

Spatial Analysis for Electoral Data: the Spread of Territorial Parties in Italy

Rosario D'Agata* and Venera Tomaselli**

Abstract To study the electoral spread of territorial political parties, we analysed the role of spatial contiguity. We therefore measured spatial autocorrelation and estimated autoregressive spatial models in order to explain the pattern of relationship among territorial units. By means of spatial models, we investigated the electoral performance of *Lega*, comparing two northern Italian regions - Lombardia and Veneto - where this party achieves its highest amount of votes.

Keywords Spatial correlation. Neighborhood effects. Spatial data analysis. Electoral data.

1 Territorial electoral data and spatial analysis

Many scholars have remarked the importance of territorial contiguity in political research. Employing measures of spatial analysis among neighbour states, some authors investigated the spread of democratic institutions in the world (O'Loughlin et al., 1998). In electoral studies, moreover, the analysis of spatial interaction among adjacent units was applied in research about electoral participation (Darmofal, 2006; Sui and Hugill, 2002), difference among electoral systems (Dow, 2001) and electoral behaviour of ethnic minorities (Çarkoğlu and Hinich, 2006; Heppen and Mesyanzhinov, 2003).

The analysis of spatial dimension takes into account the effects of interaction among different factors, by considering the role of territorial proximity. Spatial correlation measures are thus employed in order to discriminate different typologies of spatial spread, focusing on the pattern-space propagation of the ethno-regionalist

* Dipartimento di Sociologia e Metodi delle Scienze Sociali, Facoltà di Scienze Politiche, Università di Catania, Italy, rodagata@unict.it.

** Dipartimento di Sociologia e Metodi delle Scienze Sociali, Facoltà di Scienze Politiche, Università di Catania, Italy, tomavene@unict.it.

parties. This has enabled delineation of the geo-electoral spread effects of a party (Shin and Agnew, 2002).

In order to investigate the role of spatial proximity in influencing electoral spread, we tested the presence of spatial effects, observing small units such as municipalities. We began by employing global and local spatial autocorrelation indices. Finally, specifying spatial regression models, we tried to explain how the electoral party performance observed in a municipality influences the electoral outcome of neighbouring ones. After having estimated classic linear models, by means of spatial diagnostic tests, we checked for spatial effects and having rejected the null hypothesis (absence of spatial effects), we specified spatial models controlling for spatial dependence.

2 Measures and models for spatial analysis

The standard models of linear regression (*OLS*) aim to highlight a linear relationship between a dependent variable (y) and a set of explanatory variables (x) assuming the independence of observations. In territorial analysis, nevertheless, the assumption of independent observation in most cases is violated. It is very likely that a value observed in a territorial unit is influenced by the values of neighbours. In this case, *OLS* estimates are inefficient or biased and inconsistent.

The dependence of territorial observations produces two shapes of spatial dependence. In the first form the spatial autocorrelation arises from measurement errors due to the observation of a phenomenon on spatial contiguous units. In the electoral analysis of municipalities, for example, it may happen that there is not an exact correspondence between administrative boundaries and the territorial dimensions over which the electoral strength of the party spreads. The spatial effects, hence, if not prevented, affect the reliability of the model, providing biased estimates (Anselin and Tam Cho, 2002).

In order to control this typology of spatial effects, it is possible to specify a model with spatially lagged errors (Anselin *et al.*, 2004). In this way we obtained a best fitted model and, on the other hand, we located clusters of homogeneous units in relationship with some territorial features not strictly linked with administrative boundaries. If the errors are correlated, therefore, we estimate a *Spatial Error Model*, so we check the dependence among the i -territorial units as regard the x_i -independent variables. In standard notation:

$$Y = X\beta + u \quad (1)$$

$$\text{where:} \quad u = \lambda Wu + \varepsilon \quad (2)$$

The λ -parameter in the (2) is the spatial autoregressive coefficient that picks up spatial correlation of errors. If λ is non-zero and statistically significant, it means that we are in the presence of spatial effects acting on contiguous territorial units. We could consider these effects as *spin-off* effects due, for example, to the socio-economic structure of the territorial area.

