AN ANALYSIS OF METHODS FOR THE TREATMENT OF AUTOCORRELATION IN SPATIAL INTERACTION MODELS

Giuseppe Ricciardo Lamonica

1. Introduction

In regard to spatial regression models where the units subject to analysis are of territorial type, a problem of fundamental importance, often ignored, is that the observations may be dependent which entails that the value of a variable in an elementary territorial unit is determined by those observed in neighbouring localities. When this situation occurs, the phenomenon is spatially autocorrelated (or spatially dependent). The presence of spatial autocorrelation is a problem because the traditional regression models are based on the hypothesis of independence among observations in the different localities of the territory analysed. Hence, when this hypothesis is violated, the estimates of the regression model parameters are bias and/or inefficient. The spatial dependence occurs in two different forms. In the first case, it affects only the error term in the regression model and is mostly considered to be a nuisance which needs to be eliminated. In this case, the spatial error autocorrelation does not cause ordinary least squares (OLS) estimates to be biased, but it alters their efficiency. In the second case, the value of the dependent variable in a generic unit is affected by those of the neighbours units. In this case, OLS estimates are bias and no longer consistent.

The spatial econometrics that deals with treatment of spatial autocorrelation in regression models has developed various techniques to solve this problem. Among them, the most popular approaches used are the Spatial Filtering Model (Fisher and Griffith 2008; Griffith 2009; Chun and Griffith 2011), the Spatial Autoregressive Model and Spatial Error Model (LeSage 1997a; Anselin 1988; LeSage and Pace 2008). While in literature exists a voluminous series of empirical analysis to assess the utility of these methods for cross-section and panel data a scant attention was paid for interactions or flows data. In this last context, the spatial autocorrelation analysis is exacerbated because contrary to classical panel data where the sample involves n spatial unit, with each unit being an observation, the interaction spatial data involve n^2 origin-destination pairs.

Limiting the attention to the interaction data, this paper, by means of a real phenomenon, outline and compare the three methods previously quoted in order to identify their shared characteristics and those specific to each of them. It should be pointed that the literature, excluding same studies, see for example Fisher and Griffith (2008), is not particularly rich in analyses of this type. Consequently, it is interesting furnish a contribution which explores the potentialities of these methods, even if in a limited situation like the one considered by this study.

Moreover, given the particular area of inquiry, the analysis makes no claim to being exhaustive, and the results are to be considered preliminary.

2. The methods for modelling the spatial autocorrelation

In this section, the three techniques to consider the spatial autocorrelation in the interaction models previously quoted, are briefly exposed.

Suppose we have a set of n spatial unit and let **F** an n*x*n square matrix whose generic entry f_{ij} (i,j=1,.., n) is the interaction (flow) whose origin is the i-th unit whose destination is the j-th unit.

We restrict the discussion in the context of gravity models which are considered one of the most important spatial interaction model (Everett and Keller, 2002).

The interaction between two generic units depends by three type of explicative variables (covariates). Those that characterize the origin unit of flow (push), those that characterize the destination unit of flow (pull) and finally the covariates that measure the separation between the origin and destination units. Using matrix notation:

$$\mathbf{f}=\alpha\mathbf{1}+\mathbf{X}_O\otimes\mathbf{1}\boldsymbol{\beta}_O+\mathbf{1}\otimes\mathbf{X}_D\boldsymbol{\beta}_D+\mathbf{d}\gamma+\boldsymbol{\varepsilon} \quad (1)$$

where: **f** is the vector whose elements are the logarithmic of the flows observed; **1** is the n(n-1)*x*1unit vector; $\mathbf{X}_O$ and $\mathbf{X}_D$ are two n(n-1)*x*k matrices containing respectively the logarithmic of the covariates relating to the origin spatial unit and destination spatial unit of the flows; **d** is the vector whose elements are the logarithmic of the distance between two generic spatial unit; **ε** is the vector of the residual variable. Finally, ⊗ is the Kronecker product.

Referring to Fisher and Griffith (2008) and LeSage and Pace (2008) for details, the interaction phenomena may exhibit various types of autocorrelation. Autocorrelation at the origin: given a generic flow f_{ij}, also the units contiguous to the i-th unit have similar flows to the j-th unit. Autocorrelation at the destination: given a generic flow f_{ij}, the i-th unit has flows similar with respect to the units bordering on the j-th unit. Autocorrelation at both the origin and the destination: which simultaneously concerns the two previous formulations.

