Market Potential, Spatial Dependences and Spillovers in European Regions[1]

Fernando Bruna [a]
Jesus Lopez-Rodriguez [a b]
Andrés Faíña [a]

[a] Competition and Development Research Group, Faculty of Economics and Business, Department of Economic Analysis, University of A Coruña, Campus de Elviña s/n, 15071 A Coruña, Spain
[b] IPTS, European Commission DG Joint Research Centre, Inca Garcilaso s/n, 41092 Seville, Spain

Abstract

This paper reinterprets the New Economic Geography (NEG) '*wage' equation* by distinguishing two different types of spatial dependences: global spatial trend and local spatial autocorrelation. A measure of *Market Potential* in this equation is able to capture both a global core-periphery pattern and spillovers, while the standard weights matrices of Spatial Econometrics tend to be designed to capture short-distance interactions among neighbors. Using cross-sectional European regional data, the paper compares different weighting schemes to build spatial lags. The estimation of spatial models of a NEG equation for GVA per capita reveals new research challenges.

Keywords: New Economic Geography, core-periphery model, wage equation, global spatial trend, spatial autocorrelation, spatial lag

JEL codes: C21, F12, O11, R12

---

1 This is an Accepted Manuscript of an article published by Taylor & Francis in Regional Studies on 29 Jul 2015, available at: https://doi.org/10.1080/00343404.2015.1048796

INTRODUCTION

The so called '*wage' equation* of the New Economic Geography (NEG) has been widely studied in the empirical literature (REDDING, 2011). It predicts that regional *marginal costs* (BRUNA, 2015) are a function of a variable called *Market Potential* or *Market Access*, which is a weighted sum of the market size of the other regions. The weights are inversely related to bilateral trade costs, usually proxied by distances. According to Spatial Econometrics, this sum can be seen as a spatially lagged endogenous variable (MION, 2004; KOSFELD and ECKEY, 2010). This is because the measure of market size on the right-hand side of the NEG equation is closely associated with the dependent variable.

The estimation of this type of equation frequently needs to be corrected for residual spatial autocorrelation. Therefore, if Spatial Econometrics uses a spatial lag of the dependent variable to correct a regression model for spatial autocorrelation, why does Market Potential fail to capture this spatial pattern? What does a spatial lag represent in each literature? If Market Potential is a different type of spatial lag, how are the results affected if both types of spatial lags are considered jointly in the same equation?

This paper posits a new way of interpreting the NEG equation by distinguishing the two types of spatial dependences studied by the NEG and Spatial Econometrics. On the one hand, a core-periphery spatial structure, captured by Market Potential, in Geostatistics corresponds to a *'global' spatial trend*, in which variable values change systematically with the geographic space coordinates. On the other hand, a short-distance or *local spatial pattern* is the focus of most spatial econometric techniques. The paper studies the characteristics, and implications, that allow both types of spatial patterns to be considered when estimating a NEG equation for European regions.

The approach follows two strands of the empirical literature. The first uses the NEG framework to study the effects of peripherality on economic development: REDDING and SCHOTT (2003), REDDING and VENABLES (2004), LÓPEZ-RODRÍGUEZ *et al.* (2007), BOULHOL and DE SERRES (2010) or LÓPEZ-RODRÍGUEZ *et al.* (2011). It is widely acknowledged that the regional spatial distribution of economic activity and population in Europe follows a core-periphery pattern: CLARK *et al.* (1969), KEEBLE *et al.* (1982) or FAÍÑA and LÓPEZ-RODRÍGUEZ (2006), among others. The key variable in the estimation of the NEG equation, Market Potential, is able to capture this spatial structure.

However, Market Potential does not seem to adequately capture spatial autocorrelation. A second strand of the literature uses spatial econometric techniques to estimate an equation including Market Potential. See NIEBUHR (2006) or Fingleton's extensive work (e.g., FINGLETON and FISCHER, 2010). Out of the NEG framework, the research line started by BLONIGEN *et al.* (2007) considers spatial autocorrelation to estimate the effects of Market Potential on foreign direct investment. However, no previous paper has examined the particularities of these two forms of spatial interaction.

Using a cross-sectional sample of European regions, this paper analyzes alternative weighting schemes to build several types of spatial lags and compare them with HARRIS's (1954) measure of Market Potential. The simple structure of the latter measure allows a clear definition of the different elements studied by the general perspectives of Geographical Economics and Spatial Econometrics on spatial dependence. Additionally, Harris's indicator shares some of the same relevant features of other empirical definitions of Market Potential, such as those employed by REDDING and VENABLES (2004) or HANSON (2005). The paper also addresses some controversial methodological details, such as the role played by the internal markets when estimating a gross value added per capita (GVApc) to Market Potential equation or the effects of using standardized distances to model spatial autocorrelation.

The main result of the analysis shows that the estimation of a NEG equation can simultaneously capture both global and local spatial trends. However, this achievement will be qualified by several caveats, mainly caused by the endogeneity of Market Potential and the shared elements between the weighting scheme of this variable and the matrix used to collect spatial autocorrelation. Additionally, though spillovers are excluded from NEG theory, it is shown that Market Potential also captures spatial spillovers. This affects the procedures and interpretation when modeling local autocorrelation and raise new questions about the NEG agglomeration mechanisms.

The rest of the paper is organized as follows. The second section presents a short overview of the concept of Market Potential and the NEG equation. The third section introduces the data and the econometric specifications. The fourth section outlines the global and local spatial dependences in the European data. The fifth section compares the data when building spatial lags using different weighting schemes. The sixth section shows alternative estimations of the NEG equation. The final section concludes and an Appendix describes the data.

THEORETICAL FRAMEWORK: THE NEG EQUATION AND MARKET POTENTIAL

Following HARRIS (1954), the market potential of a geographical observation (region i) is defined as the summation of markets (M) accessible to i divided by their 'distances'

(d_{ij}) to that point i. Considering the $R-1$ possible markets of other j regions, the Harris's Market Potential (HMP) of region i can be broken down into its *Internal* (IMP) and *External* (EMP) components:

$$HMP_i = \sum_{j=1}^{R} \frac{M_j}{d_{ij}} = \frac{M_i}{d_{ii}} + \sum_{j \neq i}^{R-1} \frac{M_j}{d_{ij}} = IMP_i + EMP_i \tag{1}$$

where the distance to the own regional market (d_{ii}) is measured by within region distances, as discussed in the next section. This paper focuses partly on the construction and interpretation of External Market Potential. Versions of EMP have been named 'non-local' (HEAD and MAYER, 2006), 'surrounding' (BLONIGEN *et al.*, 2007) or 'foreign' (BRAKMAN *et al.*, 2009) market potential.

