Market Potential and the curse of distance in European regions

Fernando Bruna*, Andrés Faíña* and Jesus Lopez-Rodriguez * **

* Competition and Development Research Group, Economics and Business Department, University of A Coruña, Campus de Elviña s/n, 15071 A Coruña, Spain

** IPTS, European Commission DG Joint Research Centre, Inca Garcilaso s/n, 41092 Seville, Spain

Abstract

In the context of the New Economic Geography (NEG) wage equation, the 'curse of distance' is the tendency of peripheral regions to have lower income because of being far from the main markets, as captured by a variable *Market Potential*. This pattern is consistent with the core-periphery spatial distribution of the European regional economic activity. Nevertheless, during the last decades, the European Union has been implemented active transport and regional policies, which should mitigate the consequences of peripherality. This paper analyzes the changes of the cross-sectional effects of Market Potential on the European regional income per capita during the sample period 1995-2008.

The paper finds evidence that the cross-sectional elasticity of per capita income to Market Potential has been decreasing over the sample period. However, some results are sensitive to changes in the specification of the wage equation or the estimation method.

Key words:

NEG, wage equation, distance, core-periphery, regional policy, European regions

JEL codes: C21, F12, R11, R12

Corresponding author: Fernando Bruna (f.bruna@udc.es). University of A Coruña, Economics and Business Department, Campus de Elviña s/n, 15071 A Coruña, Spain. Phone number: +34 981167000. Fax number: + 34 981167070.

Acknowledgments - This paper is part of the PhD. Thesis of Fernando Bruna. We thank the helpful comments of Isabel Neira, Frank van Oort and three anonymous referees on previous drafts. We are grateful to participants in the presentation of this paper at the XXXVIII Conference of Regional Studies of the Spanish Regional Science Association (Bilbao, November 2012), specially to Francisco Requena. The views expressed are purely those of the authors and may not under any circumstances be regarded as stating an official position of the European Commission.

1. Introduction

The so-called *wage equation* of the New Economic Geography (NEG) predicts that peripheral regions tend to have lower income because of their lower access to the main international markets, as captured by a variable *Market Access* or *Real Market Potential*. Consistently with this 'curse of distance', an expression coined by Boulhol and de Serres (2010), the regional spatial distribution of economic activity in Europe follows a core-periphery pattern that has been studied by Clark *et al.* (1969) and Keeble *et al.* (1982) or Faíña and López-Rodríguez (2005).

Income per capita is negatively correlated with geographically peripherality in Europe. However, during the last decades, the European Union has been implemented active transport and regional policies, which should mitigate the consequences of peripherality. Using the NEG framework, the goal of this paper is study if the cross-sectional effects of distance on the European regional income per capita have been decreasing during the last years.

This goal is related with three strands of the empirical literature. First, the NEG wage equation appears to be empirically very successful (Redding, 2011). Inside this framework, some works have analyzed the effects of peripherality (Redding and Schott, 2003; Redding and Venables, 2004; Boulhol and de Serres, 2010). Several authors have estimated a wage equation for the European regions in different periods of time (Breinlich, 2006; López-Rodríguez and Faíña, 2006). However, different from the current paper, these works have not focused on the evolution of the cross-sectional wage equation over time or have not conducted a sensitivity analysis of their results. Second, there exists a vast literature on economic convergence in the European regions (Monfort, 2008; Borsi and Metiu, 2013), analyzing the patterns of economic growth in relation with the initial levels of income. However, if peripherality is associated with lower levels of income per capita, this literature is closely related with the debate about the curse of distance. In the present paper both economic and geographical peripherality are simultaneously considered.

A third strand of the empirical literature is the one studying the so-called 'missing globalization puzzle' or 'distance puzzle' (Disdier and Head, 2008). This debate refers to the estimation of a non-decreasing elasticity of trade to distance in spite of globalization. A number of different explanations or qualifications of the 'puzzle' has been proposed in the trade literature. The distance puzzle is about the effects of distance on trade over time while the curse of distance is a trade based prediction about the effects of distance on income. The debate about the distance puzzle directly affects the estimation of a wage equation over time when the *Market Potential* variable is built with the results of gravity equations estimated for bilateral trade data, as in Redding and Venables's (2004) methodology. However, both Breinlich (2006) and Head and Mayer (2006) find similar results when estimating a wage equation for the European regions using Redding and Venables's method or the more parsimonious measure of Market Potential defined by Harris (1954). The approach followed in this paper avoids the problems of interpretation derived by the distance puzzle debate making use of Harris's measure to successively estimate a cross-sectional wage equation for the European regions from 1995 to 2008.

Harris's (1954) *Market Potential* is an inverse distance weighted spatial lag of the income of all the others regions considered in the sample. Given that the weighting scheme of the variable is the same for any period, possible different estimates of Market Potential when the wage equation is estimated for different years can be directly interpreted as signs of a changing effect of distance. At least, the estimation of those time-varying parameters is a useful first approximation to the analysis of the curse of distance over time.

This analysis has its own limitations. In the NEG literature distances proxy trade costs, but all the empirical estimations of the wage equation using measures of distance are affected by other possible meanings of *distance* (Rodríguez-Pose, 2011; Yotov, 2012), such as informational interaction or cultural proximity. Additionally, Market Potential is a spatial lag of income, which is the dependent variable, or a closely related variable, in the wage equation. Therefore, Lopez-Rodriguez *et al.* (2014) argue that the estimation of the impact of Market Potential should be assessed through total effects, as LeSage and Pace (2009) emphasized for spatial autoregressive models. This issue has been largely ignored by the previous literature and is beyond the scope of the present work. The attention of this research is focused on analyzing the robustness of the results for different specifications of the wage equation, the inclusion or exclusion of a proxy for the internal market size, the use of instrumental variables and the estimation of standard Spatial Econometrics models.

In this paper it is presented a baseline specification of the wage equation which includes two control variables of physical and human capital. This specification allows to estimate the direct effects of Market Potential (Boulhol *et al.*, 2008; Breinlich, 2006) which are lower than when the estimation omits control variables. A baseline wage equation with control variables has three advantages. First, the control variables could partially collect the exogenous effects of the European regional and transport policies. Second, it is a prudent approach to quantify the effects of Market Potential. Third, it allows obtaining a range of estimated effects of Market Potential when the analysis is repeated omitting those control variables.

A common problem to the three lines of research previously mentioned (wage equation, convergence and distance puzzle) is that the results are sample dependent. An original contribution of the present work is to study the evolution over time of the cross-sectional effects of Market Potential for a full sample of 206 regions and for four 'regimes' of regions defined as Poor-Rich and Central-Peripheral. The focus on the curse of distance makes to pay special attention to the latter subsample.

This paper finds evidence that the cross-sectional elasticity of per capita income to Market Potential is decreasing along the period analyzed. However, some results are sensitive to changes in the specification of the wage equation or the estimation method.

The rest of the paper is organized as follows. The next section briefly introduces the theoretical framework and discusses the econometric strategy. Section 3 presents the data and the four regimes of regions. Section 4 illustrates the relation between income per capita and distance in the European regions, and the relationship between the curse of distance and economic convergence. Section 5 shows the baseline pooled OLS estimations for the sample period 1995-2008. In section 6 a cross-sectional estimation is corrected for residual spatial autocorrelations and three instrumental variables of Market Potential are studied. Section 7 presents the time-varying cross-sectional estimations of the wage equation. A final section concludes and an Appendix explains data details.

2. Theoretical and econometric framework

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

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

Following Head and Mayer (2006), the expression between brackets is called *Real Market Potential* (RMP_i) here. Redding and Venables (2004) call it *Market Access*. Krugman (1992, 1993) emphasized the similarity of this expression with Harris's (1954) measure of Market Potential: $HMP_i = \sum_{j=1}^{R} T_{ij}^{-1} E_j$. For this last indicator Harris's (1954) or Clark *et al.* (1969) carefully estimated transport cost though data restrictions frequently force to proxy trade costs with physical distances (see the data Appendix for alternatives). In Harris's index, E_j is usually a measure of the size of the market. As Combes *et al.* (2008) summarize, in order to go from RMP_i to HMP_i it is necessary to assume that the share of manufacturing goods on expenditure is the same in all regions ($\mu_j = \mu = 1$), the same that Fujita *et al.* (1999, chap. 4) did, and that $T_{ij}^{1-\sigma} = T_{ij}^{-1}$. When trade costs are proxied by geographical distances ($T_{ij} = d_{ij}$) this last assumption can be justified by the robust finding in the gravity equations literature of a trade elasticity to distance close to -1 (Head and Mayer, 2015). Therefore, the main difference of HMP_i with respect to RMP_i is the absence of an adjustment for variation in the competition index ($S_j = S = 1$), which is not directly measurable. The next section justifies why HMP_i is preferred in this work over alternative proxies of RMP_i utilized in NEG empirical literature.

