Does Spatial Correlation Matter?

An Innovative Approach to the Characterisation of the European Political Space.*

Giovanna Iannantuoni[†], Elena Manzoni[‡]
and Francesca Rossi[§]

The European political spectrum can be modeled as a two-dimensional space, whose interpretation has been investigated in the spatial voting literature by standard regression analysis. We show that data on legislators' positions display spatial clustering that is not fully explained by the standard models. We take into account correlation among legislators by explicitly modelling spatial dependence across countries, which in turn relies on new sets of linguistic, geographical, institutional and cultural metrics. We confirm the well know result that the first dimension of the European political space is mainly explained by the Members of European Parliament's ideological position on a left-right scale, although we show that spatial correlation across legislators cannot be neglected. In addition, we also show that spatial correlation plays a central role when interpreting the more controversial second dimension of the political spectrum, where the most relevant metric is based on an institutional index.

Keywords: European political space, spatial autoregressions, NOMINATE, proximity matrices, economic distances.

JEL codes: D72, C21.

*We thank John Ashworth, Marco Boretto, Michela Cella, Simon Hix, Jean-Francois Laslier, Niklas Potrafke, Leonce Roeth, Klaas Staal and Shlomo Weber for helpful comments. We also thank Keith T. Poole for his support with the NOMINATE procedure. Elena Manzoni and Francesca Rossi gratefully acknowledge financial support from Einaudi Institute for Economics and Finance (EIEF) - Research Grant 2015.

†Department of Economics, Management and Statistics, University of Milan-Bicocca.

‡Corresponding author. Department of Economics, Management and Statistics, University of Milan-Bicocca. e-mail: elena.manzoni@unimib.it

§Department of Economics, University of Verona.

1 Introduction

The European Parliament is an institution of particular interest from the point of view of economists and political scientists. It is a relatively young institution, which has increasingly gained more power in terms of the set of issues that it is called to decide upon, the number of voters represented, and the number of votes it casts. A peculiar characteristic of the European Parliament is the strong heterogeneity of its components. Members of the European Parliament (MEPs) are elected in districts that do not cross national borders, from lists chosen by national parties, and with electoral rules that, albeit proportional, are country-specific. Therefore, MEPs represent their countries and their national parties, as well as the European Political Group they belong to. Moreover, they are only accountable to their national electorate and with rules that differ from country to country. As a consequence, politics in the European Parliament are likely to be subject to more influences and effects than politics in national parliaments, as Hix et al. (2007) discuss extensively.

The understanding of what drives legislators' behaviour in the European Parliament is of a fundamental building block for the evaluation of institutional changes (e.g., changes in the electoral rules of the European Parliament, in the composition of the Parliament due to an enlargement of the EU, or to phenomena such as Brexit). Cross-influences among legislators may change the consequences of policies or institutions enhancing or weakening their effects, thus favouring or opposing the policy makers in reaching their objectives.

The heterogeneous composition of the European Parliament is likely to induce the formation of spatial networks among MEPs with respect to different similarity criteria which we aim to explore and understand. 'Space' here is intended in a broad way that includes economics/cultural characteristics. The aim of this paper is to shed light on which of these spatial networks are relevant to determine the positioning of legislators in a policy space, which is intrinsically multidimensional. Thus, the novelty of our approach compared to that of e.g. Hix, Noury and Roland, 2006 (HNR) is that the analysis is performed introducing a spatial component in the regressions by means of the spatial autoregression models (SARs).

Over the past few decades, a growing literature in political economy has focused on the analysis of the determinants of legislators' behaviour in Congress by means of records of roll call votes (Poole and Rosenthal, 1997; Rosenthal and Voeten, 2004). Multidimensional scaling methods have been designed to map information from roll call votes record to a policy spectrum, with a certain number of postulated dimensions, which contain legislators' relative ideal points (e.g. Poole, 2005; Poole and Rosenthal, 1997, and references therein). These techniques have been used recently to analyse other supranational settings such as the European Parliament (HNR; Hooghe, Marks and Wilson, 2002). More specifically, HNR's analysis of the political space and its dimensionality shows that the policy space can be described by two dimensions, where the first is related to the national party ideological positioning on a left-right scale while the second one is related to the national party's positions on European integration, although the latter interpretation is more controversial than the former. Unlike HNR we explicitly take into account possible spatial correlation across legislators' positions by means of SARs models. We confirm their interpretation of the first dimension as driven by the ideological positioning on the left-right scale. We provide further insights on the second dimension, highlighting the presence of spatial effects.