Harris's approach has been widely used in Regional Economics. One reason is that it offers a way of capturing TOBLER´s (1970) first law of Geography, which would be much quoted later by the Spatial Econometrics literature: 'Everything is related to everything else, but near things are more related than distant things'. KRUGMAN's (1993) general equilibrium setting provided microeconomic foundations to the physical analogies of Harris's indicator. The basic NEG equation predicts that regional 'wages' are a function of the size of the markets available to each region. Here it is presented following HEAD and MAYER (2006) and COMBES *et al.* (2008).

The classical NEG equation explains the equilibrium 'industrial nominal wages' of each region i (w_i) as a function of the sum of a product of two elements for all the $j = 1, \ldots, R$ regions to which industrial goods are delivered. On the one hand, it is region j's volume of demand of individual manufacturing varieties. This element is the quotient between its demand of manufacturing goods ($\mu_j E_j$) and an index capturing the level of competition in j's market (S_j), where E_j and μ_j are j's total expenditure and its manufacturing share of expenditure, respectively. On the other hand, the second element determines j's demand of the specific variety produced in region i. It is the transport cost from region i to j (T_{ij}), to the power of one minus the elasticity of substitution among the varieties of industrial goods ($\sigma > 1$). A market clearing condition defines the equation in the following way:

$$w_i = \left(\sum_{j=1}^{R} T_{ij}{}^{1-\sigma} \frac{\mu_j E_j}{S_j} \right)^{1/\sigma} = (RMP_i)^{1/\sigma} \tag{2}$$

As discussed by BRUNA (2015) the dependent variable of a *generalized 'wage' equation* is marginal costs, instead of wages (see also COMBES *et al.*, 2008, chap. 12). That is one reason to use income per capita as an empirical proxy for the left-hand side variable of equation (2), in line with the arguments put forward by REDDING and VENABLES (2004) and HEAD and MAYER (2004). Furthermore, NEG theory relies on specific mechanisms and ignores many others, such as technology, institutions or spillovers. Apart from other possible limitations of the NEG, this paper will show that spillovers are indeed implicit in the NEG equation.

REDDING and VENABLES (2004) named *Market Access* to the expression between brackets in equation (2). Here the name given by HEAD and MAYER (2006), *Real Market Potential* (RMP_i), is used in keeping with the tradition that has existed from HARRIS (1954) to FUJITA *et al.* (1999). The 'real' is added to highlight the importance of discounting expenditures by the competition or supply index $S_j = \sum_{i=1}^{R} T_{ij}{}^{1-\sigma} n_i p_i^{1-\sigma}$, where n_i is the number of manufactured goods sold in j market and produced in any region i and p_i is the mill price of those goods.

HEAD and MAYER (2006) reserved the name 'Nominal Market Potential' for the expression $\sum_{j=1}^{R} T_{ij}^{1-\sigma}\mu_j E_j$. 'Nominal' refers to the absence of an adjustment for variation in the competition index ($S_j = S = 1$). Assuming that the share of manufacturing goods on expenditure is the same in all regions ($\mu_j = \mu = 1$), as FUJITA *et al.* (1999, chap. 4) consider, and that $T_{ij}^{1-\sigma} = T_{ij}^{-1}$, the Nominal Market Potential becomes the original formulation used by HARRIS (1954): $HMP_i = \sum_{j=1}^{R} T_{ij}^{-1}E_j$, where expenditure E_j measures the size of the markets (M_j) and trade costs are usually proxied by geographical distances ($T_{ij} = d_{ij}$)[1]. Once distance is taken as a proxy for trade costs, Harris's definition of Market Potential implies $T_{ij}^{1-\sigma} = d_{ij}^{1-\sigma} = d_{ij}^{-1}$. Indeed, a trade elasticity to distance of -1 is an extremely robust empirical finding in the literature on gravity equations (HEAD and MAYER, 2014).

Therefore, the main difference between RMP_i and HMP_i is that the latter variable is not corrected by the NEG's competition index S_j, which is not directly measurable. However, the results of BREINLICH (2006) and HEAD and MAYER (2006) for regional European data are similar when using Harris's definition of Market Potential rather than a more sophisticated structural estimation of the NEG equation.

DATA AND ECONOMETRIC SPECIFICATIONS

Taking logarithms in equation (2) and proxying RMP_i by HMP_i, the econometric specification considered in this paper for a cross-sectional regression is the following:

$$\ln w_i = C + \beta \ln HMP_i + u_i \tag{3}$$

Marginal costs ('wages') are proxied by income per capita, as it is frequent in NEG literature (REDDING and VENABLES, 2004; BRAKMAN *et al.*, 2009). Gross value added per capita (GVApc)[2] is used here, as BREINLICH (2006) does. This variable confers generality to the discussion on the spatial structure of economic activity and spatial dependence. Market Potential is also constructed with GVA. Furthermore, some of the regressions estimated in this paper include a control variable of human capital, proxied by a measure of human resources in science and technology. HEAD and MAYER (2006) or BREINLICH (2006), among others, have also controlled the estimation for human capital. Details of the variables and the sample are provided in the Appendix.

The term u_i in equation (3) is supposed to collect departures from the assumptions of the theoretical model, such as the effects of omitted variables, which are assumed to be randomly distributed under OLS estimation. However, as will be shown later, the estimation of this equation using European regional data results in spatially autocorrelated residuals. When a hypothetical data generation process includes spatial dependence in the endogenous or the explanatory variables and those spatial effects are omitted, the estimator of the coefficients for the remaining variables is biased and inconsistent. In contrast, ignoring any spatial dependence in the disturbances will only cause a loss of efficiency (LESAGE and PACE, 2009, p. 156).

In order to test and model spatial autocorrelation it is necessary to choose a neighborhood criterion (who is linked with who) and to build a spatial weights matrix (W), assigning weights to the areas that are considered to be linked. The tests considered later, Moran's I and Lagrange Multipliers, use this W matrix to check whether a variable is spatially autocorrelated or a model follows a particular process of spatial dependence. These tests can detect misspecification instead of a true process of spatial autocorrelation. The most obvious reason for misspecification is the omission of relevant explanatory variables. A control variable of human capital in the estimation reduces but does not solve this issue.

The problem of misspecification is closely linked to the focus of this paper: BIVAND *et al.* (2008, p. 260) show that even a gentle global regional trend induces apparent spatial autocorrelation in Moran's index, unmasked when a correct model is fitted. As will be shown in the next section, Market Potential is able to capture a global core-periphery pattern in the data, namely a global spatial trend. This type of spatial dependence could also be termed 'polarization', 'long distance' or 'large scale' spatial structure. On the contrary, spatial autocorrelation is emphasized in this paper as an average 'local' phenomenon of spatial dependence[3].