A standard wage equation, such as equation (1), has been extended by Head and Mayer (2006) to control for human capital. A similar approach can be followed to include capital stock per worker. A version of the cross-sectional wage equation in logarithmic form for region $i = 1, \dots, n$ can be:

$$ln\, w_i = \alpha + \beta_1\, ln\, k_i + \beta_2\, ln\, h_i + \beta_3\, ln\, RMP_i \quad (2)$$

where w_i are wages, k_i is per capita capital stock, h_i is per capita human capital stock. Equation (2) has an intercept (α) derived from the parameters of the model that are assumed to be common in all regions. The control variables can be considered as proxies for exogenous regional technological differences or for exogenous effects of regional and transport policies.

Generalizing the notation, an econometric version of the cross-sectional equation (2) generalized to pooled data of T periods can be represented as:

$$y_{it} = \alpha + \beta' x_{it} + u_t + u_{it} \quad (3)$$

where $t = 1, \dots, T$ and u_t are $T - 1$ possible common shocks to all regions in each period. The term u_{it} collects the effects of omitted variables and departures from the assumptions of the theoretical model. In order to study how the coefficients of the x_{it} explanatory variables change in time, the cross-sectional equation (2) can be estimated T times to obtain a time series of β. The time-varying version of equation (3) estimated year by year can be represented as:

$$y_{it} = \alpha_t + \beta_t' x_{it} + \varepsilon_{it} \quad (4)$$

For each of the cross-sectional estimations, ε_{it} is supposed to be spatially uncorrelated in order to apply OLS. In section 6 a cross-sectional Spatial Error Model (SEM) will be estimated to correct the model for residual spatial autocorrelation, which in matrix notation is:

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

where W is the spatial weights matrix. An alternative spatial model is the Spatial Autoregressive (SAR) model: $Y = \rho WY + X\beta + u$. For reasons that will become clear below, the results of a SAR model will not be shown. However, the SAR model is particularly relevant because Market Potential is a type of spatial lag of the dependent (or a very related) variable. Therefore the impact of the explanatory variables should be calculated through total effects (LeSage and Pace, 2009, chap. 2), even when the model does not include an additional spatial lag of the dependent variable. This issue has been omitted in the previous literature and it is not further investigated in this paper in order to limit its scope. At the moment, the expression "elasticity" is used here as in standard OLS regressions with variables in logarithms. However, the results in section 6 will confirm that some problems can appear when two different types of spatial dependence are simultaneously considered in a Spatial Econometrics model including Market Potential.

Closely related to the previous issue is the endogeneity of Market Potential, which has been broadly discussed in the NEG empirical literature. See, for instance, Redding and Venables (2004), Breinlich (2006) or Head and Mayer (2006). Endogeneity is particularly severe if the variable of Market Potential includes a measure of the internal market sizes. The interpretation of possible changes over time of the cross-sectional estimate of Market Potential in a wage equation is affected by this issue. Therefore, the estimation with instrumental variables will be considered too. Additionally, the effort of the paper is oriented to analyze the robustness of the estimates with respect to different specifications, the consideration of the internal markets and the estimation of basic spatial models.

3. Data, measurement issues and samples of regions

The global sample studied in this paper consists of 206 European regions since the year 1995 to the year 2008. Table 1 summarizes the pooled means of the main variables (in levels) to be used in the later empirical analysis.

In a similar way to some other NEG empirical research (Redding and Venables, 2004; Brakman *et al.*, 2009), wages are proxied by per capita income, measured as gross value added per capita (GVA). Breinlich (2006) argues that proxying wages by GVA per worker is innocuous as long as labor's share in GVA does not vary across locations or at least not in a way systematically related to Market Potential. However, the *per capita* version is preferred here because it provides generality to the discussion. The wage equation has been broadly interpreted in terms of a relationship between Market Potential and the spatial distribution of economic activity (Redding, 2011), instead of the nominal manufacturing wages of the basic NEG models using simple definitions of marginal costs.

Human capital is proxied by the share of the population who has successfully completed education in Science and Technology (S&T) at the third level and is employed in a S&T occupation. In order to avoid jumps in the time-varying estimates due to different sample composition, missing data in this variable were imputed with a polynomial of degree 2 on the regional time trend of each region.

The Real Market Potential (RMP_{it}) of region i in time t is proxied by a Harris's (1954) measure of Market Potential, defined as the inverse distance (d_{ij}) weighted sum of the GVA of the regions j accessible to i. Given that the calculation is done on areal units, a correction for the size of the internal market of each area (self-potential) is necessary in order to measure the accessibility of its firms to the markets. Therefore, considering the $R-1$ possible markets of other j regions, the Harris's Market Potential (HMP_{it}) of region i can be decomposed into its *Internal* (IMP_{it}) and *External* (EMP_{it}) components:

$$HMP_{it} = \sum_{j=1}^{R} \frac{GVA_{jt}}{d_{ij}} = \frac{GVA_{it}}{d_{ii}} + \sum_{j \neq i}^{R-1} \frac{GVA_{jt}}{d_{ij}} = IMP_{it} + EMP_{it} \tag{6}$$

where the distance to the own regional market (d_{ii}) is measured by within region distances, as it will be discussed below. The calculation of Market Potential includes the regions of Norway and Switzerland though they are excluded from the sample because of lack of capital stock data (see Figure 2 below).

Alternatively, Redding and Venables (2004) built a measure of RMP_{it} (*Market Access*) proxying the NEG competition index (S_j) by unobserved importer fixed effects. These effects were estimated using gravity equations for bilateral trade. In order to analyze time-varying effects of Market Potential by subsamples of regions, Harris's index is preferred here because of four reasons. First, Harris's measure keeps the same weighting scheme across time and space while Redding and Venables's approach presents comparability difficulties. For instance, the calculations by these last authors are based on an estimated trade elasticity to distance which is 2.5 times the one estimated by Boulhol and de Serres (2010). Breinlich's (2006) definition of Market Access includes a measure of income absent in Redding and Venables's measure. These issues are crucial when comparing different time periods because they determine what can change in the definition of the variable and the weight of that (possibly time-varying) component.

Second, the method of Redding and Venables (2004) is based on trade data what implies two difficulties. On one hand, the time-varying estimation of a wage equation get mixed with the so-called 'missing globalization puzzle' or 'distance puzzle' (Disdier and Head, 2008). This debate refers to the estimation of a non-decreasing elasticity of trade to distance in spite of globalization. A large trade literature has been following different approaches to solve or qualify the 'puzzle'. The diversity of explanations create problems to interpret the results of a changing cross-sectional estimate of Market Access, while Harris's simple measure can offer useful initial insights. On the other hand, when working with regional data, Redding and Venables's method requires additional simplifying assumptions due to the lack of inter-regional trade data (Breinlich, 2006; Head and Mayer, 2006).

Third, in spite of the NEG interpretation of geographical distances as an indicator of trade costs, the meaning of distance is not clear. Physical distances proxy not only trade costs but 'relative' trade costs (Yotov, 2012) and capture non-trade-related barriers (Linders *et al.*, 2008) and regional characteristics, interactions and spillovers (Rodríguez-Pose, 2011) too. Even when working with trade data, the estimation of a wage equation with any measure of Market Potential based on distances is sensitive to these factors. Harris's approach makes it transparent and facilitates to focus on the direct effects of relative location on income.

Fourth, Harris's Market Potential is more parsimonious than other proxies of Real Market Potential. Both Head and Mayer (2006) and Breinlich (2006) obtain similar results for the European regions when comparing the approaches of Redding and Venables (2004) and Harris (1954).