If the autocorrelation occurs in the dependent variable in order to capture the first type of the autocorrelation, that at the origin, it is possible to consider the following model named Origin Spatial autoregressive model (also O-SAR):

$$\mathbf{f}= \alpha\mathbf{1}+\rho\mathbf{W}_O\mathbf{f}+\mathbf{X}_O\boldsymbol{\beta}_O+\mathbf{X}_D\boldsymbol{\beta}_D+\gamma\mathbf{d}+\boldsymbol{\varepsilon} \qquad (2)$$

Where $\mathbf{W}_O=\mathbf{I}\otimes\mathbf{C}^*$ and $\mathbf{C}^*$ is the $\mathbf{C}$ matrix standardized by row ($\mathbf{C}$ is the n*x*n matrix of the first-order contiguities). In this case, the vector $\mathbf{W}_O\mathbf{f}$ is such that the generic element is the mean flow from the units contiguous to i-th unit to the j-th unit, and it represents the interaction that would be observed between the two units on the hypothesis of an autocorrelation at the origin.

As regards the second type of autocorrelation, it is sufficient considers in the (2) the following $\mathbf{W}_D$ matrix instead of $\mathbf{W}_O$:

$$\mathbf{W}_D=\mathbf{C}^*\otimes\mathbf{I} \qquad (3$$

In this case, the generic element of the vector $\mathbf{W}_D\mathbf{f}$ is a mean of the flows from the i-th unit to the units contiguous with the j-th element. The corresponding spatial model (4) that reflects the destination-based dependence is named Destination-SAR (also: D-SAR). Finally, by multiplying $\mathbf{W}_O$ and $\mathbf{W}_D$ (i.e. $\mathbf{W}_{OD}=\mathbf{W}_O\mathbf{W}_D$), it is possible to consider, the third type of autocorrelation. In this last case, the spatial model is the Origin-Destination-SAR (also: OD-SAR).

On the other hand, if the dependence occurs in the error term of the (2), depending on the type of dependence, one of the following model can be utilized:

$$\mathbf{f}= \alpha\mathbf{1}+ \mathbf{X}_O\boldsymbol{\beta}_O+\mathbf{X}_D\boldsymbol{\beta}_D+\gamma\mathbf{d}+\boldsymbol{\varepsilon} \quad \boldsymbol{\varepsilon}=\lambda\mathbf{W}_j\,\boldsymbol{\varepsilon}+\mathbf{u} \text{ and } \mathbf{u}\approx N(0;\sigma^2 I) \qquad (4)$$

If j=O then the model is the Origin Spatial Error Model (O-SEM). For j=D we have the Destination Spatial Error Model (D-SEM). Finally, if j=OD we have the Origin-Destination Spatial Error Model (OD-SEM).

The previous three types of autocorrelation can be jointly considered in a regression model. Restricting, but not limiting the attention to SAR models we can take account the following specifications:

- $\mathbf{f}=\rho_O\mathbf{W}_O\mathbf{f}+\rho_D\mathbf{W}_D\mathbf{f}+\rho_{OD}\mathbf{W}_{OD}\mathbf{f}+\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon}$;
- $\mathbf{f}= \rho_O\mathbf{W}_O\mathbf{f}+\rho_D\mathbf{W}_D\mathbf{f}+\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon}$;
- $\mathbf{f}= \rho(\mathbf{W}_O+\mathbf{W}_D)\mathbf{f}+\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon}$;
- $\mathbf{f}=\rho(\mathbf{W}_O+\mathbf{W}_D+\mathbf{W}_{OD})\mathbf{f} +\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon}$;

An alternative methodology is the spatial filtering method (FS). See for example Griffith (2003) for details. It presumes that the spatial dependence in the dependent variable is due to one or more spatially autocorrelated not directly observable variables. As surrogates for the latter the method considers the Moran's I autocorrelation index:

$$I=\frac{n(n-1)}{\mathbf{1}'\mathbf{W}_j\mathbf{1}}\frac{\mathbf{f}'\left(\mathbf{I}-\mathbf{1}\mathbf{1}'\frac{1}{n(n-1)}\right)\mathbf{W}_j\left(\mathbf{I}-\mathbf{1}\mathbf{1}'\frac{1}{n(n-1)}\right)\mathbf{f}}{\mathbf{f}'\left(\mathbf{I}-\mathbf{1}\mathbf{1}'\frac{1}{n(n-1)}\right)\left(\mathbf{I}-\mathbf{1}\mathbf{1}'\frac{1}{n(n-1)}\right)\mathbf{f}} \tag{5}$$

Referring to De Jong et al. (1984) for details, if e^{min} and e^{max} respectively denote the highest and smallest eigenvalues of $\mathbf{W}_j$, the following inequality holds:

$$e^{min}\leq I \Big/ \frac{n(n-1)}{\mathbf{1}'\mathbf{W}_j\mathbf{1}} \leq e^{max} \tag{6}$$

