

Bruna, Fernando

Conference Paper

Why do empirical tests tend to accept the NEG? - An alternative approach to the 'wage equation' in European regions

55th Congress of the European Regional Science Association: "World Renaissance: Changing roles for people and places", 25-28 August 2015, Lisbon, Portugal

Provided in Cooperation with:

European Regional Science Association (ERSA)

Suggested Citation: Bruna, Fernando (2015) : Why do empirical tests tend to accept the NEG? - An alternative approach to the 'wage equation' in European regions, 55th Congress of the European Regional Science Association: "World Renaissance: Changing roles for people and places", 25-28 August 2015, Lisbon, Portugal, European Regional Science Association (ERSA), Louvain-la-Neuve

This Version is available at:

<https://hdl.handle.net/10419/124782>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Why do empirical tests tend to accept the NEG? – An alternative approach to the 'wage equation' in European regions

Fernando Bruna

University of A Coruña, Economics and Business Department, Campus de Elviña s/n, 15071 A Coruña, Spain. Phone: +34 981167000. Fax: + 34 981167070. E-mail: f.bruna@udc.es

Abstract

This paper posits a new approach to the ‘wage equation’ of the New Economic Geography (NEG) stressing the uncertain interpretation of its empirical results. It emphasizes the generality of the variable to be explained, marginal costs. Then, two artificial (no-NEG) tests are proposed in order to identify the statistical features explaining why wage-type equations tend to be accepted in tests for European data. The estimation results are shown to be similar not only when Market Potential is built for variables that do not measure market size but also when the focus of attention changes from global to local spatial patterns.

Keywords:

New Economic Geography, wage equation, Market Potential, spillovers, global trend, spatial autocorrelation

JEL codes: C21, F12, R12

Acknowledgments – The author is grateful to the participants who attended presentations of previous drafts given at the Department of Economic Theory of the Autonomous University of Madrid (2013); XVth and XVIth Conferences on International Economics of the Spanish Association of International Economics and Finance (Salamanca, 2014; San Sebastian, 2015); VIth Workshop in honor of Geoffrey Hewings (Santiago de Compostela, 2015); and 55th Congress of the European Regional Science Association (Lisbon, 2015). In particular, the author wish to thank the comments of Giuseppe Arbia, Carmen Díaz-Roldán, Andrés Faña, Eduardo Giménez, Keith Head, Geoffrey Hewings, James LeSage, Carlos Llano, Jesus Lopez-Rodriguez, Isabel Neira, Trino Níguez, Javier Perote and José Luis Zofío. Usual disclaimers apply.

Draft 1.2, September 2015

1. Introduction

The basic form of the so called ‘wage equation’ of the New Economic Geography (NEG) predicts that nominal regional manufacturing wages are a function of an index of regional accessibility to the markets, called *Market Access* or *Market Potential*. It has been widely studied in the empirical literature, which seems to confirm a ‘causal relationship’ between market access and the spatial distribution of economic activity (Redding, 2011). The problem of the observational equivalence¹ of the NEG is that ‘there are a number of other explanations that are consistent with the data and not much yet that strongly points to the explanation offered by NEG’ (Head and Mayer, 2004, 2663). A first attempt of using the wage equation to confront the NEG and Urban Economics in Europe was done by Fingleton (2006), through an artificial nesting model. He concluded that the NEG does not necessarily provide the best explanation of regional wage variations in UK. Following Fingleton’s methodology, Brakman et al. (2009a) concluded that Market Potential is more relevant at the country level, whereas population density is more relevant at the regional level.

Instead of comparing theories, the goal of the present paper is to present an alternative approach to an empirical ‘wage equation’ for European regions, which explains its empirical success and stress difficulties of interpretation in terms of the NEG framework. A *wage-type equation* is re-examined from two points of views. Firstly, from a theoretical perspective the paper focuses attention on marginal costs (Combes et al., 2008, chap. 12; Bruna, 2015b), instead of wages, for the left-hand-side of the equation. Given that marginal costs depend on factor prices and total factor productivity, this approach emphasizes the uncertainty about the phenomenon to be explained and about its determinants (specification of factors in the production function, sectoral composition, knowledge flows....). Secondly, with respect to the right-hand-side of the equation, the sensibility analysis presented in this paper provides a statistical explanation of why many of the previous empirical tests about the ‘wage equation’ tend to be accepted by the data. This is done through two novel artificial tests useful to measure our degree of ignorance when explaining the agglomeration (and development) of European regions. A first re-