Common to all the empirical methods for proxying RMP_{it} is the problem of measuring the Internal Market Potential, here defined as $IMP_{it} = GVA_{it}/d_{ii}$. Different measures of the internal market size have been proposed in the literature. A standard approach is to assume that regions are circular so the radius of region i is $r_i = \sqrt{area_i/\pi}$. In this paper internal distances are measured following Keeble *et al.* (1982), who chose $d_{ii} = 1/3 \cdot r_i = 0.188\sqrt{area_i}$ to allow for the likely clustering of economic activity in and around the regional 'centre'. This is similar to the 40% of the radius considered by Cambridge Econometrics (2014).

Table 1. Summary statistics by regime: pooled means 1995-2008 and average economic growth

	All	*Rich*	*Poor*	*Central*	*Peripheral*
GVA per capita	20,295	24,957	15,632	22,593	17,997
Market Potential	16,814	20,658	12,969	22,502	11,125
External Market Potential	14,159	16,593	11,724	18,807	9,510
Weight of the internal market in Market Potential (%)	13.1	16.5	9.7	13.5	12.6
Capital stock per capita	72,412	87,367	57,457	80,300	64,524
Human capital (core variable of S&T, % of population)	9.0	10.0	7.9	9.5	8.4
Average distance to *All* the other regions (km)	1,112	972	1,252	854	1,370
Area (km2)	15,669	12,417	18,920	9,270	22,067
Growth rate GVA per capita (annual average % 1995-2008)	1.8	1.7	1.9	1.7	2.0

Notes: 206 European regions (*All*) split into four subsamples of 103 regions each of them according to the median log of GVA per capita in 1995 (*Rich-Poo*r) and the median log of the average regional distance to the other regions (*Central-Peripheral*). The variables in the table are not log transformed. GVA and capital stock per capita are in 2000 year euros. Market Potential is in millions of 2000 euros. Human capital is proxied by the Eurostat's core variable of human resources in Science and Technology (S&T). See the data Appendix for details.

Exceptionally in the NEG empirical literature, Boulhol *et al.* (2008) analyze the weights of the internal component in their variable of Market potential. However, they do not report the absolute values of those weights. Table 1 shows the pooled mean of the weight of IMP_{it} on HMP_{it} when internal distances are calculating as $1/3$ of the radius of circular regions. For the sample of 206 regions (*All*) regions, the average share is 13.1%. However, this number is affected by a few regions with big cities. In the year 2008, the weight of Internal Market Potential in Market Potential is higher than 40% for the regions of Stockholm, Brussels, Berlin, Hamburg, Madrid, Paris, Vienna, Athens and (Inner) London. Therefore, a better indicator of the effects of the chosen methodology to measure the internal markets is the median weight of IMP_{it} on HMP_{it}. The pooled median weight of IMP_{it} is 9.9%. When the internal distances are measures as $2/3$ of a circular region, as it is frequently done in the empirical NEG research, that median is 5.2%. The approach of Keeble *et al.* (1982) to measure the internal distances as $1/3$ of the radius is preferred here to the $2/3$ alterative because it allows a higher differentiation between HMP_{it} and EMP_{it}.

The presence of IMP_{it} in the measurement of Market Potential does not only generate a huge endogeneity problem in the data of those regions (domestic GVA in both sides of the equation), but makes more difficult to interpret the time-varying estimates of Market Potential in terms of location. However, omitting the internal markets introduces measurement error by reducing the access measure of some economically larger locations (Breinlich, 2006; Head and Mayer, 2006). Therefore, the $1/3$ approach to internal distances allows establishing a broader range of results than the $2/3$ approach for the robustness analysis of an estimated wage equation with respect to the measurement of the internal markets.

Table 1 also shows the average levels of the variables for 4 subsamples of regions. The curse of distance is mainly an issue about the economic development of peripheral regions. Therefore, the sample of 206 European regions is split into four 'regimes' conceived as meaningful groups of regions according to two criteria: economic development and peripherality. On one hand, two economic regimes are defined depending of having a log of per capita GVA in 1995 over or under the sample median that year: 'Rich' or 'Poor' regions. On the other hand, two geographical regimes are defined depending of their log average distance to all the other regions being under or over the median: 'Central' or 'Peripheral' regions. Given that there is no objective dividing line about richness or peripherality, the medians are preferred over the means in order to obtain the same number of observations in each regime.

As pointed out to the authors by a referee, it can be argued that this division in regimes creates a problem of censored data because the range of variation of the dependent variable is limited with endogenous criteria. Indeed, the initial value of GVA per capita in the economic regimes is endogenous and the spatial distribution of GVA is closely related with the endogeneity problems of the Market Potential variable. However, an implicit economic criterion is always present in sample selection, particularly when the research is based on a specific geographical area. Any empirical result is the consequence of decisions about the data aggregation level, variables to study (availability) and singularities to consider (such as islands or possible 'Nordic' or 'Eastern' European regimes). Here the focus of attention is on comparing the possible different effects of Market Potential on economic development among four specific groups of regions over time. The next section provides details about their spatial structure.

4. The European spatial core-periphery pattern and regional convergence

The curse of distance is consistent with the core-periphery pattern of the spatial distribution of economic activity in the European regions. As the NEGs predicts, the data in Table 1 confirms that the European Peripheral regions tend to be poorer than then Central ones. For the cross-section of the year 2008, the following figures represent the spatial distribution of the dependent variable in the wage equation under scrutiny, the logarithm of GVA per capita. In Figure 1 this variable is plotted against the average distance from each region to all the other regions in the sample. The economically central regions (high log GVA per capita) appear to be geographically central too (low average distance). Therefore, the regression line in the plot is negatively sloped[1].

The relation between economic centrality and geographic centrality can be observed on a cloropheth map in spite of the visual distortion created by the heterogeneous size of the regions. Figure 2 shows the maps of the logarithms of GVA per capita and Market Potential in Europe in the year 2008. Their values are divided in seven quantiles, which helps to visualize their global

[1] In Figure 1 the average distance is not log-transformed to facilitate the interpretation of the horizontal axis. The estimate of log average distance on the regression of log GVA per capita is -0.59 (not shown).

spatial pattern. Darker colors are associated with higher values of the variables. The left map of Figure 2 shows that there are only a few high per capita income regions out of the geographical center of Europe, particularly those in Nordic countries. The economically central regions are mainly located around the so called *blue banana*, from West England in the North to Milan in the South. Given the spatial structure of GVA in Europe and the construction of Market Potential as a (weighted) sum of GVA, the logarithm of a Harris's measure of Market Potential shows an even more concentrated distribution and a clearer core-periphery pattern. Indeed, Table 1 shows that the pooled mean of (External) Market Potential before logs for Peripheral regions is half of the mean for the Central ones. This characteristic allows Market Potential to capture in a stylized way the global core-periphery pattern of per capita GVA.

Figure 1. Log GVA per capita (year 2008) versus average distance to the other regions

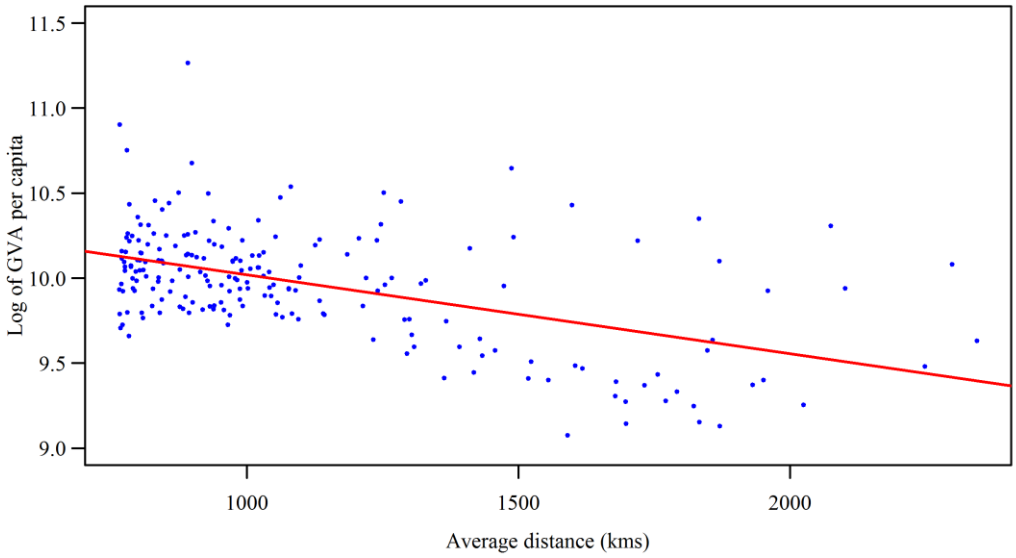


Figure 3 shows the values of the log of per capita GVA in the year 2008 for the four regimes. Again, for each regime those values are divided in seven quantiles. The key issue is the general similarity of the maps of the Poor and Peripheral subsamples, in spite of the arbitrary criteria of the medians used to classify the regimes.

