Review of Economic Studies (2017) **84**

election according to how "close" they are to her ideal. More precisely, consider a situation where a group of voters is facing a contested election with any number of candidates. Suppose that each voter has preferences (their bliss point) that can be represented by a position in some common, multi-dimensional ideological (metric) space, and each candidate can also be represented by a position in the same ideological space. According to the spatial framework, each voter will cast her vote in favour of the candidate whose position is closest to her bliss point (given the positions of all the candidates in the election).[2] In this case, we say that voters .[3]

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

Since our analysis focuses on retrieving individual-level fundamentals from aggregate data, it is related to the ecological inference problem.[5] It is also related to the vast literature on identification and estimation of discrete choice models.[6] In particular, our article is most closely related to the industrial organization literature on discrete choice models with random coefficients and macro-level data (Berry , 1995 and, more recently, Berry and Haile, 2014), and pure characteristics models (see Berry and Pakes, 2007 and references therein).[7]

In the language of the pure characteristics model, in our environment, the "consumer" (the voter) obtains utility $U^{\mathbf{t}}(C)$ $(C \quad \mathbf{t})$ $W(C \quad \mathbf{t})$ from "product" (candidate) , where $\mathbf{t}$ is a vector of individual "tastes" (the voter's bliss point), C is a vector of "product characteristics" (the candidate's position) and W is a matrix of weights. Also, the distribution of tastes depends on "market" (electoral precinct) level covariates, both observed and unobserved.[8] Whereas the distribution of tastes is typically taken to be parametric in pure characteristics models, we show that it can be - identified and estimated together with the finite dimensional components of the model (W). Our identification strategy relies on the geometric structure induced by the functional form of the utility function implied by the spatial theory of voting.

Part of the identification strategy we develop in this article is related to previous work by Ichimura and Thompson (1998) and Gautier and Kitamura (2013) on binary choice models

2. Data sets containing measures of the ideological positions of politicians based on their observed behaviour in office are widely available (see, Poole and Rosenthal, 1997 and Heckman and Snyder, 1997 for the U.S. Congress or Hix , 2006 for the European Parliament).

3. For a survey of alternative theories of voting, see Merlo (2006).

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

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

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

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

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

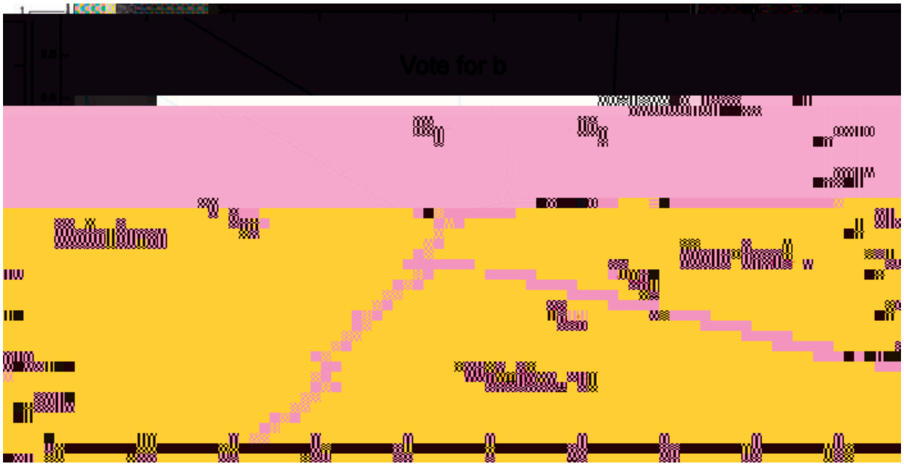

FIGURE 1

The Voronoi Tessellation for a five-candidate election in $\mathbb{R}^2$ and $W \quad I$

to the position of candidate and vice versa for the set Y_C^C. Hence, for each candidate , $V^W(\mathcal{C})$ is the intersection of the half spaces determined by the 1 hyperplanes $(H^W(C_{.}\ C$

Independent variation in characteristics is also used to identify the distributions of interest in Ichimura and Thompson (1998) and Gautier and Kitamura (2013). We also note that, except for in Ichimura and Thompson (1998), which can be directly applied to our setting to establish identification for the two-candidate case. The argument in the proof generalizes the simple insight that for two-candidate elections the Voronoi tessellation is given by an affine hyperplane. One can then sweep the space looking for an affine hyperplane that delivers different election outcomes for two distinct preference type distributions. That such an affine hyperplane exists is guaranteed by the Cramér–Wold device.[15] In fact, even for the case where there are more than two candidates, as