Section 2 describes the methodology and the proximity matrices, Section 3 presents the results and Section 4 concludes.

2 Methodology

In spatial voting models legislators' preferences are characterized by their ideal points. The existing literature (see HNR) shows that about 90% of legislators' choices are correctly classified if a bi-dimensional space is postulated, i.e. if legislators' positions are bi-dimensional vectors. The interpretation of these underlying dimensions of the policy space is the main object of our analysis.

In order to determine the bi-dimensional vector of the legislators' positions, we rely on the methodology known as NOMINATE (Poole and Rosenthal, 1997, and references therein), which in a nutshell is a scaling method for roll call votes data. The NOMINATE technique has been designed to estimate the position of each legislator's bliss point in the policy space. The procedure relies on a criterion of similarity across legislators, based on the number of times they vote in the same way in roll calls. Hence, NOMINATE estimates relative positions across legislators rather than their actual ideal points along the two aforementioned dimensions.[1] NOMINATE estimates are informative only to the extent of the relative positions of legislators, but do not convey any insight about the economic and political meaning of the dimensions themselves, which in turn have to be investigated using regression analysis (e.g. in HNR).

The novelty of our approach is the introduction of a spatial component in these regressions in order to account for the possible transnational networks across legislators, by means of SARs.[2] In this context, 'space' is defined in broader terms compared to the standard geography literature, and it relies on the definition of a general economic distance. The standard notion of geographical space is thus covered as a special case. In general, SARs offer a useful framework for describing data which are recorded across

[1] Our results are derived by applying the weighted dynamic version of NOMINATE, DW-NOMINATE, to take full advantage of a larger dataset. As we only deal with five legislatures, the standard static model would most likely deliver similar results.

[2] For exhaustive surveys of SARs and applications see for instance Elhorst (2014). Examples of the application of SAR models to political economy can be found in Williams (2015) and Böhmelt et al. (2016).

'space', and are thus irregularly spaced, without a natural ordering and possibly without geographical interpretation, such as legislators' coordinates. In SAR models the cross-correlation across agents is embodied in a weight matrix, denoted W, which needs to be chosen by the practitioner and cannot be estimated. Thus, the spatial structure of data is assumed to be known up to one (or few) parameter(s) that defines the strength of the spatial correlation. Conventionally, the spatial interaction of each legislator with itself is set to zero. As in many empirical settings we normalise W so that the entries in each row sum to one.

Let y and X be standard sets of observable variables, indicating respectively dependent and independent variables, while ϵ indicates a vector of independent and identically distributed normal random variables, with mean zero and unknown variance σ^2. We adopt the variant of SAR known as Spatial Durbin model, defined as

$$y = \lambda W y + X\beta + W X \gamma + \epsilon, \tag{1}$$

for some unknown parameters β and γ and a scalar unknown parameter λ. The significance and magnitude of the estimates of λ and γ define the spatial effects. According to eq. (1), the dependent variable of each unit is not only explained by its own vector of characteristics X, but it is also related to a weighted average of the features of neighbouring units. We refer to the spatial terms Wy and WX as 'endogenous' and 'exogenous' spatial effects, respectively, in order to differentiate the channels through which ideal points' of legislators are related to those of their neighbours. The endogenous component Wy allows each agent's ideal point to be potentially related to a weighted average of his/her neighbor's ideal points, while WX captures the explicit relationship between one agent's ideal point and a weighted average of his/her neighbor's characteristics.

The model of eq. (1) is a very parsimonious method of describing spatial dependence, conveniently based only on economic distances rather than actual locations. Although a drawback of SARs is the ex ante specification of W, eq. (1) has been widely used in practice because of its flexibility and indeed allows us to investigate the effects of multiple sources of interactions among legislators.