Some of the (darker) relatively rich regions in the regime Poor are regions with high economic growth during the sample period. Table 1 shows the distribution of the average annual growth rate 1995-2008 of per capita GVA by regimes. Poor regions had higher average economic growth than Rich ones, which implies absolute convergence. Peripheral regions had higher growth than Central ones. This means that Peripheral regions are escaping the curse of distance, though the spatial distribution of economic activity in Europe continued to present a core-periphery pattern in the year 2008.

Both issues, regional convergence and the curse of distance, are closely related in the European case. The time-varying estimation to be presented below allows studying how these issues affect the estimation of a European regional wage equation.

Figure 2. Cloropleth maps of the logs of GVA per capita and Market Potential (year 2008)

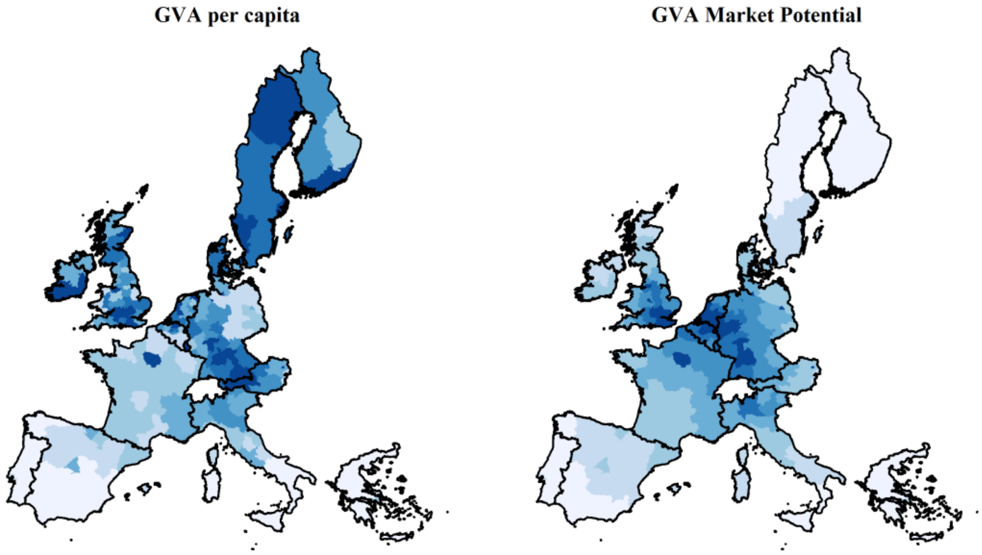
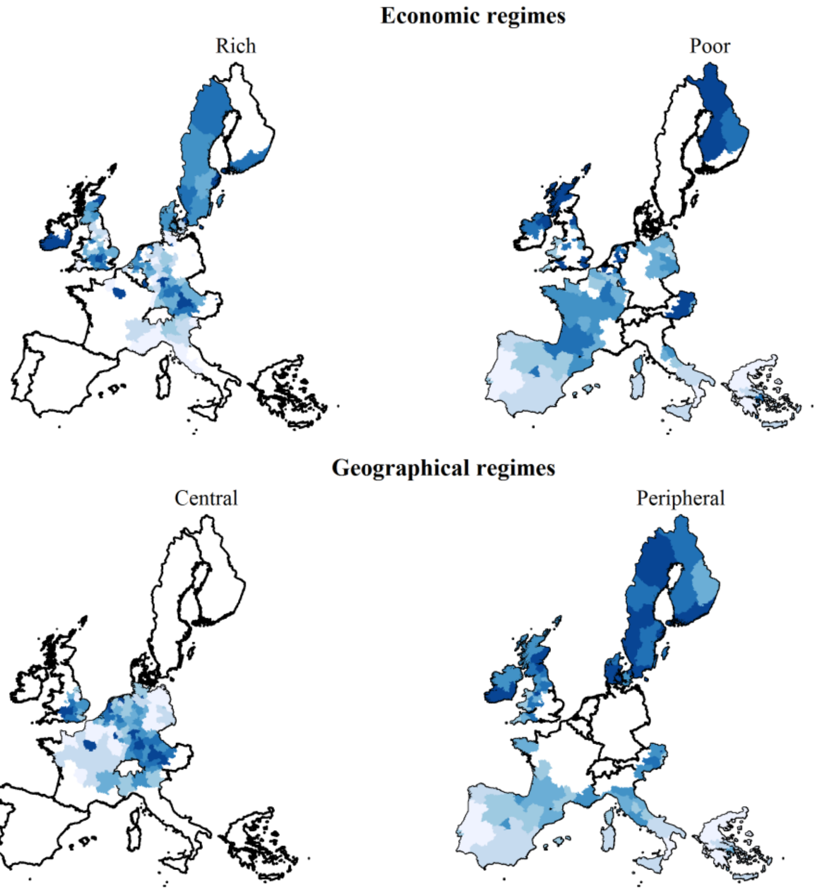


Figure 3. Cloropleth maps of the log of GVA per capita by regimes (year 2008)

(Rich/Poor by year 1995 values)



5. Baseline pooled estimation: global sample and regimes

Table 2 analyzes the pooled equation (3) by parts[2]. Each row reports two standard errors: those estimated by OLS (above) and clustered standard errors (below) that allow for heteroskedasticity and serial correlation of arbitrary form (Arellano, 1987). Columns (1) and (2) of Table 2 show that capital stock per capita and Market Potential by themselves produce an adjusted coefficient of determination of 0.73 and 0.42, respectively. If human capital is the only explanatory variable that coefficient is 0.34 (not shown). The estimated elasticity of GVA per capita to Market Potential decreases from 0.42 when the latter is the only explanatory variable to 0.15-0.18 when control variables are considered. The inclusion of time effects is supported by Lagrange Multiplier tests but it does not have great influence on the pooled estimates. Column (7) shows that imputing missing data in the variable of human capital (see the Appendix) does not alter the results. The clustered standard errors in columns (6) and (7) show that the human capital variable loses significance when country dummies are introduced in the regression: the impact of human capital is partially due to country characteristics. However, the estimate of Market Potential is not very sensitive to the inclusion of country dummies. For the baseline model to be presented in Table 3 the specification with human capital is preferred.

In columns (5) and (6) of Table 2 Market Potential is replaced by its value lagged one period. Using lagged values has been done in the literature as a way of reducing endogeneity problems (Redding and Venables, 2004). However, this simple test with Market Potential lagged one year reveals that the results do not change in spite of losing the year 1995. The reason is that the pooled estimates are dominated by the cross-sectional relative values of the variables in levels, which are similar from one year to another one. The robustness of the results with respect to the inclusion of a proxy for the internal markets in Market Potential will be analyzed in the following sections.

Table 3 shows the baseline pooled models that will be the reference for the time-varying estimation below. The estimations by regimes reveal some differences. For instance, human capital in the regime 'Peripheral' might be able to collect some North-South differences of per capita GVA. Focusing on Market Potential, the benchmark estimate in the first column of Table 3 is 0.16[3]. According to the clustered standard errors Market Potential is not significant in the Rich regime, probably because this regime includes some rich Peripheral regions, which have low Market Potential. For the other three regimes the elasticity of GVA per capita to Market Potential ranges from 0.21 to 0.31. Therefore, considering all the results, it is possible to conclude that a rough pooled OLS estimate of the direct cross-sectional 'effect' of Market Potential is around 0.2.