¹ Duranton and Puga (2004) called it *Marshallian equivalence*. The difficulty to discern between alternative theories of location has also been mentioned by Overman (2004), Rosenthal and Strange (2004), Brakman et al. (2009b, chap. 5) or Puga (2010).

examination of a wage-type equation is done by building Market Potential with variables different from the ones suggested by NEG theory. A second test is the repeated estimation of that equation when ‘Market Potential’ is built for the first nearest neighbor, for the two nearest neighbors and so on, instead of using the whole sample of regions. The role of local spillovers and the internal market size in the estimation results is also analyzed.

The empirical analysis is mainly based on cross-sectional data, though some preliminary panel data evidence is also shown. The dependent variable is proxied by gross value added per capita (GVApc) and Market Potential is proxied by Harris’ (1954) indicator of accessibility to the markets, built with GVA. European regional data is used in order to explain Breinlich (2006) and Head and Mayer’s (2006) similar empirical results when using Harris’ indicator or a more sophisticated structural estimation of the NEG equation.

It is shown that the results of an empirical wage-type equation are similar when the key explanatory variable, Market Potential, is built in ways different from what NEG theory suggests. The structure of a Market Potential variable, as a sum for all the spatial observations in the sample, weighted by distances, makes irrelevant the variable considered in the summation, as long as the possible candidates present comparable patterns of local spatial autocorrelation. This result might indicate that Market Potential is not necessary measuring the accessible market size. Moreover, Market Potential mainly captures a global pattern in the spatial distribution of the dependent variable (Bruna et al., 2015) but the estimation results are similar when a version of Market Potential is built to capture local spillovers. Putting all together, the paper highlights the high degree of uncertainty when interpreting the results of an empirical wage-equation as a confirmation of the trade related interaction channels described by the NEG.

The remaining of the paper is structured as follows. Section 2 sets up the theoretical framework. Sections 3 describes the empirical strategy and the statistical properties of the data. Section 4 shows the results of testing the wage-type equation under different specifications for the right-hand-side of the equation. Section 5 concludes.

2. A generalized NEG wage-type equation

Since the contributions of Krugman (1991, 1992) the so called ‘wage equation’ is usually presented as an explanation of wages. The theoretical NEG model presented empha-

sizes that, without additional restrictive assumptions about the production function and factor mobility, the dependent variable of that equation is marginal costs, which has implications to interpret the empirical results presented later. The derivation is based on Fujita et al. (1999, chap. 14) and Redding and Venables' (2004) models, thought omitting intermediate inputs, as Breinlich (2006) and Head and Mayer (2006) do. Additional details about this version are provided by Bruna (2015a).

The world is composed by $i = 1, \dots, R$ regions and the focus here is on the sector producing differentiated goods by firms exhibiting internal increasing returns to scale and operating under a market structure of monopolistic competition (M sector). The representative consumer of region j decides the quantity of consumption for each M variety via utility maximization of a Dixit-Stiglitz CES utility function:

$$\begin{aligned} \max_{x_{ij}} U_j &= \left[\sum_{i=1}^R n_i x_{ij}^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}} \\ \text{s. t. } \sum_{i=1}^R n_i p_{ij} x_{ij} &= E_j^M \end{aligned} \tag{1}$$

where $\sigma > 1$ is the elasticity of substitution between any pair of varieties, x_{ij} is the amount of consumption in j of the variety produced in i , p_{ij} is the delivery price of that variety and E_j^M is the expenditure of region j in all the varieties of the M good. If $\mu_j = E_j^M / E_j$ is the share of M consumption in total expenditure of j , solving the optimization problem, the demand facing a firm i from location j is:

$$x_{ij}^d = p_{ij}^{-\sigma} \frac{E_j^M}{\sum_{i=1}^R n_i p_{ij}^{1-\sigma}} = p_{ij}^{-\sigma} \frac{\mu_j E_j}{S_j^M} \tag{2}$$

where S_j^M is called here 'competition index' to emphasize that measures the level of competition between M varieties in j market given the characteristic tastes of consumers.² Firms of the same region are assumed to have the same *free-on-board* price. Trade costs are assumed to be borne by consumers, so firms follow a mill pricing policy. Therefore, the delivered price in market j for a of a good produced in region i is assumed to be

² S_j^M is called 'supplier access' by Redding and Venables (2004) and 'supply' index by Head and Mayer (2006). The term $S_j^{M^{1/(1-\sigma)}}$ is equivalent to the 'true' price index of Krugman (1992), which is the unit cost of utility for the consumer.