Moreover, these extreme eigenvalues also coincide with those of the following matrix:

$$\left(\mathbf{I}-\mathbf{1}\mathbf{1}'\frac{1}{n(n-1)}\right)\mathbf{W}_j\left(\mathbf{I}-\mathbf{1}\mathbf{1}'\frac{1}{n(n-1)}\right) \tag{7}$$

As Griffith (2009), reported the eigenvector associated with the largest eigenvalue of (7) is the one possessing the highest autocorrelation, and it is orthogonal to the other eigenvectors. Likewise, the eigenvector associated with the second largest eigenvalue is the one possessing the second highest autocorrelation. The remaining eigenvectors can be considered in similar fashion.

The set of all the eigenvectors of (7) can be regarded as distinct and uncorrelated spatial maps, each exhibiting a certain degree of autocorrelation coincident with the corresponding eigenvalue. These spatial configurations are therefore likened to proxy variables depicting all the possible forms of autocorrelation which, starting from matrix $\mathbf{W}_j$, are latent in the variable subject to analysis. The spatial filtering method uses these artificial indicators as surrogates for the unobservable variables that cause the spatial autocorrelation. For empirical purposes, it is not reasonable to add the full set of eigenvectors as spatial proxy variables to the model (1), but only the dominant eigenvectors. The latter may be chosen in various way. For example in a stepwise procedure regressing the dependent variable **f** on the set of the eigenvectors and using the conventional R^2 maximization criterion. As a consequence, the regression model with spatial filtering can be formalized in the following way, where **A** is the matrix of the dominant eigenvectors of (7):

$$\mathbf{f}=\alpha\mathbf{1}+\mathbf{X}_O\otimes\mathbf{1}\boldsymbol{\beta}_O+\mathbf{1}\otimes\mathbf{X}_D\boldsymbol{\beta}_D+\mathbf{d}\gamma+\mathbf{A}\boldsymbol{\rho}+\boldsymbol{\varepsilon} \tag{8}$$

Whereas the regression model with spatial filtering does not require particular techniques except for the usual least squares method, in the cases of the SAR and SEM it is necessary to use iterative procedures such as that developed by LeSage (1997).

3. Data and experimentation results

To compare the methods previously exposed, the migration flows for change of residence between the twenty Italian regions collected by the ISTAT for the year 2006 were used and analysed by a gravity model.

In according with the literature, the version applied this paper considers not only the classic determinants of migration (i.e., the size of populations and the distance between places), but also the effects attributable to a set of push and pull variables explaining the economic and demographic differences between the Italian regions. To this regard, we consider 17 indicators (see Appendix). As a preliminary examination of the indexes showed the presence of correlations that rendered them unsuitable for use in a regression model, they have been synthesized by means of factor analysis

The results of this analysis are set out in extreme synthesis in Appendix. The high and positive coefficients of correlation between the first factor and all the variables of economic nature suggest identification of this factor as a complex index of the economic structure, while the close correlations of the second factor with the remaining indexes suggest its identification as a complex index of the demographic structure. Them, the following 18 models were considered:

