MARKET POTENTIAL AND SPATIAL AUTOCORRELATION IN THE EUROPEAN REGIONS

Andres Faina *

Jesus Lopez-Rodriguez *

Fernando Bruna *

Abstract

The concept of market potential developed by Harris (1954) has several shortcomings but it has proved its practical utility even when compared with more sophisticated versions derived from the wage equation of the New Economic Geography. We stress the ability of this variable to summarize regional markets accessibility and their spatial distribution. We specially highlight a feature of the market potential/market access concept in its foreign or external form (without computing own region's market potential). Namely, External Market Potential is a non-standardized inverse distance spatial lag of the regional internal markets. Considered together, internal and external market potential, a variable of Harris's Market Potential is able to capture the core-periphery pattern of economic activity and per capita Gross Value Added in the European regions. This long-distance spatial dependence is different from the short-distance spatial dependence emphasized by Spatial Econometrics, with its standardized spatial weights for a reduced number of neighbours. But capturing these spatial patterns in a regression with a variable of Market Potential comes at a cost of spatial autocorrelation. We show how the Spatial Econometrics procedures allow a correction of the spatial autocorrelation keeping the interpretation of Market Potential in terms of the accessibility to the markets and as an indicator of peripherality. So, both types of spatial dependences should be considered together.

JEL Code: C33, F12, R12

Andres Faina (fai@udc.es)
Jesus Lopez-Rodriguez (jelopez@udc.es)
Fernando Bruna (f.bruna@udc.es)
University of Coruna
Economics and Business Department
Campus de Elvina s/n
15071 - A Coruna, Spain
Phone number: +34 981167000
Fax number: + 34 981167070

C+D Research Group

April 2013

1 Introduction[1]

Since 1954's seminal article of Chauncy Harris the concept and measures of "market potential" has been widely used in Regional Economics, especially after the theoretical support given to it in the nineties by the New Economic Geography (NEG).

The nominal wage equation of the NEG uses a concept related with the one developed by Harris, with some differences that will be commented bellow. The empirical specifications closer to NEG approach use gravity equations to estimate market potential, in a way certainly different from Harris's measure. Data restrictions force to additional assumptions when working at regional level. For instance, in their estimations for the European regions Breinlich (2006) or Head and Mayer (2006) have to assume that the export behaviour of each region is the same than the export behaviour of its country. However, at the end Breinlich (2006) obtains the same explanatory power using the gravity equation approach than using the simple Harris's approach.

In spite of its theoretical limitations, Harris's approach continues to be useful and two of the authors of this article have used it widely in previous work[2]. Here we focus on the relation of the concept of Market Potential with Spatial Econometrics in a way that has not been sufficiently highlighted in the previous literature. When we distinguish internal and external market potential, called foreign market potential by Brakman et al. (2009b), there is a close relation between the spatial structure of economic activity and the measure of external market potential. Using European regional data we show how a Market Potential captures the long distance spatial patterns while Spatial Econometrics procedures are complementary to correct for spatial autocorrelation. We focus in the similarities and differences between both approaches to spatial dependence. In a parallel work, Blanco (2012) uses a distinction between types of spatial interaction similar to ours, but her concept of surrounding market potential loses the interpretation of the variable in terms of market size.

In section 2 we remind the original approach developed by Harris. Section 3 deals with the role of market potential in the nominal wage equation of the NEG, and with the differences with respect to Harris's measure. In section 4 we comment on some contributions of the Spatial Econometrics literature to test and model spatial dependence. Section 5 sets our

---

[1] Paper presented at the XIII Conference on International Economics (Granada, Spain, June 2012) and at the VI World Conference of the Spatial Econometrics Association (Salvador, Brazil, July 2012). We thank to Adrian Gutierrez, Vanessa Mato and Paulino Montes for helpful research assistance. We thank to the R project community for making available the procedures for spatial data, especially to Bivand (2012). Too, we thank to David G. Rossiter for helpful insights.

[2] See, for instance: Faina and Lopez-Rodriguez (2006a, b and c; 2008); Lopez-Rodriguez (2007); Lopez-Rodriguez and Faina (2004, 2006, 2007); Lopez-Rodriguez, Faina and Lopez-Rodriguez (2007); Lopez-Rodriguez, Faina and Garcia (2007).

empirical framework and data. In section 6 we discuss the spatial structure of our variables and in section 7 we estimate some spatial models. Section 8 concludes.

2 Regional Economics: Harris's Market Potential

The concept of "market potential" (MP), widely used in Regional Economics, was developed by Chauncy Harris (1954). The market potential of a point (region i) is defined as the summation of markets (M) accessible to i divided by their "distances" to that point i. Considering the $N-1$ possible markets of other j regions, Harris's Market Potential is defined as:

$$HMP_i = \sum_{j=1}^{N} \frac{M_j}{d_{ij}}$$

where the distance to the own regional market (d_{ii}) is measured by within region distances[3]. Therefore, for later discussion in section 5, we already distinguish between the Internal Market Potential (IMP) and External Market Potential (EMP) of region i:

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

Harris claims that the term market potential, suggested by Colin Clark, is analogous to that of population potential as proposed by Jonh Q. Stewart (1947). The concept is derived ultimately from physics, as in similar formulas for the strength of a field, whether electrical, magnetic or gravitational. Therefore the indicator is inspired by the gravity equation. In fact, a later paper by Clark has the title "Demographic Gravitation", but there have been applications of Newton's theory of gravitation in "social physics" since the nineteenth century, including those in Economics during the first third of twentieth century.

Harris´s main interest was to develop and map a measure of the market relevant for the location of industries. He collected retail sales data on counties to give the values of M in the previous equation. His measure of the Market Potential for each city "is an abstract index of the intensity of possible contact with markets" for goods:

> *"Market potential appears to gauge the possible spatial interaction between producers and markets, of the likely flow of goods from a point to accessible regions. A number of studies indicate that freight movement as well as many other types of relationships between any two points varies directly with their size and inversely with their distance*

[3] Harris developed his measures for cities (points). As we will see in the empirical part, later literature working with areal data summarizes the data in a point, usually the regional centroid, and evaluates the distance to the own regional market through an adjustment based in the area of the region.

apart. Actually there is a complex hierarchy of distribution areas from any given city; some products may have national or international distribution, others regional, and many local only. The aggregation of these various distribution areas results in a large volume of local and nearby movement with amounts decreasing with distance (...)"

Harris' approach has been widely used in Regional Economics. One reason is that it offers a way of capturing Tobler´s (1970)[4] 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". It is particularly useful as an index of accessibility in core-periphery studies, as in Keeble's report to the European Commission (Keeble et al., 1988). Too it has been very used in planning studies for the location of plants or facilities. And in agglomeration studies it proposes a measure for Smith's idea of the size of the market as a determinant of the division of labour and increasing returns. In fact, increasing returns will be a key feature in the models of Krugman (1991) and Fujita et al. (1999) setting the micro-foundations for Harris's approach. We will review it in next section about the New Economic Geography (NEG).

There are a number of methodological issues related with Harris's measure. Later we will mention some of them. Here we just remark two ideas about Harris's original approach.

First, Harris was interested in firm's locational decisions considering the size of its markets. In order to measure the over-all market for goods (the M values) he used figures for retail sales. His paper has many references to demographic studies, including some about the relationship of population and distance in retail trade. Population has been widely used in later studies but Harris was trying to get a general description of the market potential of each city and he used data on retail sales in a particular year. He did not estimate any equation using time series, so he did not discuss how to deal with changing prices to measure changing market potentials. Given the type of data he used, many year later, the New Economic Geography will call Harris's measure Nominal Market Potential (NMP) to distinguish it from the Real Market Potential (RMP) derived in its wage equation, which includes a price index. We will come back to this below.

A second remark about HMP is related to the measure of "distance". Already in his seminal paper, Harris claims that for his purpose transport cost is superior to sheer miles. He avoids tabulating individual routes assuming that the shortest distances on a map are proportional to actual route miles. But he estimates total transport cost carefully, considering railroad terminal costs, truck delivery costs, water transportation terminal costs and running costs per ton-mile. Given data restrictions, later literature usually proxy trade costs with physical

[4] See the discussion in the 2004 issue of Annals of the Association of American Geographers 94:2.

distances[5]. This assumption means homogeneous trade cost per unit among goods or sectors, observations and distances. It is worthy to make this assumption explicit and to remember Harris's effort to avoid using only physical distances. With multi country estimations it is hard to work with good data about trade costs but the assumption of identical trade costs is strong (Anderson and van Wincoop, 2004). The problem appears too when using trade trade equations to estimate NEG's approach to Market Potential. Bosker and Garretsen (2010) call for a much more careful treatment of trade costs in future empirical NEG studies, because of the specification of trade costs can matter a lot for the conclusions reached.

We do not focus here in this issue[6]. After this call to attention, later we will use physical distances to proxy trade costs with our European regional data. A support for this assumption is given by Ahlfeldt and Feddersen (2008) sensitivity analysis of the wage equation, which cannot reject that the application of straight-line distances and road travel times yield the same results[7].

3 New Economic Geography: wage equation and Real Market Potential

The concept of market potential has been received a strong theoretical foundation within the models developed in the New Economic Geography (NEG) literature (Krugman, 1991; Fujita et al., 1999). Krugman's general equilibrium setting provides microeconomic foundations to the physical analogies of Harris's Market Potential function[8].

The most estimated prediction of the NEG is the wage equation. Different versions the theoretical framework has been shown many times[9] and here we just write the final equation, following Brakman et al. (2009b) and Combes et al. (2008).

The NEG's wage equation explains the equilibrium industrial nominal wages of each region i (W_i) as a function of the sum, for all the j regions to which industrial goods are exported[10], of

[5] See Boulhol et al. (2008) for the construction of an aggregate index of transportation costs for an OECD sample.

[6] Too there is a promising research line on economic, institutional or social distances. In our context these distances would be related with trade barriers but they can be barriers for knowledge or political spillovers too. See, among others, Beck et al. (2006), Arbia et al. (2007), Linders et al. (2008) or Rodriguez-Pose (2011).

[7] Too Breinlich (2006) obtains similar results using travel times and geographical distances.

[8] See Brakman et al. (2009a) or Combes et al. (2008). Redding (2011) offers a survey with emphasis in the empirics of the NEG.

[9] See, for instance: Head and Mayer (2011); Lopez-Rodriguez et al. (2007); Breinlich, (2006); Hanson (2005); Head and Mayer (2004); Redding and Venables (2004); Redding and Schott (2003).

the product of two elements. On one hand, their respective volume of demand to region i ($\mu_j E_j$, being E_i the expenditure and μ_j the share of income spent on manufacturing goods in region j), weighed respectively by their prices index (P_j) properly adjusted. On the other hand, it is the transport costs from region i to the destiny j (T_{ij}), to the power of one minus the elasticity of substitution among the varieties of industrial goods (σ) or range of product differentiation:

$$W_i = \left[\sum_{j=1}^{R} \mu_j E_j P_j^{\sigma-1} T_{ij}^{1-\sigma}\right]^{1/\sigma} = [RMP_i]^{1/\sigma}$$

The term RMP stands for the real market potential (RMP). In order to maintain continuity with prior work (from Harris, 1954, to Fujita et al., 1999), Head and Mayer (2006) employ the term Real Market Potential[11]. The "real" is added in order to contrast it with an alternative formulation that the authors refer to as "Nominal Market Potential" or $NMP_i = \sum_{j=1}^{R} \mu_j E_j T_{ij}^{1-\sigma}$. The "nominal" refers to the absence of an adjustment for variation in the price index term P_j. The authors conclude that under certain assumptions and using aggregate measures of demand such as GDP or retail sales, the NMP is proportional to the original formulation of market potential used by Harris (1954) and in subsequent work of geographers. We comment below on this terminology.

As summarized by Combes et al. (2008, page 305), to derive the expression proposed by Harris from the RMP defined above, we must make three additional assumptions:

1) The term of trade costs becomes $T_{ij}^{1-\sigma} = d_{ij}{}^{-\lambda}$ and $\lambda = 1$. This last assumption can be justified by estimations of gravity equations given values of λ close to 1.
2) The share of income spent on manufacturing goods (μ_j) is the same across regions. Regarding the expenditure on intermediate goods, this implies the strong assumptions that either all sectors use the same amount of each factor, or regional sectorial compositions are the same.

---

[10] In the nominal wage equation the whole set of regions is named R while in the HMP function we have used the notation N. With this we try to differentiate that the equilibrium wage equation is given by trade, but in Harris formulation the channel of influence is the potential demand affecting firm's location decision. So R and N could not be the same. This is related with the sample problems of omitting zero trade relationships when estimating gravity equations (Baldwin and Harrigan, 2011). However, we will not address more the possible differences of R and N.

[11] Redding and Venables (2004) use the expression "Market Access" (MA) instead of Market Potential and distinguish the "market access" of the exporting region (E_j as income in consumer markets) from the "supply access of the importing region (E_j as expenditure in intermediate inputs). They find that at the aggregate level, both measures are highly correlated, so the distinction would be more useful when using sectorial data. See Boulhol et al. (2008) for a similar result.

3) The price index (P_j), present in RMP, disappears in Harris formulation. An increase in the number of competitors located in a given destination fragments demand, which in turns implies a decrease in the corresponding RMP. Harris's market potential does not take into account this effect to explain a particular site's profitability.