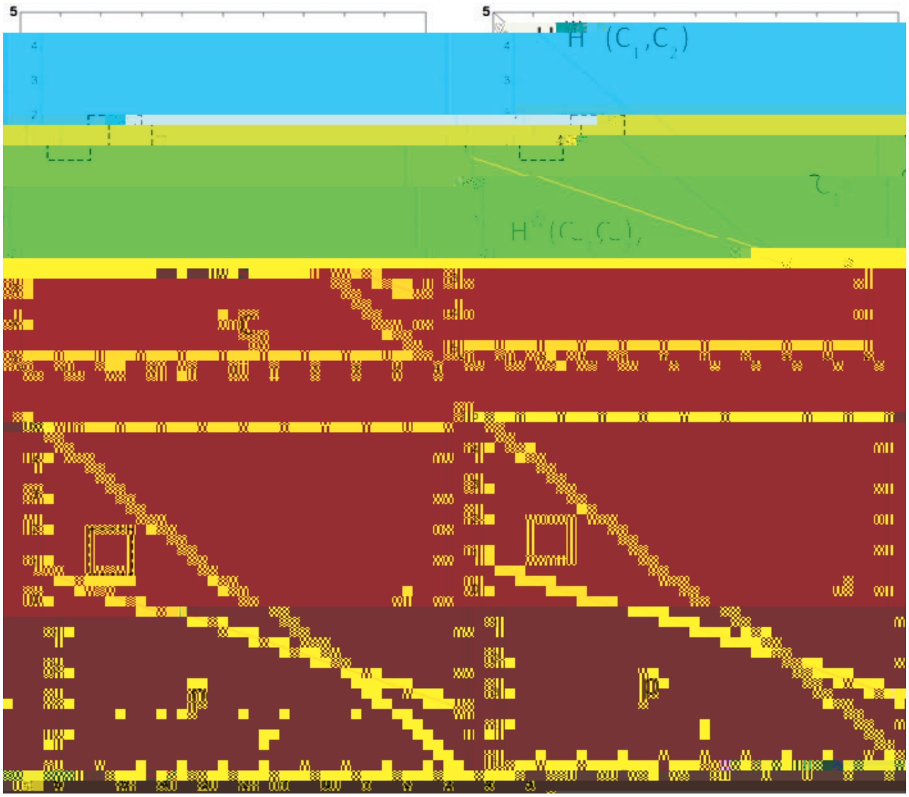

FIGURE 2

Voronoi tessellations for candidates C_1, C_2

are two candidates are separated by the line

$$H^W(C_1, C_2) \equiv \{ \mathbf{t} \in \mathsf{R}^2 : \underbrace{C_1' W C_1 \; C_2' W C_2 \; 2(C_2 \; C_1)' W \mathbf{t}}_{\equiv \; {}^W(\mathbf{t}, C_1)^2 \; {}^W(\mathbf{t}, C_2)^2} \; 0 \}. \tag{2}$$

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

Note also that the two lines $H^W(C_1, C_2)$ and $H^{\overline{W}}(C_1, C_2)$ intersect at the mid-point $(C_1 C_2)/2$. If the two tuples (P_{T}, W) and $(\mathsf{P}_{\overline{T}}, \overline{W})$ are observationally equivalent, the two candidates C_1 and C_2 should obtain the same shares of votes under (P_{T}, W) as they would under $(\mathsf{P}_{\overline{T}}, \overline{W})$. Denote by the vote share of candidate C_2. As indicated in the two lower panels in Figure 2, this is the probability of the area $w H^W(C_1, C_2)$ under P_W and the area $w H^{\overline{W}}(C_1, C_2)$ under $\mathsf{P}_{\overline{W}}$.