[2] R's plm package (Croissant and Millo, 2008) has been used.

[3] As it will be shown in Table 6 below, if Table 3 is repeated using only the external component of Market Potential, the estimate for the sample with all the regions is 0.14.

Table 2. Pooled OLS estimation of alternative specifications (1995-2008, 206 EU regions)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(Intercept)	0.759	5.834	0.284	1.371	1.418	-0.088	-0.020
	(0.103)***	(0.087)***	(0.095)**	(0.117)***	(0.123)***	(0.133)	(0.115)
	(0.392)	(0.443)***	(0.334)	(0.392)***	(0.409)***	(0.671)	(0.604)
Capital stock per capita	0.818		0.709	0.669	0.670	0.764	0.756
	(0.009)***		(0.010)***	(0.011)***	(0.011)***	(0.011)***	(0.010)***
	(0.035)***		(0.034)***	(0.033)***	(0.034)***	(0.064)***	(0.060)***
Human capital				0.162	0.171	0.085	0.075
				(0.009)***	(0.009)***	(0.010)***	(0.009)***
				(0.030)***	(0.030)***	(0.034)*	(0.029)**
Market Potential		0.420	0.183	0.158	0.154	0.174	0.175
		(0.009)***	(0.006)***	(0.007)***	(0.007)***	(0.008)***	(0.007)***
		(0.045)***	(0.027)***	(0.026)***	(0.026)***	(0.052)***	(0.051)***
Year dummies?	No	No	Yes	Yes	Yes	Yes	Yes
Country dummies?	No	No	No	No	No	Yes	Yes
R-squared	0.730	0.425	0.798	0.831	0.824	0.929	0.929
Adj. R-squared	0.730	0.424	0.793	0.825	0.819	0.918	0.919
F	7804	2128	754	793	767	1110	1238
Sum sq. errors	88.86	189.53	66.66	52.39	50.38	20.20	23.50
N	2884	2884	2884	2605	2472	2472	2884

Notes: Table displays coefficients and two standard errors between brackets, those estimated by OLS (above) and Arellano's (1987) clustered standard errors (below). The coefficients are * significant at 10% level; ** at 5% level; *** at 1% level. The variables are in logarithmic form. The dependent variable is gross value added per capita. In columns (5) and (6) Market Potential is replaced by its values lagged one year for each region. In Column (7) missing data in human capital were imputed. See the data Appendix.

Table 3. Baseline pooled OLS estimation by regimes (1995-2008, 206 regions)

	All	*Rich*	*Poor*	*Central*	*Peripheral*
(Intercept)	1.523	4.212	2.079	-0.653	1.789
	(0.111)***	(0.171)***	(0.170)***	(0.174)***	(0.166)***
	(0.404)***	(0.855)***	(0.690)**	(0.698)	(0.539)***
Capital stock per capita	0.648	0.488	0.538	0.675	0.583
	(0.010)***	(0.014)***	(0.015)***	(0.013)***	(0.014)***
	(0.032)***	(0.053)***	(0.061)***	(0.051)***	(0.042)***
Human capital	0.151	0.176	0.078	0.004	0.232
	(0.008)***	(0.010)***	(0.011)***	(0.012)	(0.011)***
	(0.028)***	(0.033)***	(0.038)*	(0.036)	(0.035)***
Market Potential	0.163	0.079	0.205	0.315	0.238
	(0.006)***	(0.008)***	(0.008)***	(0.013)***	(0.010)***
	(0.025)***	(0.044)	(0.035)***	(0.068)***	(0.031)***
R-squared	0.820	0.648	0.768	0.778	0.838
Adj. R-squared	0.815	0.641	0.759	0.769	0.829
F	816	164	295	312	462
Sum sq. errors	59.32	23.13	23.50	20.67	30.64
N	2884	1442	1442	1442	1442

Notes: See notes of Table 1 and Table 2. The specification includes year dummies. The proxy of human capital includes imputed missing values. For the whole sample of regions, column *All* repeats column (7) of Table 2 but without country dummies.

6. Cross-sectional interactions: spatial models and instrumental variables

The time-varying estimations in the next section will be done for the cross-section of each year. Before that, it is convenient to analyze two possible problems in the cross-sections that have been pooled in the models of Table 3. First, residual spatial autocorrelation would violate the OLS assumption, calling for the estimation of spatial econometric models. Second, the endogeneity of the Market Potential variable would bias the OLS results. These issues are analyzed in this section for the cross-section of a particular year, 2008.

Table 4 and Table 5 show the estimation of equation (2) by OLS and the estimation of a Spatial Error Model, as in equation (5), by maximum likelihood, as well as the second and first stages of three instrumental variables estimations[4]. The first table uses the full variable of Market Potential, including the proxy for the internal markets, while the second one uses the external component.

The significant p-values of Moran's I in column (1) of the tables show that the OLS residuals are spatially autocorrelated. Non-reported Lagrange multiplier tests reveal that the simple versions of the tests for a SEM and a SAR model, including the spatially lagged dependent variable, are significant. Only the robust test for the SEM model is significant, pointing out to an erroneously omitted spatial process in the disturbances. If the true model is a SEM, OLS estimates are not biased and very different OLS and SEM estimates would indicate problems of specification. Column (2) of Table 4 shows that the OLS and SEM estimates of Market Potential are similar, 0.16 and 0.18, respectively. On the contrary, the SEM estimate of External Market Potential in Table 5 is zero. The Lagrange multiplier tests for the specification with the external component of Market Potential actually select the SAR model. However, the estimation of a SAR model (not show) is not able to correct the model for residual spatial autocorrelation. Additionally, the estimated total effect of External Market Potential in a SAR model is 0.08, which is similar to the estimate of 0.03 in column (2) of Table 5.

As it was discussed in section 2 these results are due to the simultaneous inclusion of two types of spatial dependence, a short-distance spatial autocorrelation and a core-periphery long-distance spatial pattern in the dependent variable, captured by External Market Potential. The result of the SEM model in Table 5 invalidates a possible estimation of that specification by OLS for different years. However, non-reported estimations of the cross-sectional models in Table 5 excluding the control variables show that the OLS estimate of External Market Potential is 0.35 while the SEM estimate is 0.26, both of them being significant. Therefore, the results are very sensitive to the inclusion of control variables. The next section will compare the time-varying results when OLS is used for different specifications of the equation

The comparison of results when the internal component of Market Potential is included and omitted is relevant to analyze the issue of endogeneity too. As mentioned in section 3, Internal Market Potential introduced strong endogeneity problems. However, similarly to the spatial lag of the dependent variable in SAR models, External Market Potential is endogenous. Therefore columns (3) to (5) of the tables show the second stage of instrumental variables estimations, while the first stage is shown in columns (6) to (8).

Apart from the control variables of physical and human capital, which are considered as 'exogenous', the instruments are the Market Potential variables in the year 1991, the average distance of each region to all the other regions and the regional area. The first instrument uses data lagged 16 years while the other two instruments use purely geographic data. These instruments present shortcomings. The lagged values of Market Potential do not exclude the possibility of endogeneity in a long run relationship. The average distance implicitly determines a European

[4] The spatial estimations were done with R's spdep package (Bivand, 2014) and the instrumental variables estimation with R's AER package (Kleiber and Zeileis, 2008).

center (Head and Mayer, 2006), which happens to be located in the European *blue banana* of rich regions[5].

The regional area can potentially extract from Market Potential the endogenous component of internal markets. Indeed, column (8) of Table 4 shows a negative relationship between area and Market Potential. The rationality would be that a bigger area increases internal distances and reduces GVA density, reducing Market Potential. However, the same negative relation appears in column (8) of Table 5, which does not consider the internal markets. This is probably due to the fact that peripheral regions tend to have bigger size and lower External Market Potential than central regions (see Table 1)[6]. One of the criteria to classify regions as NUTS 2 is population. Peripheral countries tend to be less densely populated and the variable area contains information about density, which is an endogenous factor.

However, in spite of the shortcomings, this type of instruments has been used in the literature and provides both a first approach to the issue of endogeneity of Market Potential and knowledge about the characteristics of this variable.