Therefore, Harris's market potential is at best a rough approximation of the NEG´s RMP. We have mentioned Breinlich's (2006) result of a similar explanatory power estimating the wage equation with Harris's Market Potential (HMP) or with a model-derived measure of RMP. Given that we and other authors use Harris's formulation for practical reasons, it is worthy to extend the discussion about these three assumptions to go from RMP to HMP:

1) It is said that when using HMP instead of RMP we are assuming $T_{ij}^{1-\sigma} = d_{ij}^{-\lambda} = d_{ij}^{-1}$. This could be translated as trade costs as a constant proportion of distance for all regions and a elasticity of substitution among varieties $\sigma = 2$. There is a long tradition in several fields using the inverse of physical distances, and Harris himself analyses this argument for his topic, but he chooses to measure "distance" with trade costs. We have discussed this in section 2, so now we keep the physical distance approximation to emphasize the assumption of a distance decay parameter $\lambda = 1$. The argument in Head and Mayer (2004) or Combes et al. (2008) about findings in the literature of gravity equations of a value of $\lambda = 1$ is under debate. As already noted, Bosker and Garretsen (2010) remark how the specification of trade costs changes the results. For instance, an exponential specification, such as $eMP_i = \sum_{j=1}^{R} M_j e^{-\lambda d_{ij}}$, produces a strong distance decay (see Hanson, 2005, or Brakman et al., 2009b, page 212). This has been criticized on the basis of argument raised in the transport economics literature. But comparing cross sections analysed in different studies Niebuhr (2006) provides an alternative explanation to the one about functional form. The pronounced differences in the geographic extent of demand linkages could be a consequence of the regional system under consideration, in particular of the average size of the observational units. The size of the estimated distance decay seems to increase with the declining size of regions. This might imply that the nature of spatial effects detected in studies working at different NUTS level with EU regional samples would not be the same. This is a fresh field of study and we focus on these issues in a different work so here we will continue the tradition of assuming $\lambda = 1$ in order to get a broad picture of the access of each region to the European markets.

2) To go from the direct demand effect of exports on the wages of regions i to the Harris's function of market potential we need the strong assumption of equal share of income spent on manufacturing across regions. Normalizing $\mu_j = \mu = 1$, we have: $RMP_i = \sum_{j=1}^{R} E_j P_j^{\sigma-1} T_{ij}^{1-\sigma}$. Therefore, in the NEG's wage equation, nominal wages of region i become a function of its "potential" (total) demand (expenditure) from all regions. We have here the issue of the sectorial composition of each region (Karahasan and Lopez-Bazo, 2011) and its relation with what it was called above "supply market potential". Besides, we

have the familiar issue in trade literature with non-tradable goods, and its effects on the purchasing power parity of each currency. We do not study here these issues. Again, the use of a Harris's-type market potential approach can provide useful insights, but we have to be careful in order to interpret them in terms of NEG's nominal wage equation.

3) The absence of prices in Harris's formulation has been considered as its main difference with the "real" market potential in the New Economic Geography. But the term "Nominal Market Potential" (NMP) to describe Harris's formulation could be misleading. As we emphasized in section 2, Harris used a nominal variable, retail sizes, to measure the potential size of the each regional market. But he used data on just one year; he did not work on how changes of prices would affect firm's location decisions. Contrary, when using regional data, price indices are typically not available. For instance, working with Cambridge Econometrics data, as we do, the literature uses data on regional real GDP or regional Gross Value Added. It is frequent to test NEG's wage equation using real per capita income instead of nominal wages and to measure HRM with real income (Reading and Venables, 2002; Brakman et al., 2009). In this sense, we distinguish Harris's market potential as:

$$HMP_i = \sum_{j=1}^{R} \frac{M_j}{d_{ij}} = \sum_{j=1}^{R} \frac{E_j/P_j}{d_{ij}} = \sum_{j=1}^{R} E_j\, P_j^{-1} d_{ij}^{-1}$$

where the price index makes a clear difference from the nominal market potential (with $\mu_s = 1$): $NMP_i = \sum_{j=1}^{R} E_j T_{ij}^{1-\sigma}$

When estimating for a single year, as Harris does, the price indexes would refer to that year so HMP_i becomes exactly the same than the one proposed by Harris, deflating does not alter nominal values in the base year. But when using time series with regional data on real income or value added (at the base year prices), we are working with HMP_i, not with NMP_i. Comparing HMP_i with its equivalent in the NEG:

$$RMP_i = \sum_{j=1}^{R} E_j P_j^{\sigma-1} T_{ij}^{1-\sigma}$$

in HMP_i we are making the assumption $P_j^{\sigma-1} = P_j^{\ -1}$ so it is assumed that the constant elasticity of substitution among industrial goods is $\sigma = 2$. With this last assumption, too we are assuming $T_{ij}^{-1} = d_{ij}^{-1}$.

We make three additional remarks to finish this section:

4) The strength of NEG is that it is a general equilibrium setting. So the price indexes in the nominal wage equation are endogenous. In fact, location is endogenous, as Harris intuited, ruled by increasing returns in what it is called "the home market effect" (see the estimation of Davis and Weinstein, 2003, or Head and Mayer, 2006). The literature shows how a region's

RMP depends on the location choices of agents (migration, FDI...). We do not address here the issue of the endogeneity of RMP but focus on the preeminent role that literature gives to Market Potential as an explanatory variable. Again we note that taking as given the regional market potential is a poor test of the NEG, particularly when working with long time series.

5) Another issue is the election of the endogenous variable to estimate the wage equation. As we said before, a number of works use real GDP per capita instead of compensation per employee. We will comment on this in section 5 about our empirical framework.

6) A final point is about the explanatory variables in a wage equation in terms of the NEG´s RMP or in terms of HMP. On one hand, there are other possible determinants of wages. For instance, human capital density (Breinlich, 2006; Head and Mayer, 2006). Again the problem of endogeneity appears (Hering and Poncet, 2009; Lopez-Rodriguez et al., 2007). Too, it would be necessary to control for other factors correlated with Market Potential as the sectorial composition of employment (Karahasan and Lopez-Bazo, 2011) or non-pecuniary externalities. There is a large literature about the estimation of border effects[12]. In this sense, our focus is to compare the approach to spatial dependence taken by the Market Potential tradition and by Spatial Econometrics. In section 8 we will control for human capital or innovation capacity with a variable of human resources in science and technology and we will introduce fixed effects in a panel estimation to control for omitted regional variables. But our main point is about spatial dependence not about the magnitude of the estimated elasticity of Market Potential.

With all the previous reservations, our main theoretical equation based in NEG's wage equation is:

$$\ln W_i = \frac{1}{\sigma} \ln RMP_i$$

And considering the omission of other possible determinants of the nominal wages, and the assumptions needed to proxy RMP_i by HMP_i we estimate (see a full derivation in Brakman et al., 2009b, including a theoretical justification for the constant parameter β_0):

$$\ln W_i = \beta_0 + \beta_0 \ln HMP_i + \epsilon_i$$

In the rest of the paper we stress how the concept of Market Potential is related with the spatial structure of the economic activity and, therefore, with the analysis of spatial dependence in regression models.

[12] See, for instance, Niebuhr (2006), Huber et al. (2011) or Head and Mayer (2011).

4 Spatial Econometrics and spatial dependence

The spatial structure of the economic space, as studied in the Harris-NEG tradition is closed linked to the concept of spatial dependence studied by Spatial Econometrics. The term Spatial Econometrics was used by first time by Jean Paelinck (1967) and was consolidated by Paelinck and Klaassen (1979). Some major contributions have been the books by Cliff and Ord (1981), Anselin (1988) and LeSage and Pace (2009). It was not until the last years of the past century that mainstream econometrics expressed a growing interest in spatial statistical methods but the integration of spatial methods with econometrics nevertheless remains in an early phase (Arbia, 2006). Therefore, Spatial Econometrics is a young discipline which has been developed many years later than Harris's contribution. The relation between both has not been properly stressed.

Spatial Econometrics deals with spatial interactions (autocorrelation) and with the spatial structure (heterogeneity) in regression models. We focus here in spatial autocorrelation. A variable is spatially autocorrelated when its values at different observational units are correlated with those of their neighbours (clustering). Violation of the independence of observations makes OLS estimates of parameters and variance of perturbation inefficient and inconsistent even if still unbiased. The sampling variances are biased and in most cases significantly underestimated. As a consequence the coefficient of determination as well as the test statistics t and F tend to be inflated leading to accept the model more frequently than it should (Arbia, 2006, quoting Maddala). As we will mention again in section 7, when the 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 spatial dependence in the disturbances, if present, will only cause a loss of efficiency (LeSage and Pace, 2009, page 156).

Durbin (2008) claims that while autocorrelation in a time series context is well understood, and researchers routinely test and correct for this problem, the same cannot be said of autocorrelation in a cross-sectional context. As Getis (2007) says, the special effort to discern spatial autocorrelation makes identifying temporal autocorrelation a relatively simple job compared to the same task for spatial autocorrelation. The multidimensional nature of spatial autocorrelation, i.e., the need to search in all directions as opposed to the one-way temporal direction, is one complicating factor.

That is why in order to test and model spatial autocorrelation it is necessary to choose a neighbourhood criterion (who is linked with who) and to create a spatial weights matrix (W) assigning weights to the areas that are linked. It is the use of this matrix what links directly Spatial Econometrics with the concept of (External) Market Potential and its matrix of inverse distances, as we see below.

LeSage and Pace (2009, pages 25-33) provide five separate motivations for regression models with spatial effects:

- A time-dependent motivation: a cross-sectional autoregresive model relation can arise from time-dependence of decisions by economic agents located at the various points in space when decisions depend on those of neighbours.
- An omitted variables motivation: unobservable factors influencing the dependent variable can exhibit spatial dependence (ex.: highway accessibility, political boundaries…).
- Spatial heterogeneity motivation: with cross-sectional data it is not possible to estimate separate intercepts for each region. If observational units in close proximity exhibit effects levels that are similar to those from neighbouring units, spatial heterogeneity can motivate spatial error dependence.
- An externalities-based motivation: the spatial average of neighbouring characteristics could play a role in determining the endogenous variable.
- A model uncertainty motivation: spatial effects of the dependent and explanatory variables can be justified when the specific character of spatial dependence in the underlying data generation process is uncertain.

Taken together all this five motivations, Spatial Econometrics proposes a variety of models to capture spatial interaction. Our purpose here is not to estimate sophisticated spatial models but to study how Harris's Market Potential captures spatial interactions when compared with the most common approach taken by Spatial Econometrics. In order to do this, we focus here in just two simple spatial models.

We mention here two of them. The Spatial Lag Model, also known as Spatial Autoregressive or SAR Model, captures an endogenous interaction effect (Elhorst, 2011):

$$Y = \rho WY + X\beta + u$$

An extension of it, the Spatial Durbin Model can be motivated by the omitted variables argument. In section 7 we will mention it again. This model, defended by LeSage and Pace (2009) and Elhorst (2010), includes exogenous interactions effects ($WX\theta$):

$$Y = \rho WY + X\beta + WX\theta + u$$

The second basic spatial model is the Spatial Error Model (SEM)[13], which captures interaction effects among the error terms because omitted variables or spatially autocorrelated unobserved shocks:

[13] The literature uses a confusing mix of abbreviations. The model named SEM here is named SAR (Simultaneous Autoregresive) by Bivand et al. (2008) and Bivand (2012), while those authors reserve the name of Spatial Lag Model for what here it was called SAR. Sometimes the SAR abbreviation is used for the whole family of autoregresive models: spatial error = SAR_{err}, lagged = SAR_{lag} and mixed = $SAR_{mix.}$

$$Y = X\beta + u$$

$$u = \lambda W u + \varepsilon$$

There are many ways of selecting W for describing an unknown structure of global spatial interaction, if such a global structure exists[14]. Generally the weights w_{ij} are (row) standardized with the sum of all weights for region i. This allows creating proportional weights when the chosen criterion of neighbourhood implies unequal number of neighbours. Row-standardization allows dealing with data which has been aggregated, as our regional data is. Additionally, row standardization is useful to compare spatial parameters across different data sets with different connectivity structures. Lesage and Pace (2009) argue in favour of row-standarization. But too it is a common practice because after standardization the spatial lag of a variable X for each region can be interpreted as the weighted average of X for its "neighbours" (however defined): $(WX)_i = \sum_j \frac{w_{ij}}{\sum_j w_{ij}} X_j = \overline{wX}_i$.

When the weights are inverse distances, standardization might correct some biases (edge effects, distance between centroids in regions with different area). But, as we will insist in the empirical part, standardization emphasizes what region are neighbours, not at what distance they are. Contrary, External Market Potential is an index of centrality or peripherality. If we represent the size of the market through Gross Value Added, External Market Potential is a non-standardized spatial lag of the internal market (GVA) of the regions:

$$EMP_i = \sum_j w_{ij} GVA_j = \sum_j \frac{1}{d_{ij}} GVA_j$$

The focus of our work is to show how the variable of Market Potential captures a different type of spatial dependence from the spatial models. We are not conscious of a similar focus in the previous literature. However, Blanco (2012) makes a similar distinction. In her model of the determinants of FDI in Latin America, she distinguishes two forms of spatial interdependence of FDI: (1) surrounding market potential, and (2) spatial autocorrelation. Surrounding market potential refers to the impact that economic conditions in other countries have on the amount of capital inflows that a specific country receives. She measures this variable with the sum of all inverse distance weighted GDPs per capita (in constant US dollars) of all countries in the sample. Spatial autocorrelation refers to the impact that capital inflows in other countries have on capital inflows in a specific country, measured through a standardized W matrix. Though the idea is similar to ours, her definition of surrounding market potential loses the meaning of market size present in Harris proposal and in the NEG wage equation.

[14] See Harris et al. (2011) or Mur et al. (2011) for revisions of some of the methods for selecting W.

Some studies apply Spatial Econometrics techniques to study the NEG equation (Fingleton, 2009; Kosfeld and Eckey, 2010). But spatial models are not just used to correct for spatial autocorrelation but to estimate externalities too (Anselin, 2004). Certainly a measure of market potential might be collecting a spatial structure of the data induced by multiple generating processes: spillovers (non-pecuniary externalities) or other factors correlated with market potential. This adds to the need of being cautious when interpreting the results of the estimation in terms of the NEG's core model. Too it means that there are opportunities to enrich our understanding of different channels of spatial influence. For instance, Ertur and Koch (2007) build and estimate a neoclassical growth model with human and physical capital externalities and technological interdependence among countries[15]. However, empirically it is not easy to disentangle the channels of spatial influence (Tappeiner et al., 2008. See Chasco et al., 2012, for some results). The NEG wage equation offers a solid framework to interpret some spatial effects which can be complemented with other theoretical or empirical approaches. In any case, the key point of the empirical sections bellow will be to show the ability of a Market Potential variable, grounded in the literature of Economic Geography, to capture spatial structures in a different way than using a standardized spatial weight matrix.