Spatial Econometrics uses the W matrix to specify a variety of spatial models. LESAGE and PACE (2009, chap. 2) discuss some of their motivations. Two simple models are studied in this paper. Firstly, the Spatial Autoregressive (SAR) Model, also known as Spatial Lag Model, includes an endogenous interaction effect. A cross-sectional SAR model can arise from time-dependence of decisions when decisions depend on those of neighbors. It takes the form in equation (4):

$$Y = \rho WY + X\beta + u \tag{4}$$

LESAGE (2014) argues that the SAR specification involves '*global' spatial spillovers* because the feedback effects imply a new steady-state equilibrium, to be studied through *total impacts*. That is a useful though 'somewhat artificial' distinction from 'local' spill-overs due to the possible inclusion of WX in other spatial models not studied here.

The second basic spatial model is the Spatial Error Model (SEM), which captures interaction effects between the error terms. A cross-sectional SEM model can be motivated by spatially autocorrelated omitted variables. Its specification is as follows:

$$\begin{aligned} Y &= X\beta + u \\ u &= \lambda W u + \varepsilon \end{aligned} \tag{5}$$

There are many ways of selecting W for describing an unknown structure of average spatial interaction in a particular sample of data, if such a structure exists. For technical reasons[4], the w_{ij} elements of this matrix are usually row-standardized (normalized) with the sum of all the weights for region i. Therefore, the spatial lag of a variable X can be interpreted as the weighted average of X in the 'neighbors' (however defined): $(WX)_i = \sum_j \frac{w_{ij}}{\sum_j w_{ij}} X_j$. When W is a binary matrix, with 1 if two regions are considered as neighbors and 0 otherwise, standardization by rows implies that the spatial lag of a variable is the mean value of the variable for the neighbors.

A rule of parsimony in Spatial Econometrics recommends not imposing a strong structure on the W matrix when trying to capture an unknown distribution of spatial dependence (GRIFFITH, 1996). Moreover, in order to capture local spatial patterns it is useful to apply a restricted neighborhood criterion. The baseline spatial matrix considered in this paper is a row-standardized binary weights matrix to the 5 nearest neighbors. This W matrix allows checking possible average interactions of each region with its surroundings and is sufficient to distinguish global and local spatial patterns. For reasons summarized by LESAGE (2014), the main results of the paper do not depend on choosing a lower or higher number of neighbors.

As pointed out by ANSELIN (1988, pp. 23–24), when w_{ij} represents a distance decay, scaling the rows so that the weights sum to one leads to the loss of the economic interpretation of that distance decay. In contrast, Market Potential is an index of accessibility to the markets, which requires absolute distances to be considered[5]. However, both approaches to spatial dependence are related. The external component of Harris's Market Potential in equation (1) is a non-standardized inverse distance spatial lag of the regional internal markets. The next section discusses the specific differences between Market Potential and a number of other spatial lags frequently used in Spatial Econometrics.

Given that Market Potential can be viewed as a spatial lag of the dependent variable, it is an endogenous variable, capturing 'global' spillovers, which biases the OLS estimation of a NEG equation. Therefore, this equation has been estimated by instrumental variables (HEAD and MAYER, 2006; BRUNA *et al.*, 2014). Likewise, the estimation of a SAR model with the procedure of KELEJIAN and PRUCHA (1998) instruments the spatial lag of the dependent variable with the spatial lags of the explanatory variables. This latter procedure is problematic in a NEG equation because one of the explanatory variables is similar to a spatial lag of the dependent variable (see endnote 8).

Indeed, this issue has major consequences. When calculating the total effects of a SAR model, as will be carried out in the sixth section of this paper, the simultaneous endogenous effects of GVApc and GVA Market Potential should be considered. Moreover, a Spatial Error Model of equation (3) also contains a form of spatial lag of the dependent variable. It is actually very similar to a Spatial Autocorrelation (SAC) Model, which includes spatial dependence in both the dependent variable and the errors. Therefore, the elasticities of the variables should also be calculated through total effects, although this discussion is beyond the scope of this paper. The empirical part of the paper will test the robustness of the results to instrumental variable estimation.

The inclusion of the internal markets in HMP_i aggravates the general endogeneity problem of Market Potential. Nonetheless, excluding the IMP_i component of equation (1) introduces measurement error by reducing the access measure of some economically larger locations (BREINLICH, 2006; HEAD and MAYER, 2006), such as capital cities. A practical issue when proxying the own regional market is the measurement of internal distances (d_{ii}). The standard method assumes that regions are circular so the radius of region i is $r_i = \sqrt{area_i/\pi}$. In this paper internal distances are measured following KEEBLE *et al.* (1982), who chose $d_{ii} = 1/3 \cdot r_i = 0.188\sqrt{area_i}$, to allow for the likely clustering of economic activity in and around the 'center'. This is similar to the 40% of the radius considered by CAMBRIDGE ECONOMETRICS (2014). Calculating d_{ii} as $1/3$ of the radius increases the weight of IMP_i in HMP_i when compared with the $2/3$ used by some authors. Therefore, it helps to study the sensibility of the results to the inclusion of IMP_i.

EUROPEAN REGIONAL SPATIAL DEPENDENCES

As discussed in the introductory section, previous literature has shown that the European regional spatial distribution of economic activity follows a core-periphery pattern, with just a few high income regions outside the geographical center of Europe, particularly those in Nordic countries. The economically central regions (with high GVApc) are mainly located around the so called *blue banana*, from West England in the North to Milan in the South. They are also geographically central regions.

Figure 1. Market Potential and residuals of GVA per capita on Market Potential (logs, year 2008)

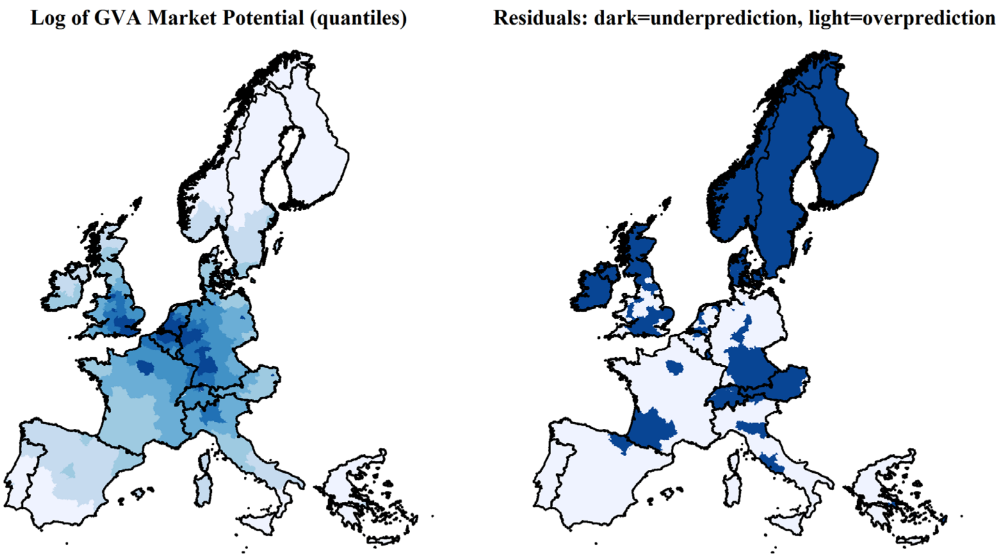


Figure 1 shows maps of the regional Market Potential (HMP_i) in Europe in 2008 and the residuals of a regression of GVApc on Market Potential. Both variables are in natural logarithmic form, as this is the form in which they are studied in the NEG equation. The values of the log of Market Potential are divided into seven quantiles, which helps to visualize their global spatial pattern. Darker colors are associated with higher values of the variables. Alternatively, and in order to simplify the visualization of spatial autocorrelation, the right map of the figure only distinguishes two types of values depending on the sign of the residuals. Positive residuals, shown in a dark color, indicate that Market Potential under-predicts the regional GVApc, while negative residuals point to over-predictions.