The second shape of spatial dependence is, perhaps, more interesting for the electoral studies. In this case, the y -dependent variable is affected by the values of the dependent variable in neighbouring places. The percentage of votes obtained by the party A in a municipality, for example, is related to the percentage of votes collected by

the same party in a nearby municipality. Some scholars define it as *spillover effect* (Cracolici *et al.*, 2009). Also in this case, we cannot assume independent observations and uncorrelated errors. To avoid this problem, in analogy to time series models, a lagged dependent variable (WY) is introduced in the model. This is known as the *Spatial Lag Model*. In standard notation

$$Y = \rho WY + \beta X + u \quad (3)$$

where:

ρWY measures the *spillover* effect among the i -units arising from the correlation with y -dependent variable.

In electoral data analysis, *Spatial Lag Model* could allow to measure the spread power of a party. This model could be useful to check and evaluate territorial expansion of a party controlling for some covariates related to a specific electoral trend from a theoretic point of view.

In an explorative step of our analysis, we measure the effects of spatial interdependence, computing Moran's Index (Moran, 1948):

$$I = \frac{N \sum_i \sum_j W_{ij} (X_i - \bar{X})(X_j - \bar{X})}{(\sum_i \sum_j W_{ij}) \sum_i (X_i - \bar{X})^2} \quad (4)$$

where:

- X_i is the variable observed in i -th municipality
- $\bar{X}$ is the average calculated on all of units
- N is the number of units
- W_{ij} are the weights of first order contiguity matrix W

$$W = \begin{bmatrix} 0 & w_{12} & \dots & w_{1N} \\ w_{21} & 0 & \dots & w_{2N} \\ \dots & \dots & \dots & \dots \\ w_{N1} & w_{N2} & \dots & 0 \end{bmatrix}$$

where:

- w_{ij} is equal to 1 if the territorial i -unit boundaries with the j -territorial unit and it equals 0 otherwise.

In order to examine the units with spatial autocorrelation among adjacent areas, we use local indicators of spatial interdependence, including Local Indicator of Spatial Association (*LISA*):

$$I_i = \frac{Z_i}{\sum_i Z_i^2 / N} \sum_{j \in J_i} w_{ij} Z_j \quad (5)$$

where:

- Z_i is the normalized value given by the variable within the i -th spatial unit
- J_i is the set of spatial units contiguous with i -th unit
- w_{ij} are the weights of the contiguity matrix.

This measure is proportional to a similar global measure and is very useful in order to disaggregate the global autocorrelation. Anselin defines *LISA* measure able to satisfy two conditions: a) for each observation, it calculates the value of spatial aggregation of similar values around the observation space; and b) the sum of these measures for all the observations is proportional to a global spatial association indicator (1995, 94).

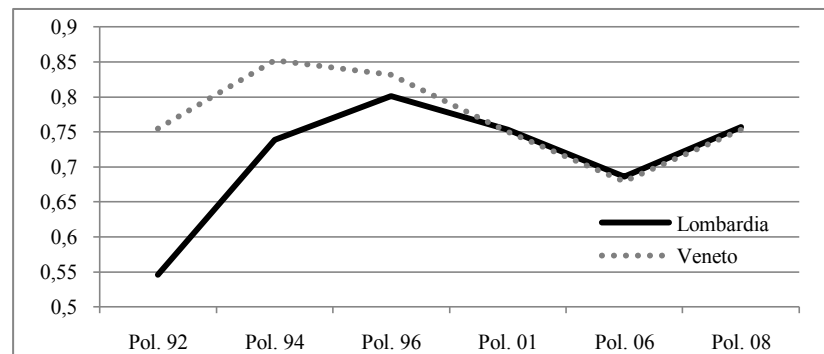
Territorial units with LISA positive values show a homogeneous spatial cluster. This provides information to add to Moran's I . The latter, in fact, represents the values observed of the random variable and the weighted average of the values of each set of units as standardized values. Thus, in addition to identifying homogeneous local areas, the relationship between the LISA values and global spatial autocorrelation measures provides additional benefits to analyse these global measures.