$p_{ij} = T_{ij}p_i$, where $T_{ij} \geq 1$ are ‘iceberg’ transport or trade costs and p_i is the mill price in i . Hence, the effective demand from j -market is:

$$x_{ij} = T_{ij}x_{ij}^d = T_{ij}^{1-\sigma}p_i^{-\sigma}\frac{\mu_j E_j}{S_j^M} \quad (3)$$

Given equation (3), total demand to a representative M firm in region i will be the sum of what it sells to the world markets:

$$x_i \equiv \sum_j^R x_{ij} = p_i^{-\sigma} \sum_j^R \mu_j T_{ij}^{1-\sigma} \frac{E_j}{S_j^M} = p_i^{-\sigma} RMP_i \quad (4)$$

where RMP_i stands for *Real Market Potential* (RMP_i), as was named by Head and Mayer (2006), or *Market Access*, in the vocabulary of Redding and Venables (2004). The latter authors used equation (3) to derive a ‘trade equation’ reflecting bilateral trade flows in an Anderson and van Wincoop’s gravity-type equation. The estimation of this equation allows proxying the term E_j^M/S_j^M by the estimates for importing country dummies. Breinlich (2006) followed a variant of this approach. This strategy to proxy RMP_i will be discussed in section 3.1.

Production is assumed to involve a fixed cost f , defined in units of output. The production function of the M firms in region i is:

$$x_i = -f + A_i I_i = A_i \left(-\frac{f}{A_i} + I_i \right) \quad (5)$$

where I_i is a compound input and A_i is a Ricardian technology. Therefore $I_i = c_i(f + x_i)$, being $c_i = 1/A_i$ the marginal input requirement: the fixed input requirement, $c_i f$, is allowed to vary across regions. If, q_i is the price index of I_i , the cost of producing x_i is $q_i c_i(f + x_i)$. Marginal cost, the price of the compound input in efficiency units, is $m_i = q_i c_i$. Firms, facing given factor prices in m_i , maximize the following profit function with respect to their mill prices p_i :

$$\pi_i = p_i x_i - m_i(f + x_i) \quad (6)$$

Considering the effective demand in equation (4), if each firm takes the competition index S_j^M in RMP_i as given, profit maximization implies that firms choose price as a mark-up over marginal costs:

$$p_i = \frac{\sigma}{\sigma - 1} m_i \quad (7)$$

At these optimum mill prices, profits are:

$$\pi_i = m_i \left(\frac{1}{\sigma - 1} x_i - f \right) \quad (8)$$

Taking into account equations (4) and (7), the ‘profit equation’, similar to the one derived by Combes et al. (2008, chap. 12), is the following:

$$\pi_i = \frac{(\sigma - 1)^{\sigma-1}}{\sigma^\sigma} m_i^{1-\sigma} RMP_i - f m_i \quad (9)$$

Free entry assures that long run profits will be zero so, from equation (8), $x_i = (\sigma - 1)f = \bar{x}$. Therefore, from the effective demand equation (4), active firms at location i attain this level of output and break even if and only if the mill price they charge satisfies $p_i^\sigma = \frac{1}{\bar{x}} RMP_i$. From equation (7) this price has also to verify the relationship $m_i = \frac{\sigma-1}{\sigma} p_i$. Consequently, the maximum value of marginal costs that that each firm in region i can afford to pay is a function of its Real Market Potential:

$$m_i = \frac{\sigma - 1}{\sigma} \left(\frac{1}{\bar{x}} RMP_i \right)^{\frac{1}{\sigma}} = \frac{\sigma - 1}{\sigma} \left(\frac{1}{\bar{x}} \sum_j^R \mu_j T_{ij}^{1-\sigma} \frac{E_j}{S_j^M} \right)^{\frac{1}{\sigma}} \quad (10)$$

The ‘market-clearing condition’ (Baldwin et al., 2003, 19) in equation (10) is called ‘generalized wage-type equation’ here, emphasizing that the dependent variable is not wages, but marginal costs. Bruna (2015a) shows that this equation can encompass many of the ‘wage equations’ previously derived in the literature. For Redding and Venables (2004), the dependent variable of equation (10) is the price of the composite immobile factor of production, which they interpret as labor. Alternatively, Head and Mayer (2004) reinterpreted its logarithmic form as a cost-share weighted sum of logged primary factor prices.

