rank is zero (by the rank-nullity theorem). Since the only matrix with zero rank is a matrix of zeros, $W - \alpha\overline{W} = \mathbf{0}$. Then $\|W\| = |\alpha|\|\overline{W}\|$, which implies that $|\alpha| = 1$ (since $\|W\| = \|\overline{W}\| = \sqrt{k}$). Given that $W \neq \overline{W}$, this means that $W = -\overline{W}$ and hence $\mathbf{t}^\top W\mathbf{t} > 0$ for any $\mathbf{t} \neq 0 \Leftrightarrow \mathbf{t}^\top \overline{W}\mathbf{t} = -\mathbf{t}^\top W\mathbf{t} > 0$ for any $\mathbf{t} \neq 0$. Then, either W or $\overline{W}$ cannot be positive definite.

Since the events $\cup_{i=1,\ldots,k}\{(C_{11}, \ldots, C_{1k}, C_{21}, \ldots, C_{2k}) \in \mathbb{R}^{2k} : C_{1i} = C_{2i}\}$ and $\{(C_1, C_2) \in \mathbb{R}^{2k} : W(C_1 - C_2)$ and $\overline{W}(C_1 - C_2)$ are linearly *dependent*$\}$ both have zero (Lebesgue-)measure, so does their union. Consequently, the complementary event that elections such that $C_{1i} \neq C_{2i}$ for every ith coordinate ($i = 1, \ldots, k$) in C_1 and C_2 and $W(C_1 - C_2)$ and $\overline{W}(C_1 - C_2)$ are linearly independent has probability one.

Step 1: (There is more than one set of candidates that generates the same partition of voters for a given weighting matrix W.) Consider an election where $C_{1i} \neq C_{2i}$ for every ith coordinate ($i = 1, \ldots, k$) in C_1 and C_2 and such that $W(C_1 - C_2)$ and $\overline{W}(C_1 - C_2)$ are linearly independent. (Note that this event happens with probability one given Assumption 2.) Take the set of vectors $\mathbf{t}$ such that

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

The above equation can be explicitly written as

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

The solution set for this equation contains at least one element as long as $(C_1 - C_2)^\top W$ is different from the zero vector. Since W is positive definite and, consequently, has full rank, its nullspace is a singleton (comprised of the vector zero). Hence, $(C_1 - C_2)^\top W \neq \mathbf{0}$ (on the event that $C_{1i} \neq C_{2i}, i = 1, \ldots, k$). In this case, let P denote one solution to the equation above such that $(C_1 - C_2)^\top \overline{W} P \neq 0$ (to be used in Step 3) and let $\mathcal{C}'$ be such that

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

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

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

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

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

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

<u>**Step 2:**</u> **(For $\mathcal{C}, \mathcal{C}'$ defined above, $H^W(C_1, C_2)$ and $H^{\overline{W}}(C_1, C_2)$ are different hyperplanes and $H^{\overline{W}}(C_1, C_2)$ and $H^{\overline{W}}(C_1', C_2')$ are parallel.)** With only two candidates, Voronoi cells are simply half-spaces of $\mathbb{R}^k$ defining the nearest-neighbor sets for each can-

didate. Consider $\mathcal{C}$ and $\mathcal{C}'$ such that their Voronoi tessellations under W coincide, i.e., $V^W(\mathcal{C}) = V^W(\mathcal{C}')$ where $\mathcal{C}$ and $\mathcal{C}'$ are obtained as in Step 1.

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

$$H^W(C_1, C_2) \equiv \{\mathbf{t} \in \mathbb{R}^k : C_1^\top W C_1 - C_2^\top W C_2 + 2(C_2 - C_1)^\top W \mathbf{t} = 0\}, \tag{1}$$

is the solution set to a linear equation. Since W is positive definite, $2(C_2 - C_1)^\top W$ is nonzero on the event that $C_{1i} \neq C_{2i}, i = 1, \ldots, k$. Then, with probability one the solution set to the equation defining $H^W(C_1, C_2)$ above has dimension $k-1$, which is the dimension of the nullspace of $2(C_2 - C_1)^\top W$. The same holds for $H^{\overline{W}}(C_1, C_2)$, which is defined as in (1) using $\overline{W}$ instead of W.

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

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

This solution set has dimension $k-2$ as long as $2(C_2 - C_1)^\top W$ and $2(C_2 - C_1)^\top \overline{W}$ are linearly independent. This is because in this case the nullspace for the matrix of coefficients (which stacks these two row vectors) has dimension $k-2$. Consequently, because the dimension of their intersection is smaller than the dimension of either $H^W(C_1, C_2)$ or $H^{\overline{W}}(C_1, C_2)$, these two sets are different.