One can then obtain a translation of the candidates, say (C_1', C_2'), such that $C_1 \; C_2 \; C_1' \; C_2'$, and the same original Voronoi diagram is generated under W, as illustrated in the upper-left panel in Figure 3. The line characterizing the $\overline{W}$-Voronoi cells for the new pair (C_1', C_2') is parallel to

that the addition of more information with a larger number of candidates would still allow for identification. This is indeed so. As in Lemma 1, this is established if one can sample candidates function as in the previous case.[17] It is nevertheless true that for a given election:

$$\mathsf{E}\left[\int \mathbf{1}_{\mathbf{t}\in V^{W}(\mathcal{C})}\ T\ X(\mathbf{t}\ \mathbf{X})\ \ \mathbf{t} \qquad \middle|\ \mathbf{X}\right] \quad 0 \qquad \in\ 1 \quad .$$

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

In a parametric context, this structure suggests searching for parameters characterizing W and

of J and J and using the optimal values as starting parameters for higher orders. The program is executed in `Fortran` using a High Performance Computing cluster. In our estimation, we follow Gallant and Tauchen (1989) and rescale the covariates (see Section 5 for further details).

To establish consistency we rely on the following assumptions:

Assumption 3. *( ) E* ; ( ) $(\mathbf{X})$ *w* - ; ( ) $\mathbf{X}$ *w* *0.*

Assumption 4. *( ) T* $\mathsf{E}\ {}^J(\mathbf{X})\ {}^J(\mathbf{X})$ *w* J; ( ) $(\cdot)$ *w* $\mathsf{E}\ (\mathbf{X})^2\ \infty$, ${}^J(\mathbf{X})$ $\mathsf{E}\ (\mathbf{X})\ \left|\ {}^J(\mathbf{X})\ \right|^2$ (1).

Assumption 5. *( )* $\left|\ \widehat{}(\mathbf{X})\ \ (\mathbf{X})\right|$ (1) $(\mathbf{X})$; ( ) $\left|\ (\mathbf{X})\right.$ $(\mathbf{X})$.

Assumption 6. *L* (J) $(\cdot)$ (J) (J) *(8). A* , (W) W. *T* , ($\in$.

under the proportional representation system and typically with closed party lists. This means that voters in each electoral precinct do not vote for specific candidates, but for parties, and the total fraction of votes received by a party across all electoral precincts determines its proportion of seats in the Parliament. The identity of the politicians elected to Parliament is then determined by the parties' lists (if a party obtains three seats, the first three candidates on its list are elected).[27] Hence, in this context, the electoral candidates in an election are the parties competing in the election. As pointed out by Spenkuch (2015), among others, under proportional representation "it is in practically every voter's best interest to reveal his true preferences over which party he wishes to gain the marginal seat by voting for said party" (p. 1). In other words, in elections with proportional representations, voters have no incentives to behave strategically, and the maintained assumption that voters vote ideologically is particulary well suited for the European Parliament elections.

Our data consist of ideological positions of the candidates/parties competing in the election, electoral outcomes, and demographic and economic characteristics, for each electoral precinct. Since data on all demographic and economic variables are not available at the electoral precinct level for Austria, Belgium, Denmark, and Italy, we exclude these countries from the empirical analysis. Hence, our data set is a cross-section of elections for the European Parliament in the 693 electoral precincts of Finland, France, Germany, Greece, the Netherlands, Ireland, Portugal, Spain, Sweden, and the U.K. in 1999.[28]

The ideological positions of the parties were obtained from Hix (2006), who used roll-call data for the 1999–2004 Legislature of the European Parliament to generate two-dimensional ideological positions for each member of parliament along the lines of the NOMINATE scores of Poole and Rosenthal (1997) for the U.S. Congress.[29] As indicated in Heckman and Snyder (1997), ideological positions are obtained essentially through a (nonlinear) factor model with a large number of roll-call votes and parliament members. Given the magnitude of these dimensions, we follow the empirical literature on "large N and large T" factor models and take these scores as data (see, Stock and Watson, 2002; Bai and Ng, 2006a or Bai and Ng, 2006b).