In spite of the general visual limitations of cloropheth maps (see endnote 5), the left-hand map of Figure 1 shows that the Market Potential variable is able to capture a global core-periphery spatial trend. However, this attractive feature of Market Potential also has a drawback. The map on the right of Figure 1 shows that the residuals of a GVApc-Market Potential equation present strong local spatial dependence. If Market Potential under-predicts a region, it tends to do the same with its neighbors. Therefore, the residuals of the regression will be positive in close under-predicted regions and negative in close over-predicted regions. OLS merely distributes the under and over predictions in order to obtain zero mean residuals.

This local clustering of residuals is formally tested in Table 1. Moran's I is calculated for the variables and the residuals of OLS regressions of GVApc on Market Potential. The zero p-values of Moran's test reject the null hypothesis of absence of positive spatial autocorrelation at short distances.

Table 1. Spatial autocorrelation of variables (logs) and OLS residuals

	Moran's I statistic	p-value
GVA per capita (GVApc)	0.617	0.000
Market Potential	0.854	0.000
External Market Potential	0.921	0.000
Residuals Market Potential	0.587	0.000
Residuals External Market Potential	0.494	0.000

Note: Cross-section of 220 regions for the year 2008. The residuals are those of the regression of the log of GVApc on the logs of the Market Potential variables. Moran's tests use the randomization assumption for the variables and the normality assumption for the residuals. The alternative hypothesis for the p-values is that Moran's I is greater than expected under the null hypothesis of absence of spatial autocorrelation. The weights matrix is a row-standardized binary matrix to the 5 nearest neighbors.

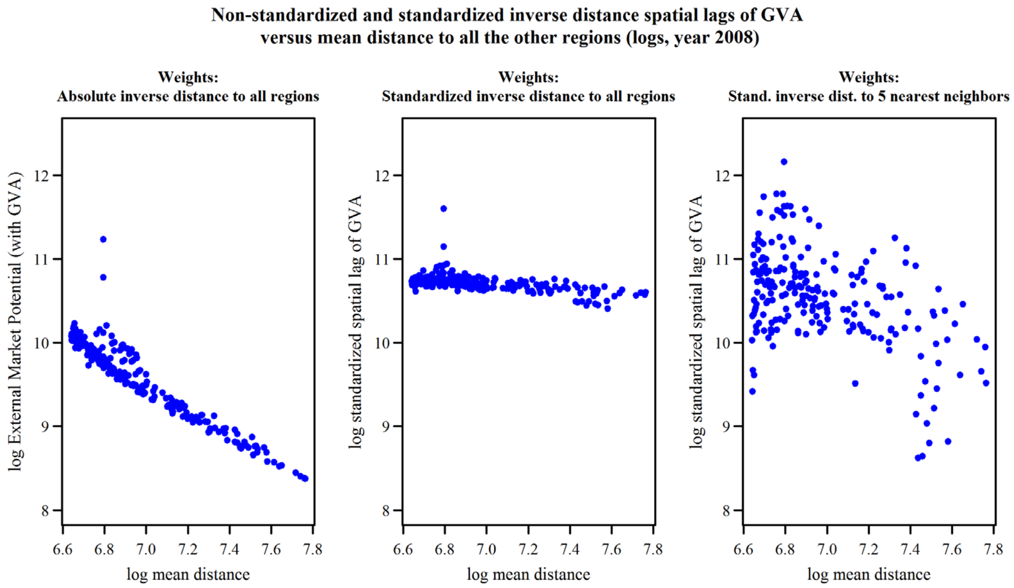
Market Potential fails to correct an empirical NEG equation for residual spatial autocorrelation because is constructed as a weighted sum for all the regions in the sample, what produces a smoothed spatial distribution of values. The summation effect makes Market Potential to be more spatially autocorrelated than GVApc. Therefore, capturing the global core-periphery pattern of the European GVApc through a Market Potential variable tends to come at the cost of spatially autocorrelated residuals at short distances. Before studying some spatial models that could solve the violation of the OLS assumptions, the following section emphasizes four differences between Market Potential and other types of spatial lags frequently used in Spatial Econometrics.

MARKET POTENTIAL AND OTHER SPATIAL LAGS: THE ROLE OF DIFFERENT WEIGHTING SCHEMES

Figure 2 illustrates the two main differences between External Market Potential and some other types of spatial lags: the use of absolute distances and the number of regions considered in the sum. The plots show the variation of the log of three spatial lags of GVA with respect to the log of the mean distance of each region to all the other regions in the sample.

The values of EMP, on the left plot[6], decrease steadily with the distance from the European geographical center. The inverse distance weights enable the highest values of Market Potential to be geographically centered on the *blue banana* (see Figure 1). However, Market Potential exaggerates the global (core-periphery) regional trend in GVApc (not shown), as is expected when trying to capture a stylized property of the data. The smoothing effects of the sum for all the regions in Harris's accessibility index and the use of absolute distances are common to other proxies of Market Potential, such as those of REDDING and VENABLES (2004)-BREINLICH (2006)-HEAD and MAYER (2006) or HANSON (2005)-NIEBUHR (2006), though they are not studied here.

Figure 2. The effects of standardizing distances and reducing the number of neighbors



These smoothing effects can also be observed in the central plot of Figure 2. However, the weights are now standardized, which reduces the dispersion of the variable in the distance dimension. The discount factor of distance and the economic interpretation of accessibility are lost. The sum of the spatial weights is 1 for all regions in the standardized version of EMP, while it ranks from 0.1 to 0.7 in EMP for this sample, according to regional peripherality.

On the contrary, attending a rule of parsimony and in an attempt to capture an average local spatial pattern, Spatial Econometrics tends to restrict the criterion of neighborhood.

The right plot of Figure 2 shows the standardized version of the log of EMP for the 5 nearest neighbors only. Now the log of the spatial lag of GVA displays a high dispersion in the distance dimension, according to the GVA level of the neighbors in the nearby distances. Indeed, the role of the standardized distances in this variable is not that important. A standardized *binary* matrix to the 5 nearest neighbors produces a plot (not shown) very similar to the right plot of Figure 2, because a criterion of distance is implicit when using a few nearest neighbors. However, with a binary matrix some outliers disappear, which might affect the estimation results of spatial models.