3 Results

The application focused on the electoral performance of the most important territorial party in Italy: *Lega*. In detail, we considered the percentage of the votes received by *Lega* in all the municipalities of Lombardia and Veneto, in the last 6 national elections (1992, 1994, 1996, 2001, 2006, 2008).

Firstly we employed Moran's I in order to analyse the global clustering and capture prospective neighboring effects. Figure 1 shows the trend of spatial autocorrelation coefficient in Lombardia and Veneto. Comparing the two regions, we note how the Moran's I differs in the national election of 1992, being higher in Veneto (0.759) than in Lombardia (0.546). This emphasizes the importance of territory in influencing *Lega* electoral performance in Veneto that appears as a 'cradle' of this party's votes (Shin and Agnew, 2002). Observing all the elections, moreover, we note effects such as a convergence of the two coefficients towards an index value of 0.75 found in 2008 national election.

Figure 1: National Election 1992 - 2008. Moran's I calculated in Lombardia and Veneto.



SOURCE: Ministry of the Interior.

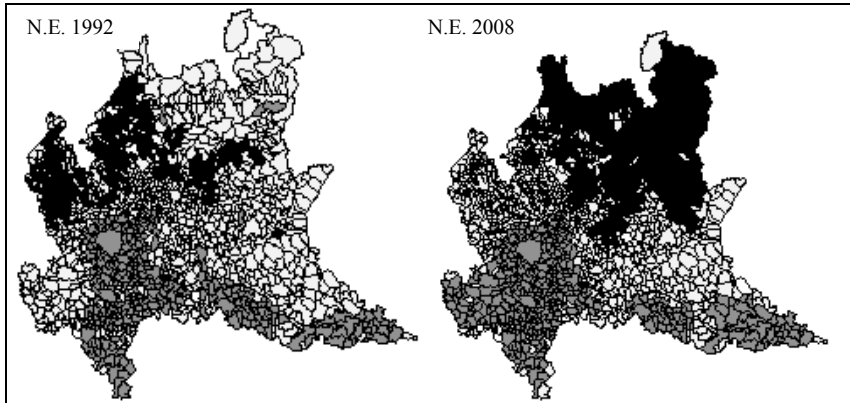
After having analysed the global spatial autocorrelation in the two regions, we tried to spot the clusters of municipalities with an high spatial interdependence. As we argued previously, Local Indicator of Spatial Association (*LISA*) allowed us to assign a value of spatial association for each municipality in relationship with its neighbourhood.

By means of *LISA* cluster map we located geographically the municipalities with highest spatial association measures. In order to simplify the map reading (Figure 2 and 3), we plotted just the statistically significant values ($p < 0.05$) distinguishing between high-high correlation (in black) and low-low correlation (in gray). We then

mapped clusters of adjacent municipalities sharing an increasing (decreasing) value of *Lega* votes.

Observing *LISA* cluster maps of Lombardia (Fig. 2), we note how the localisation of *Lega* electoral performance, from 1992 to 2008, moved easterly. In other word, while in the 1992 national election, *Lega* localised its votes chiefly in the west municipalities, in the last election we observed an increase of *Lega* electoral success in those municipality of the North-East, at the border with Veneto. Moreover, all the grey areas of the map, in both observed elections, are localised in the South of the Region.

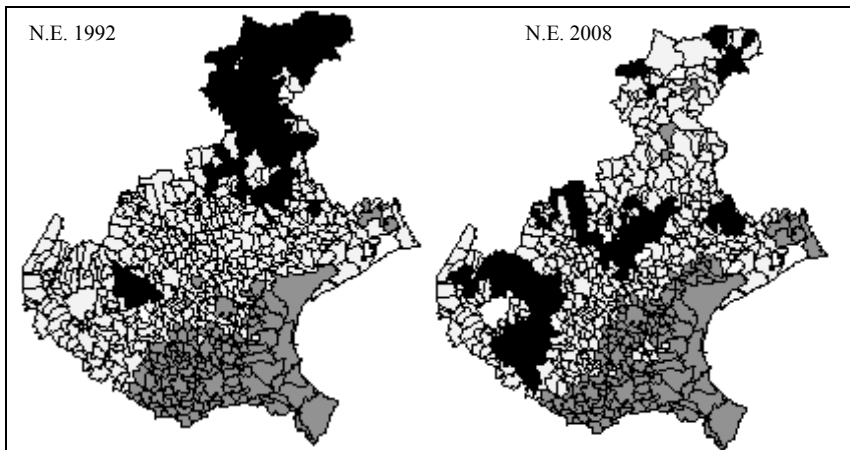
Figure 2: National Elections 1992 and 2008. *LISA cluster Map* - Lombardia.