We finish this section summarizing some ideas about the evaluation of spatial autocorrelation. There are a number of test for spatial autocorrelation. In spite of its shortcomings (Arbia, 2006, page 90), we focus here in the simplest and most commonly used test statistics in the Spatial Econometrics literature: Moran's I. Moran's I compares the value of a variable Y at any location with the value at its W neighbors. It is calculated as a ratio of the product of the variable of interest and its spatial lag, with the cross-product of the variable of interest, and adjusted for the spatial weights used:

$$I = \frac{n}{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \ (y_i - \bar{y}) \ (y_i - \bar{y})}{\sum_{i=1}^{n} \ (y_i - \bar{y})^{2}}$$

where y_i is the ith observation, $\bar{y}$ is the mean of the variable of interest, and w_{ij} is the spatial weight of the link between i and j. Too the Moran's I for the variable Y can be interpreted as the slope of the regression: $WY = \alpha + YI + u$ (Anselin, 1996). This is the base for the Moran Scatterplots we will draw later. When Moran's I has a p-value higher than 0.05, the null hypothesis of no-spatial autocorrelation for the chosen W is accepted.

Centering on the mean is equivalent to asserting that the correct model has a constant mean, and that any remaining patterning after centering is caused by the spatial relationships encoded in the spatial weights. Bivand et al. (2008, page 60) show that even a gentle global regional trend induces apparent spatial autocorrelation in the Moran's indicator, unmasked

[15] See Basile and Usai's (2012) survey about regional endogenous growth.

when a correct model is fitted. That is one of the reasons why we later estimate the global geographical trends of our variables.

It is important to note that this is a measure of global spatial autocorrelation. Global tests for spatial autocorrelation assume that the spatial process is the same everywhere. Though this is a strong assumption, local tests are problematic too. They are conditional by the possible influence of spatial data generating processes at different range of scales (Bivand et al., 2008, page 270). They depend one on another; they are subject to spatial autocorrelation itself because many observations used to test one neighbour will be used to test its neighbours (Getis, 2010), though there are procedures to deal with this last issue. However, the focus of our work is in global spatial structures so we do not make any attempt here of calculating Local Indicators of Spatial Association (LISA). We will emphasize that Market Potential can capture the long-distance global spatial pattern of the European economic activity.

Finally, it is necessary to remind that the Moran's I of the regression residuals is sensitive to other factors, apart from a global trend. Moran's I can detect misspecification instead of a true spatial interaction (autocorrelation) in the data generating process. For instance, multicollinearity or heterokedasticity can make Moran's I to detect misspecification (Bivand et al., 2008, pages 276 and 281). The obvious reason for misspecification is the omission of relevant explanatory variables. This will be the case in the econometric exercise we carry on section 7. Therefore, we stress that the goal of that exercise will be to check if the way in which spatial dependence is captured by an explanatory variable of Market Potential is compatible with a correction for spatial autocorrelation using Spatial Econometrics models.

5 Data and empirical framework

We estimate a wage equation type of relation using Harris's Market Potential as the main explanatory variable. We use Cambridge Econometrics regional data at NUTS level 2 for the EU countries plus Norway and Switzerland. We have 275 regions after excluding Cyprus[16], the Outermost French regions and the Spanish Ceuta and Melilla, in the African Coast. We mainly consider here this big sample of 275 regions to show the spatial structure of the variables, though the concept of market potential should be better applied to areas with similar degree of market integration. We focus the discussion about spatial interactions using

[16] Cyprus joined the European Union in 2004. Apart from this broad sample, later we will use a reduced sample excluding the countries from the 2004 and 2007 EU enlargements. Given that we never consider Cyprus, the countries excluded in the reduced sample are Malta and the East European countries. On the other hand, we keep Greece in both the broad and the reduced sample, in spite of its particular geographic and economic characteristics (Bivand and Brunstad, 2006).

cross-sectional data for one year (2008) but we confirm it using a balanced panel from 1991 to 2008. In some regressions we additionally use Eurostat data on human resources in science and technology (S&T), which covers less years and does not allow for a balanced panel with the same sample.

In order to measure the size of the markets we use Gross Value Added (GVA), given that the variable GVA is the basis for the regional accounts[17]. But we find not significant differences when Gross Domestic Product (GDP) is used, what is confirmed by Ahlfeldt and Feddersen (2008). As it was discussed in section 3 we use income data in real terms. Cambridge Econometrics GDP or GVA are given at year 2000 prices. Their deflators are regional in the sense that use the sectorial deflators published in the Annual Macro-economic Database of the European Commission (AMECO), so deflators vary according to the size of the respective sectors in a given region. Then these estimates of real regional GVA are scaled to the national estimates[18]. However, Cambridge Econometrics does not provide their deflators.

We use real per capita GVA as our endogenous variable. NEG's wage equation is referred to nominal wages, and we could use Cambridge Econometrics nominal compensation per employee instead. There are several reasons for choosing real GVA. First, real per capita income is more comparable with other studies proxying nominal wages with per capita income because of lack of data on wages (Reading and Venables, 2002; Brakman et al., 2009b, among others). Second, real income could be a more robust variable than nominal compensation, especially when East European high inflation countries are included. Third, real per capita GVA is related with real GVA per worker which has been used as a proxy of productivity in European regional studies at NUTS 2 level (Le Gallo and Kamarianakis, 2011; Vieira et al., 2011). This allows some comparability too. GVA per worker could be a better proxy for wages than per capita GVA, but the latter variable is more general and we focus in a general description of the spatial structure of the European economic activity. Finally, our tests using nominal compensation per worker produce similar qualitative results but the regressions have lower explanatory power.

We reduce the areal data to points so each region is represented by its centroid, the geographical centre of the region. We use great circle distances to measure market potential. Geographical distances among NUTS 2 centroids create some distortions. For instance, the small area of UK NUTS 2 regions makes the somewhat arbitrary great circle distance among Inner and Outer London centroids to be very small too. Their inverse distance External Market Potential is very affected by this, so those two regions are outliers in all later analysis. We study aspects related with the regional areas in a different paper so here we just show the data as it is. But following we mention some practical problems.

---

[17] See Breinlich (2006) for other arguments in favour of using GVA instead of GDP.

[18] We thank helpful clarifications of Jon Stenning, from Cambridge Econometrics.

An important practical issue to calculate Harris's Market Potential is how to measure internal distances to calculate the internal market component. It is usually assumed that regions are circular so their radius is: $r = \sqrt{area/\pi}$. Stewart (1947) had assumed that the self-potential of a uniform circular disc at its centre is equal to the mass divided by half the radius. And Rich (1980) stresses that the more concentrated is the mass around the zone's centroid, the greater is the self-potential created, so internal distances could be taken to be less than half of the zone's radius. Keeble et al. (1982) 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". But Head and Mayer-Thisse (2000)[19] provide a different argument: If all production concentrates in the centre of a disk and consumers are randomly distributed throughout the rest of the area, the average distance between a producer and a consumer is $d_{ii} = 2/3 \cdot r_i = 0.376\sqrt{area_i}$. Too this was one of the criteria used by Redding and Venables (2004) and it is becoming the standard way of measuring internal distances.

For internal distances in Internal Market Potential (IMP), we use the approach of Keeble et al. (1982), with $d_{ii} = 1/3 \cdot r_i$ instead of the now most common $d_{ii} = 2/3 \cdot r_i$. Keeble et al.'s (1982) measure, similar to the 40% of the radius used by Cambridge Econometrics to calculate internal distances[20], improves the significance of market potential in our regressions. The reason is that taking $1/3$ of the radius as internal distance gives more weight to the domestic GVA in HMP, so more regional specific GVA is introduced into the regression. However for the broad European sample the main component of HMP is External Market Potential (EMP). In the year 2008, for the average region, IMP is a 14% of HMP when $1/3$ of the radius is taken for internal distances, while it is just a 7.5% when $2/3$ is used. But this is calculation for the average region.

Our estimations of spatial models in section 7 will reveal that the consideration of the internal markets in the measure of Market Potential is crucial. The theoretical argument in favour of $d_{ii} = 2/3 \cdot r_i$ is based in a uniform distribution of consumers around the centre of a circular area. Given the wide use of this criterion is convenient to remind that regions are not circular but, more important, that at least in the coastal regions of the European Union, we do not expect the population to be uniformly distributed around the centre of each region. With the $1/3$ criterion, more concentration around the centre would be assumed, but our argument is empirical, it is not based in a theoretical distribution of the population.

---

[19] Head and Mayer (2000) find that the measure of internal distances affects the estimation of border effects.

[20] http://www.camecon.com/Europe/Regional_Local_Cities/KnowledgeBase/Appendices/EuropeanEconomicModel/ModelOverview.aspx

The role of the internal market for each region is going to depends on many factors, among them the size of the region. This is related not just with the NUTS level of the data but with the average area or population of that NUTS level for each country. The discussion exceeds the scope of the present research and we treated more in detail in a different work. However we note that Boulhol et al. (2008) calculate a higher total Market Potential for Canada than for the United States given their specific capital-to-capital measure of distance, which is a similar measure than the distance among centroids. This feature disappears when the distance measure takes into account not only the capital but also the biggest cities in each country, which it is harder with our regional European data. Additionally, their calculations show that the domestic (Internal) Market Potential for the United States is only 30% greater than the one for the Netherlands, even though its GDP is 20 times bigger. This is because the internal distance of the United States is 15 times bigger and what matters is the domestic market size relative to the internal distance. This affects the levels of Market Potential calling for estimation with panel data and individual fixed effects, as we will do at the end of section 7. In any case, and apart from a more careful treatment in a different work, we belief that all these possible distortions do not affect the big picture we are going to show about the spatial structure of our data.

Separating Internal Market Potential from External Market Potential is equivalent to regress per capita GVA on a kind of GVA density. In this case, we have GVA in both sides of the equation and EMP loses its significance in a cross-sectional OLS estimation. But considering alone EMP does not properly collect the accessibility to markets, including the internal one. Excluding the own regional market in HMP introduces measurement error by reducing the access measure of some economically larger locations (Breinlich, 2006; Head and Mayer, 2006), as the capital cities tend to be. But including it increases the endogeneity problem of having GVA in both sides of the equation. A Harris's type measure of market potential takes a compromise solution of giving some weight to the own market, and we choose the option of Keeble et al.'s (1982) for internal distances. However, given that the focus now is the spatial structure, bellow we compare some results using EMP alone instead of HMP.

The problem of endogeneity is intrinsic when dealing with spatial interactions. The direct effects of an exogenous variable in a region are measured through the estimated parameters of the variable, while for the indirect effects it is necessary to consider how the variable affects the endogenous variable of the neighbour regions and this in turns affects the endogenous variable of the region through the spatial interactions. Apart from the internal market issue, an explanatory variable such as HMP has a more general problem of endogeneity, given that the GVA market size of a region increases when the per capita GVA of that regions benefits from the GVA market size of its neighbours, so the market size of neighbours is reciprocally explained (transmission of exogenous shocks). This is one reason for expecting high spatial autocorrelation in a HMP measure.

We are conscious of all these endogeneity problems. We do not study them more in detail here for two reasons. First, exogenous instruments for market potential are problematic, at least for a European sample, given its strong structure of core-periphery, as we will see. Instruments usually are the distance to Luxembourg (Breinlich, 2006) or the distance to Brussels (Brakman et al., 2009b). This tries to be an exogenous instrument because of its geographical nature. However, given that the economic centre of Europe matches its geographical centre, around the so called "blue banana" or "hot banana", this is not just a geographical measure but a measure of the distance to the economic centre of Europe. Considering the sum of distances does not solve the problem. With a measure based in sums, as pointed by a referee to Head and Mayer (2006), the restriction to European regions implicitly continue to determine a centre. In Europe this centre tends to be around the blue banana too. In fact, we use the mean distance of each region to the other regions in the sample to show the core-periphery spatial pattern of our variables.

Head and Mayer (2006) build an exogenous measure of global centrality, distance to the centre of every inhabited 1° by 1° cell in the world population grid. Bouhol et al. (2008) consider a similar less sophisticated measure of global centrality for the OECD countries. This is out of the scope of our present research, what take us to the second and more important reason to avoid the analysis of the endogeneity problem now. As we said at the end of section 3, here we are not that interested in studying the size of the estimated elasticities, but in showing how different academic traditions approach the spatial structure, and how their procedures behave in a context of economic core-periphery. At this point we want to emphasize this last spatial pattern more than control for it.

To conclude this empirical framework, let´s examine our baseline regression based in the wage equation of the New Economic Geography. The endogenous variable (Y) is per capita Gross Value Added ($\frac{GVA}{H}$) and the explanatory variable (X) is Harris's GVA market potential. Separating the latter into their internal and external components and writing down EMP as a spatial lag of the internal GVA, we have:

$$log\left(\frac{GVA}{H}\right) = \left[log\left(\frac{GVA}{d_i} + WGVA\right)\right]\beta + u$$

Compare this equation with a model with exogenous interactions effects ($WX\theta$) when $X = GVA$[21]:

$$Y = X\beta + WX\theta + u \Rightarrow \frac{GVA}{H} = GVA\beta + WGVA\theta + u$$

Or with the Spatial Durbin Model when $X = GVA$:

---

[21] In the next equations we omit the specification in logarithms of the previous baseline equation in order to remark the similarities and differences in the role of the spatial lag of the variables. Too this is useful to introduce the "lag of log" or "log of lag" issue that we will discuss later.

$$Y = \rho WY + X\beta + WX\theta + u \Rightarrow \frac{GVA}{H} = \rho W \frac{GVA}{H} + GVA\beta + WGVA\theta + u$$

With these brief comparisons we just want to show how the model with Market Potential collects spatial interactions in a different but similar way to some Spatial Econometrics models, but under the framework of Economic Geography and with a particular weight matrix. Following we explain this idea more in detail using a cross-section of European regions. Given that we are conscious of the omission of other determinants of per capita GVA, including those common features of the regions in the same country, in section 7 we control for individual fixed effects using panel data.

Apart from the weights used in External Market Potential, our baseline weights matrix for Spatial Econometrics techniques is a standardized inverse distance matrix of the five nearest neighbours, following Griffith's (1996) rules-of-thumb for the specification of W.