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

The members of the European Parliament (MEPs) organize themselves into ideological party groups (EP groups) as in traditional national legislatures. Each EP group contains all the MEPs representing the parties that belong to that group. Within each country, it is typically the case that parties that belong to the same EP groups form electoral coalitions, where all the parties in the same EP group run a common electoral campaign based on a unified message representing the ideological positions of their group. Often, these positions vary across electoral constituencies within a country, representing regional differences in policy stances.[30] Since the closed-list proportional representation system induces strong party cohesion (see

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

28. We only have complete data on one electoral precinct in Ireland, Dublin, which is included in our analysis.

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

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

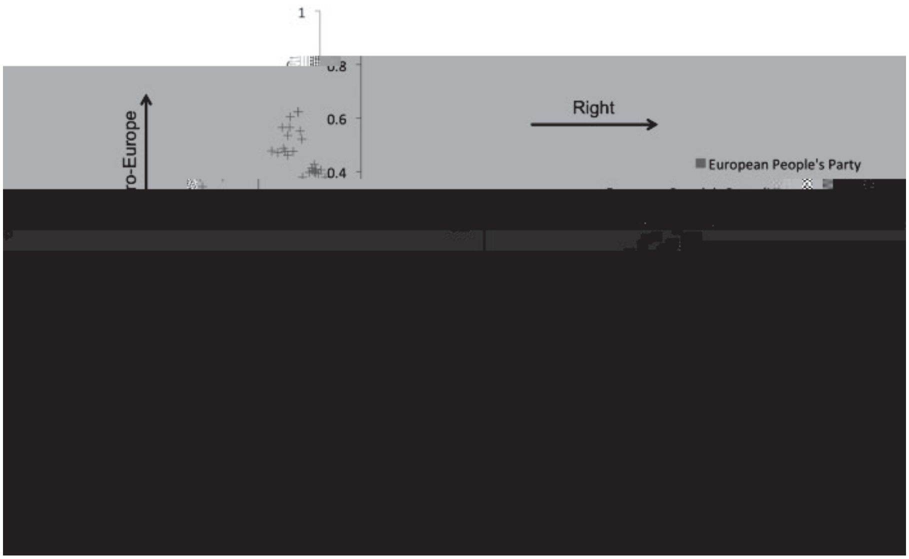

FIGURE 4
Candidate Positions, 1999

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

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

To further illustrate the data on ideological positions, in Figure 5 we also plot the ideological positions of a few notable politicians who ran in the 1999 European Parliament elections as front runners on their parties' lists. On the left-wing/pro-Europe quadrant, for example, we can locate François Hollande, current president of France, at coordinates (0 372, 0 609), whereas in the South-west quadrant (left, anti-Europe integration), we find Claudia Roth, leader of the German Green Party, at coordinates (0.715, 0.663). In the right-wing/anti-Europe quadrant, we find Nicholas Clegg, leader of the U.K. Liberal Democrat Party, at (0 123, 0 049); Jean-Marie Le

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

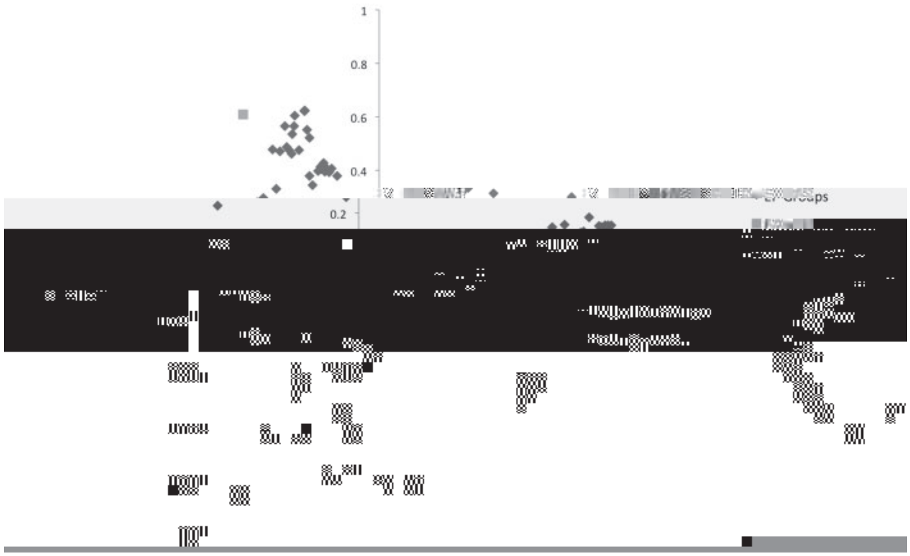

FIGURE 5
Individual politician positions, 1999

Pen, founder and former leader of the French National Front Party, at (0 576, 0 816); and Nigel Paul Farage, leader of the U.K. Independence Party, at (0 566, 0 825).[32]

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

Each electoral precinct corresponds to a different tessellation of the ideological space, and we measure the proportion of voters in each cell using the proportion of votes obtained by each of the candidates in that electoral unit. Figure 7 combines the Voronoi tessellations for all the elections in our data. It is apparent from the figure that these tessellations cover the ideological space and provide sufficient variation that allows us to identify and estimate the distribution of voter types (see our discussion of the conditions for identification in Section 3 above).