From a geo-electoral point of view, Lombardia, appears split in two areas. Territorial spread of *Lega* votes seems to hallmark the North municipalities, whereas in the South, local indicators of spatial association suggest the presence of clusters with internal spatial interdependency concerning the decrease of *Lega* votes.

Looking at the Veneto *LISA* cluster maps some mirror-like remarks can be made (Fig. 3).

Figure 3: National Elections 1992 and 2008. *LISA cluster Map* - Veneto.



In this case, we registered a shifting of *Lega* electoral localisation from the North-East of the Region to the South-West. Also in this case, we note as the municipalities with a low-low correlation are localised in the South of the Region, both in the national election of 1992 and 2008. Finally, we tried to investigate spatial effects of territorial structure in influencing *Lega* electoral performance. So, according to the theoretical literature (Bonomi, 1997; Diamanti, 1996), we choose 8 variable related to some economical¹ and socio-demographic² dimensions employed in territorial electoral studies. Data were collected from 1991 Census and 2001 Census in order to estimate two different regression models according two national elections, 1992 and 2008.

Trying to catch territorial effects in influencing *Lega* electoral performance, we specified two spatial models, comparing the classic model (OLS) and the spatial model (Tab. 1).

Table 1: National elections 1992 and 2008. *Classic and spatial models* - Lombardia

Lombardia	N. E. 1992 - Dip: <i>LEGA</i> 92				N. E. 2008 - Dip: <i>LEGA</i> 08			
	Classic Model		Spatial <i>Error</i>		Classic Model		Spatial <i>Lag</i>	
Adjusted R^2	0.252				0.214			
Log likelihood <i>LIK</i>	-4761.71		-4570.18		-5383.66		-4712.92	
<i>AIC</i>	9541.43		9158.36		10785.3		9445.84	
Schwarz Criterion <i>SC</i>	9589.52		9206.45		10833.4		9499.27	
Variable	β	<i>p-value</i>	β	<i>p-value</i>	β	<i>p-value</i>	β	<i>p-value</i>
<i>W_LEGA</i>							0.764	0.000
CONSTANT	32.268	0.000	26.132	0.000	22.711	0.000	4.741	0.000
AGRICULTURE	-0.214	0.000	-0.125	0.012	-0.218	0.000	-0.029	0.250
BUILDING	0.069	0.001	0.035	0.125	0.286	0.000	0.117	0.000
TRADE	-0.132	0.002	0.016	0.673	-0.190	0.000	-0.048	0.006
BROKING	-1.026	0.000	-0.265	0.017	-0.321	0.029	-0.205	0.017
UNEMPLOYM.	-0.082	0.150	-0.099	0.067	-0.616	0.000	-0.268	0.006
LESS_5	0.301	0.036	0.298	0.021	1.208	0.000	0.394	0.000
NOQUALIF	-0.438	0.000	-0.162	0.000	0.112	0.057	0.085	0.014
FOREIGNERS	1.084	0.000	0.672	0.000	-0.172	0.105	-0.099	0.114
<i>LAMBDA</i>			0.669	0.000				
<i>MCN</i>	25.133				17.806			
Moran's I	0.354		0.000		0.624		0.000	
Robust LM <i>Lag</i>	43.851		0.000		179.850		0.000	
Robust LM <i>Error</i>	97.834		0.000		136.575		0.000	