We generate several versions of the proximity matrices. These are built from the pairwise distance d_{ij} between home countries' characteristics of legislators i and j.[3] Each choice of W is built using $w_{ij} = \frac{1}{d_{ij}}$, where w_{ij} denotes the $i-j$th entry of W corresponding to legislators i and j. We set w_{ij} equal to zero if legislators i and j belong to the same country, as we are interested in the implication of transnational correlations across legislators. We define matrices based on geographical, linguistic, institutional and cultural distances.

Geographical proximity is based on the distance in kilometres between capitals of

[3] In order to define W, national rather than individual characteristics are used as a more interesting set of distances is available.

European member states of legislators i and j, measured as the average of the shortest outbound and inbound routes suggested by Google Maps.
Linguistic proximity is based on a the lexiscostatistical distance between languages of legislators' home countries (Dyen et al., 1992).[4] Linguistic proximity has an effect on economic and political outcomes such as trade, immigration and voting behaviour, as shown by Ginsburgh and Weber (2011).
Institutional proximity is based on the distance between legislators' institutional background, measured by their home country score in terms of the Parliamentary Power Index by Fish and Kroenig (2009). The authors identify 32 possible powers that a legislature may have and compute the Parliamentary Power Index as the fraction of such powers that a legislature has. The distance based on PPI is defined as $|PPI_i - PPI_j|$.
Cultural proximity is based on one of the cultural indexes by Hofstede et al. (2010) These indexes describe differences in national cultures along orthogonal dimensions. In the paper we report results based on the **masculinity** index (**MAS**) which is the most relevant.[5] MAS classifies societies based on the distinction (or absence of distinction) of emotional roles by gender. High MAS scores are correlated with preferences for large organizations, with the tendency of resolving conflicts by letting the strongest win and with low participation of women in politics and management. The distance based on MAS is $|MAS_i - MAS_j|$.[6]

3 Results

The estimates of legislators' bliss points on the two dimensions of the policy space, i.e., the first step of our analysis, have been obtained by implementation of the DW-NOMINATE. As in HNR, we focus on the first five legislatures and construct our dependent variable, y_d, where $d = 1, 2$ indicates dimension, as the averages of legislators' positions belonging to the same national party. Therefore, our analysis focuses on national party characteristics rather than on individual legislators' features. In order to take advantage of a larger dataset, we stack data for five legislatures. Data pertaining to roll calls, national parties and European political groups have been obtained from http://personal.lse.ac.uk/hix/HixNouryRolandEPdata.HTM. Our dataset consists of a total of 347 data points. We refer to HNR for an exhaustive descriptive analysis of roll call votes data and of legislators' respective positions in the political space.

We begin our analysis by looking for the presence of spatial correlation across y_d along

[4] For the construction of the linguistic matrix, French Belgium and Flemish Belgium were considered as separate countries. This lexicostatistical distance is not available for pairs which involve legislators from Finland, as their official language is not Indo-European. We set all these values to 0 (minimal proximity, or maximal distance).

[5] Results based on additional cultural metrics by Hofstede et al. (2010) are available upon request.

[6] MAS is available separately for French Belgium and Flemish Belgium, which have been treated as separate countries.

the two dimensions. We perform a Moran I test (Moran, 1950), which is designed to detect spatial clustering of data according to the measure of proximity implied by the choice of W. If the value of the Moran I (MI) statistic exceeds the suitable χ^2 critical value we reject the null hypothesis of no spatial correlation. Results of MI test on y_d, $d = 1, 2$ (Table 1) reveal a strong spatial effect, especially on the second dimension.

Table 1: Moran I test based on raw data.

	First Dimension	**Second Dimension**
Km	3.64*	120.68***
Lang.	1.22	48.56***
Inst.	6.95***	102.38***
MAS	0.91	53.99***