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

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

Then see that

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

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

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

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

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

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

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

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

Let

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

for $j \ne i$.

We note that $\Delta_{ij} \ne 0$ (on the event that $C_{1i} \ne C_{2i}, i = 1, \ldots, k$) when P is chosen so that $(C_i - C_j)^\top \overline{W} P \ne 0$. To see this, note first that, if $\Delta_{ij} = 0$, from the expressions above for $V_i^{\overline{W}}(\mathcal{C})$ and $V_i^{\overline{W}}(\mathcal{C}')$, we would have that $V_i^{\overline{W}}(\mathcal{C}) = V_i^{\overline{W}}(\mathcal{C}')$. This is in turn means that

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

Then, note that

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

Given that $C_i' = C_i + P$, where P is defined in Step 1 to attain $V_i^W(\mathcal{C}) = V_i^W(\mathcal{C}')$, we also have that

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

These then imply that $(C_i - C_j)^\top \overline{W} P = 0$, which contradicts the criterion used to select P from the solution set to $d^W(P, C_1) - d^W(P, C_2) = 0$.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Furthermore, note that

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

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

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

countable subadditivity implies that

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

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

Proof of Theorem 1

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

Proof of Proposition 1.

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

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

with $\mathbf{t}_s = b + A\mathbf{x} + R\mathbf{z}_s$ is Euclidean as defined in Pakes and Pollard (1989). This is because the first factor is essentially a polynomial in $\mathbf{z}_s$ with bounded coefficients and the second factor is an indicator for $\mathbf{z}_s$ belonging to a Voronoi cell, which is an intersection of halfspaces. Both are Euclidean classes (see Example 2.9 in Pakes and Pollard (1989) for the first factor and Lemmas 2.4, 2.5 and the discussion before Lemma 2.8 also in Pakes and Pollard (1989) for the second factor). Finally, the product of two functions in Euclidean classes forms an Euclidean class (Lemma 2.14). Lemma 2.8 in Pakes and Pollard (1989) then shows that $|\rho_{iS}(W, f) = \rho_i(W, f)|$ converges almost surely to zero uniformly in the parameters.

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

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

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

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

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

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

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

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

which in turn delivers

$$\begin{aligned} |Q_n(W_1,f_1) - Q_n(W_2,f_2)| &= \left|\textstyle\sum_{i=1}^{n-1}\left(\frac{1}{E}||P\rho_i(W_1,f_1)||^2 - \frac{1}{E}||P\rho_i(W_2,f_2)||^2\right)\right| \\ &\leq \textstyle\sum_{i=1}^{n-1}\left|\frac{1}{E}||P\rho_i(W_1,f_1)||^2 - \frac{1}{E}||P\rho_i(W_2,f_2)||^2\right|, \end{aligned} \tag{3}$$

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

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

each of the terms in the sum in expression (3) is bounded by

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

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

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

we have

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

This in turn gives

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

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

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

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

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

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

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

for any $\epsilon > 0$.

Let then

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

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

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

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

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

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

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

Online Appendix: Monte Carlo Experiments

In this Appendix, we examine the small sample performance of the suggested estimation strategy in a few Monte Carlo experiments. We investigate models without covariates with three potential distribution of voter types. We use the distributions suggested by Ichimura and Thompson (1998) and summarized in Table 1 and Figure 1. For each of these, we postulate two different weighting matrices W for the weighted distance function. The first one has $W_{1,2} = W_{2,1} = 0$ and $W_{2,2} = 2$, and the second $W_{1,2} = W_{2,1} = 0.5$ and $W_{2,2} = 2$. Both matrices are normalized to have $W_{1,1} = 1$. We assume that the analysis has 100 observations in each set of Monte Carlo experiments.[2] 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.

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

Table 1: Data Generating Processes

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

Table 6: Data Generating Processes (Continued)