Table 1 – *Models considered in the analysis.*

SAR1: $\mathbf{f}=\mathbf{X\beta}+\rho\mathbf{W}_O\mathbf{f}+\boldsymbol{\varepsilon}$	SAR2: $\mathbf{f}=\mathbf{X\beta}+\rho\mathbf{W}_D\mathbf{f}+\boldsymbol{\varepsilon}$	SAR3: $\mathbf{f}=\mathbf{X\beta}+\rho\mathbf{W}_{OD}\mathbf{f}+\boldsymbol{\varepsilon}$
SAR4: $\mathbf{f}=\mathbf{X\beta}+\rho(\mathbf{W}_O+\mathbf{W}_D)\mathbf{f}+\boldsymbol{\varepsilon}$	SAR5: $\mathbf{f}=\mathbf{X\beta}+\rho(\mathbf{W}_O+\mathbf{W}_D+\mathbf{W}_{OD})\mathbf{f}+\boldsymbol{\varepsilon}$	SAR6: $\mathbf{f}=\mathbf{X\beta}+(\rho_O\mathbf{W}_O+\rho_D\mathbf{W}_D)\mathbf{f}+\boldsymbol{\varepsilon}$
SAR7: $\mathbf{f}=\mathbf{X\beta}+(\rho_O\mathbf{W}_O+\rho_D\mathbf{W}_D+\rho_{OD}\mathbf{W}_{OD})\mathbf{f}+\boldsymbol{\varepsilon}$	SAR8: $\mathbf{f}=\mathbf{X\beta}+(\rho_O\mathbf{W}_O+\rho_D\mathbf{W}_D+\rho_O\rho_D\mathbf{W}_{OD})\mathbf{f}+\boldsymbol{\varepsilon}$	
SEM1: $\mathbf{f}=\mathbf{X\beta}+\boldsymbol{\varepsilon}$ and $\boldsymbol{\varepsilon}=\lambda\mathbf{W}_O\boldsymbol{\varepsilon}+\mathbf{u}$	SEM2: $\mathbf{f}=\mathbf{X\beta}+\boldsymbol{\varepsilon}$ and $\boldsymbol{\varepsilon}=\lambda\mathbf{W}_D\boldsymbol{\varepsilon}+\mathbf{u}$	SEM3: $\mathbf{f}=\mathbf{X\beta}+\boldsymbol{\varepsilon}$ and $\boldsymbol{\varepsilon}=\lambda\mathbf{W}_{OD}\boldsymbol{\varepsilon}+\mathbf{u}$
SEM4: $\mathbf{f}=\mathbf{X\beta}+\boldsymbol{\varepsilon}\ \boldsymbol{\varepsilon}=\lambda(\mathbf{W}_O+\mathbf{W}_D)\boldsymbol{\varepsilon}+\mathbf{u}$	SEM5: $\mathbf{f}=\mathbf{X\beta}+\boldsymbol{\varepsilon}\ \ \boldsymbol{\varepsilon}=\lambda(\mathbf{W}_O+\mathbf{W}_D+\mathbf{W}_{OD})\boldsymbol{\varepsilon}+\mathbf{u}$	
SF1: $\mathbf{f}=\mathbf{X\beta}+\mathbf{A}_1\boldsymbol{\rho}+\boldsymbol{\varepsilon}$	SF2: $\mathbf{f}=\mathbf{X\beta}+\mathbf{A}_2\boldsymbol{\rho}+\boldsymbol{\varepsilon}$	SF3: $\mathbf{f}=\mathbf{X\beta}+\mathbf{A}_3\boldsymbol{\rho}+\boldsymbol{\varepsilon}$
SF4: $\mathbf{f}=\mathbf{X\beta}+\mathbf{A}_4\boldsymbol{\rho}+\boldsymbol{\varepsilon}$	SF5: $\mathbf{f}=\mathbf{X\beta}+\mathbf{A}_5\boldsymbol{\rho}+\boldsymbol{\varepsilon}$	

where, $\mathbf{X\beta}=\beta_0+\beta_1\mathbf{p}_i+\beta_2\mathbf{p}_j+\beta_3\mathbf{d}+\beta_4\mathbf{f1}_i+\beta_5\mathbf{f1}_j+\beta_6\mathbf{f2}_i+\beta_7\mathbf{f2}_j$, and: $\mathbf{f}$ is the vector that contain the annual flows between two generic regions; $\mathbf{p}_i$ and $\mathbf{p}_j$ the vectors that contains the logarithmic amounts of the populations resident respectively in the origin and destination of the flows; $\mathbf{d}$ the vector that contain the logarithmic geographical distances between two regions; $\mathbf{f1}_i$, $\mathbf{f1}_j$ and $\mathbf{f2}_i$, $\mathbf{f2}_j$ the vectors that contains the values of first two factors extracted by the factor analysis in the regions of the origin and destination of the flows; $\mathbf{A}_j$ (for j=1,..,5) are the matrices of the predominant eigenvectors respectively of the $\mathbf{W}_O$, $\mathbf{W}_D$, $\mathbf{W}_{OD}$, $(\mathbf{W}_O+\mathbf{W}_D)$ and

($\mathbf{W}_O$+$\mathbf{W}_D$+$\mathbf{W}_{OD}$) matrices. In this enquiry, the predominant eigenvectors were chosen in a stepwise procedure, regressing, the observed logarithmic of flows on the set of the eigenvectors of the previously **W** matrices, using the conventional R^2 as the maximisation criterion. Finally, $\boldsymbol{\varepsilon}$ and **u** two residual vectors.

It should be noted that, given the nature of the dependent variable (count data), the linear regression model was chosen instead of the Poisson regression model, because: the data showed the presence of the well-known problem of overdispersion. the investigation was intended to be explanatory, not predictive, the mean flow take a value such that as Baltagi (2011) and Lejenne (2010) reported, the Poisson variable can be well approximated by the normal variable. Finally, the use of this model is in full accordance with the literature. See: Black (1992), Egger (2005), Lewer and Van den Berg (2008), Griffith (2009), Kim and Koen (2010), Mayda (2010), Ludo (2012) and LeSage and Agnan (2015).