Head and Mayer (2004) sets one of the problems of testing a theory³ in terms of the statistical Error Type II: failing to reject a false null hypothesis. The issue to be analyzed in the present paper is not whether the NEG is a false theory but whether the way in which that theory is frequently tested allows an unambiguous confirmation of its assumptions. The formulation presented here emphasizes its lack of specificity with respect to what is supposed to be studied: Manufacturing wages? Total factor productivity? The spatial distribution of economic activity (Redding, 2011)? The phenomenon to be ex-

³ The methodology for testing a theory has been discussed by Popper (1959, 95) or Leamer and Levinsohn (1995), among others.

plained is marginal costs ($m_i = q_i/A_i$) so the effects of RMP_i might operate through the prices of the compound factor (q_i) or through total factor productivity (A_i). That will be relevant to interpret the possible meanings of the empirical results to be presented below.

3. Empirical approach to the wage-type equation

3.1. Econometric strategy

Starting from equation (10) the benchmark specification considered in this paper for a European regional cross-sectional regression is the following:

$$\ln m_i = C + \beta \ln HMP_i + u_i \quad (11)$$

where C is an intercept and Real Market Potential is proxied by Harris' (1954) index of Market Potential. If the share of manufacturing goods on expenditure is assumed to be the same in all regions ($\mu_j = \mu = 1$), as Fujita et al. (1999, chap. 4) do, and trade costs are proxied by physical distances, the Real Market Potential defined in equations (4) and (10) is the following:

$$RMP_i = \sum_j^R d_{ij}^{1-\sigma} \frac{E_j}{S_j^M} \quad (12)$$

In contrast, Harris' (1954) indicator of accessibility to the markets is defined as (names of variables without italics to emphasize the lack of microfundamentals):

$$HMP_i = \sum_{j=1}^R d_{ij}^{-1} E_j = d_{ii}^{-1} E_i + \sum_{j \neq i}^{R-1} d_{ij}^{-1} E_j = IMP_i + EMP_i \quad (13)$$

where E_j stands for a measure of market size and the Internal Market Potential (IMP) is distinguished from the External Market Potential (EMP) of region i . This distinction is absent from the theoretical definition of RMP_i but affects any of its possible proxy variables. Excluding the own regional market 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 aggravates the general endogeneity problem of Market Potential. Therefore, both the full measure HMP_i and its external component, EMP_i , will be used in later empirical tests.

For Head and Mayer (2006) the adjective 'real' in the name of RMP_i underlines the importance of discounting expenditures by the competition index S_j^M , which is not pre-

sent in HMP_i . As discussed when setting equation (4), the empirical strategy of Redding and Venables (2004) allows building a proxy of RMP_i after the estimation of the exponent of distance in a bilateral trade equation and proxying S_j^M by the estimates of importing country dummies. However, the results of Breinlich (2006) and Head and Mayer (2006) estimating a wage equation for European regions are similar when using Harris' definition of Market Potential rather than Redding and Venables' (2004) methodology.

The reasons for this result are the same explaining why NEG theory usually is accepted in empirical tests, at least under the restrictions of scope of the present paper. First, the variable is built as a summation for all the regions in the sample, which produces a spatially smoothed distribution of values. Second, with an exponent of distance close to -1, Market Potential captures in an stylized way the core-periphery spatial pattern present in the European income per capita. That -1 trade elasticity to distance, as in Harris' index, is an extremely robust empirical finding in the literature on gravity equations (Head and Mayer, 2014). For instance, Breinlich's (2006) estimates for different periods rank from -0.6 to -1.0. Third, the way in which Market Potential is built makes of little relevance for the wage-type of equation the variable used to measure the size of the markets, E_j or E_j/S_j^M . Indeed, the estimation results of wage-type equations are similar if Market Potential is built with spatially autocorrelated variables that are not suggested by NEG theory. Four, given that many European regional variables are spatially autocorrelated, what matters for a wage-type equation is the locational information captured by Market Potential. However, the estimation results are similar when that information is captured by a few neighbors, instead of evaluating Market Potential for the whole sample of regions.