There are two additional differences between Market Potential and some of the spatial lags most frequently used in Spatial Econometrics. On the one hand, there is what can be termed the 'lag of log' versus the 'log of lag' issue. In a NEG equation for GVA per capita, such as ln GVApc $= C + \beta$ ln $MP + u$, when (External) Market Potential is built with GVA the explanatory variable is the log of a spatial lag of GVA: ln (W_1·GVA). However, spatial econometric models often include spatial lags of the variables, which are already in log form. Therefore, a spatial lag of Market Potential would have the form W_2·ln (W_1·GVA). Under the SAR model of equation (4), the spatially lagged dependent variable is also the spatial lag of a log: W_2·ln GVApc. On the other hand, Market Potential is built with GVA while the latter spatial lag is built with GVA *per capita*.

Table 2 provides a first decomposition of the empirical effects of the four differences mentioned above[7], through the correlations among 10 variables using four W matrices. The dependent variable [ln GVApc] is called variable (1) and the Market Potential variables are called (2) and (3). Alternative logs of the spatial lags of GVA [ln (W·GVA)], numbered as (4) to (6), are shown to compare their weighting schemes with that of variable (3), External Market Potential, which is also constructed as ln (W·GVA). The three plots in Figure 2 correspond to variables (3), (4) and (5) here. The similar correlations in the columns of variables (5) and (6) confirm the limited role of the inverse distance weighting scheme when the weights to a few nearest neighbors are standardized. The comparison of variables (6) and (7) shows the 'lag of log' versus 'log of lag' issue for the same binary weights matrix: as might be expected, variable (7) [W·ln GVA] is more closely correlated to the dependent variable than variable (6) [ln (W·GVA)].

Table 2. Cross-sectional correlations (year 2008)

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Variable	log GVA per capita [ln GVApc]	log Market Potential [ln MP]	log External MP [ln EMP = ln (W · GVA)]	log spatial lag of GVA [ln (W·GVA)]			Spatial lag of log GVA [W·ln GVA]	Spatial lag of log GVA per capita [W·ln GVApc]		
W	Neighbors		All	All	All	5 nearest	5 nearest	5 nearest	All	All	5 nearest
	Weights		Absolute inverse distance	Absolute inverse distance	Standardized inverse distance	Standardized inverse distance	Standardized binary	Standardized binary	Absolute inverse distance	Standardized inverse distance	Standardized binary
(1)	ln GVApc	1.00	0.56	0.48	0.29	0.25	0.23	0.33	0.48	0.70	0.75
(2)	ln MP	0.56	1.00	0.96	0.64	0.50	0.47	0.59	0.93	0.65	0.53
(3)	ln EMP	0.48	0.96	1.00	0.69	0.54	0.52	0.63	0.96	0.70	0.58
(4)		0.29	0.64	0.69	1.00	0.77	0.71	0.72	0.54	0.59	0.45
(5)	ln (W·GVA)	0.25	0.50	0.54	0.77	1.00	0.98	0.92	0.39	0.47	0.46
(6)		0.23	0.47	0.52	0.71	0.98	1.00	0.93	0.37	0.43	0.44
(7)	W·ln GVA	0.33	0.59	0.63	0.72	0.92	0.93	1.00	0.48	0.53	0.51
(8)		0.48	0.93	0.96	0.54	0.39	0.37	0.48	1.00	0.67	0.54
(9)	W·ln GVApc	0.70	0.65	0.70	0.59	0.47	0.43	0.53	0.67	1.00	0.91
(10)		0.75	0.53	0.58	0.45	0.46	0.44	0.51	0.54	0.91	1.00

Note: The baseline W matrix used in the following tables corresponds to the weighting scheme of variables (6), (7) and (10).

Variables (8) to (10) are different spatial lags of the log of GVA *per capita*. Variable (10) will be added as explanatory variable in the SAR models of the next section. Comparing it with variables (6) [ln (W·GVA)] and (7) [W·ln GVA], it could be said that 0.23 of the 0.75 correlation of variable (10) [W·ln GVApc] with the dependent variable [ln GVApc] would be 'explained' by using GVA, 0.10 additional correlation by the 'lag of log' issue, while the remaining 0.42 would be given by using GVApc instead of GVA. However, the 0.96 correlation between variables (8) and (3) reveals that the spatial lag of the log of GVA per capita [W·ln GVApc] built with absolute inverse distances to all the regions captures the same information than the log of EMP [ln (W·GVA)]. Therefore, when an inverse distance W matrix includes all the regions in the sample the smoothing effects of the summation make irrelevant both the 'lag of log' issue and the utilization of GVApc instead of GVA.

The correlation of $\ln EMP$, (3), with the dependent variable (1) is considerable, 0.48. That correlation is reduced to 0.25-0.23 when the weights of 5 nearest neighbors are standardized, as invariables (5) and (6). The correlation of variable (3) and (10) is quite high, 0.58, which might create multicollinearity problems in the SAR models of Table 5 below. Nevertheless, a correlation of 0.58 might be qualified as not being very severe when considering the importance of capturing two different spatial patterns, namely global and local spatial dependences.

MODELING SPATIAL DEPENDENCES: THE CASE OF EUROPE

The final stage of this discussion is to model the global core-periphery spatial structure of the European regional economic activity as the same time as the short-distance spatial autocorrelation studied above.

Table 3 shows the OLS cross-sectional estimation of four alternative specifications of equation (3) and the IV estimation of the specifications with EMP. Three conclusions can be drawn from the results. Firstly, columns (3) and (4) show that External Market Potential and Human Capital jointly 'explain' around half of the dispersion of GVApc, each in approximately the same proportion. Consistently with the correlations in Table 2, the improvement of the R-squared is negligible when the internal component of Market Potential is included in columns (1) and (2). Secondly, as in Table 1, the p-values of Moran's I show that the residuals of all the OLS estimations are spatially autocorrelated. Thirdly, the endogeneity of External Market Potential is rejected when this variable is instrumented by the mean distance of each region to all the other regions. As any contextual test, the Wu-Hausman test is conditional to the specification and the quality of the instruments. Therefore, these IV estimations do not try to be a definitive evidence of exogeneity but a support to use the OLS estimates as a benchmark for the spatial models below.