6 Spatial structure and autocorrelation in the European regions

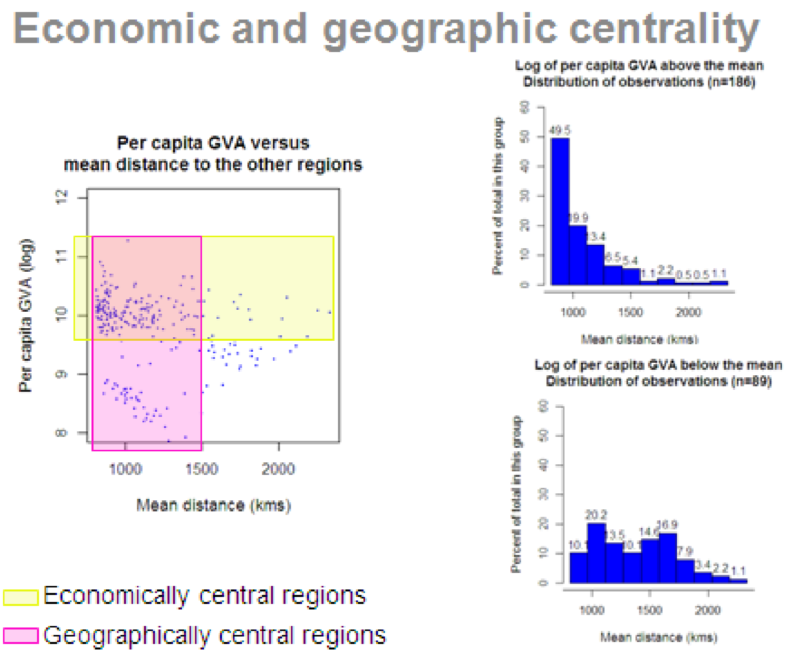
In order to fix ideas, we initially approach the spatial structure of our endogenous variable, the logarithm of per capita Gross Value Added, from the point of view of Geography, using the mean distance of each region to the other regions considered in our study.

The following figure shows the relation between economic and geographical centrality in the European (NUTS 2) regions using data of the year 2008. The left plot shows the distribution of the logarithm of per capita GVA on the mean distance from each region to the rest of the regions in our broad sample. We can see that the majority of high per capita GVA regions are geographically central. The bottom left cluster of points contains the geographically central but poorer regions. They are mainly the East European regions.

We split this broad sample in two groups based in having a logarithm of per capita GVA above ("rich" regions) over below ("poor" regions) the sample mean. The right histograms in the figure show the percentage of observations in each group at each decile of mean distance. The economically central regions, with high per capita GVA, tend to have a low mean distance, i.e., tend to be at the geographical centre of the sample.

The following table tests this hypothesis[22]. Given that the East European regions have specific characteristics, we repeat the test for both the broad European Union sample and the sample excluding the regions of the countries joining the EU in its 2004 and 2007 enlargements, which mainly are the East European countries. When these regions are excluded, the mean logarithm of per capita GVA increases and a higher number of regions is below this mean.

[22] We thank to Xose Manuel Martinez Filguera for some suggestions.



A Welch t-test of differences of means for each group is useful to show the average mean distance for rich and poor regions. In both samples, a p-value of zero allows to reject the null hypothesis of equal distances in favour of the alternative hypothesis: richer regions have lower mean distances to the rest of the regions. But the t-test assumes normal populations. To avoid this assumption we use a nonparametric test, the Wilcoxon test (Mann-Whitney *U* test) to arrive to identical conclusion[23].

Comparing mean distance to the other regions among regions with log of per capita GVA above or below the sample mean					
Samples	**Groups by log of per capita GVA**		**Welch t-test**		**Wilcoxon rank sum test**
	Groups	**Number of observations (n)**	**Average mean distance (kms)**	**p-value**	**p-value**
Broad sample	Above the mean	186	1070	0	0
	Below the mean	89	1374		
	All	275	1168		
Excluded countries of the last EU enlargements	Above the mean	116	1017	0	0
	Below the mean	104	1230		
	All	220	1118		
Alternative hypothesis: The mean distance of the regions with log of per capita GVA below its sample mean tend to be higher than the mean distance of the richer regions.					

[23] These tests assume that the two groups are independent. Spatial autocorrelation might violate this hypothesis. Later we study the spatial autocorrelation of the logarithm of per capita GVA. Now the goal is to provide a first description of the distribution of this variable in the geographical space.

Let’s characterize more this relation between economic and geographical centrality. We start showing a quintile map of our main variables in the year 2008, in which darker colours are associated with higher values of the variables. In these choropleth maps the number of quantiles and the selected sample determine the cut-off values. Therefore, the spatial patterns more economically meaningful are not always clear in these maps. Nevertheless, the maps are enough two distinguish the two spatial patters in which we are going to focus. First, the European distribution of the logarithm of per capita GVA follows a core-periphery pattern, with just a few high per capita income regions out of the geographical centre of Europe, especially those in Nordic countries. Second, given the spatial structure of GVA in Europe, the logarithm of GVA Market Potential shows an even more concentrated distribution and a clearer core-periphery pattern.

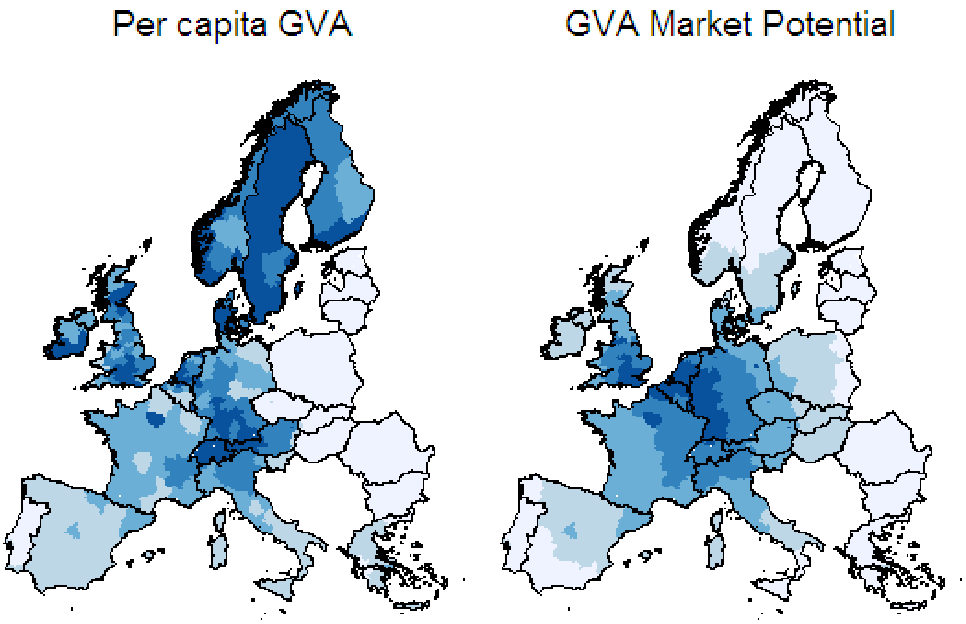
In order to stylize these spatial patters, we build two different maps. They show the predictions of these variables on a grid using an estimated third degree polynomial of the variables on the geographical coordinates of the regional centroids[24] (out of range predictions in white), plotted as a contour map. This type of geostatistical graphical representation is not common in Economics, so we provide four related reasons to justify its utility in our context.

On one hand, we are interested in isolating the core-periphery spatial pattern to provide a stylized image of it, more that in describing the levels of the variables for particular regions or their distribution by quantiles. Additionally we are studying here global spatial patterns so they are better represented with a contour map based in global geographical trends, instead of a contour map representing locally on a grid similar levels of the variables. Now the focus is on the European spatial pattern as a whole, not in local clusters neither in a representation of hot spots.

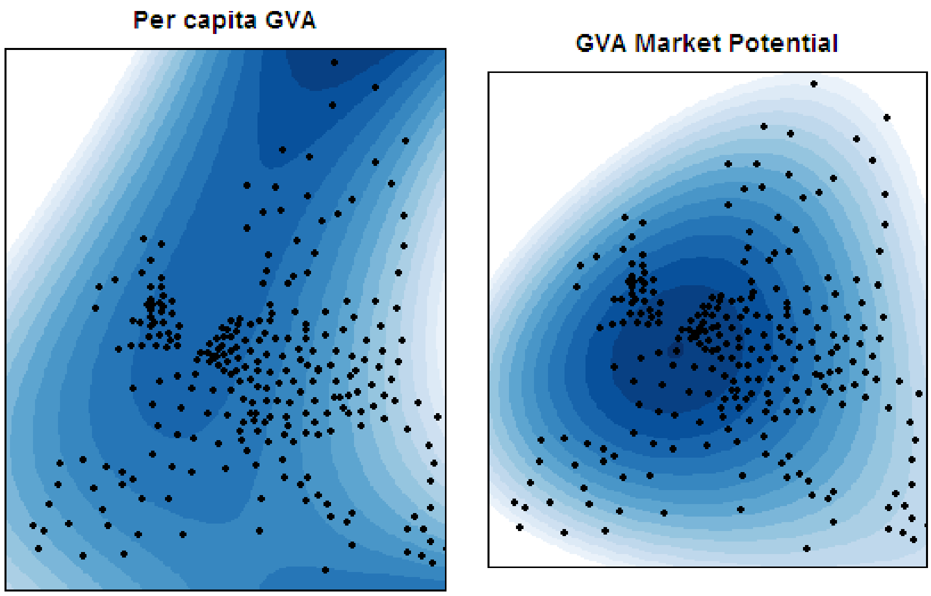
Second, as it was pointed out, a measure of global autocorrelation as Moran’s I is sensitive to a global (geographical) trend in the variable so we begin the analysis of global autocorrelation visualizing the spatial pattern derived from that trend. Given that these plots show that the assumption of a constant mean used by Moran´s I is wrong, we already expect that the Moran’s I of these variables will reveal spatial autocorrelation. Therefore the maps show a free of W first view of global spatial autocorrelation, the co-variation of values with the values of the neighbour regions, in this case by contour lines.

[24] If N and E are the South-North and West-East planar coordinates of the regional centroids, the predictions on a grid from a second degree polynomial offer a qualitatively similar view of our variables (Y): $Y = N\beta_1 + S\beta_2 + N^2\beta_3 + S^2\beta_4 + (N \cdot S)\beta_5$. In order to obtain more rotation, we estimated the terms $N^3\beta_6 + S^3\beta_7 + (N^2 \cdot S)\beta_8 + \ (N \cdot S^2)\beta_{10}$ too. For the coordinates we have used Lambert Azimuthal Equal Area projections (EPSG 3035).

Quantiles (year 2008, logs)



Predictions from a third-order trend surface (logs, year 2008)



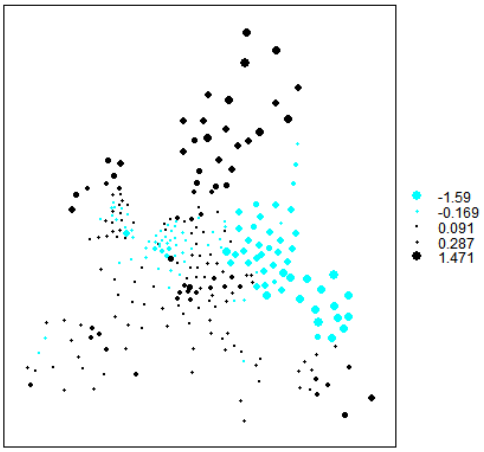
Regional centroids overprinted

The maps show this idea of global long distance spatial dependence, central in the concept of Market Potential, which in the case of Europe follow a core-periphery pattern. So we were able to visualize what information is not considered when the analysis of spatial autocorrelation is done with a (usually standardized) W matrix under the assumption that the correct model has constant mean and any remaining pattering is caused by the spatial relationships encoded in the spatial weights. In both cases we speak of "global" spatial dependence is the sense that we use information from the whole sample, but Market Potential emphasizes long distance dependence while Spatial Econometrics usually emphasizes shorter distance covariation. As we will discuss it later, with this we mean that there are two different spatial structures in our data. On one hand we have a correlation of the values with the values of the neighbour regions. That is shown, for instance, in the contour levels with the same colour and it is the main concern of Spatial Econometrics. On the other hand we have a core-periphery (long distance) structure of spatial dependence, which has been the main concern of Regional Economics.

Apart from the issue of spatial autocorrelation and its test, a third reason for showing the global trend surface maps is that they provide a clear picture to analyse the characteristics of the (long distance) core-periphery pattern of our variables. In the case of the log of per capita GVA, the predictions from the global trend show the economic centre of Europe. Too it detects a North-South trend, that the reasons why the predictions based in geography are higher for the Nordic countries. Additionally, it is detected a slighter geographical trend coming from the South, which is show with higher (over-)predictions around Sicily. This is due to the fact that the geographical trend of the log of per capita GVA detects the blue banana going from North West England in the North to Milan in the South. Therefore, the global trend fine-tunes the main North to South trend by adding a South to North component. However, the key idea here is the configuration of geographical centripetal forces in the trend of per capita GVA. The trend surface maps show that the core-periphery pattern of this variable can be captured in a simple way regressing it into the even stronger core-periphery pattern of the log of Market Potential. Of course, we would have to add a number of control variables to explain the more complex distribution of per capita GVA, but that is not the focus of our work here. What we are arguing is that GVA Market Potential captures part of the global spatial trend of per capita GVA stylized in the plots, the core-periphery structure of that geographical trend.

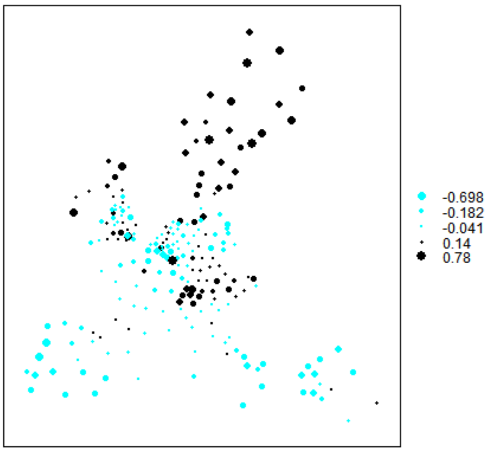
Too the plots show a West to East trend starting in England, which takes us to the fourth reason why these maps are useful. Both the quantile maps and the trend surface maps allow us to improve our understanding of the spatial distribution of residuals when Market Potential is used in a cross-sectional regression to explain the European per capita income (in logarithms). Given the strong North-South global trend in per capita GVA, we expect that this regression will under-predict Nordic countries per capita GVA. And this is what happens. But given that the West-East trend is present in both per capita GVA and market potential variables, the question is how GVA Market Potential is going to predict the per capita GVA for the geographically central but poorer East European countries. For the whole sample the answer is that the OLS zero mean residuals are distributed in a way that the regression over-predicts mainly just two zones, the economic centre of Europe and the Eastern European countries.