Endogeneity tests are sensitive to heteroskedasticity error terms, so the Eicker-Huber-White covariance estimator is used in the IV estimations. The weak instrument tests confirm that the instruments are not considered weak. However, as any contextual test, the Wu-Hausman tests for the exogeneity of Market Potential are conditional to the quality of these instruments as exogenous variables. Under the cautionary remarks presented above, in this analysis endogeneity is accepted in columns (3) and (4) of Table 4. However, the estimates of Market Potential in columns (1) to (5) of Table 4 are very similar, with values around 0.15-0.18 and a slightly lower value of 0.11 in column (4). The endogeneity of External Market Potential tends to be rejected in Table 5, so the 0.14 OLS estimate in column (1) would be consistent.

In conclusion, the analysis does not reveal strong endogeneity problems provoking relevant biases in the OLS estimates of the Market Potential variables. The estimation of Spatial Econometrics models confirms Lopez-Rodriguez *et al.*'s (2014) about the important role of internal markets. If the internal markets are not considered in the estimation, External Market Potential has significance problems when the specification of a spatial model includes control variables. Added to this shortcoming is the general problem of calculating total effects when Market Potential is built with a variable closely related with the dependent variable. At the moment, the strategy followed in this paper is to focus on the robustness of the time-varying estimations with respect to the control variables and the inclusion of the proxy for the internal market size.

[5] In this sample the two NUTS 2 regions with lower average distance to the other regions are Trier, in Germany, and Luxembourg. Extending the sample to 274 regions, the geographical centre of Europe is Darmstadt, in German Hesse state. In this sense, it would be possible to say that a European regional index of Harris's Market Potential captures the peripherality with respect to the seats of the European Central Bank and the German Federal Bank, which are located in Frankfurt, Darmstadt.

[6] The negative effects or the regional areal are robust to the simultaneous inclusion of the average distance (not shown). Aside from this, Breinlich (2006) finds a positive significant effect of the region's home country area, which would capture the advantage conferred to large national markets by the trade-reducing effects of national borders. With the data bank used in the present paper, the country size does not produce positive significant effects in different specifications of the Market Potential variables.

Table 4. Cross-sectional estimations with Market Potential (year 2008, 206 regions)

	OLS (1)	ML-SEM (2)	IV second stage (3)	(4)	(5)	IV first stage (6)	(7)	(8)
(Intercept)	0.900	1.374**	0.921	1.006	0.909	1.443***	21.060***	6.638***
	(0.490)	(0.520)	(0.542)	(0.551)	(0.562)	(0.152)	(1.047)	(1.362)
Capital stock per capita	0.706***	0.646***	0.711***	0.735***	0.708***	-0.062***	0.001	0.484***
	(0.042)	(0.039)	(0.043)	(0.045)	(0.050)	(0.013)	(0.065)	(0.104)
Human capital	0.195***	0.205***	0.196***	0.203***	0.195***	0.097***	0.222***	0.162
	(0.038)	(0.038)	(0.039)	(0.038)	(0.041)	(0.012)	(0.058)	(0.099)
Market Potential	0.158***	0.183***	0.150***	0.116***	0.154**			
	(0.023)	(0.035)	(0.027)	(0.027)	(0.051)			
λ		0.715***						
		(0.054)						
Market Potential 1991						0.980***		
						(0.007)		
Average distance (km)							-1.555***	
							(0.062)	
Area (km^2)								-0.222***
								(0.025)
R-squared	0.806		0.805	0.802	0.805	0.992	0.813	0.445
Adj. R-squared	0.803		0.802	0.799	0.803	0.992	0.810	0.437
Log likelihood	104.14	156.56				343.70	16.83	-95.05
Moran's I residuals	0.534	-0.021						
Moran's I p-value	0.000	0.649						
Weak inst. F test			10252.7	836.7	61.8			
Weak inst. p-value			0.000	0.000	0.000			
Wu-Hausman F test			12.941	7.357	0.012			
Wu-Hausman p-value			0.000	0.007	0.913			
Sum squared errors	4.39	2.35	4.39	4.47	4.39	0.43	10.24	30.35

Note: Table displays coefficients and standard errors. Columns (3) to (5) include Eicker-Huber-White standard errors. Moran's tests use the normality assumption for the residuals. The alternative hypothesis for the p-values is that Moran's I is greater than expected under the null hypothesis of absence of spatial autocorrelation. The weights matrix (W) for Moran's test and the SEM estimation is a row-standardized binary matrix to the 5 nearest neighbors. The Stock and Yogo's (2005) critical value for the first-stage F-statistic weak identification test for 1 endogenous regressor, 1 instrumental variable and 10% of desired maximal size of a 5% Wald test is 16.38.

Table 5. Cross-sectional estimations with External Market Potential (year 2008, 206 regions)

	OLS (1)	ML-SEM (2)	IV second stage (3)	(4)	(5)	IV first stage (6)	(7)	(8)
(Intercept)	0.874	2.311***	0.889	0.953	0.725	1.129***	22.456***	6.546***
	(0.510)	(0.595)	(0.576)	(0.575)	(0.615)	(0.108)	(0.744)	(1.394)
Capital stock per capita	0.733***	0.703***	0.735***	0.748***	0.704***	-0.056***	-0.117*	0.428***
	(0.043)	(0.040)	(0.044)	(0.046)	(0.054)	(0.054)	(0.046)	(0.107)
Human capital	0.215***	0.259***	0.215***	0.217***	0.211***	0.054***	0.107**	0.052
	(0.039)	(0.039)	(0.039)	(0.039)	(0.042)	(0.042)	(0.041)	(0.101)
External Market Potential	0.136***	0.032	0.132***	0.111***	0.185**			
	(0.024)	(0.043)	(0.027)	(0.027)	(0.065)			
λ		0.709***						
		(0.055)						
External Market Pot. 1991						0.995***		
						(0.005)		
Average distance (km)							-1.622***	
							(0.044)	
Area (km2)								-0.186***
								(0.025)
R-squared	0.792		0.792	0.790	0.787	0.996	0.892	0.338
Adj. R-squared	0.788		0.788	0.787	0.784	0.995	0.891	0.328
Log likelihood	97.01	144.18				415.07	87.15	-99.86
Moran's I residuals	0.500	-0.038						
p-value Moran's I	0.000	0.790						
Weak inst. F test			14041.1	2760.2	39.2			
Weak inst. p-value			0.000	0.000	0.000			
Wu-Hausman F test			3.941	2.707	0.957			
Wu-Hausman p-value			0.048	0.101	0.329			
Sum squared errors	4.70	2.65	4.70	4.73	4.80	0.21	5.18	31.80

7. Time-varying estimations by year

If the possible effects of endogeneity and spatial autocorrelation in the estimated model remain constant over time, any temporal change of the cross-sectional elasticity of income per capita to Market Potential would be attributable to a change of the relative importance of location. Under this framework it is worthy to make the simple exercise of estimating the cross-sectional wage equation for different years.

Figure 4 presents the time-varying estimated elasticities of per capita GVA to Market Potential for the specifications in the pooled models of Table 3. The estimates remain pretty stable around the values of Table 3 for the five samples of regions, close to the 0.2 benchmark estimate. In spite of the convergence process discussed in section 4, this is probably due to the fact that the variables are in (log-transformed) levels and their cross-sectional dispersion (coefficient of variation) only had a slight reduction during the sample period (not shown). In particular, when stock variables are used as controls the estimates of Market Potential are not expected to present a very sloped trend. However, a slight declining trend is present in the lines of Figure 4. This is shown in Table 6.