Firstly, we tried to detect the presence of multicollinearity by means of Multicollinearity Condition Number (*MCN*) (Jobson, 1992). Since in all the models the value is less than 30, we supposed the models do not affected by serious multicollinearity problems (Anselin, 2005). Test significance being equal (*p-value*=0.000), since robust LM *Error* provided an higher value (97.834) than LM *Lag* (43.851), we estimated a spatial error model introducing so, spatial dependence in the error term. Comparing the measure of fit (*LIK*, *AIC*, *SC*) between the two model, we notice an improvement of fit for the spatial error model, strengthened by *lambda* value (0.669) highly significant. From a substantial point of view, we observe spatial effects due to the shared socio-economic structure (such a '*spin-off*' effect) and not yet linked

¹ Rate of workers in agriculture sector (AGRICULTURE), building sector (BUILDING), trade sector (TRADE) and broking sector (BROKING).

² Unemployment rate (UNEMPLOYMENT), rate of residents less than five years (LESS_5), rate of residents without school diplomas (NOQUALIF), rate of foreign residents (FOREIGNERS).

to the electoral ability of *Lega*. By contrast, the diagnostic of the classic model for the 2008 suggested to specify a spatial *lag* model. The Moran's I test (0.624) highly significant indicates the presence of spatial effects and the large value of LM specific test (179.850) highly significant, moreover, suggests that these effects could be due to the dependent variable performing in adjacent municipalities. Also in this case, introducing spatial effects entails an improvement of model fit as remarked by the value of spatial autoregressive coefficient (0.764). We note an increasing of Log likelihood (from - 5383.66 to -4712.92) and, at the same time, a decreasing of *AIC* and *SC*. Specifying a spatial lag highlighted how the magnitude of all the coefficients estimated with classic model were affected, all of them show a decrease in absolute value and in significance. The large and positive value of *W_Lega*, therefore, indicates that municipalities where *Lega* obtains many (few) votes are clustered and *Lega* performance, observed in a municipality, is function of percentage of votes obtained by *Lega* in neighbouring municipalities.

A similar remark could be done observing models in Veneto (Tab. 2), where, in both national elections, diagnostic tests for spatial effects suggested to specify a spatial lag model.

Table 2: National elections 1992 and 2008. *Classic and spatial models* - Veneto

Veneto	National Election 1992				National Election 2008			
	Classic Model		Spatial Lag		Classic Model		Spatial Lag	
Adjusted R^2	0.346				0.277			
Log likelihood <i>L/K</i>	-1842.37		-1642.04		-1959.11		-1672.58	
<i>AIC</i>	3702.74		3306.07		3936.22		3367.13	
Schwarz criterion <i>SC</i>	3742.03		3354.08		3975.50		3415.17	
Variable	β	<i>p-value</i>	β	<i>p-value</i>	β	<i>p-value</i>	β	<i>p-value</i>
<i>W_LEGA</i>			0.744	0.000			0.822	0.000
CONSTANT	25.970	0.000	4.440	0.001	24.185	0.000	2.333	0.035
AGRICULTURE	-0.234	0.000	-0.09	0.002	0.353	0.111	0.355	0.004
BUILDING	0.374	0.000	0.145	0.000	0.106	0.025	0.075	0.005
TRADE	-0.356	0.000	0.026	0.572	-0.253	0.001	-0.074	0.018
BROKING	-0.412	0.000	-0.153	0.003	-0.553	0.017	-0.522	0.011
UNEMPLOYM.	-0.138	0.018	-0.099	0.009	-0.444	0.006	-0.079	0.381
LESS_5	1.301	0.000	0.428	0.043	1.478	0.000	0.411	0.009
NOQUALIF	-0.731	0.037	-0.148	0.012	0.138	0.214	0.232	0.000
FOREIGNERS	1.308	0.002	0.645	0.023	0.890	0.000	0.074	0.416
<i>MCN</i>	24.082				15.708			
Moran's I	19.693		0.000		22.905		0.000	
Robust LM (<i>Lag</i>)	113.57		0.000		104.611		0.000	
Robust LM (<i>Error</i>)	0.543		0.461		6.381		0.011	