Residuals of the regression of per capita GVA on GVA Market Potential (Baseline broad European sample)



Variables in logs, year 2008

Residuals of the regression of per capita GVA on GVA Market Potential (Excluded countries of the 2004 and 2007 EU enlargement)



Variables in logs, year 2008

We can see that in a plot of the residuals over the regional centroids, with the size of the circles proportional to the magnitude of the residuals by quintiles. Positive residuals in a darker colour mean under-prediction of the log of per capita GVA made by the log of Market Potential, while a clear colour means over-prediction. On the left plot we can see that for the broad sample, given the high concentration of Market Potential in the geographical centre of Europe, the OLS regression over-predicts the per capita GVA of these central regions. Additionally, even considering their lower GVA, the regression with GVA Market Potential over-predicts Eastern European regions. They are poorer that what they should be according with the European core-periphery pattern of Market Potential, which it is related with their recent transition to the market economy. Considering the whole sample, with many countries with a long history of economic integration, the relation between per capita GVA and GVA market potential gets distorted by the inclusion of these countries. Therefore, the OLS distribution of errors makes European Southern countries to be under-predicted too. Excluding the countries of the last European Union enlargements, on the right plot, we find a different pattern of over and under-predictions. Bellow we come back to this issue but we chose to continue working with the whole sample in order to show the spatial properties of the broad European regional sample.

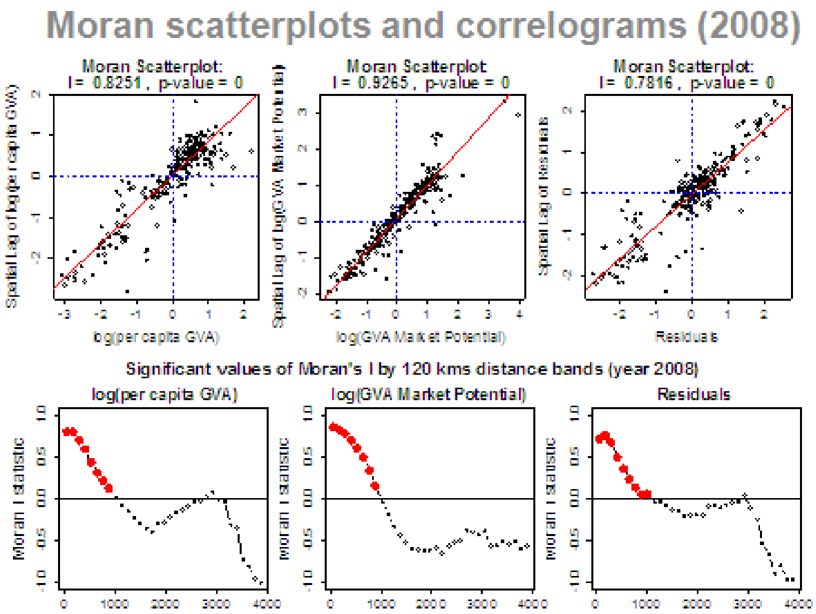
All this reflects the characteristics of the European history and the European spatial distribution of economic activity. In order to understand the behaviour of Market Potential as explanatory variable we remark the issue of sample selection, and we comment on the

North-South and West-East global trends of the European per capita GVA. But these aspects are not our research focus either. Using a broad sample of European regions, we claim that the strong core-periphery structure of Market Potential, as shown by the predictions of the global trend, can help to explain the somewhat less accused but still strong core-periphery structure of the European per capita GVA. However, the plot of the residuals allows us to introduce another issue. The market potential variable is globally capable of capturing the core-periphery structure of per capita GVA but it does not solve the different, though related, problem of spatial autocorrelation. What the plot shows is that the residuals are highly spatially autocorrelated, violating the hypothesis of independence assumed by the OLS estimation and justifying the spatial models we estimate in section 7.

Spatial autocorrelation is an "unavoidable feature" (Le Gallo and Kamarianakis, 2011) in the (NUTS) European regions (Ertur and Le Gallo, 2003; Ertur et al., 2006; Fingleton and Lopez-Bazo, 2006; Bivand and Brunstad, 2006; Chasco et al., 2012). Our previous analyses already allow us to anticipate that both the variables and the regression residuals are going to be highly spatially autocorrelated. This is shown in next figure. For the year 2008 and using our baseline weight matrix for the standardized inverse distance to the five nearest neighbours, the figure shows the Moran scatterplot[25] of the log of per capita GVA, the log of market potential and the residuals of the regression of the first variable into the second one. The three variables have high values of Moran's I[26] (the slope of the line in the plots; Anselin, 1996) and zero p-values, which points to the rejection of the null hypothesis of no global spatial autocorrelation. Additionally, each Moran scaterplot allows us to study its quadrants. The top right quadrants show regions with a high value for the variable and with a high value for its spatial lag, i.e., with high values for the weighted average of the variable for their (five nearest) neighbours. On the top left plot we can see that observations with positive spatial autocorrelation of per capita GVA concentrate at high levels of per capita GVA which we saw that generally are around the geographical centre of Europe, but the plot shows too the positive spatial autocorrelation among Eastern European Countries in the bottom left quadrant. Contrary, and apart from the two top right outliers of Inner and Outer London, positive spatial aucorrelation of market potential appears more evenly distributed along all the values of the variable, given the almost concentric structure we saw before. Finally, as expected, OLS residuals concentrate around zero values for both the residual and its spatial lag. But the positive spatial autocorrelation of the residuals is clear too.

[25] As Anselin (2007), we show the Moran Scatterplots with standardized variables, in the sense of having a mean of zero and a standard deviation of one, so the axis of the plots can be interpreted in units of standard deviations.

[26] We apply Moran's one way tests under the randomization null hypothesis.

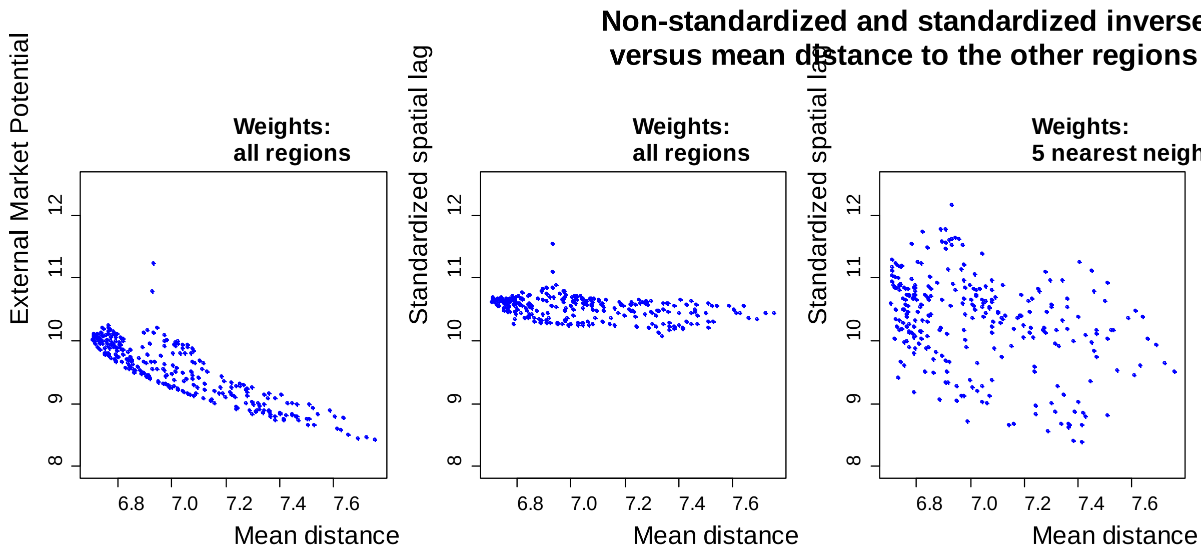


Additionally, we show the correlograms of the variables with Moran's I calculated for pairs of regions at distance bands of 120 kilometres: 0-120, 120-240…. Moran's I at each distance band is calculated with a standardized binary W matrix identifying neighbours with centroids at that distance band. Therefore, here we do not weight by inverse distances but compare previous results taking the distance in the form of distance bands. If spatial autocorrelation is related with inverse distances, we will expect decreasing Moran's I as distance increases. So this is a way of checking previous results without using our baseline W matrix, which standardizes absolute distances. Indeed, the meaningful values of Moran's I, with filled circles, show how spatial autocorrelation decreases with the distance among regions. They are still meaningful until a pretty large distance of around 1000 kilometers, according to our previous discussion of the European spatial structure in Europe. So absolute distances matter for spatial autocorrelation. According to the core-periphery spatial pattern, the spatial autocorrelation of the European economic activity is pretty spatially persistent, which somewhat was shown before in the plots of the predictions with a geographical trend.

7 Standardization of weights and corrections of spatial autocorrelation

Gross Value Added External Market Potential (EMP), the main component of GVA market potential for the average region in our sample, is a non-standardized inverse distance spatial lag of GVA. To understand better the role of row-standardization we build a standardized version of EMP. Note that the spatial weights of the standardized form are similar to our baseline weight matrix, but in our W we just use the five nearest neighbours and in EMP all regions in the sample are considered. Therefore, we additionally analyse the role of the definition of "neighbourhood" (in our baseline W matrix) by comparing the two previous variables with the standardized version of EMP for only the five nearest neighbours.

The two first plots in the following figure compare the logarithm of External Market Potential and the logarithm of its standardized version when they are plotted against the logarithm of the mean distance of each region to the rest of the regions in the broad European sample. Apart from the two mentioned UK outliers in the top of the plot, we see again how External Market Potential has a core-periphery structure, with values decreasing with distance from the geographical centre of Europe, as measured through mean regional distances.



Logs of spatial lags of GVA (year 2008) and log of me

Contrary to this pattern of the logarithm of EMP, the standardization of weights makes the (logarithm of the) spatial lag to lose the distance as a discount factor, which is the key element for the Economic Geography interpretation of market potentials. The logarithm of the (row-)standardized inverse distance spatial lag of GVA has lower dispersion because the sum of the spatial weights is one for all regions, while in EMP ranks from 0.13 to 0.73 in our broad sample, according to the degree of peripherality of each region. Standardization makes that only the neighbour's relative distance matters, which emphasizes a local aspect of spatial dependence: the relevant information is the list of neighbours and their relative distance but not the absolute distance, which is the main point of an index of accessibility as Market Potential. As a consequence, the log of the standardized spatial lag of GVA behaves more or less uniformly along the log of the mean distance.

If the dispersion of the standardized (logarithm of) the spatial lag of GVA is so low, how is it possible to correct for spatial autocorrelation using a standardized W matrix?. Here it is the role of the neighbour criterion chosen to build W. In the third plot of the previous figure we construct the spatial lag of GVA using our baseline W matrix, a standardized inverse distance matrix of the five nearest neighbours. Now, the logarithm of the spatial lag has high dispersion, which is the key to collect the short distance spatial dependence in the analysis of spatial autocorrelation. The reason for this higher dispersion with respect to the standardized spatial lag using all regions (in the centre of the figure) is that in both cases, the sum of the weights is one, but only the (five) nearest regions are used in the left plot. With standardization, the more regions are considered as “neighbours”, the lower is the weight of the nearest neighbours, and the lower is the dispersion of the spatial lag.

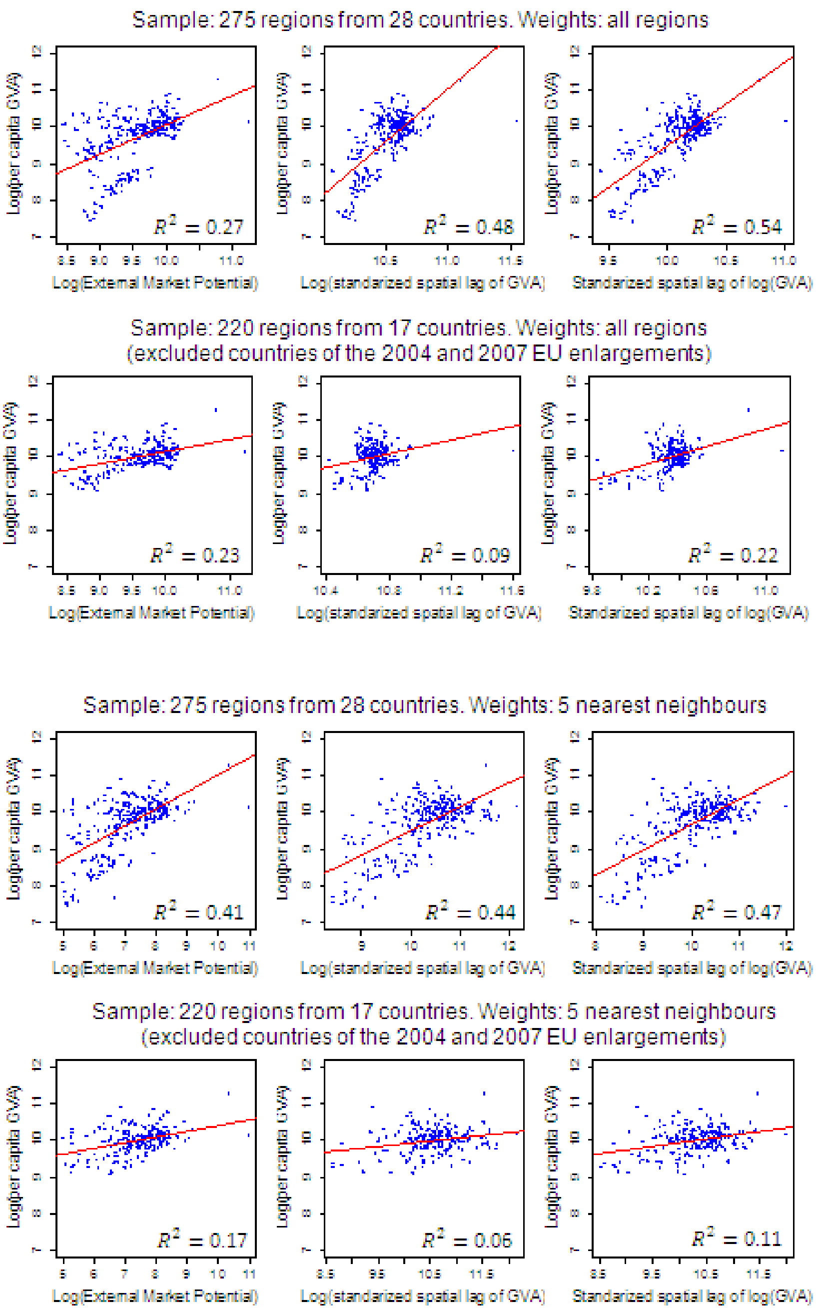
The Spatial Econometrics literature emphasizes the role of the neighbourhood criterion, the sensitivity of spatial autocorrelation test and models to the construction of W and the need of checking how the results are robust to different W. Imposing lower structure on W is preferred (Griffith, 1996), so a standardized spatial lag as in the second plot would not be useful to correct for spatial autocorrelation in a spatial model. Therefore, the restricted neighbourhood criterion is the second difference between the market potential approach to spatial dependence and the approach of Spatial Econometrics. And this is why later we will use our baseline W matrix to correct a wage-type equation for spatial autocorrelation.