Figure 4. Time-varying cross-sectional elasticities of GVA per capita to Market Potential (based on Table 2)

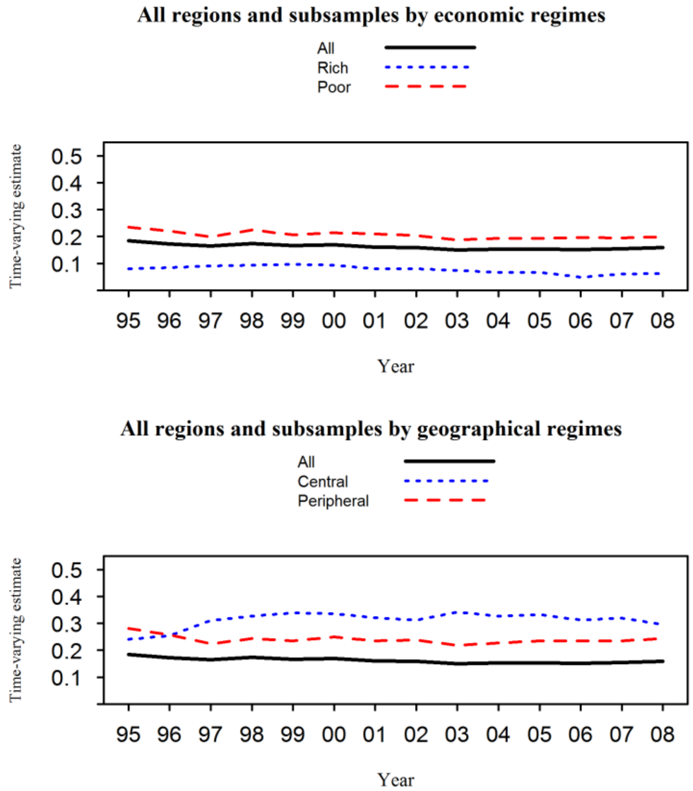


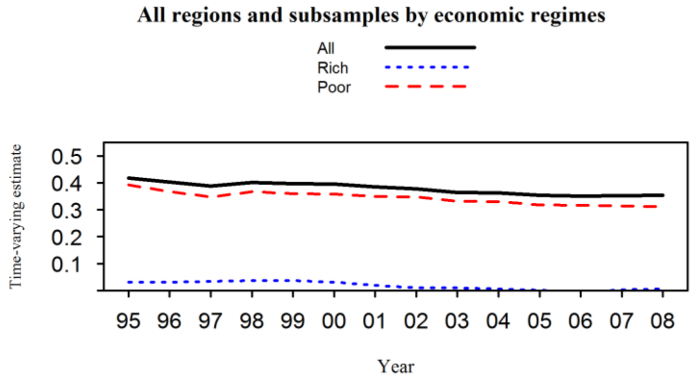
Table 6 shows a robustness analysis for different specifications when the control variables and the proxy for internal market size are omitted. The prudent benchmark pooled estimate of 0.2 is preferred in this paper to the 0.4 or 0.5 estimates reported in some columns of Table 6. However, the time-varying estimates are calculated for different specifications to check if there are contradictory results. Additionally, Table 6 shows the percentage change of the time-varying estimates of the Market Potential variables since the first to the last year of the sample period. Given that the first two or three years sometimes present extreme values, this indicator must be supplemented by the whole series of estimates, as plotted in the accompanying figures.

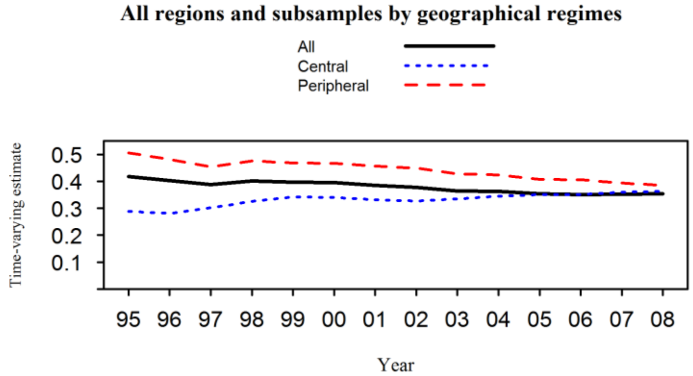
Table 6. Alternative pooled estimates of Market Potential and their cross-sectional change

Specification	Indicators	*All*	*Rich*	*Poor*	*Central*	*Peripheral*
Market Potential with capital stock per capita and human capital	Estimate	0.163***	0.079	0.205***	0.313***	0.240***
	% change	-14.4	-22.0	-15.4	22.8	-13.5
Market Potential with human capital	Estimate	0.323***	0.112	0.305***	0.522***	0.367***
	% change	-16.9	-29.9	-24.8	-13.9	-19.3
Market Potential	Estimate	0.407***	0.112	0.365***	0.499***	0.461***
	% change	-10.8	-9.1	-16.6	7.0	-16.8
External Market Potential with capital stock per capita and human capital	Estimate	0.138***	0.037	0.182***	0.222**	0.243***
	% change	-19.9	-9.4	-23.5	136.3	-19.1
External Market Potential with human capital	Estimate	0.294***	0.046	0.283***	0.280	0.373***
	% change	-18.8	-39.5	-27.6	-15.9	-19.5
External Market Potential	Estimate	0.379***	0.018	0.344***	0.333	0.443***
	% change	-15.5	-84.0	-20.5	25.4	-24.0

Notes: 'Estimate' is the pooled estimate of Market Potential or External Market Potential when using the explanatory variables in the first column (in log form) and time dummies, as in Table 3. '% change' is the percentage of change of the cross-sectional estimates for the years 2008 and 1995. The stars mean significance levels of the pooled estimates using clustered standard errors (see notes of Table 2).

Figure 5. Time-varying elasticities of External Market Potential as unique explanatory variable





Comparing the specifications with the same control variables, the omission of the internal market sizes does not change too much the estimates of Market Potential when both the full variable and its external component are significant. The pooled effects of External Market Potential in Table 6 tend to be more relevant for the regime Peripheral than for the Central one. A possible reason for this is the inverse distance weighting in External Market Potential. The GVA level of the nearest neighbors would be a more discriminatory variable for Peripheral regions that for regions located around the European geographical center, which have more similar values of Market Potential.

The cross-sectional estimate of Market Potential decreases a14,4% for the sample of *All* regions when the years 1995 and 2008 are compared, though this reduction is not obvious in Figure 4. The evidence of Table 6 shows that this decreasing trend appears for all the specifications and in all the subsamples except the Central regime.

The time-varying estimates with External Market Potential generally show more pronounced downward trends than their analogous including a measure of internal markets. Figure 5 shows the time-varying estimates with more pronounced declining trend for the Peripheral regime, which corresponds to the specification using only External Market Potential (last row of Table 6).

In summary, contrary to the distance puzzle in the trade literature, measures of Harris's Market Potential allow to identify signs of a decreasing role of distance in the determination of the cross-regional dispersion of the European per capita production or income. The evidence is not totally conclusive because of the limitations emphasized in the previous sections. However, with the exception of the regime Central, the finding of a negatively sloped trend of the cross-sectional effects of Market Potential on GVA per capita is robust to the alternatives analyzed in Table 6. Moreover, the estimates of the variable with a more direct interpretation in terms of location, External Market Potential, seem to present clearer declining trends.

For Peripheral regions this trend implies that their relative GVA per capita tends to be less related with location. This is consistent with the data of economic growth shown in Table 1. The nearest neighbors of Peripheral regions tend to be Peripheral too and have relatively low Market Potential. As Peripheral regions converge to the levels of economic development of the geographically Central regions, the GVA level of their nearest neighbors becomes less discriminant to explain their cross-sectional differences of GVA per capita. The results confirm that Peripheral regions are slowly escaping the curse of distance.

8. Conclusions

This paper analyzes the evolution of the cross-section elasticity of GVA per capita to Market Potential during the period 1995-2008 in a sample of 2006 European regions and in four subsamples ('regimes') characterized as Poor-Rich and Central-Peripheral regions. This is done under the framework of the NEG wage equation and using a Harris's (1954) measure of Market Potential.

The empirical exercise shares some of the limitations of the previous literature, such as the possible different interpretations of the meaning of distance or ignoring that Market Potential is closely related with a spatial lag of the dependent variable. However, the exercise is considered useful to study the possible changing effects of distance after several decades of European efforts on regional and transport policies. The paper focuses on analyzing the robustness of the results for different specifications of the wage equation, the inclusion or exclusion of a proxy for the internal market size, the use of instrumental variables and the estimation of Spatial Econometrics models.