Notes. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

We next replicate the analysis in HNR to test whether the aforementioned spatial correlation can be fully explained by an appropriate set of regressors or if a spatial model is needed. Let LR and $EUint$ be indexes of left-right political orientation and EU integration propensity, respectively. Let D be a set of dummy variables containing country-specific and European political group-specific controls, as well as dummy variables to indicate whether the national party was in power during each legislature, and whether it had a European Commissioner during such period of time (taking value one if it had a Commissioner for the whole period, 0.5 if it had a Commissioner for at least half of the period, and zero otherwise). LR and $EUint$ have been obtained from expert judgement data in Marks and Steenbergen (2004), while the dummy variables have been obtained from information contained in the European Parliament and European Commission websites. We estimate the parameters of the following regression

$$y_d = \beta_0 + \beta_1 LR + \beta_2 EUint + \gamma D + \epsilon \tag{2}$$

and perform an MI test on the obtained ordinary least squares residuals. The values of MI statistics (Table 2) indicate that residuals from regressions along both dimensions display severe spatial correlation for almost all the choices of proximity measures. Thus, the exogenous regressors are not able to account for spatial patterns in the dependent variables.

Outcomes of previous tests motivate the inclusion of explicit spatial components into the regression equations. Our main specification is

$$y_d = \lambda W y_d + \beta_1 LR + \beta_2 EUint + \beta_3 W * LR + \gamma D + \delta P + \epsilon, \tag{3}$$

Table 2: Moran I test based on the specification of Hix et al (2006).

	First Dimension	**Second Dimension**
Km	11.45***	24.54***
Lang.	2.42	25.78***
Inst.	7.14***	13.01***
MAS	14.41***	4.71**

Notes. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

where W is the proximity matrix, P is a set of dummy variables that controls for the legislature[7] and everything else is defined as in (2). Although data pertaining to different legislatures are pooled, W is constructed so that spatial correlation across observations only affects units within the same legislature. Thus, all our choices of W have a block diagonal structure where each block reflects interactions of national parties rather than agents within each legislature. As previously mentioned, $w_{ij} = 0$ if national parties i and j belong to the same country. We only include the exogenous effect $W * LR$ to avoid inflated standard errors, as a preliminary analysis reveals that $W * LR$ and $W * EUint$ are highly correlated. Results are reported in Tables 3 and 4.

The inclusion of the set of legislature dummies P allows us to control for time trends across European countries as we expect global political trends in Europe to generate ex-ante correlations across legislators along unobservable characteristics. Our specification isolates the effects of genuine cross-correlations within each legislature from the effects of the time-varying composition of the Parliament. Without the inclusion of P, if unobserved time trends led spatially close countries to elect candidates who are similar to each other, the spatial parameter estimates would be spuriously inflated. Indeed, the magnitude of the estimates of the spatial parameters would erroneously account for both these time trends and the genuine within-parliament spatial correlation generated by the various notions of proximity.

Estimation is performed by maximum likelihood. In addition to the standard estimates and t-statistics (in brackets), we report the estimated values of average marginal direct and indirect effects of both LR and $EUint$ on y_d, indicated as MDE_{LR}, MIE_{LR}, MDE_{EUint} and MIE_{EUint} in the Tables. $MDE_{LR/EUint}$ measures the effect of a change of $LR/EUint$ in unit i on y_d of unit i itself, averaged across all i. In addition to the standard effect captured by β_1/β_2, this measure contains the feedback generated by the network structure, i.e. the effect of a change in $LR/EUint$ induced on all other

[7] The first European Parliament is considered as reference group.

regions by a change in $LR/EUint$ of region i, and in turn their effect on y_d of region i itself. $MIE_{EU/EUint}$ indicates the change of y_d in region i for a shift of $LR/EUint$ in all other regions, averaged across i (e.g. LeSage et al., 2013).[8] For each table we also report the value of MI statistic to test the null hypothesis that the residuals computed are free from spatial correlation. If MI is not significant we can then conclude that all sources of spatial interactions have been controlled for.

Table 3 shows that the first dimension, as in HNR, is essentially interpreted as the ideological position on the left-right scale. Spatial correlation has a significant exogenous effect through $W * LR$. The effect of spatial correlation differs across proximity matrices. In particular, direct and indirect effects of LR reinforce each other when we consider institutional or MAS proximities, while this is not true for the ones based on geography or language. This suggest that legislators are influenced by the ideological positioning of those who are more affine to them in terms of policy-relevant background.[9] Spatial correlation based on geographical distance, instead, induces a total effect of LR ($MDE_{LR} + MIE_{LR}$) that is not significantly different from zero.