First, let’s focus in the role of standardization. We mentioned in section 5 some reasons for it. Here the point is that the standardized spatial lag of a variable can be nicely interpreted as the weighted average of the values of the variable for the neighbours (however defined) but standardization emphasized “local” covariation. Given that the terminology can be confusing, in the sense that a global measure of autocorrelation is the one using equally the whole set of data in the sample, we say that the standardization of inverse distances emphasizes short term spatial patterns, as opposite to the long term spatial pattern reflected by a measure of accessibility like Market Potential. Or in other words, Market Potential captures a structure of spatial polarization (Arbia, 2001), which is a form of spatial dependence different from the one captured with a standardized weight matrix.

Given that both spatial lags measure different things we compare their behaviour in a simple regression. But before that, we have to consider the “lag of log” or “log of lag” issue. According to the NEG’s wage equation, this type of regressions is estimated in logarithmic form. Therefore, with external market potential, the explanatory variable is the log of a non-standardized inverse distance spatial lag of GVA. But Spatial Econometrics applies spatial lags to the variables in the forms that are introduced in the regression, so the spatial lag of the log of the variables is taken. When the spatial lag of the endogenous variable is introduced into the regression, as in the spatial lag model, the explanatory variable will be the spatial lag of the logarithm of per capita GVA.

Therefore the following figure compares cross-sectional OLS regressions for one year using the three weighting schemes for GVA, using all the regions in the sample to make them comparable with External Market Potential. The log of per capita GVA is plotted against the log of GVA External Market Potential, the log of the standardized spatial lag of GVA and the standardized spatial lag of the log of GVA. We add this last comparison in order to approach how GVA would be considered in a spatial lag model, though in this model per capita GVA would be used and a W with less neighbours would be considered.

OLS, year 2008: log(per capita GVA) on spatial lag of GVA by weighting scheme



We are conscious of the limitations of these regressions but take the squared R as an indicator of the behaviour of each variable to explain the log of per capita GVA. The result for the broad sample, in the top row plots of the figure, is that the regression with External Market Potential has the lower explanatory power. But this result is driven by the Eastern European Countries, the cluster of points in the button of the top left plot. When we repeat the exercise excluding the countries of last the EU enlargements, the results are different. The log of the standardized inverse distance spatial lag of GVA loses almost all explanatory power. The squared R using the log of External Market Potential in the reduced sample (bottom left plot) is similar to the one using the spatial lag of the log of GVA (bottom right plot). But in this last plot we see how concentrated are the points around the centre and the stronger influence of outliers on the regression results.

In the last two rows of the figure we focus on the role of the definition of “neighbourhood”. We repeat the exercise but using just the five nearest neighbours to define the spatial lags of GVA. For comparability we keep the name “External Market Potential” in the left plots but this variable would be a restricted EMP because not all the regions in the sample are considered. Surprisingly, the squared R of the regression with this variable increases to 0.41 but this is again an effect from the East European countries, clustered under the regression. When they are excluded from the sample, the squared R of the regression is 0.17, somewhat lower than the 0.23 when all the regions are considered in the reduced sample (second and fourth rows of the figure). This result is relevant. Given that Market Potential weights GVA with inverse distances, the nearest neighbours have the higher weights. Considering the five nearest neighbours instead of the 220 regions in the restricted sample reduces the squared R in just 0.05. The reduction of squared R is more important when the explanatory variable is the standardized inverse distance spatial lag of the log of GVA: from 0.22 to 0.11. As it was said before, the consideration of just 5 nearest neighbours instead of all the regions increases the dispersion of this last explanatory variable. But it does not improve the results obtained with External Market Potential.

This is a simple exercise. But it allows us to make several remarks:

- In order to capture the core-periphery spatial pattern of per capita GVA in Europe, we need to consider absolute (long-)distances as in the concept of Market Potential, avoiding the standardization of weights.
- But we saw that a variable as Market Potential is highly spatially autocorrelated in Europe. This is the cost of capturing the core-periphery pattern of per capita GVA, so a correction for spatial autocorrelation in the residuals is needed. Spatial Econometrics corrects this focusing on short-distance spatial relations through a standardized W matrix and considering a reduced number of neighbours.
- When comparing the logarithm of External Market Potential with the standardized spatial lag of the logarithm of GVA as explanatory variables of the logarithm of per capita GVA, we see that they are competing specifications in terms of explanatory power.

- We know the economic meaning of Market Potential and there are a long tradition and a theory behind it (here we just took the external component for illustrative purposes). Even if the results using the standardized spatial lag of the log of GVA were similar we are not so sure about the channels of economic interaction collected in a variable like this one.
- However, in order to correct for spatial autocorrelation we still need the Spatial Econometrics techniques. Keeping a framework of simple models, we need to introduce those more diffuse mechanisms of dependence among regions: spillovers through unspecific channels, effects of omitted variables, spatially autocorrelated unobserved shocks…
- The exercise was comparing spatial lags of GVA. The standardized spatial lag of the log of GVA is not necessary the variable considered in spatial models. But it illustrates the differences between long-and-short distance types of spatial dependence and how the standardized spatial lag of the logarithm of a variable can have an explanatory power similar to the logarithm of External Market Potential.

Therefore, the following question is what happens when both long-and-short distance spatial dependences are considered together, how significant is going to be a Market Potential variable when controlling for spatial effects to correct autocorrelation. For instance, Blanco (2012), in the study parallel to ours quoted in section 4, finds that surrounding market potential (inverse distance weighted GDP per capita) has a positive significant effect on net world FDI in Latin America, but the spatially lagged dependent variable is statistically insignificant. Contrary, when considering only FDI inflows from the US, she finds that surrounding market potential is no longer statistically significant once the spatial autocorrelation of FDI is account for.

Using our baseline weights matrix, we estimate the Spatial Lag Model (SAR)[27] and the Spatial Error Model (SEM) under four different variants. On one hand we estimate separately the models for Market Potential and for External Market Potential. As we said in section 5, when Internal Market Potential is excluded, there are a lower endogeneity problem, but at the cost of introducing measurement error in the evaluation of the markets accessible to each region. However, given our focus in the ways of capturing spatial dependencies, it is worthy to check the behaviour of External Market Potential by itself. On the other hand, for both previous specifications we control by the percentage of population who has successfully

[27] Too we have estimated the spatial lag model using the generalized spatial two stage least squares procedure by Kelejian and Prucha's (1998). The problem of instrumenting the spatial lag of the endogenous variable with the spatial lag of market potential is that the instrumented variable shows very high correlation with market potential, so the estimated equation has multicollinearity.

completed education in Science and Technology (S&T) at the third level and are employed in a S&T occupation[28].

For each specification, Lagrange Multiplier statistics test for a spatial lag spatially lagged dependent variable or error dependence in the OLS regression. The decision rule (Anselin and Florax, 1995; Anselin et al., 1996) is based in a comparison of the Lagrange Multiplier tests for the same specification. We follow it as described by Florax et al. (2003). Adapting it to the terminology in R spdep package (Bivand, 2012), the tests are named LMerr and LMlag, and their robust versions are RLMerr and RLMlag[29]. The decision rule is:

1. Estimate the initial model by means of OLS.
2. Test the hypothesis of no spatial dependence due to an omitted spatial lag or due to spatially autoregressive errors, using LMlag and LMerr respectively.
3. If both tests are not significant (the “no spatial dependence” is accepted), the initial estimates from step 1 are used as the final specification. Otherwise proceed to step 4.
4. If both tests are significant (the alternative hypothesis is accepted), estimate the specification pointed to by the more significant of the two robust tests. For example, if RLMlag > RLMerr then estimate the Spatial Lag (or Autoregresive) Model (SAR). If RLMlag < RLMerr then estimate the Spatial Error Model (SEM). Otherwise, proceed to step 5.
5. If LMlag is significant but LMerr is not (the “spatial error dependence” is rejected”), estimate the SAR model. Otherwise proceed to step 6.
6. Estimate the SEM model.

The null hypothesis for LMlag and RLMlag is $\rho = 0$, while for LMerr and RLMerr is $\lambda = 0$. Under all the specifications, the p-values of LMerr and LMlag reject this null hypothesis, pointing to significant spatial effects. The p-values for the robust versions allow choosing the Spatial Lag Model when Market Potential is considered alone and the Spatial Error Model when External Market Potential is considered alone. But when controlling by the effects of the variable of human resources in science and technology (S&T), RLMlag and RLMerr point to significant spatial dependence and RLMlag < RLMerr both using Market Potential or External Market Potential, so the decision rule recommends to estimate the SEM.

[28] This variable has been used by Bivand and Brunstad (2006) in a regional European cross-section regression, but they used it as a percentage of active population while here is in percentage of total population.

[29] Instead of this Specific-to-General strategy, later literature defend a General-to-Specific approach, or a combination of both (Mur and Angulo, 2009; Elhorst, 2010). The first strategy seems suitable for our purpose of showing how the Market Potential variable behaves under simple spatial models. For the same reason, we use our baseline W matrix, and do not check if the results are robust to the choice of W.

Lagrange Multiplier tests and Maximum Likelihood estimates
Y = log of per capita GVA (year 2008)
Variables in logs (probabilities in brackets), ML eigen method

	Spatial Lag Model (SAR) $Y = \rho WY + X\beta + u$				**Spatial Error Model (SEM)** $Y = X\beta + u, \quad u = \lambda Wu + \varepsilon$			
LM test (LMlag for SAR model and LMerr for SEM model)	362.54 (0.000)	289.40 (0.000)	378.80 (0.000)	303.19 (0.000)	409.68 (0.000)	**370.79 (0.000)**	357.85 (0.000)	345.77 (0.000)
Robust LM test (RLMlag and RLMerr)	0.04 (0.850)	13.09 (0.000)	10.94 (0.001)	13.63 (0.000)	47.17 (0.000)	**84.49 (0.000)**	0.00 (0.986)	56.21 (0.000)
Rho (ρ)	0.8161 (0.000)	0.7512 (0.000)	0.8523 (0.000)	0.7798 (0.000)				
Lambda (λ)					0.8923 (0.000)	**0.8986 (0.000)**	0.8843 (0.000)	0.9043 (0.000)
GVA Market Potential (HMP)	0.2140 (0.000)	0.1509 (0.000)			1.0084 (0.000)	**0.6297 (0.000)**		
External GVA Market Potential (EMP)			0.1055 (0.016)	0.0685 (0.081)			-0.0533 (0.772)	-0.0068 (0.964)
Population (%) with third level education in S&T and employed in S&T occupation		0.5032 (0.000)		0.5275 (0.000)		**0.5359 (0.000)**		0.6901 (0.000)
Log likelihood	-91.3	-47.3	-101.9	-54.0	-57.8	**-14.1**	-104.6	-36.8
ML residual variance	0.0933	0.0707	0.0979	0.0729	0.0684	**0.0494**	0.0969	0.0579
AIC	190.60	104.62	211.82	117.96	123.67	**38.26**	217.18	83.50
Moran's I of residuals	-0.052 (0.899)	0.081 **(0.013)**	-0.121 (0.999)	0.043 (0.113)	-0.077 (0.973)	**-0.038 (0.816)**	-0.130 (0.999)	-0.071 (0.962)

Though the decision rule recommends the Spatial Error Model, in order to check the consequences of both spatial models in the estimated parameters for each equation, we show anyway the Maximum Likelihood (ML) estimation of all the specifications. The magnitudes of the coefficients in the Spatial Lag and the Spatial Error Models cannot be compared directly. In the SEM the coefficient estimates represent the marginal effects of the explanatory variables, but not in the SAR. In this last model, the marginal effects of the explanatory variables is given by the sum of the direct effects given by the coefficient estimated and the indirect effects of the explanatory variables in the log of per capita GVA of the neighbour regions. Here we do not present estimates of the total effects (LeSage and Pace, 2009) under the SAR model.

Our previous analysis showed how the standardized spatial lag of the logarithm of GVA could compete in explanatory power with the logarithm of External Market Potential. Following that previous exercise about the weighting schemes, our concern right now is to check the significance of the Market Potential variables under the Spatial Lag and the Spatial Error Models, using our baseline weight matrix to construct the spatial lags with standardized inverse distances to the five nearest neighbours and constructing the Market Potential variables with the non-standardized inverse distances to the rest of the regions in the broad sample. Therefore, the SAR model includes a (five nearest neighbour) standardized spatial lag of the logarithm of per capita GVA. The result is that External Market Potential (EMP) tends to be not meaningful when the spatial models are estimated, and its parameter is close to zero. The spatial information of the neighbours seems to be better collected by the spatial models than by the non-standardized inverse distance spatial lag of GVA. But the full Harris's Market Potential variable (HMP) keeps its significance and it is robust to the inclusion of the S&T variable under the estimation of the spatial models.