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

We combine the data on the ideological positions of electoral candidates with electoral outcomes in the 1999 elections and demographic and economic variables at the electoral precinct level from the 2001 European Census.[33] The election outcomes data were obtained from the CIVICACTIVE European Election Database.[34] The demographic and economic data were obtained from EUROSTAT and we extracted four variables at the electoral precinct level: the

32. Note that Le Pen and Farage are remarkably aligned in the ideological space. This may not come as a surprise after Marine Le Pen, daughter of Jean-Marie Le Pen, tweeted “congratulations” to the U.K. Independence Party after their recent success in local elections.

33. Since the European Census is conducted every 10 years, we use data from the 2001 census, which is the closest to 1999.

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

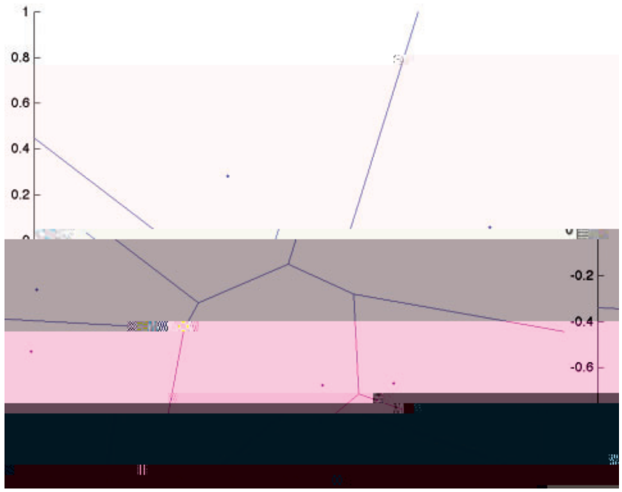

FIGURE 6
Voronoi diagram for Paris (France), 1999

FIGURE 7
Superimposed Voronoi diagrams, 1999

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

Using these data, which as noted above contain a cross-section of 693 elections, we estimate our model. Following Gallant and Tauchen (1989), we re-scale the data to avoid situations

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

TABLE 1

C ( )

	Dimension 1		Dimension 2	
	Min	Max	Min	Max
Finland	0 802	0 572	0 597	0 474
France	0 834	0 569	0 792	0 280
Germany	0 885	0 690	0 438	0 622
Greece	0 815	0 587	0 550	0 551
Ireland	0 874	0 547	0 376	0 564
Netherlands	0 856	0 577	0 518	0 461
Portugal	0 846	0 580	0 632	0 475
Spain	0 916	0 629	0 400	0 603
Sweden	0 833	0 571	0 591	0 274
U.K.	0 868	0 899	0 855	0 521

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

TABLE 2

S

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

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

in which extremely large (or small) values of the polynomial part of the conditional density are required to compensate for extremely small (or large) values of the exponential part. We transform the data so that $\tilde{\mathbf{X}}\ \ \mathbf{S}^{\,1\ 2}(\mathbf{X}\quad\overline{\mathbf{X}})$ where $\mathbf{S}\ \ (1\ E)\sum^{E}_{\ \ 1}(\mathbf{X}\quad\overline{\mathbf{X}})(\mathbf{X}\quad\overline{\mathbf{X}})\ .\ \overline{\mathbf{X}}$ $(1\ E)\sum^{E}_{\ \ 1}\mathbf{X}$ and $\mathbf{S}^{\,1\ 2}$ is the Cholesky factorization of the inverse of $\mathbf{S}$. The estimates for $(\cdot)$ as defined in equation (5) are linear projections on covariates. We use Hermite polynomials of