Figures in Table 4 show that institutional and MAS weight matrices generate a strongly significant endogenous effect (λ) along the second dimension, thus confirming the intuition that these are the most relevant matrices that we consider. In this case, however, direct and indirect effect move in opposite directions. Our interpretation is that legislators' positions along second dimension is mostly driven by national issues (e.g., agricultural policies), where legislators of each country are likely to maintain a tendency to differentiate themselves from those of other countries. Thus, direct and indirect effects have opposite signs, reflecting a negative spatial correlation. However, a different type of analysis is required in order to validate this interpretation, and to identify the peculiar issues for which the second dimension is relevant. For this purpose it is necessary to consider individual level observations, and to perform the NOMINATE scaling followed by a regression analysis issue by issue.

The relevance of the proximity measure based on MAS also suggests that gender composition of national parties may play an explicit role in interpreting the European policy space. If we explicitly include a measure of gender composition as a regressor in our model we can show that this is indeed true, but results in Tables 3 and 4 are robust to this modification.[10]

[8] t-statistics are reported in brackets and the corresponding critical values, as well as standard errors of $MDE_{LR/EUint}$ and $MIE_{LR/EUint}$ have been obtained by bootstrap.

[9] Recall that MAS is correlated with characteristics that affect institution and the policy-making process, such as the way in which conflicts are resolved.

[10] Results are available upon request.

Table 3: First dimension.

First Dimension	Km (1)	Lang. (2)	Inst. (3)	MAS (4)
λ	-0.2723	-0.0339	0.0850	0.0421
	(-1.08)	(-0.26)	(0.48)	(0.30)
LR	1.0567	1.0591	1.0499	1.0667
	(14.65)***	(14.59)***	(14.43)***	(14.68)***
EUint	0.0125	0.0124	0.0113	0.0115
	(1.40)	(1.39)	(1.27)	(1.28)
$W * LR$	-1.0252	-0.1782	0.3767	0.2264
	(-2.62)***	(-1.72)*	(2.19)**	(1.06)**
EP	Yes	Yes	Yes	Yes
EPG	Yes	Yes	Yes	Yes
Const.	0.1449	-0.1244	-0.5087	-0.3001
	(0.73)	(-0.88)	(-3.16)***	(-2.54)***
MDE_{LR}	1.0631	1.0595	1.0510	1.0672
	(14.49)***	(16.29)***	(21.33)***	(14.52)***
MIE_{LR}	-1.0384	-0.2075	0.5082	0.2827
	(-4.21)***	(-1.64)	(1.72)**	(1.29)*
MDE_{EU}	0.0125	0.0120	0.0110	0.0115
	(1.43)	(1.27)	(1.29)	(1.30)
MIE_{EU}	-0.0027	-0.0040	0.0010	0.0005
	(-0.82)	(-0.25)	(0.44)	(0.28)
MI	1.18	0.96	0.17	0.26
N	347	347	347	347

Notes. t-statistics in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

4 Conclusions

In this paper we highlighted how understanding spatial correlations among legislators is fundamental to understand the determinants of the European policy space. Following the literature (Hicks et al., 2006) we postulate a bi-dimensional policy space and investigate the substantive meaning of its two dimensions. The novelty of our approach rests in both the introduction of spatial autoregressive models in the analysis, and the discussion of some relevant proximity matrices.

Table 4: Second dimension

Second Dimension	Km (1)	Lang. (2)	Inst. (3)	MAS (4)
λ	-0.4420	0.1083	-0.6915	-0.3830
	(-1.78)*	(0.79)	(-3.23)***	(-2.40)**
LR	-0.7307	-0.7183	-0.7293	-0.7230
	(-5.50)***	(-5.39)***	(-5.53)***	(-5.49)***
EUint	0.0347	0.0331	0.0336	0.0331
	(2.13)**	(2.02)**	(2.08)**	(2.03)**
$W * LR$	0.1232	-0.1124	0.2865	0.4519
	(0.17)	(-0.55)	(0.93)	(1.18)
EP	Yes	Yes	Yes	Yes
EPG	Yes	Yes	Yes	Yes
Const.	-0.0178	0.1204	-0.0706	-0.1210
	(-0.05)	(0.47)	(-0.24)	(-0.55)
MDE_{LR}	-0.7340	-0.7194	-0.7417	-0.7350
	(-5.81)***	(-6.53)***	(-5.96)***	(-5.49)***
MIE_{LR}	0.3127	-0.2122	0.4799	0.5389
	(0.66)	(-0.96)	(3.23)***	(2.17)*
MDE_{EU}	0.0349	0.0331	0.0340	0.0333
	(2.18)*	(1.85)*	(2.12)**	(1.84)*
MIE_{EU}	-0.0108	0.0040	-0.0141	-0.0094
	(-1.62)	(0.98)**	(-1.69)*	(-1.42)
MI	1.42	0.13	0.13	0.57
N	347	347	347	347