Table 3. Cross-sectional estimations for 220 European regions (year 2008)

	OLS				IV			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(Intercept)	6.450***	8.411***	6.803***	8.696***	6.612***	20.694***	8.470***	20.729***
	(0.356)	(0.339)	(0.397)	(0.354)	(0.527)	(0.258)	(0.404)	(0.261)
Market Potential (MP)	0.365***	0.285***						
	(0.037)	(0.030)						
External MP			0.333***	0.270***	0.353***		0.292***	
			(0.041)	(0.033)	(0.054)		(0.038)	
Human Capital		0.527***		0.569***			0.564***	0.033
		(0.048)		(0.049)			(0.053)	(0.033)
Mean distance						-1.596***		-1.590***
						(0.037)		(0.037)
R-squared	0.313	0.556	0.229	0.521	0.228	0.895	0.520	0.896
Adj. R-squared	0.310	0.552	0.225	0.517	0.224	0.895	0.516	0.895
AIC	73.6	-20.3	99.1	-3.7	98.8	-181.0	-6.8	-180.0
p-value Moran's I	0.000	0.000	0.000	0.000		0.000		0.000
Moran's I residuals	0.587	0.563	0.494	0.509		0.493		0.507
Sum squared errors	17.52	11.33	19.67	12.21	19.69	5.50	12.24	5.48
Weak inst. F test					4390		5554	
Weak inst. p-value					0.000		0.000	
Wu-Hausman F test					0.952		2.025	
Wu-Hausman p-value					0.330		0.156	

Note: Table displays coefficients: * significant at 10% level; ** at 5% level; *** at 1% level. Standard errors are in brackets. Variables are in log form. The dependent variable is GVApc. Columns (5) and (6) show the second stage IV estimations, with Eicker-Huber-White standard errors, for the specifications in columns (3) and (4), respectively. Columns (6) and (7) show their first stage regressions. The Stock and Yogo's critical value for the first-stage F-statistic weak identification test for 1 endogenous regressor, 1 instrumental variable and 10% of desired maximal size of a 5% Wald test is 16.38.

In order to correct these regressions for spatial autocorrelation, the proper spatial model can be selected using Lagrange Multiplier diagnostics for the presence of a spatially lagged dependent variable or error dependence. The decision rule, following FLORAX *et al.* (2003), is based on a comparison of the LM tests for the same specification. Adapting it to the terminology in R *spdep* package (BIVAND, 2014) the simple tests are named LMerr and LMlag, while RLMerr and RLMlag are their versions robust to the possible presence of the other type of spatial dependence. The null hypothesis for LMlag and RLMlag is $\rho = 0$ in equation (4), while for LMerr and RLMerr is $\lambda = 0$ in equation (5). If the p-value of a test is very close to zero the alternative hypothesis of an erroneously omitted spatial process of the type under consideration is accepted.

KOSFELD and ECKEY (2010) prefer a SEM specification for the NEG equation in order to avoid introducing two spatial lags of the dependent variable, as occurs in a SAR model including Market Potential. The LM tests in Table 4 show that the SEM is chosen in all cases except in the specification of External Market Potential without human capital. However, Table 5 presents the maximum likelihood estimation of both spatial models for three reasons: the specifications and spatial models studied here are simple; each model has different motivations; and the SAR model is apparently more critical for a Market Potential variable.

The estimates obtained with SAR and SEM models are not comparable. Therefore, Table 6 shows the total effects (LESAGE and PACE, 2009, chap. 2) of the variables in the SAR specifications. Nevertheless, as was discussed earlier, both the estimates of the SEM models and the total effects of the SAR models should be considered as rough estimations of elasticities, given that they fail to consider the simultaneous relationship between GVApc and GVA Market Potential. However, it should be noted that when controlling for human capital the OLS estimates of the two Market Potential variables in Table 3 have a similar magnitude to the SEM estimates in Table 5 and to the total effects in Table 6[8].

Market Potential has a significant positive effect in all the specifications but the external component is only significant at 5% level in some spatial specifications (see also BRUNA *et al.*, 2014). However, both ρ and EMP are significant at 1% level in the SAR specification of column (4), in spite of the 0.58 correlation shown in Table 2.

Four additional robustness analyses were carried out and are available upon request. Firstly, all the calculations in this paper were repeated using a broader sample including 54 additional regions from Central and Eastern Europe (CEE). Secondly, all the estimations for both samples were repeated using the standardized inverse distance weights matrix to the 5 nearest neighbors, as in variable (5) of Table 2. Thirdly, the spatial models were also estimated including country dummies. Fourthly, in order to check the specific role of CEE, the models were re-estimated omitting some other peripheral countries[9].

The significance of External Market Potential is sensitive to these changes. For instance, focusing on the SEM specification in column (8) of Table 5, EMP is not significant in the sample of 274 regions. In the sample of 220 regions it is only significant at the 10% level when the standardized inverse distance W matrix is used. However, in both samples and with both weights matrices, EMP becomes significant at the 1% level when country dummies are included in the regression. Finally, if four Mediterranean countries are excluded from the sample with CEE, EMP continues to be insignificant in the SEM specification but becomes significant at 1% level under the SAR model. This last result reinforces the interaction channel of global spillovers, maybe due to common institutions and history among close neighbors.

Table 4. Lagrange Multiplier diagnostics for spatial dependence on the OLS residuals

	Without Human Capital		With Human Capital	
	Statistic	p-value	Statistic	p-value
Columns (1) and (2) with Market Potential				
LMerr	213.924	0.000	196.428	0.000
LMlag	159.486	0.000	108.188	0.000
RLMerr	60.867	0.000	88.310	0.000
RLMlag	6.429	0.011	0.070	0.791
Columns (3) and (4) with External Market Potential				
LMerr	151.171	0.000	160.512	0.000
LMlag	152.924	0.000	103.469	0.000
RLMerr	0.790	0.374	57.077	0.000
RLMlag	2.543	0.111	0.034	0.853

Note: LM tests on the residuals of the OLS estimations in Table 3.

Table 5. ML estimations of spatial models

	Spatial Lag Model (SAR)				Spatial Error Model (SEM)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ρ	0.629***	0.513***	0.674***	0.540***				
	(0.060)	(0.055)	(0.058)	(0.057)				
λ					0.819***	0.746***	0.682***	0.703***
					(0.038)	(0.049)	(0.057)	(0.055)
(Intercept)	1.878***	4.181***	2.185***	4.492***	4.133***	7.509***	7.596***	9.617***
	(0.449)	(0.520)	(0.510)	(0.569)	(0.545)	(0.542)	(0.748)	(0.622)
Market Potential	0.187***	0.161***			0.607***	0.364***		
	(0.037)	(0.029)			(0.057)	(0.050)		
External MP			0.111**	0.109***			0.249**	0.179**
			(0.036)	(0.031)			(0.079)	(0.063)
Human Capital		0.393***		0.418***		0.460***		0.596***
		(0.044)		(0.048)		(0.049)		(0.049)
AIC	-20.4	-96.2	4.4	-75.0	-75.3	-144.5	6.1	-104.0
p-value LMerr resid	0.267	0.000	0.334	0.015	0.521	0.627	0.329	0.479
p-value Moran's I	0.107	0.000	0.806	0.005	0.704	0.647	0.809	0.726
Moran's I residuals	0.045	0.143	-0.039	0.097	-0.026	-0.019	-0.039	-0.028
Residual variance	0.047	0.034	0.052	0.038	0.034	0.025	0.052	0.031
Sum squared errors	10.42	7.56	11.48	8.26	7.44	5.60	11.53	6.87

Note: SAR and SEM estimations of the four OLS specifications in Table 3.