The short-distance spatial patterns collected by the spatial effects models make not meaningful the long-distance (core-periphery) spatial patterns collected by EMP but not by the HMP. Comparing the regression diagnostics of the similar specifications but using HMP or EMP, we see that the estimations with HMP are generally better, but not very different. The estimated spatial parameters rho and lamda are similar too. What it is very different in those estimations is the coefficient estimate and significance of the market potential variables. Given that EMP suffers from measurement error because the market accessible to each region is better captured when the internal market is included, as in HMP, the cross-sectional results confirm the need of considering the internal markets. The cost of doing this is a bigger problem of endogeneity. That it is out of the scope of the present study, and we deal with it in other works.

Considering both Market Potential (HMP) and the S&T variable, the residuals of the Spatial Lag Model are still spatially autocorrelated (the p-value of Moran's I is 0.013 < 0.05), but not under the Spatial Error Model. Additionally, the previous robust LM tests and the other regression diagnostics recommend the SEM specification, which presents higher Log likelihood and less information loss, as the Akaike Information Criterion (AIC) reveals[30].

We repeated this exercise excluding the countries of the last EU enlargements (not reported). In this case, under the specification with Market Potential and the S&T variable, the robust LM test recommends estimating the SAR model. But again the residuals of this model are spatially autocorrelated and the regression diagnostics recommend the SEM specification. It is worthy to mention that the estimated elasticity of Market Potential under the Spatial Error Model in the reduced sample is almost half of the one with the broad sample of European regions (0.3711).

---

[30] All the eight specifications reduce AIC when comparing with their analogous OLS estimation. The estimation of the selected SEM specification by Maximum Likelihood using full matrix methods produces identical results.

All these results might be due to the limitations of the exercise. We do not control for the endogeneity of Market Potential, for border effects and other explanatory variables, for outliers, for sample selection, different W matrix… Additionally, we have already mentioned the Specific-to-General versus General-to-Specific debate about the proper modeling strategy. More generally, as we mentioned in section 4, Lesage and Pace (2009) argue that the cost of ignoring spatial dependence in the endogenous or the explanatory variables (Elhorst, 2011) is relatively high if relevant independent variables are omitted: the estimator of the coefficients for the remaining variables is biased and inconsistent. In contrast, ignoring spatial dependence in the disturbances, if present, will only cause a loss of efficiency. Additionally the spatial Durbin model allows collecting spillovers and distinguishing the direct impact and the indirect impact of a change in an explanatory variable[31]. Contrary, Glass et al. (2012) argue in favour of the SEM model.

However, given the previous results, for our purposes, the estimation of the Spatial Error Model with a cross-section of European regions for just one year and using just our baseline W matrix permit us to conclude that it is possible to correct for spatial autocorrelation a wage-type equation keeping the Economic Geography interpretation of Market Potential and the capacity of this variable to collect the core-periphery spatial pattern of the European per capita GVA.

We provide additional evidence about this. One of the motivations of spatial models presented in section 4 was spatial heterogeneity. In a cross-section context it is not possible to estimate separate intercepts for each region. If observational units in close proximity exhibit effects levels that are similar to those from neighbouring units, spatial heterogeneity can motivate spatial error dependence. Therefore, it is convenient to check if panel estimation with individual effects changes our previous main result about the Market Potential variable. Additionally, in section 5 we have mentioned the need of estimating with fixed effects to avoid some problems of levels derived from the construction of the Market Potential variable. Apart from this, and the unobserved (omitted) effects argument, panel

[31] We have estimated the Durbin model (not reported) adding to the SAR model the spatial lags of Market Potential and the variable of human resources in S&T. All the variables are significant and the residuals are not spatially autocorrelated but the parameters of the spatial lags of the explanatory variables are negative. It is difficult to explain why the presence of neighbours with a high percentage of population educated and occupied in S&T would depress per capita GVA. In a similar way, we would expect that the higher GVA Market Potential of the neighbours, the higher the indirect market size of each region. But there are an issue of multicollinearity. And given that Market Potential weights GVA with the absolute distances to the neighbours, it is not clear the meaning of a spatial lag of Market Potential using standardized inverse distances to the (five nearest) neighbours. The internal GVA is part of the Market Potential variable, so a spatial lag of Market Potential is probably to contain again the internal GVA of each region, aggravating the endogeneity problem. Contrary, in the specification with External Market Potential instead of Market Potential, both External Market Potential and its spatial lag are not significant under the Durbin model.

data models are more informative. Too, they have some drawbacks among them, the data availability associated with selectivity of the sample (Arbia, 2006, page 148).

Therefore, we finish the econometric exercise checking the behaviour of the Market Potential variable in a spatial panel data setting under our two simple static Spatial Lag and Spatial Error Models, though the possible variety of spatial models is very rich in a panel data framework, including the growing literature about dynamic spatial panel data. For per capital GVA and GVA Market Potential we have a complete panel for the years 1991-2008 and the 275 regions. But data of human resources in S&T is not available for all regions in such a period. Therefore we would have to work with an unbalanced panel, which is not currently supported in R project for spatial panel models and even for some procedures of standard panel specifications. However, preliminary non-spatial unbalanced panel estimations using different variables of human resources in S&T reveal that they tend to lose its significance when regional fixed effects are considered. These fixed regional effects do not solve the spatial dependence of the residuals (Anselin and Arribas-Bel, 2011), as checked with a global or local Pesaran test or with Moran's I, what justifies the need of estimating spatial models.

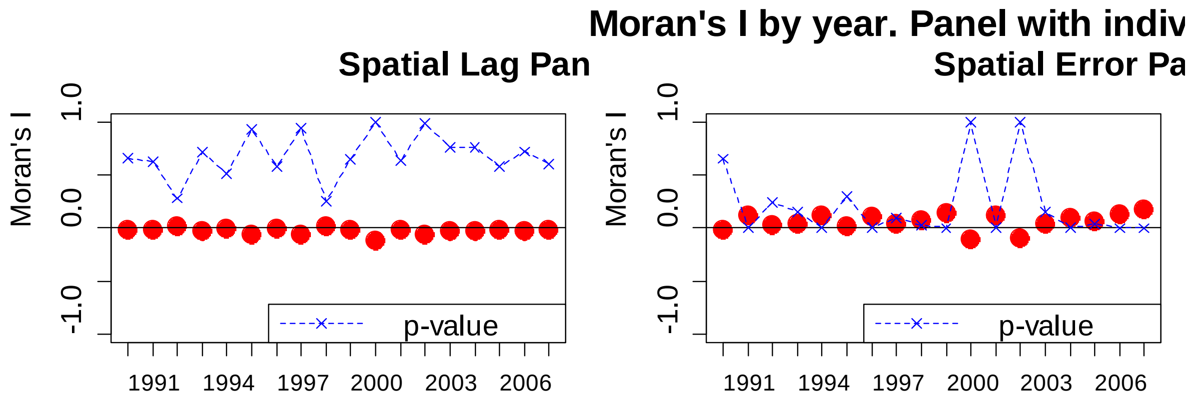
Pesaran's CD tests for cross-sectional dependence in a panel; Hausman test for random effects estimators in spatial panel data and Baltagi, Song and Koh conditional LM test for SEM

Formula: log(per capita GVA) on log(GVA Market Potential). Time effects not considered.

	Statistic Value	p-value	Alternative Hypothesis
Pesaran global fixed effects	53.5722	0.000	Cross-sectional dependence
Pesaran local fixed effects	51.129	0.000	Cross-sectional dependence
Hausman SEM	11.0113	0.000	Fixed individual effects
LM*-lamda conditional	82.6361	0.000	Spatial error dependence

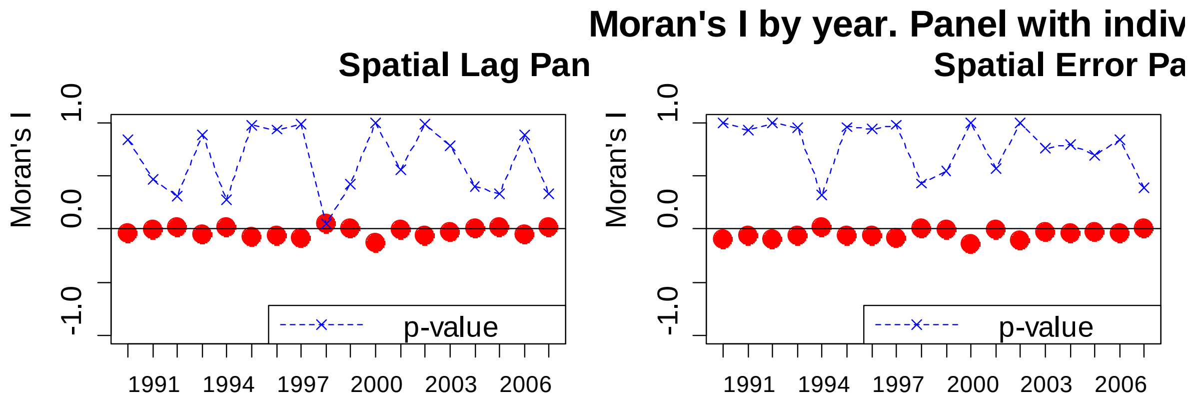
Following we present some preliminary results of the ML spatial panel estimations with fixed regional effects and the log of GVA Market Potential as the explanatory variable of the log of per capita GVA[32]. For each year, we show the plots of the Moran's I of the residuals, with the values of the estimated I as circles and their p-values in a dotted line. An upper limit of 1 for Moran's I is considered in the plots, which is the case for row-standardization of W.

[32] We thank to Giovanni Millo for his advices with R splm package (Millo and Piras, 2012). Here we keep the previous notation for the parameters rho and lambda, with is the opposite of the one used in splm.



Rho	0.6332 (0.000)
GVA Market Potential	0.3866 (0.000)

Lambda	0.7283 (0.000)
GVA Market Potential	1.2610 (0.000)



Rho	0.5938 (0.000)
GVA Market Potential	1.9146 (0.000)

Lambda	0.6994 (0.000)
GVA Market Potential	3.7340 (0.000)

The preliminary tests in the table before accept the SEM panel model with fixed individual effects. But the p-values of Moran's I with the Spatial Error Model are lower than 0.05 in 8 of the 18 years, so spatial autocorrelation is not totally corrected[33]. Too, we estimate the SAR panel model and repeat the estimations including time fixed effects[34]. When the time effects are included the p-value of Moran's I of the residuals is higher than 0.05 in both spatial

[33] A similar result for the Spatial Error Model is found by Arbia and Piras (2005) estimating with panel data a convergence growth regression in per capita GDP across European regions.

[34] When both the Spatial Lag and the Spatial Error Models are estimated together, the parameter of the spatially lagged dependent variable becomes negative.

models for all years. In this case, the estimated parameter of Market Potential increases dramatically in both models, though again the magnitude is not comparable between them[35]. We do not have an explanation for these high estimated parameters in our simple setting with just one explanatory variable, as we said at the end of sections 3 and 4. But our argument is about the significance of the explanatory variable. Market Potential is robust to the introduction of individual and time fixed effects in a spatial panel data context. This just reinforces our conclusion: it is possible to keep the Economic Geography interpretation of Market Potential, capturing long-distance spatial patterns, and at the same time capturing short-distance spatial patterns through corrections for spatial autocorrelation.

8 Conclusions

Harris's measure of market potential has been widely used in Regional Economics during half a century as an indicator of market accessibility. The New Economic Geography and its wage equation provide the micro-economic foundations of Harris's approach, which in spite of its shortcomings appears to be a practical way of getting to similar results than more sophisticated model derived measures.

We show how Market Potential, or more exactly, its component of External Market Potential, is really a non-standardized inverse distance spatial lag of whatever indicator be used for the market size, real Gross Value Added in our case. Therefore it is intimately related with the Spatial Econometrics literature. But there are two important differences. On one side, Spatial Econometrics most frequently works with a standardized weights matrix in order to capture spatial interactions. On the other side, Spatial Econometrics focuses more in short distance spatial structures, both through the standardization of the weight matrix and choosing a criterion of neighbourhood which normally considers just a few (continuity or nearest) neighbours.

This paper emphasized that both approaches to spatial dependence are different. Using a broad sample of European regions, we show the core-periphery pattern of the European per capita income. We claim that Market Potential captures this spatial pattern. The distance discount implied by a core-periphery pattern is eliminated when the inverse distance weights are standardized. Therefore standardization loses the key issue for the Economic Geography interpretation of market potentials.

A Market Potential variable captures long distance spatial structures, which in the European case correspond to a core-periphery representation. But capturing this structure comes at a

---

[35] If External Market Potential is used instead of Market Potential, when time effects are included the parameter becomes negative in the SEM but it is positive and meaningful in the SAR.

cost. The European Market Potential is very highly spatially autocorrelated, and the residuals after regressing per capita GVA on Market Potential are autocorrelated too.

Therefore the clear economic interpretation of Market Potential as an explanatory variable of per capita income does not preclude the need of a spatial model in order to correct for spatial autocorrelation. Here is where Spatial Econometrics techniques can be used to model short distance spatial dependences with a standardized weight matrix and a restricted criteria of neighbouhood. Therefore, both the Regional Economics-NEG and the Spatial Econometrics approaches are complementary in order to capture spatial dependencies.

We show that the Spatial Econometrics procedures allow a correction for spatial autocorrelation keeping the interpretation of market potential in terms of the accessibility to the markets and as an indicator of peripherality. For this, it is crucial to take into consideration the internal markets, according to the economic theory. So, both types of spatial dependences should be considered together.