A preliminary analysis was conducted with the goal to verifying the presence of spatial correlation in the data. To this regard in the following Table 2 are set out the results of the non-spatial interaction model (i.e. $\mathbf{f}=\mathbf{X}\boldsymbol{\beta}$).

Table 2 – *Results of classical log gravity model.*

Parameters	Estimates	p-value
Intercept	-21.06	<.001
$\mathbf{p}_i$	0.96	<.001
$\mathbf{p}_j$	0.98	<.001
d	-0.41	<.001
$\mathbf{f1}_i$	-0.11	<.001
$\mathbf{f1}_j$	0.15	<.001
$\mathbf{f2}_i$	-0.09	0.013
$\mathbf{f2}_j$	0.003	0.934
Statistics	Value	
F-Fisher of the model	292.34	<.001
R^2	0.84	
Breush-Pagan test of heteroscedasticity	17.69	0.01
Kolmogorov test of normality	0.046	0.05
Statistics	Value	p-value
Moran (using $\mathbf{W}_O$)	9.54	<.001
Moran (using $\mathbf{W}_D$)	9.63	<.001
Moran (using $\mathbf{W}_{OD}$)	10.02	<.001

As is possible to shows the estimates of the constant and the parameters associated with the population size of the regions, as well as the parameter relative to the distance, were highly significant. The estimates of the parameters associated with the economic factor in the regions of origin and in the destination regions of flows ($\mathbf{f1}_i$ and $\mathbf{f1}_j$ respectively) were also significant. When we looked at the demographic factor of the places of origin (**f2**$_i$) and of the places of destination (**f2**$_j$), the estimates of the associated parameters were found to have been non-significant especially for the **f2**$_j$ factor. Finally, the tests of Breush-Pagan, Kolmogorov-Smirnov, and the Moran Index showed that the regression residuals were respectively: homoscedastic, normally distributed and spatially correlated at the origin and the destination of flows. Consequently, the previously spatial interaction models were considered and estimated. In Appendix are set out the results of this analysis.

First to be noted a particular coincidence in the estimation of the parameters of the various models. However, the estimates are more stable in the various spatial filtering and spatial error models respect to the spatial autoregressive models. Relating the economic factor, to be noted that this always has a negative effect in the origin regions of the flows (**f1**$_i$). It should be noted that this factor is always significant in the spatial filtering models and in some spatial autoregressive models, while it is never significant in the spatial error models. In the destination regions (**f1**$_j$), this factor is significant only in the spatial filtering models.

Similar considerations apply to the demographic factor, which in the spatial filtering models, always has a negative estimate in the origin regions of flows (**f2**$_i$) and a positive estimate in the destination ones (**f2**$_j$). On the contrary, in the remaining models, **f2**$_j$ has a negative estimate. However, the effect of the demographic factor is not uniform in the various models. This complex variable is never significant in the destination regions, whilst in the origin regions of the flows, it is generally significant.

To recapitulate, it emerges quite clearly that the three approaches considered and the various types of autocorrelation, do not exhibit either among themselves, substantial differences in the estimates and signs of the parameters. All the models verifies the hypothesis of normal distribution, the SAR models and SEM, generates heteroscedastic residuals. All the models considered are able to capture the effects of the spatial autocorrelation present but the Moran test of autocorrelation shows that the residuals of some SAR models (SAR1, SAR2, SAR4 and SAR5) are still correlated, while those of the SEM and the SF models are still uncorrelated.

The Root Mean Square Error (RMSE) has values between 0.38 (SEM 3) and 0.63 (SAR 7) and the Spatial Filtering models are those with the lowest value in the index considered. They are followed by the Spatial Error Models (SEM) and then by the SAR models. Moreover, the SAR models that involves separately more than one weight matrix (SAR6, SAR7 and SAR8) have a more goodness of fit respect the models that involves a single weight matrix (SAR1, SAR2 and SAR3) and respect those involving cumulative weight matrices (SAR4 and SAR5). On the contrary, the SEM do not shows substantial differences between the various alternatives considered. Finally, the spatial filtering models that involves more than one weight matrix (SF4 and SF5) have a more goodness of feet. Particularly, the SF4 is the model with the highest goodness of fit.