Table 6. Effects (impacts) of variables in the SAR models

	Direct	Indirect	Total
(1) Market Potential	0.207	0.298	0.505
(2) Market Potential	0.171	0.160	0.331
Human Capital	0.416	0.390	0.806
(3) External Market Potential	0.125	0.215	0.340
(4) External Market Potential	0.116	0.120	0.236
Human Capital	0.446	0.462	0.908

Note: Impacts of the explanatory variables in columns (1) to (4) of Table 5.

These exercises show that the empirical validity of the NEG must address several complications when a NEG equation is corrected for spatial autocorrelation, to be summarized in the next section.

CONCLUSIONS

Under the theoretical approach of the NEG and by means of simple spatial econometric techniques, this paper shows the features that allow a 'wage' equation to capture both the global core-periphery spatial structure of the European regional economic activity and the short-distance interactions among close neighbors.

Using HARRIS's (1954) measure of Market Potential, it has been shown that the combination of a sum for all the regions in the sample with a weighting scheme based on absolute distances allows Market Potential to capture a global spatial trend. However, the smoothing effects of the sum make the residuals of an estimated NEG equation spatially autocorrelated at short distances. These characteristics are common to other empirical measures of Market Potential used in the literature, although they are not studied here.

Harris's Market Potential is a non-standardized inverse distance spatial lag of any indicator used to measure market size, in this case gross value added. For the first time in the literature, its comparison with other spatial lags frequently used in Spatial Econometrics revealed four differences: the number of neighbors considered in the summation; the standardization of weights; the reference to the dependent variable in SAR models (instead of market size); and the 'lag of log' versus 'log of lag' issue. Those differences were empirically analyzed. Additionally, SAR and SEM models were estimated for a cross-sectional GVA per capita equation of the European regions. The results confirmed the possibility of capturing both global and local spatial dependences. However, they were shown to be sensitive to specification, coverage, model selection, country dummies and weights matrices. It remains an open issue for future research whether this sensitivity is due to the limitations of NEG or to misspecification.

If Market Potential is seen as a spatial lag of the dependent variable, an empirical NEG equation implicitly captures spillover effects, in spite of being absent from NEG theory. This issue will have to be part of the future debate between the NEG and Urban Economics literatures. Additionally, the analysis revealed that simultaneously capturing two different types of spatial dependence in a simple equation implies a number of pitfalls. The Market Potential variable induces endogeneity problems that affect the calculation of total effects in both SAR and SEM specifications and expands the research agenda about instrumental variables in spatial models of the NEG equation. Multicollinearity is sensitive to the selection of the spatial weights matrix. Furthermore, the problematic consideration of the internal markets can be critical when estimating spatial models.

These challenges do not hide the major achievement of capturing a global and a local spatial pattern in the same equation. However, they have been only partially addressed in the literature and raise new questions about the NEG agglomeration mechanisms.

Acknowledgments – We wish to thank Tomasz Mickiewicz and two anonymous referees for very useful contributions. We are grateful to participants in the presentations of previous drafts at the XV Applied Economic Meeting (A Coruña, Spain, 2012), the XIII Conference on International Economics (Granada, Spain, 2012) and the VI World Conference of the Spatial Econometrics Association (Salvador, Brazil, 2012). The first author thanks the financial support of the University of A Coruña for a research visit at the Università Cattolica del Sacro Cuore in Rome and the conversations with Giuseppe Arbia there. Jesus Lopez-Rodriguez acknowledges the support received from the Spanish Ministry of Science and Innovation (project ECO2011-28632) and Xunta de Galicia (project EM2014/051). The views expressed are purely those of the authors and may not under any circumstances be regarded as stating an official position of the European Commission. We thank Eduardo Giménez, Coro Chasco, Julie Le Gallo, Robert Bivand, Frank van Oort, David G. Rossiter, Jose-María Montero, Laura Varela, Nicolas Devarsy and Xosé Manuel Martínez Filguera for helpful insights on particular aspects of previous drafts. Jon Stenning made helpful clarifications about the Cambridge Econometrics's data. The valuable research assistance of Adrián Gutiérrez, Vanessa Mato and Paulino Montes is highly recognized. Usual disclaimers apply.

APPENDIX: DATA DESCRIPTION

The disaggregation level for the regional data is NUTS 2 (2006 version). The main sample includes 220 regions from 17 European countries. The following NUTS 2 regions are excluded: the Atlantic islands (Canary Islands, Madeira and the Azores), the Spanish Ceuta and Melilla and the French Departments Guadeloupe, Guiana, Martinique and Reunion.

All the variables refer to year 2008. Gross value added, in 2000 year euros, and population are taken from Cambridge Econometrics. Human capital is proxied by Eurostat's share of the population who has successfully completed education in Science and Technology (S&T) at the third level and is employed in a S&T occupation.

Geographical distances are measured as great circle distances between regional centroids calculated, in kilometers, using GISCO's shape files (© EuroGeographics for the administrative boundaries). Internal regional distances are based on regional areas calculated from these files after an EPSG 3035 projection.

REFERENCES

ANSELIN L. (1988) *Spatial Econometrics: Methods and Models*, Kluwer, Dordrecht.

BIVAND R. (2014) *spdep. Spatial dependence: weighting schemes, statistics and models*, R package. http://CRAN.R-project.org/package=spdep

BIVAND R., PEBESMA E. J. and RUBIO V. G. (2008) *Applied Spatial Data Analysis with R*, Springer, Berlin.

BLONIGEN B. A., DAVIES R. B., WADDELL G. R. and NAUGHTON H. T. (2007) FDI in space: Spatial autoregressive relationships in foreign direct investment, *European Economic Review* **51(5)**, 1303–1325.

BOULHOL H. and DE SERRES A. (2010) Have developed countries escaped the curse of distance?, *Journal of Economic Geography* **10(1)**, 113–139.

BRAKMAN S., GARRETSEN H. and VAN MARREWIJK C. (2009) Economic Geography within and between European nations: The role of Market Potential and density across space and time, *Journal of Regional Science* **49(4)**, 777–800.

BREINLICH H. (2006) The spatial income structure in the European Union—what role for Economic Geography?, *Journal of Economic Geography* **6(5)**, 593–617.

BRUNA F. (2015) A generalized NEG wage-type equation, in DÍAZ-ROLDÁN C. and PEROTE J. (Eds) *Advances on International Economics*, pp. 63–82. Cambridge Scholars Publishing, Newcastle.

BRUNA F., FAÍÑA A. and LOPEZ-RODRIGUEZ J. (2014) *Market Potential and the curse of distance in European regions*. MPRA Paper 56747. University Library of Munich, Munich.