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

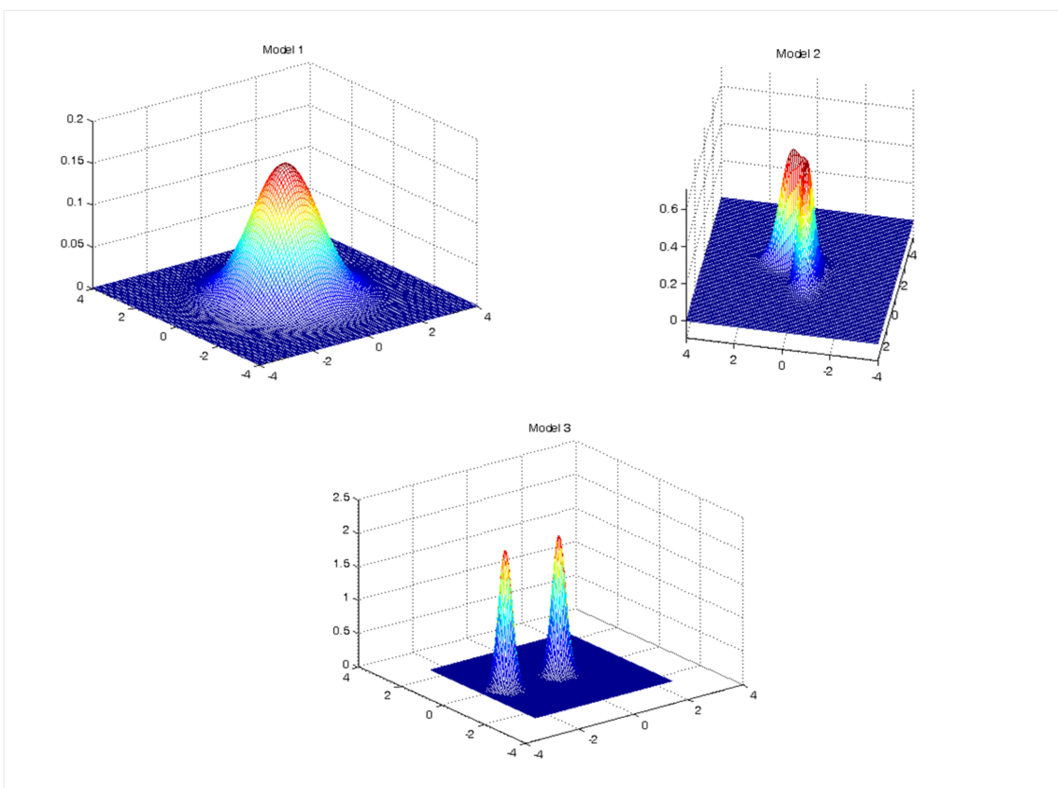

Figure 1: DGP Densities

The estimation follows the guidelines prescribed in the previous section. For the estimation of $m(\cdot)$ we use linear splines (with cross-products) for Models 1 and 2 and simple linear projections for Model 3. The estimation weighting matrix ($\tilde{\Sigma}$) is the identity. In Tables 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 $\overline{f}(\mathbf{t}) = \sum_{i=1}^{MC} \hat{f}_i(\mathbf{t})/MC$. The pointwise squared bias is then defined as $\left(\overline{f}(\mathbf{t}) - f(\mathbf{t})\right)^2$ and the pointwise variance is $\sum_{i=1}^{MC} \left(\hat{f}_i(\mathbf{t}) - \overline{f}_i(\mathbf{t})\right)^2 /MC$. We report squared bias, variance and MSE integrated over a grid of 100×100 points.

Table 2: Monte Carlo Results: Model 1

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

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

Table 3: Monte Carlo Results: Model 2

$(W_{1,2}, W_{2,2}) = (0, 2)$			
Bias^2	Variance	MSE	J_t
$(0.0023, 0.4658, 0.0025)$	$(0.2206, 0.2212, 5.5267 \times 10^{-4})$	$(0.2228, 0.6871, 0.0031)$	1
$(0.0010, 0.0853, 0.0016)$	$(0.1215, 0.2108, 4.6720 \times 10^{-4})$	$(0.1224, 0.2961, 0.0020)$	2
$(0.0001, 0.0201, 0.0012)$	$(0.0912, 0.1316, 4.3440 \times 10^{-4})$	$(0.0913, 0.1517, 0.0016)$	3
$(0.0006, 0.0120, 9.3694 \times 10^{-4})$	$(0.0693, 0.0952, 3.9928 \times 10^{-4})$	$(0.0699, 0.1072, 0.0013)$	5
$(0.0001, 0.0088, 8.4013 \times 10^{-4})$	$(0.0556, 0.0900, 3.6408 \times 10^{-4})$	$(0.0557, 0.0988, 0.0012)$	5
$(W_{1,2}, W_{2,2}) = (0.5, 2)$			
Bias^2	Variance	MSE	J_t
$(0.2346, 2.5381, 0.0042)$	$(0.2391, 0.8908, 0.0042)$	$(0.4737, 3.4289, 0.0042)$	1
$(0.2435, 2.0193, 0.0038)$	$(0.2473, 1.0363, 0.0007)$	$(0.4908, 3.0556, 0.0046)$	2
$(0.2005, 1.9740, 0.0037)$	$(0.2497, 1.0579, 8.1940 \times 10^{-4})$	$(0.2005, 3.0319, 0.0045)$	3
$(0.1958, 1.0874, 0.0036)$	$(0.2458, 1.0874, 8.2810 \times 10^{-4})$	$(0.4416, 3.0167, 0.0044)$	4
$(0.1937, 1.9216, 0.0036)$	$(0.2439, 1.0867, 8.7007 \times 10^{-4})$	$(0.0180, 0.5403, 0.0045)$	5