4. Conclusions

Spatial interaction models of the gravity type are widely used to model origin destination flows. This, model is misspecified if the residuals are spatially correlated. This problem arises when the observations are dependent over the space. To solve the problem same methods were proposed in literature. The most popular approaches are the Spatial Filtering Model, the Spatial Autoregressive Model and Spatial Error Model. Using real data based on entries and cancellation for change of residence between the twenty Italian regions, corresponding to the second level of the Nomenclature of Territorial Units for Statistics (NUTS 2) this enquiry has illustrated and compared the previously three methods. As is obvious the results of the inquiry lay no claim to exhaustiveness indeed, they are to be regarded as only preliminary but they seem very interesting.

The most important results obtained highlights that the three approach do not exhibit substantial differences in the estimation of the parameters and in the goodness of fit with the reality.

Consequently, from this point of view the choice of one or other method is indifferent. On the contrary, analysing the regression residuals, substantial differences are clear. If on the one hand the three approaches generates normally distributed and homoscedastic residuals on the other hand, the residuals of the SAR, in several situations are heteroscedastic and still correlated. The analysis was repeated also for the years 1995 to 2005. The results perfectly coincides with those of the present study.

Appendix

Table 3 – *Factorial analysis results*.

	Factor 1	Factor 2
Variance explained	0.60	0.19
Correlations between the variables and the first two factors.		
Variables	Factor 1	Factor 2
Employment rate	0.946	0.002
Per capita added value	0.420	0.409
Per person employed added value	0.965	-0.174
GDP per capita	0.895	-0.366
GDP per person employed	0.965	-0.198
% of employed in industry	0.862	-0.416
% of employed in agriculture	-0.836	0.055
% of employed in other activities	0.816	-0.240
Consumption per capita	-0.853	0.215
Income per capita	0.674	-0.607
Units of labour per inhabitant	0.935	0.014
Size of unit of labour	0.275	0.793
Age dependency ratio	-0.930	-0.186
Index of turnover in the active population	0.589	0.707
% of persons aged 65 and over	0.514	0.796
Old-age dependency ratio	-0.010	-0.465
Index of active population structure	0.854	0.340

Table 4 – *Results of the SAR Models*.

	SAR 1: $\rho\mathbf{W}_O$	SAR2: $\rho\mathbf{W}_D$	SAR3: $\rho\mathbf{W}_{OD}$	SAR4: $\rho(\mathbf{W}_O+\mathbf{W}_D)$	SAR5: $\rho(\mathbf{W}_O+\mathbf{W}_D+\mathbf{W}_{OD})$	SAR6: $\rho_O\mathbf{W}_O+\rho_D\mathbf{W}_D$
Var.			Estimates			
Int.	-20.52*	-20.39*	-23.88*	-19.38*	-24.16*	-20.80*
$\mathbf{p}_i$	0.67*	1.08*	1.02*	0.80*	0.98*	0.97*
$\mathbf{p}_j$	1.08*	0.65*	1.03*	0.80*	0.99*	0.93*
$\mathbf{d}$	-0.36*	-0.36*	-0.39*	-0.36*	-0.36*	-0.46*
$\mathbf{f1}_i$	-0.06	-0.13*	-0.10*	-0.04	-0.08*	-0.11*
$\mathbf{f1}_j$	0.06	0.11*	0.11*	0.09*	0.08	0.14*
$\mathbf{f2}_i$	-0.07	-0.08*	-0.08	-0.12*	-0.10*	-0.09*
$\mathbf{f2}_j$	-0.01	-0.02	-0.01	-0.06	-0.03	-0.01
ρ	0.31*	0.35*	0.18*	0.47*	0.38*	
ρ_O						0.06
ρ_D						0.01
Statistics				Value		
B-P	19.39	22.09	19.21	21.63*	20.83*	5.88
K-N	0.041	0.04	0.04	0.04	0.03	0.04
log lik	-341.18	-337.30	-360.44	-333.22	-348.65	-942.97
RMSE	0.58	0.58	0.62	0.57	0.60	0.62
Moran	3.43*	3.14*	6.28*	8.26*	0.17*	0.83

*Legend: B-P= Breush-Pagan test of heteroscedasticity; K-N= Kolmogorov test of normality; log-lik=Log-likelihood; Moran=Moran test of autocorrelation on residuals; RMSE=Root Mean Square Errors; AIC=Akaike Information Criterium; * p-value<0.01.*