CAMBRIDGE ECONOMETRICS (2014) *European Regional Database*. http://www.camecon.com/SubNational/SubNationalEurope/RegionalDatabase.aspx

CLARK C., WILSON F. and BRADLEY J. (1969) Industrial location and economic potential in Western Europe, *Regional Studies* **3(2)**, 197–212.

COMBES P.-P., MAYER T. and THISSE J.-F. (2008) *Economic geography: the integration of regions and nations*, Princeton University Press, Princeton, N.J.

FAÍÑA A. and LÓPEZ-RODRÍGUEZ J. (2006) European Union Enlargement, European Spatial Development Perspective and Regional Policy: Lessons from Population Potentials, *Investigaciones Regionales* **9**, 3–21.

FINGLETON B. and FISCHER M. M. (2010) Neoclassical theory versus new economic geography: competing explanations of cross-regional variation in economic development, *The Annals of Regional Science* **44(3)**, 467–491.

FLORAX R. J., FOLMER H. and REY S. J. (2003) Specification searches in spatial econometrics: the relevance of Hendry's methodology, *Regional Science and Urban Economics* **33(5)**, 557–579.

FUJITA M., KRUGMAN P. and VENABLES A. J. (1999) *The spatial economy: cities, regions and international trade*, The MIT Press, Cambridge.

GRIFFITH D. A. (1996) Some Guidelines for Specifying the Geographic Weights Matrix Contained in Spatial Statistical Models, in ARLINGHAUS S.L. and GRIFFITH D.A. (Eds) *Practical handbook of spatial statistics*, pp. 65–82. CRC Press, Boca Raton.

HANSON G. H. (2005) Market potential, increasing returns and geographic concentration, *Journal of International Economics* **67(1)**, 1–24.

HARRIS C. D. (1954) The Market as a Factor in the Localization of Industry in the United States, *Annals of the Association of American Geographers* **44(4)**, 315–348.

HEAD K. and MAYER T. (2004) The empirics of agglomeration and trade, in HENDERSON J.V. and THISSE J.-F. (Eds) *Handbook of regional and urban economics 4*, 4, pp. 2609–2669. North Holland, Amsterdam.

HEAD K. and MAYER T. (2006) Regional wage and employment responses to market potential in the EU, *Regional Science and Urban Economics* **36(5)**, 573–594.

HEAD K. and MAYER T. (2014) Gravity Equations: Workhorse, Toolkit, and Cookbook, in GOPINATH G., HELPMAN E., and ROGOFF K. (Eds) *Handbook of International Economics 4*, pp. 131–195. Elsevier, Amsterdam.

KEEBLE D., OWENS P. L. and THOMPSON C. (1982) Regional accessibility and economic potential in the European Community, *Regional Studies* **16**(**6**), 419–432.

KELEJIAN H. H. and PRUCHA I. R. (1998) A Generalized Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances, *The Journal of Real Estate Finance and Economics* **17**(**1**), 99–121.

KELEJIAN H. H. and PRUCHA I. R. (2010) Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances, *Journal of Econometrics* **157**(**1**), 53–67.

KOSFELD R. and ECKEY H.-F. (2010) Market access, regional price level and wage disparities: the German case, *Jahrbuch für Regionalwissenschaft* **30**(**2**), 105–128.

KRUGMAN P. (1993) First Nature, Second Nature, and Metropolitan Location, *Journal of Regional Science* **33**(**2**), 129–144.

LESAGE J. P. (2014) What Regional Scientists Need to Know about Spatial Econometrics, *The Review of Regional Studies* **44**(**1**), 13–32.

LESAGE J. P. and PACE R. K. (2009) *Introduction to Spatial Econometrics*, CRC Press, Boca Raton.

LÓPEZ-RODRÍGUEZ J., FAÍÑA A. and COSMIN-GABRIEL B. (2011) Economic Remoteness And Wage Disparities In Romania, *Tijdschrift voor economische en sociale geografie* **102**(**5**), 594–606.

LÓPEZ-RODRÍGUEZ J., FAÍÑA A. and LÓPEZ-RODRÍGUEZ J. (2007) Human capital accumulation and geography: Empirical evidence from the European Union, *Regional Studies* **41**(**2**), 217–234.

MION G. (2004) Spatial externalities and empirical analysis: the case of Italy, *Journal of Urban Economics* **56**(**1**), 97–118.

NIEBUHR A. (2006) Market access and regional disparities, *The Annals of Regional Science* **40**(**2**), 313–334.

REDDING S. J. (2011) Economic Geography: A Review of the Theoretical and Empirical Literature, in BERNHOFEN D., FALVEY R., GREENAWAY D., and KREICKEMEIER U. (Eds) *Palgrave Handbook of International Trade*, pp. 497–531. Palgrave Macmillan, London.

REDDING S. J. and SCHOTT P. K. (2003) Distance, skill deepening and development: will peripheral countries ever get rich?, *Journal of Development Economics* **72**(**2**), 515–541.

REDDING S. J. and VENABLES A. J. (2004) Economic geography and international inequality, *Journal of International Economics* **62**(**1**), 53–82.

TOBLER W. (1970) A Computer Movie Simulating Urban Growth in the Detroit Region, *Economic Geography* **46**, 234–240.

NOTES

[1] Breinlich (2006), as well as other authors, finds that using travel times does not alter the results significantly.

[2] GVApc is preferred to wages for theoretical and empirical reasons. Its correlation with nominal remuneration per worker is 0.81 in the sample under study but the data of the latter variable has lower quality.

[3] The words 'global' and 'local' are not always used with this meaning in Spatial Econometrics. A 'global' measure of spatial autocorrelation is that which applies a common weights matrix to the space of observations, even if that matrix is designed to capture (average) local dependence.

[4] Row-standardization guarantees that the maximum eigenvalue of W is 1 and the invertibility of a linear combination of W. It prevents the estimated spatial parameter from implying explosive models with unknown properties. Alternatively, KELEJIAN and PRUCHA (2010) propose dividing each element of W by the spectral radius of W, the maximum absolute value of its eigenvalues. This method is not discussed here.

[5] The heterogeneous size of the observational units (and, therefore, sample selection) is always an issue when modeling space (see Figure 1). That is related to the large literature on the modifiable areal unit problem and is beyond the scope of this paper.

[6] The two outliers at the top of the first two plots of Figure 2 correspond to Inner and Outer London, which have a somewhat arbitrary small distance between centroids.

[7] We appreciate the suggestions of an anonymous referee to study correlations and improve the explanation of Table 2.

[8] On an experimental basis, the SAR model in column 4) of Table 5 was estimated using the spatial lags of the explanatory variables (X, WX, WWX) as instruments for the spatially lagged dependent variable. This last variable becomes not significant and the estimates of External Market Potential and Human Capital get close to the OLS ones in column 4) of Table 3. The additional instrumentation of External Market Potential by mean distances does not change the qualitative results.

[9] We thank Tomasz Mickiewicz for raising this point.