Notes. t-statistics in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

We find that the first dimension is essentially explained by the left-right ideological positioning, consistently with the existing literature. We show, however, that LR has both a direct and an indirect effect on legislators' positioning, and that these effects reinforce each other when we consider the institutional and cultural proximity matrices.

Results on the second dimension support the observation that the institutional and cultural matrices are the most relevant. However, along the second dimension, direct and indirect effects have opposite signs suggesting that national parties have a tendency to differentiate themselves from parties of neighboring countries, counterbalancing their possible movements in the political space. This peculiarity is likely to depend on the type of issues on which the second dimension is relevant, which in turn could have national characteristics (e.g. agricultural policies). In order to fully confirm our interpretation we would need estimates of legislators' positions obtained from roll call votes on separate issues, and a regression analysis performed at individual rather than at national party's level. This is currently under investigation in separate work.

References

[1]Böhmelt T., L. Ezrow, R. Lehrer, H. Ward (2016). Party Policy Diffusion, *American Political Science Review*, 110(2), 1-14.

[2]Dyen, I., J.B. Kruskal and P. Black (1992). An Indo-European classification: A lexicostatistical experiment, *Transactions of the American Philosophical Society*, 82 (5), 1-132.

[3]Elhorst, J.P. (2014). *Spatial Econometrics From Cross-Sectional Data to Spatial Panels*. Springer.

[4]Fish M.S. and M. Kroenig (2009). *The Handbook of National Legislatures. A Global Survey.* Cambridge University Press.

[5]Ginsburg V. and S. Weber (2011). *How Many Languages Do We Need? The Economics of Linguistic Diversity.* Princeton University Press.

[6]Hix, S., A. Noury and G. Roland (2006). Dimensions of Politics in the European Parliament, *American Journal of Political Science* 50(2) 494-511.

[7]Hix, S., A. Noury and G. Roland (2007). *Democratic Politics in the European Parliament*. Cambridge University Press, Cambridge, UK.

[8]Hofstede G., Hofstede G.J. and M. Minkov (2010).*Cultures and organizations: software of the mind: intercultural cooperation and its importance for survival.* New York: McGraw Hill.

[9]Hooghe L., G. Marks and C.J. Wilson (2002). Does Left/Right Structure Party Positions on European Integration?, *Comparative Political Studies*, 35, 965-989.

[10]LeSage J.P. and R.K. Pace (2013). Interpreting Spatial Econometric Models, in *Handbook of Regional Science*, 1535-1552.

[11]Marks, G. and M. Steenbergen (2004). *Marks/Steenbergen Party Dataset*, Chapel Hill: University of North Carolina Chapel Hill, available at: http://www.unc.edu/∼gwmarks/data.htm.

[12]Moran, P.A.P. (1950). A test for the serial dependence of residuals. *Biometrika*, 37, 178-81.

[13]Poole K.T. (2005). *Spatial Models for Parliamentary Voting.* Cambridge University Press.

[14]Poole K.T. and H.Rosenthal (1997). *Congress: A Political-Economic History of Roll Call Voting*. New York: Oxford University Press.

[15]Rosenthal H., E. Voeten (2004). Analyzing Roll Calls with Perfect Spatial Voting: France 1946-1958, *American Journal of Political Science*, 48(3), 620-632.

[16]Williams L.K. (2015). It's all relative: spatial positioning of parties and ideological shifts, *European Journal of Political Research*, 54, 141-159.