Table 5 – *Results of the SEM. Year 2006.*

Var.	SEM1: $\lambda\mathbf{W}_O$	SEM2: $\lambda\mathbf{W}_D$	SEM3: $\lambda\mathbf{W}_{OD}$	SEM4: $\lambda\,(\mathbf{W}_O+\mathbf{W}_D)$	SEM5: $\lambda\,(\mathbf{W}_O+\mathbf{W}_D+\mathbf{W}_{OD})$
			Estimates		
Int.	-20.34*	-20.14*	-20.08*	-19.50*	-19.79*
$\mathbf{p}_i$	0.96*	0.94*	0.94*	0.96*	0.95*
$\mathbf{p}_j$	0.95*	0.96*	0.94*	0.94*	0.94*
d	-0.65*	-0.70*	-0.60*	-1.01*	-0.73*
$\mathbf{f1}_i$	-0.04	-0.06	0.004	-0.01	0.01
$\mathbf{f1}_j$	0.12*	0.22*	0.17*	0.17	0.15
$\mathbf{f2}_i$	-0.11	-0.11*	-0.11*	-0.11	-0.11*
$\mathbf{f2}_j$	-0.05	-0.03	-0.06	-0.05	-0.07
λ	0.53*	0.59*	0.64*	0.91*	0.87*
Statistics			Value		
B-P	22.27*	24.43*	23.25*	34.39*	23.09*
K-N	0.06	0.04	0.04	0.04	0.05
log lik	-318.84	-311.23	-335.04	-255.81	-294.15
RMSE	0.53	0.52	0.57	0.43	0.51
Moran	-0.69	-0.56	-3.42*	0.01	-0.003

*Legend: B-P= Breush-Pagan test of heteroscedasticity; K-N= Kolmogorov test of normality; log-lik=Log-likelihood; Moran=Moran test of autocorrelation on residuals; RMSE=Root Mean Square Errors; AIC=Akaike Information Criterium; * p-value<0.01.*

Table 6 – *Results of the Spatial Filtering Models. Year 2006*

Var.	SF1: $\mathbf{W}_O$	SF2: $\mathbf{W}_D$	SF3: $\mathbf{W}_{OD}$	SF4: $\mathbf{W}_O+\mathbf{W}_D$	SF5: $\mathbf{W}_O+\mathbf{W}_D+\mathbf{W}_{OD}$
			Estimates		
Int.	-21.06*	-21.06*	-21.06*	-21.06*	-21.06*
$\mathbf{p}_i$	0.96*	0.96*	0.96*	0.96*	0.96*
$\mathbf{p}_j$	0.98*	0.98*	0.98*	0.98*	0.98*
d	-0.41*	-0.41*	-0.41*	-0.41*	-0.41*
$\mathbf{f1}_i$	-0.11*	-0.11*	-0.11*	-0.11*	-0.11*
$\mathbf{f1}_j$	0.15*	0.15*	0.15*	0.15*	0.15*
$\mathbf{f2}_i$	-0.09*	-0.09*	-0.09*	-0.09*	-0.09*
$\mathbf{f2}_j$	0.00	0.00	0.00	0.003	0.003
			Dominant eigenvectors		
	a_{3} 1.40*	a_{11} -2.03*	a_{4} 3.34*	a_{1} -1.14*	a_{1} -1.15*
	a_{8} 1.96*	a_{18} -1.28*	a_{5} 2.96*	a_{2} 0.97*	a_{4} 3.56*
	a_{11} 1.85*	a_{22} -1.32*	a_{14} -3.04*	a_{4} 3.76*	a_{5} -2.62*
	a_{15} 1.67*	a_{27} -3.38*	a_{18} -4.58*	a_{8} 3.38*	a_{7} 2.08*
	a_{27} 2.50*	a_{38} -1.52*		a_{9} -1.23*	a_{9} 1.30*
	a_{35} 1.28*	a_{44} -1.69*		a_{12} -1.54*	a_{10} -1.46*
	a_{37} -1.34*	a_{45} 1.71*		a_{14} -2.94*	a_{14} -3.72*
	a_{38} 2.45*	a_{49} 1.76*		a_{18} 2.85*	a_{17} -2.59*
	a_{44} 1.46*	a_{51} -1.52*		a_{20} 0.88*	a_{18} 3.71*
	a_{45} -2.24*	a_{55} -1.31*		a_{23} 3.76*	a_{28} 2.31*
	a_{49} 1.30*	a_{57} -1.42*		a_{26} 2.45*	a_{39} -2.22*
	a_{51} 1.56*	a_{60} -1.21*		a_{28} 1.46*	
	a_{55} 1.26*	a_{62} -1.34*		a_{33} -1.11*	
	a_{57} -1.75*	a_{70} -2.36*		a_{37} 1.45*	
	a_{73} 2.07*	a_{71} 1.38*		a_{42} -1.981*	
	a_{88} 1.53*	a_{72} 1.79*		a_{52} 1.05*	
		a_{73} 2.52*		a_{59} -1.28*	

		a_{75} -1.92*		a_{87} -1.48*	
		a_{77} 1.39*		a_{107} 1.60*	
		a_{85} -1.22*		a_{120} 1.66*	
		a_{93} 1.33*			
Statistics			Value		
F-Fisher	90.85*	105.4*	291.20*	183.7*	233.8*
R^2	0.91	0.92	0.89	0.94	0.92
B-P	67.64	42.84	48.84*	64.37*	23.41
K-N	0.03	0.03	0.05	0.05	0.04
log-lik	-243.42	-222.17	-288.78	-172.04	-238.31
Moran	1.61	1.63	0.54	-0.018	-0.02
RMSE	0.46	0.43	0.52	0.38	0.45