Also in these models, we observed a *MCN* below 30. Controlling for spatial effects, the model fit improves as the proper measures of fit show. The coefficient of spatially lagged dependent variable in both elections captures the spatial effects due to the nearby locations. In 1992 introducing spatial effect entails a decrease of all covariates values in the model. In the case of TRADE, moreover, the value becomes not significant assuming a threshold *p*-value equal to 0.05 and the LESS_5 value decreases but it is always significant. In generally, however, we note a decreasing of significance that drops in so far as the covariate is affected by spatial effects. In 2008, instead, not all the covariates are affected by spatial effects, because AGRICULTURE and NOQUALIF become significant in the lag model and, on the contrary, UNEMPLOYMENT and FOREIGNERS, not significant.

4 Conclusions

The analysis of global spatial autocorrelation, by means of Moran's Index, highlighted an increasing trend of territorial expansion of *Lega* votes inside of both observed regions. Employing *LISA* measures, furthermore, we noted in Lombardia a sort of territorial polarization between North and South municipalities whereas in Veneto a more homogeneous geo-electoral distribution of *Lega* votes appeared. Again, the plot of Local Indicators of Spatial Association shows a geo-electoral convergence towards an area composed by eastern municipalities of Lombardia and western municipalities of Veneto. Finally, in order to analyse the linear relationship between some structural dimensions and the percentage of votes obtained by *Lega*, we specified several spatial models controlling the two different spatial dependence shapes, estimating so a spatial error model and three spatial lag models. As previously described, the lag model, unlike spatial error, captures spatial effects due to the dependent variable. These effects, that usually entail troubles for a correct modelling caused by biased estimates, could prove an useful tool to investigate socio-economic structure of territorial parties.

References

1. Anselin, L.: Local Indicators of Spatial Association—*LISA*. *Geographical Analysis*. 27 (2), 93–115 (1995).
2. Anselin, L., Florax, R., Rey, S. (eds): *Advances in Spatial Econometrics: Methodology, Tools and Applications*. Springer-Verlag, Berlin Heidelberg (2004).
3. Anselin, L., Tam Cho, W.T.: Spatial Effects and Ecological Inference. *Political Analysis*. 10(3), 276–297 (2002).
4. Bonomi, A.: *Il capitalismo molecolare*. Einaudi, Torino (1997).
5. Çarkoğlu, A., Hinich, M.: A spatial analysis of Turkish party preferences. *Electoral Studies*. 25, 369–392.
6. Cracolici, F., Cuffaro, M., Nijkamp, P.: A spatial analysis on Italian unemployment differences. *Statistical Methods and Applications*. 18, 275–291 (2009).
7. Darmofal, D.: The political geography of macro-level turnout in American political development. *Political Geography*. 25, 2, 123–150 (2006).
8. Diamanti I., *Il male del Nord*. Donzelli, Roma (1996).
9. Dow, J.: A comparative spatial analysis of majoritarian and proportional elections. *Electoral Studies*. 20, 109–125 (2001).
10. Heppen, J., Mesyanzhinov, D.: Political geography and regionalism in Louisiana: the impact of the French influence in the post-World War II era. *Political Geography*. 22, 519–533 (2003).
11. Jobson, J.D.: *Applied Multivariate Data Analysis*, Vol. 2. Springer-Verlag, New York (1992).
12. Moran, P.: The interpretation on statistical maps. *Journal of the Royal Statistical Society*, B 10, 243–251 (1948).
13. O'Loughlin, J., Ward, M.D., Lofdahl, C. H., Cohen, J. S., Brown, D. S., Reilly, D., Gleditsch, K. S., Shin, M.: The Diffusion of Democracy 1946–1994. *Annals of the Association of American Geographers*. 88, 4, 545–574 (1998).
14. Shin, M., Agnew J.: The Geography of Party Replacement in Italy, 1987–1996. *Political Geography*. 21, 221–242 (2002).
15. Sui, D., Hugill, P.: A GIS-based spatial analysis on neighborhood effects and voter turn-out: a case study in College Station, Texas. *Political Geography*. 21, 159–173 (2002).