References

- Ahlfeldt, Gabriel M. and Arne Feddersen (2008): Determinants of spatial weights in spatial wage equations: A sensitivity analysis; Hamburg contemporary economic discussions 22.
- Anderson, James E. and Eric van Wincoop (2004): Trade costs; Journal of Economic Literature 42:3, 691-751.
- Anselin, Luc (1988): Spatial Econometrics; Methods and Models; Kluwer Academic Publishers; Dordrecht, The Netherlands.
- Anselin, Luc and Raymond J. G. M Florax (1995): Small sample properties of tests for spatial dependence in regression models: some further results; in New Directions in Spatial Econometrics, ed. by L. Anselin, and R. Florax, pp. 2-74. Springer-Verlag, Berlin.
- Anselin, Luc (1996): The Moran Scatterplot as an ESDA Tool to Assess Local Instability in Spatial Association in Spatial Analytical Perspectives on GIS: GISDATA 4, pp. 111-125, edited by M. Fischer, H. J. Scholten and D. Unwin.
- Anselin, L., A. K. Bera, R. Florax and M. J. Yoon (1996): Simple diagnostic tests for spatial dependence; Regional Science and Urban Economics 26:1, 77-104.
- Anselin, Luc (2003): Spatial Externalities; Spatial Multipliers and Spatial Econometrics; International Regional Science Review 26:2, 153-166.
- Anselin, Luc (2007): Spatial Regression Analysis in R. A Workbook; Center for Spatially Integrated Social Sciences (CSISS), University of California Santa Barbara.
- Anselin, Luc and Daniel Arribas-Bel (2011): Spatial Fixed Effects and Spatial Dependence; GeoDA Center for Geospatial Analysis and Computation working paper 04, Arizona State University.
- Arbia, Giuseppe (2001): The role of spatial effects in the empirical analysis of regional concentration; International Journal of Geographical Systems 3:3, 271-281.
- Arbia, Giuseppe and Gianfranco Piras (2005): Convergence in per-capita GDP across European regions using panel data models extended to spatial autocorrelation effects; Istituto di Studi e Analisi Economica working paper 51, Rome
- Arbia, Giuseppe (2006): Spatial Econometrics: Statistical foundations and applications to regional convergence; Springer-Verlag, Berlin.
- Arbia, G., M. Battisti and G. Di Vaio (2007): Do Institutions Matter More Than Geography? Empirical Tests of Spatial Growth Regressions for the Regions of Europe; Luiss Lab of European Economics Working Document 57; Libera Università Internazionale degli Studi Sociali Guido Carli, Rome.
- Basile, Roberto and Stefano Usai (2012): Analysis of regional endogenous growth, Centro Ricerche Economiche Nord-Sud (CRENOS) working paper 11, Cagliari and Sassari, Italy.

- Baumont, C., C. Ertur and J. Le Gallo (2003): Spatial Convergence Clubs and the European Regional Growth Process 1980-1995. In Fingleton, B. (ed): European Regional Growth, Springer Verlag, Berlin, 131-158.
- Beck, N., K. S. Gleditsch and K. Beardsley (2006): Space Is More than Geography: Using Spatial Econometrics in the Study of Political Economy; International Studies Quarterly 50:1, 27-44.
- Bivand, Roger and Rolf J. Brunstad (2006): Regional growth in Western Europe: detecting spatial misspecification using the R environment; Papers in Regional Science 85:2, 277-297.
- Bivand, Roger (2012): Spdep: Spatial dependence: weighting schemes, statistics and models; http://CRAN.R-project.org/package=spdep.
- Bivand, R.S., E.J. Pebesma and V. Gómez-Rubio (2008): Applied Spatial Data Analysis with R, Springer, New York.
- Blanco, Luisa R. (2012): The Spatial Interdependence of FDI in Latin America; World Development 40:7, 1337-1351.
- Bosker, Maarten and Harry Garretsen (2010): Trade costs in empirical New Economic Geography; Papers in Regional Science 89:3, 485–511.
- Boulhol, H., A. de Serres and M. Molnar (2008): The Contribution of Economic Geography to GDP per Capita: OECD Journal Economic Studies 9:1, 1–37.
- Brakman, S., H. Garretsen and C. Van Marrewijk (2009a): The New Introduction to Geographical Economics; Cambridge University Press, NY.
- Brakman, S., H. Garretsen and C. Van Marrewijk (2009b): 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, Holger (2006): The Spatial Income Structure in the European Union. What Role for Economic Geography? ; Journal of Economic Geography 6:5, 593-617.
- Chasco, C., A. Lopez & R. Guillain (2012): The Influence of Geography on the Spatial Agglomeration of Production in the European Union; Spatial Economic Analysis 7:2, 248-263.
- Cliff, A. D. and J. K. Ord (1981): Spatial processes. Models and applications; Pion Limited, London.
- Combes, P.-P., T. Mayer and J.-F. Thiesse (2008): Economic Geography; Princeton University Press, New Jersey.
- Davis, Donald R. David E. Weinstein (2003): Market access, economic geography and comparative advantage: an empirical test; Journal of International Economics 59, 1-23.
- Durbin (1998): Spatial Autocorrelation: A Primer; Journal of Housing Economics 7:4, 304-327.
- Elhorst, J. Paul (2010): Applied Spatial Econometrics: Raising the Bar; Spatial Economic Analysis 5:1, 9-28.

- Elhorst, J. Paul (2011): Spatial Panel Models; Department of Economics and Related Studies, University of York, seminar November 2011.
- Ertur, Cem and Julie Le Gallo (2003): An Exploratory Spatial Data Analysis of European Regional Disparities, 1980-1995; Papers Reg. Sci. 82, 175–201. Too, in Fingleton, B. (ed): European Regional Growth, Springer Verlag, Berlin, 55-97.
- Ertur, C., J. Le Gallo and C. Baumont (2006): The European Regional Convergence Process, 1980-1995: Do Spatial Regimes and Spatial Dependence Matter?; International Regional Science Review January 29:1, 3-34.
- Ertur, Cem and Wilfried Koch (2007): Convergence, Human Capital and International Spillovers; Journal of Applied Econometrics 22:6, 1033-1062.
- Faíña, Andrés and Jesús López-Rodríguez (2006a): Renta per Capita, Potencial de Mercado y Proximidad: El Caso de España; Papeles de Economía Española 107, 268-276.
- Faina, Andres and Jesus Lopez-Rodriguez (2006b): Market access and human capital accumulation: The European Union case; Applied Economics Letters 13, 563-567.
- Faíña, Andrés and Jesús López-Rodriguez (2006c): EU enlargement, European Spatial Development Perspective and Regional Policy: Lessons from Population Potentials; Investigaciones Regionales 9, 3-21.
- Faíña, Andrés and Jesús López-Rodríguez (2008): Perificidad, desarrollo y capital humano en Europa: ¿Qué cuenta la Nueva Geografía Económica?; Claves de la Economía Mundial, Pags. 71 y ss.; Instituto Español de Comercio Exterior- Univ. Complutense, Madrid.
- Fingleton, Bernard (2009): Testing the NEG model: further evidence from panel data; Région & Développement 30, 141-158.
- Fingleton, Bernard and Lopez-Bazo, Enrique (2006): Empirical growth models with spatial effects; Papers in Regional Science 85:2, 177-198.
- Florax, R. J. G. M., H. Folmer and S. J. Rey (2003): Specification searches in spatial econometrics: the relevance of Hendry's methodology; Regional Science and Urban Economics 33:5, 557-579.
- Fujita M., P. Krugman and A. Venables (1999): The Spatial Economy; MIT Press, Cambridge MA.
- Getis, Arthur (2007): Reflections on spatial autocorrelation; Regional Science and Urban Economics 37:4, 491-496.
- Getis, Artur (2010): Spatial autocorrelation; in Fischer, M. M. and A. Getis (eds): Handbook of Applied Spatial Analysis; Springer Verlag, Berlin, ch. 3.
- Griffith, Daniel A. (1996): Some Guidelines for Specifying the Geographic Weights Matrix Contained in Spatial Statistical Models; in Airlinghaus, Sandra L. (ed.): Practical Handbook of Spatial Statistics, CRC Press, Florida, ch. 4.

- Hanson, Gordon H. (2005): Market potential, increasing returns and geographic concentration; Journal of International Economics 67:1, 1 -24.
- Harris, Chauncy 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.
- Harris, R., J. Moffat and V. Kravtsova (2011): In Search of ‘W’; Spatial Economic Analysis, 6:3, 249-270.
- Head, Keith and Thierry Mayer (2000): Non-Europe: The magnitude and causes of market fragmentation in the EU; Weltwirtschaftliches Archiv. 136:2, 284-314.
- Head, Keith and Thierry Mayer (2004): The Empirics of Agglomeration and Trade; Handbook of Urban and Regional Economics 4, chapter 59; edited by V. Henderson and J.F. Thisse.
- Head, Keith and Thierry Mayer (2006): Regional wage and employment responses to market potential in the EU; Regional Science and Urban Economics 36:5, 573-594.
- Head, Keith and Thierry Mayer (2011): Gravity, market potential and economic development; Journal of Economic Geography 11:2, 281-294.
- Hering, Laura and Sandra Poncet (2009): The impact of economic geography on wages: Disentangling the channels of influence; China Economic Review 20:1, 1-14.
- Huber, P., M. Pfaffermayr and Y. Wolfmayr (2011): Are there border effects in the EU wage function? ; DANUBE Law and Economics Review 2, 23-41.
- Karahasan, Burhan Can and Enrique López-Bazo (2011): The Spatial Distribution of Human Capital: Can It Really Be Explained by Regional Differences in Market Access?; IREA Working Papers 02, University of Barcelona, Research Institute of Applied Economics.
- Keeble, D., P. L. Owens and C. Thompson (1982): Regional accessibility and economic potential in the European community; Regional Studies 16:6, 419-432.
- Keeble, D., J. Offord and S. Walker (1988): Peripheral Regions in a Community of Twelve Member States; European Commission.
- Kelejian, H. H. and I. R. Prucha (1998). A generalized spatial two stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. Journal of Real Estate Finance and Economics 17, 99-121.
- Kosfeld, R. and H.-F. Eckey (2010): Market access, regional price level and wage disparities: the German case; Jahrbuch für Regionalwissenschaft 30:2, 105-128.
- Krugman, Paul (1991): Increasing returns and economic geography; Journal of Political Economy 99:3, 483-99.
- Le Gallo, Julie. and Yiannis Kamarianakis (2011): The Evolution of Regional Productivity Disparities in the European Union from 1975 to 2002: A Combination of Shift–Share and Spatial Econometrics, Regional Studies 45:1, 123-139.
- LeSage, J.P. and R.K. Pace (2009): Introduction to Spatial Econometrics; Taylor & Francis, Boca Raton, FL, USA.

- Linders, G.-J. M., M. J. Burger and F. G. Van Oort (2008): A rather empty world: the many faces of distance and the persistent resistance to international trade; Cambridge Journal of Regions, Economy and Society 1:3, 439-458.
- Lopez-Rodriguez, Jesus (2007): Is market potential robust in explaining regional educational levels in Europe?; Applied Economics Letters, 14:7, 489-492.
- Lopez-Rodriguez, Jesus and Andres Faina (2006): Does distance matter for determining regional income in the European Union? An approach through the market potential concept; Applied Economics Letters 13:6, 385-390.
- López-Rodríguez, Jesús and Andrés Faíña (2007): Regional Wage Disparities in the European Union: What role for Market Access; Investigaciones Regionales 11, 5-23.
- Lopez-Rodriguez, J., A. Faina and J. Lopez Rodriguez J. (2007): Human Capital Accumulation and Geography: Empirical Evidence from the European Union; Regional Studies, 42, 217-234.
- Lopez-Rodriguez, Jesus and Andres Faina (2004): The Economic Geography of EU Income: Evolution since the Eighties, Research papers in Environmental and Spatial Analysis, december.
- Lopez-Rodriguez, J., A. Faina and A. Garcia Lorenzo (2007): The Geographic Concentration of Population and Income in Europe: Results for the Period 1984-1999; Economics Bulletin 18:7, 1-7.
- Millo, Giovanni and Gianfranco Piras (2012): Splm: Spatial Panel Data Models in R. Journal of Statistical Software, 47:1, 1-38. http://www.jstatsoft.org/v47/i01/.
- Mur, Jesús and Ana Angulo (2009): Model selection strategies in a spatial setting: Some additional results; Regional Science and Urban Economics 39:2, 200-213.
- Mur, J., M. Herrera and M. Ruiz (2011): Selecting the W Matrix: Parametric vs Non Parametric Approaches; European Regional Science Association conference papers 1055.
- Niebuhr, Annekatrin (2006): Market access and regional disparities. New economic geography in Europe; The Annals of Regional Science 40:2, 313-334.
- Paelinck, J.H.P. (1967): L'efficacité des mesures de politique économique régionale; Rapport introductif; Faculté des Sciences Economique, Centre de Recherches, Namur, Belgium.
- Paelinck, J.H.P. and L.H. Klaassen (1979): Spatial Econometrics; Saxon House, Farnborough, UK.
- Redding, Stephen and Anthony J. Venables (2004): Economic geography and international inequality; Journal of International Economics 62:1, 53-82.
- Redding, Stephen and Peter K. Schott (2003): Distance, Skill Deepening and Development: Will Peripheral Countries Ever Get Rich? ; Journal of Development Economics 72:2, 515-41.

- Redding, Stephen (2011): Economic Geography: a Review of the Theoretical and Empirical Literature, in Bernhofen, D., R. Falvey, D. Greenaway and U. Kreickemei (eds.): Palgrave Handbook of International Trade; Palgrave Macmillan.
- Rich, D. (1980): Potential models in Human Geography, Geoabstracts 26, Norwich.
- Rodriguez-Pose, Andres (2011): Economists as geographers and geographers as something else: on the changing conception of distance in geography and economics; Journal of Economic Geography 11:2, 347-356.
- Stewart, Jonh Q. (1947): Empirical Mathematical Rules Concerning the Distribution and Equilibrium of Population; Geographical Review 37, pp. 461-485.
- Tappeiner, G., C.H. Hauser and J. Walde (2008): Regional knowledge spillovers: Fact or artifact?; Research Policy 37:5, 861-874.
- Tobler, Waldo (1970): A computer movie simulating urban growth in the Detroit region Economic Geography 46, 234–40.
- Vieira, E., I. Neira and E. Vazquez (2011): Productivity and innovation economy: comparative analysis of European nuts II, 1995–2004; Regional Studies 45:9, 1269-1286.