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

Table 4: Monte Carlo Results: Model 3

$(W_{1,2}, W_{2,2}) = (0, 2)$			
Bias^2	Variance	MSE	J_t
$(0.0015, 0.0002, 0.0274)$	$(0.0442, 0.1275, 0.0014)$	$(0.0457, 0.1277, 0.0287)$	1
$(0.0008, 0.0111, 0.0274)$	$(0.0221, 0.0633, 0.0015)$	$(0.0229, 0.0633, 0.0289)$	2
$(0.0007, 0.0164, 0.0136)$	$(0.0938, 0.0317, 0.0125)$	$(0.0946, 0.0481, 0.0260)$	3
$(0.0036, 0.0064, 0.0149)$	$(0.0775, 0.0258, 0.0136)$	$(0.0812, 0.0323, 0.0285)$	4
$(0.0008, 0.0389, 0.0073)$	$(0.0244, 0.2360, 0.0131)$	$(0.0252, 0.2749, 0.0204)$	5
$(W_{1,2}, W_{2,2}) = (0.5, 2)$			
Bias^2	Variance	MSE	J_t
$(0.0021, 0.0208, 0.0279)$	$(0.0752, 0.4363, 0.0019)$	$(0.0773, 0.4571, 0.0019)$	1
$(0.0002, 0.0056, 0.0274)$	$(0.0186, 0.0226, 0.0016)$	$(0.0187, 0.0282, 0.0289)$	2
$(0.0004, 0.0561, 0.0133)$	$(0.1189, 0.1552, 0.0138)$	$(0.1193, 0.2113, 0.0271)$	3
$(0.0010, 0.0099, 0.0140)$	$(0.0880, 0.0226, 0.0139)$	$(0.0890, 0.0326, 0.0279)$	4
$(0.0001, 0.0301, 0.0071)$	$(0.0097, 0.1467, 0.0115)$	$(0.0098, 0.1768, 0.0186)$	5

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

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

Online Appendix: Random Intercept

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

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

where δ_i is a candidate-specific valence term and $u^{\mathbf{t}}(\cdot)$ is a decreasing function as in Section 2 above. For this "linear-quadratic" specification of voter preferences, which is widely used in the political economy literature (see, e.g., Enelow and Hinich (1984)), Degan and Merlo (2009) have shown that the set of nearest neighbors to a given candidate is still given by an intersection of halfspaces as when utility functions are given by expression (1). Specifically, the Voronoi cells for each candidate are now given by:

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

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

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

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

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

Under these assumptions, we can then establish that voter preference distributions and other parameters of interest (which include the valence parameter vector $\delta \equiv (\delta_1, \ldots, \delta_n)$) can be identified from aggregate electoral data. The result is demonstrated along the same lines of our previous results and we provide further details on the necessary modifications to our arguments below.

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

Proof of Theorem 2

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

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

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

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

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

Consider P in this set such that $P \neq aC_1 + (1-a)C_2$ and such that $2(C_1 - C_2)^\top \overline{W}(P - aC_1 - (1-a)C_2) \neq 0$, where

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

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

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

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

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

Since both P and $aC_1+(1-a)C_2$ belong to $H^W(C_1, C_2, \delta_1, \delta_2)$, the last term in the right-hand side is zero. Consequently, $C_1'^\top WC_1' - C_2'^\top WC_2' = C_1^\top WC_1 - C_2^\top WC_2$. This in turn implies that

$$2(C_1 - C_2)^\top W\mathbf{t} = (C_1^\top WC_1 - C_2^\top WC_2) + \delta_1 - \delta_2$$
$$\Leftrightarrow$$
$$2(C_1' - C_2')^\top W\mathbf{t} = (C_1'^\top WC_1' - C_2'^\top WC_2') + \delta_1' - \delta_2',$$

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

We note that while in the proof of Theorem 2 the profile of valence terms $(\delta_1, \ldots, \delta_n)$ are assumed to be fixed across elections (which is consistent, for example, with our empirical application, where it would correspond to the valence of parties or party groups), in principle, they could also be allowed to vary across elections.

References

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

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

Blundell, R., X. Chen, and D. Kristensen (2007): "Semi-Nonparametric IV Estimation of Shape-Invariant Engel Curves," *Econometrica*, 75, pp.1613–1669.

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

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

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

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

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

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

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

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