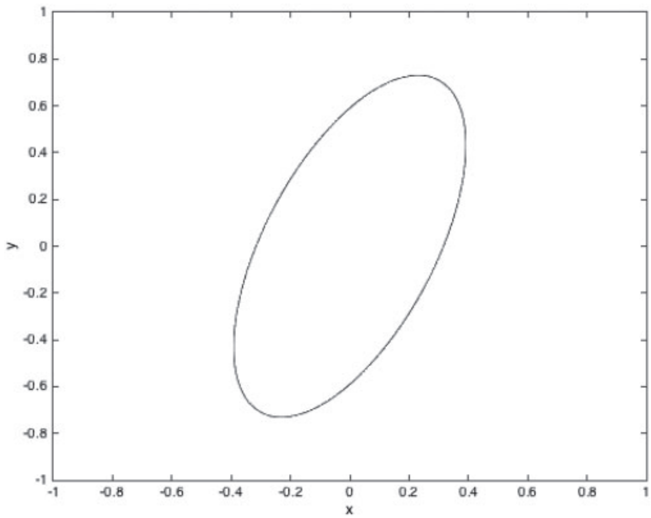

FIGURE 8

Indifference curve for $^{W}(\mathbf{0}\ C)\quad 1$

order J 2 (types) and J 2 (covariates).[36] Finally, we use the identity matrix as our estimation weighting matrix ( ).

The estimates of the weighting matrix W we obtain are $W_{2\,2}$ 0 287 and $W_{1\,2}$ $W_{2\,1}$ 0 316. Bootstrap standard errors for $W_{2\,2}$ and $W_{1\,2}$ are equal to 0 052 and 0 049, respectively. Given J and J , the estimator is essentially an (overidentified) GMM estimator. We compute the standard errors from estimates obtained from 200 bootstrap samples (after recentring the targeted moments as recommended by standard practice, see Horowitz, 2001).[37] Bootstrap standard errors are also presented for functionals of the estimated distributions of voter types.

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

Turning attention to the estimates of the distribution of the ideological positions of voters, $\mathsf{P}_{T\,X}$, Figure 9 plots level curves for the voter type distribution for electoral precincts at the 75th percentile of the female-to-male ratio (approximately 1.06 in our data) and the 25th percentile of the proportion of residents above 35 years old (approximately 0.58 in the data) and various percentile combinations for the other two variables (per-capita GDP and the unemployment

36. In total, we have 78 parameters, including the parametric component ($(J\ )$ $(J\ )$ (W)). Since we have up to seven candidates per election (1 6atio n and thtD(n)Tjp7(the)-224.7 TD(n)Tjfercentile86nPnš emal0es)didates

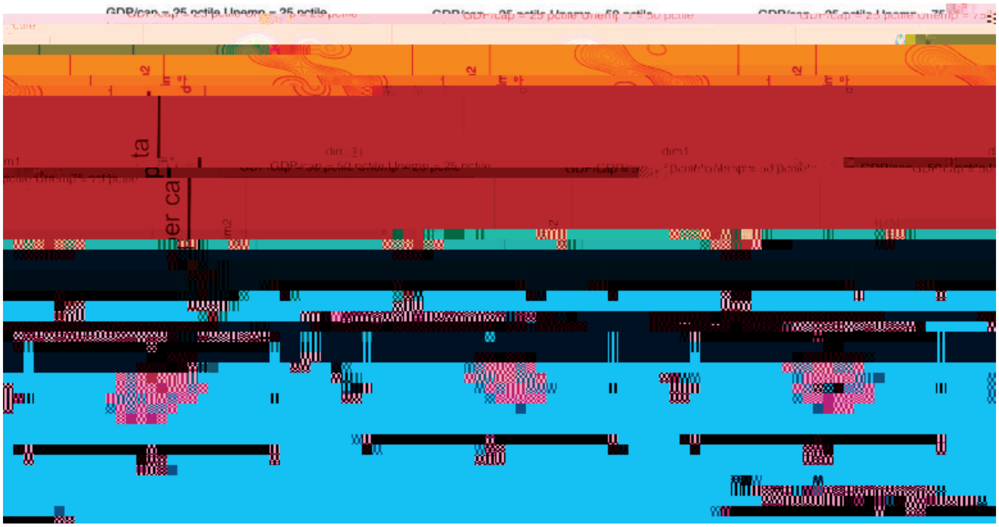

FIGURE 9

Results at percentiles of conditioning variables (female/male: 75 percentile and percentage / 35 Years old: 25 percentile)

rate).[38] As we can see from the figure, multi-modality and non-concavity are pronounced features of the recovered distribution of voter preferences. These findings represent a potential challenge for theoretical research in political economy, which systematically assumes that the distribution of voters' preferences is uni-modal and/or log-concave (see, Austen-Smith and Banks, 2000; Persson and Tabellini, 2000 and Austen-Smith and Banks, 2005).