*Legend: B-P= Breush-Pagan test of heteroscedasticity; K-N= Kolmogorov test of normality; log-lik=Log-likelihood; Moran=Moran test of autocorrelation on residuals; RMSE=Root Mean Square Errors; AIC=Akaike Information Criterium; * p-value<0.01.*

References

ANSELIN L. 1988. *Spatial Econometrics: methods and model*. Dondrecht, Kluver Academic Publishers.

BALTAGI H.B. 2011. *Econometrics 5th edition*. Berlin/Heidelberg: Springer.

BLACK W.R. 1992. Network autocorrelation in transport network and flow system, *Geographical Analysis*, Vol. 24, No. 3, pp. 207-222.

CHUN Y., GRIFFITH D.A. 2011. Modeling network autocorrelation in space-time migration flow data: an eigenvector spatial filtering approach, *Annals of the association of American Geographer*, Vol. 101, No. 3, pp. 523-536.

DE JONG P., SPRENGER C., VAN VEEN F. 1984. On the extreme value of Moran's I and Gary's c, *Geographical Analysis*, Vol. 16, No. 1, pp. 17-24.

EVENETT S.J., KELLER W. 2002. On the Theories Explaining the Success of the Gravity Equation, *Journal of Political Economy*, Vol. 110, pp. 281-316.

FISCHER M.M., GRIFFITH D.A. 2008a. Modelling spatial autocorrelation in spatial interaction data: An application to patent citation in the European Union, *Journal of Regional Science*, Vol. 48, No. 5, pp.969-89.

FORREST R., MURIE A. 1990. Moving strategies among home owners. In JOHNSON J. H. and SALT J. (Eds.) *Labour Migration: the internal geographical mobility of labour in the developed world*, London David Fulton Publisher, pp. 191-209.

GRIFFITH D.A. 2003. *Spatial autocorrelation and spatial filtering*. Springer-Verlang, Berlin.

GRIFFITH D.A. 2009. Modelling spatial autocorrelation in spatial interaction data: Empirical evidence from 2002 Germany journey-to-work flows, *Journal of Geographical System*, Vol. 11, pp. 117-40.

KIM K., KOEN E.J. 2010. Determinants of international migration flows to and from industrialized countries: A panel data approach beyond gravity, *International Migration Review*, Vol. 44, No. 4, pp. 899-932.

LEJENNE M. 2010. *Statistique la théorie e ses application*. Paris, Springer.

LESAGE J.P. 1997a. Regression analysis of spatial data, *Regional Analysis and Policy*, Vol. 27, No. 2, pp. 83-94.

LESAGE J.P., PACE R.K. 2008. Spatial econometric modeling of origin-destination flows, *Journal of Regional Science,* Vol. 48, No. 5, pp. 941-967.

LESAGE J.P., AGNAN C.T. 2015. Interpreting spatial econometric origin-destination flow models, *Journal of Regional Science*, Vol. 55, No. 2, pp. 188-208.

LEWER J.J., VAN DEN BERG H. 2008. A gravity model of immigration, *Economic Letters*, Vol. 99, pp. 164-167.

LUDO P. 2012. Gravity and spatial structure: The case of interstate migration in Mexico, *Journal of Regional Science*, Vol. 52, No. 5, pp. 819-856.

MAYDA A.M. 2010. International migration: A panel data analysis of the determinant of bilateral flows, *Journal of Population Economics*, Vol. 23, pp. 1249-1274.

SUMMARY

An analysis of methods for the treatment of autocorrelation in spatial interaction models

Using real data, this paper sets out the results of an analysis of the methods developed in the literature for the treatment of spatial autocorrelation in the spatial interactions models. In particular, the inquiry compares the autoregressive method (SAR), the method with autoregressive errors (SEM), and the spatial filtering method (SF). The results shows a substantial uniform behaviour among the considered approaches but the residuals of the SAR in same situations remains spatial correlated.

Giuseppe RICCIARDO LAMONICA, Polytechnic University of Marche – Faculty of Economics “G. Fuà”, g.ricciardo@staff.univpm.it