A negatively sloped trend on the cross-sectional estimated effects of Market Potential on gross value added per capita seems to be a pretty robust finding. Moreover, the estimates of the variable with a more direct interpretation in terms of location, External Market Potential, present clearer declining trends. However, the evidence is not totally conclusive because is highly sensitive to the inclusion of control variables. In particular, the cross-sectional effects of External Market Potential disappear when the estimation is controlled for physical and human capital and for spatial autocorrelation. With other specifications the evidence is more solid towards a decreasing role of location to explain the relative GVA per capita of Peripheral regions. This is consistent with the peripheral regions being slowly escaping the curse of distance.

This research can be extended in several directions. The wage equation has been estimated with unobserved individual effects. Preliminary tests show that this extension requires further discussion and is current research. The exercise can be repeated using measures of Market Potential derived from trade data and considering the alternative explanations to the distance puzzle proposed by the literature. The time-varying models can be estimated with different methods, using other sets of variables or for other samples of European regions and periods.

References

Ahlfeldt, G. and A. Feddersen (2008). Determinants of Spatial Weights in Spatial Wage Equations: A Sensitivity Analysis within a Consistent Framework. *Chair for Economic Policy Working Paper* 22. University of Hamburg.

Arellano, M. (1987). Computing Robust Standard Errors for Within-Groups Estimators. *Oxford Bulletin of Economics and Statistics* 49(4), 431–434.

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

Borsi, M. T. and N. Metiu (2013). The evolution of economic convergence in the European Union. *Deutsche Bundesbank Discussion Paper* 28/2013. Deutsche Bundesbank Research Centre, Frankfurt am Main.

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

Boulhol, H., A. de Serres and M. Molnár (2008). The contribution of economic geography to GDP per capita. *OECD Journal: Economic Studies* 2008(1), 1–37.

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

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

Cambridge Econometrics (2014). *European Regional Database*. Web page. On line: http://www.camecon.com/SubNational/SubNationalEurope/RegionalDatabase.aspx

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

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

Croissant, Y. and G. Millo (2008). Panel Data Econometrics in R: The plm Package. *Journal of Statistical Software* 27(2), 1–43.

Disdier, A.-C. and K. Head (2008). The Puzzling Persistence of the Distance Effect on Bilateral Trade. *The Review of Economics and Statistics* 90(1), 37–48.

Faíña, A. and J. López-Rodríguez (2005). The economic geography of EU income: evolution since the eighties. In: Trívez, F.J., J. Mur, A. Angulo, M. Ben Kaabia, and B. Catalán (eds.), *Contributions in spatial econometrics*, ch. 9. Copy Center Digital, Zaragoza, Spain.

Fujita, M., P. Krugman and A. J. Venables (1999). *The spatial economy: cities, regions and international trade*. The MIT Press, Cambridge.

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

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

Head, K. and T. Mayer (2015). Gravity Equations: Workhorse,Toolkit, and Cookbook. In: Helpman, E., K. Rogoff, and G. Gopinath (eds.), *Handbook of International Economics* 4, pp. 131-195 (forthcoming). North Holland. Amsterdam; New York.

Honaker, J., G. King and M. Blackwell (2011). Amelia II: A Program for Missing Data. *Journal of Statistical Software* 45(7), 1–47.

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

Kleiber, C. and A. Zeileis (2008). *Applied Econometrics with R*. Springer, New York.

Krugman, P. (1992). A Dynamic Spatial Model. *NBER Working Paper* 4219. National Bureau of Economic Research, Inc., Cambridge, Massachusetts.

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

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

Linders, G.-J., M. Burger and F. 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.

López-Rodríguez, J. and A. Faíña (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.

Lopez-Rodriguez, J., A. Faíña and F. Bruna (2014). Market Potential and Spatial Dependencies in the European Regions. *Regional Studies* (submitted).

Monfort, P. (2008). Convergence of EU regions. Measures and evolution. *European Union Regional Policy Working Papers* 01/2008. Directorate-General for Regional Policy, European Commission, Brussels.

Redding, S. J. (2011). Economic Geography: A Review of the Theoretical and Empirical Literature. In: Bernhofen, D., R. Falvey, D. Greenaway, and U. Kreickemeier (eds.), *Palgrave Handbook of International Trade*, 1st edition, pp. 497-531. Palgrave Macmillan, London.

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

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

Rodríguez-Pose, A. (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.

Stock, J. H. and M. Yogo (2005). Testing for weak instruments in linear IV regression. In: Andrews, D.W.K. and J.H. Stock (eds.), *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, pp. 80–108. Cambridge University Press, Cambridge.

Yotov, Y. V. (2012). A simple solution to the distance puzzle in international trade. *Economics Letters* 117(3), 794–798.

Appendix. Data description

Sample

The disaggregation level for the regional data is NUTS 2 (2006 version), which involves the basic regions for the application of regional policies. The sample includes 206 regions from 15 countries of the European Union: Austria, Belgium, Spain, Finland, France, Greece, Ireland, Italy, Luxembourg, Netherlands, Portugal, Sweden and United Kingdom. The following NUTS 2 regions are excluded: the Atlantic islands (the Spanish Canary Islands and the Portuguese Madeira and the Azores), the Spanish Ceuta and Melilla in the North African coast and the French Departments Guadeloupe, Guiana, Martinique and Reunion. Oil related regions are not excluded. Norway and Switzerland are omitted because of the lack of capital stock data but their 14 regions are included to compute Market Potential.

Variables

All the variables used in the models are in logarithmic form. Cambridge Econometrics data is used for gross value added (GVA), capital stock and population. GVA and capital stock per capita are in 2000 year euros. Market Potential is built with GVA and it is in millions of 2000 euros. Cambridge Econometrics's deflators are regional in the sense that are based on 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 each region. Cambridge Econometrics scales these estimates of real variables to the national estimates.

GVA excludes value added taxes but includes subsidies linked to production. Eurostat calculates regional gross domestic product on the basis of GVA, using approximations to distribute national tax income to regions. Thus GVA is the more direct indicator of regional economic activity (Breinlich, 2006).

Human capital stock (H_{it}) is proxied by the following Eurostat variable: share of the population who has successfully completed education in Science and Technology (S&T) at the third level and is employed in a S&T occupation. 9.7% of the observations 1995-2008 were missing and imputed using R's Amelia II package (Honaker *et al.*, 2011). The imputed data is the average of 5 multiple imputations with a small ridge prior predicting with a polynomial of degree 2 on the time trend of each region and including lags and leads: $H_{it} = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 H_{it-1} + \beta_4 H_{it+1}$. This method allows imputing this control variable in mainly seven regions with high degree of missingness.

Geographical distances (d_{ij}) are measured as great circle distances among regional centroids calculated using GISCO's shape files (© EuroGeographics for the administrative boundaries). Regional areas and internal regional distances are calculated from these files after an EPSG 3035 projection.

This paper uses a Harris's (1954) measure of Market Potential. The External Market Potential of region $i = 1, \dots, R$ is defined as the inverse distance (d_{ij}) weighted sum of the GVA of all the other regions in the sample (time subscripts are omitted for simplicity):

$$EMP_i = \sum_{j \neq i}^{R-1} \frac{GVA_j}{d_{ij}}$$

An apparent improvement is to use travel times instead of physical distances. However, Breinlich (2006) or Ahlfeldt and Feddersen (2008) obtain similar results using travel times or geographical distances with European regional data.

The full variable of Harris's (1954) Market Potential is calculated correcting this last measure by a proxy of the internal market size: $HMP_i = IMP_i + EMP_i$. The Internal Market Potential is defined here as: $IMP_i = GVA_i / d_{ii}$. To allow for the likely clustering of economic activity in and around the regional centre, internal distances are defined as $1/3$ of the radius (r_i) of circular regions: $d_{ii} = 1/3 \cdot r_i = 0.188\sqrt{area_i}$. This is the approach is similar to the 40% of the radius considered by Cambridge Econometrics. When estimating with MP_i or EMP_i, $1/3$ of the radius allows obtaining a broader range of results than the $2/3$ used by part of the literature. However, with any of these proportions, the resulting weight of IMP_i on HMP_i depends on the number of regions included in the sample.

Geographical distances (d_{ij}) are measured as great circle distances among regional centroids calculated using GISCO's shape files (© EuroGeographics for the administrative boundaries).