Another summary of our estimates is presented in Table 3, where we present the average coordinates of the estimated distribution of voter preferences and the correlation between coordinates for each country in our sample (see the table's notes for the exact construction). For purposes of comparison, Table 3 also reports the average coordinates of the distribution of candidate positions in the data and the correlation between coordinates for each country. According to our estimates there is a positive correlation between candidates' (average) coordinates and voters' (average) coordinates, which is equal to 0.76 for dimension 1 and 0.23 for dimension 2.[39] With respect to the correlation between (average) coordinates, the signs of the correlation for voters and for candidates are the same for six out of the ten countries.

To investigate the relationships between demographic and economic variables and the distribution of voters' preferences, Tables 4 and 5 report the fraction of voters who are on the right of the left–right socio-economic policy dimension and the fraction of voters who are pro-Europe, respectively, for electoral precincts at the 25th, 50th, and 75th percentiles of each covariate and average levels for all other covariates.[40] As we can see from the tables, electoral precincts with a relatively larger female-to-male ratio, precincts with a relatively larger share of the population above the age of 35 years, and precincts with a relatively higher level of GDP per-capita are

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

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

40. Loosely speaking, the table reports the "marginal effects" of each covariate.

employment status of European citizens and their sentiments towards European policies.[41] In particular, according to the 1999 survey, relatively fewer women (37.04%) and relatively fewer people who are unemployed (29.71%) locate themselves on the “right” of the political spectrum than men (39.48%) and people who are employed (38.09%), respectively.[42] Moreover, according to the 1995 EUROBAROMETER survey, relatively fewer women (14.92%) and relatively more people who are unemployed (18.17%) consider their country’s membership of the European Union “a bad thing” than men (15.87%) and people who are employed (15.43%), respectively.[43] On the other hand, according to the same EUROBAROMETER surveys, relatively more people older than 35 years locate themselves on the “right” of the political spectrum (39.39%) and consider their country’s membership of the European Union “a bad thing” (16.57%) than their younger counterparts (36.35% and 12.92%, respectively), which is somewhat at odds with our findings.

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

6. DISCUSSION

In this article, we have addressed the issue of non-parametric identification and estimation of voters’ preferences using aggregate data on electoral outcomes. Starting from the basic tenets of one of the fundamental models of political economy, the , and building on the work of Degan and Merlo (2009), which represents elections as Voronoi tessellations of the ideological space, we have established that voter preference distributions and other parameters of interest can be retrieved from aggregate electoral data. We have also shown that these objects can be consistently estimated using the methods by Ai and Chen (2003), and have provided an empirical illustration of our analysis using data from the 1999 European Parliament elections.

One potential extension of interest allows for electoral candidates to differ not only with respect to their locations in the ideological space, but also with respect to (non-spatial) individual characteristics related to their quality. These quality characteristics, which are commonly referred to as “valence” in the literature (see, 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. The identificaton of such a model may be demonstrated along the same lines of our previous results and we provide further discussion in the Online Appendix.

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

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

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

To conclude, it may be useful to cast our model into the broader context of a general spatial model of preferences with generic products, where the "consumer" obtains utility $U^{\mathbf{t}}(C\ )$ $(C\quad \mathbf{t})\ \ W(C\quad \mathbf{t})$ from "product" , $\mathbf{t}$ is a vector of individual "tastes",

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

CHEN, X. and POUZO, D. (2009), "Efficient Estimation of Semiparametric Conditional Moment Models with Possibly Nonsmooth Residuals", *J E*, **152**, 46–60.

CHESHER, A. and SILVA, J. M. C. S. (2002), "Taste Variation in Discrete Choice Models", *T R w E S*, **69**, 147–168.

NEWEY, W. K. and POWELL, J. L. (2003), "Instrumental Variable Estimation of Nonparametric Models", *E* ,
71, 1565–1578.
OKABE, A., BOOTS, B., SUGIHARA, K. and CHIU, S. N. (2000), *S* *T*