In order to test the previous hypothesis, the empirical analysis below describes the spatial patterns of the variables and propose the following steps of successive variations on the right-hand-side of the equation:

- 1) Estimating benchmark equations using GVA to build Market Potential;
- 2) Repeating the estimation with 'Market Potential' built different variables;
- 3) Repeating it after defining 'Market Potential' for 1, 2, 3... nearest neighbors.

The focus of attention will be on Harris' External Market Potential (EMP). The artificial variables built for the sensibility analysis in steps 2 and 3 will be named using quotation marks ('EMP'). The scope of the paper is limited to cross-sectional estimations, though some panel data results will also be shown in section 4.3.

Turning to equation (11), the term u_i is supposed to collect the effects of omitted variables and departures from the assumptions of the theoretical model, which are assumed to be randomly distributed under ordinary least squares (OLS) estimation. However, as will be shown below, the OLS residuals will result to be spatially autocorrelated, which calls for the estimation of spatial models. A detailed analysis of spatial wage-type equations is out of the scope of the present paper (Niebuhr, 2006; Ahlfeldt and Feddersen, 2008; Fingleton and Fischer, 2010; Bruna et al., 2015). Table 2 below will show estimations of a simple Spatial Error Model (SEM), in which the error term is assumed to follow a spatial autocorrelation process: $u = \lambda Wu + \varepsilon$. W is a spatial weights matrix determining what regions are considered to be linked and the weights given to those links. The W matrix studied in this paper for the SEM models, Moran's tests and Lagrange Multiplier tests is a standardized binary matrix to the five nearest neighbors. As emphasized by LeSage and Pace (2014) and LeSage (2014), changing the number of neighbors does not alter the main conclusions of the present research.

The estimate of Market Potential by OLS would be unbiased if the 'true' data generation process would follow a SEM. Possible strong differences between the OLS and SEM estimates would point to misspecification in the equation or in the spatial process. Moreover, Given that Market Potential is similar to a spatial lag of the dependent variable, Kosfeld and Eckey (2010) prefer to estimate a wage-type equation using the SEM. Bruna et al. (2015) show that 'global' spillovers (LeSage, 2014) are implicit in a wage-type equation, in spite of being absent from NEG theory. A model of 'local' spillovers, such as the SEM, will be useful to clarify the lack of identification between local spillovers and global spatial trends appearing in the tests below.

Market Potential will not be instrumented, contrary to what it is usually done in the NEG empirical literature (Head and Mayer, 2006; Doran and Fingleton, 2015). The problem discussed here is whether the way in which Market Potential is constructed allows an unambiguous confirmation of the NEG interaction channels. The primary focus of this paper is the structure of a Market Potential variable, while the validity of its possible exogenous instruments depends on that structure.

3.2. Data description

Bruna (2015b) showed that the empirical results of a cross-sectional wage-type equation for European regions do not depend crucially from the election of the dependent variable, due to the high correlation between the different dependent variables that has been used in the NEG literature. Gross value added per capita (GVApC) is used here,⁴ as Breinlich (2006) does. The benchmark Market Potential variable is also built with GVA, though alternative measures are tested in section 4.2. Some estimations are controlled by human capital (Head and Mayer, 2006), proxied by a variable of human resources in science and technology. Details about the sample (220 regions) and the variables are provided in the Appendix.

Internal transport costs can be crucial in NEG theory (Behrens et al., 2006) and the problem of its measurement has been discussed since the work of Stewart (1947). The standard approach is to assume that regions are circular so the radius of region i is $r_i = \sqrt{area_i/\pi}$. 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 regional ‘center’. Based on geometrical arguments about the random distribution of consumers when production concentrates in the center of a disk, Head and Mayer (2000)-Thisse proposed to use $d_{ii} = 2/3 \cdot r_i = 0.376\sqrt{area_i}$. The discussion about these type of methods is frequently omitted in the NEG literature,⁵ but ‘estimating the average intra-zonal trip length is still an ongoing challenge in spatial models’ (Kordi et al., 2012). The area-based approximation can lead to problems of interpretation because of its relationship with the measurement of density (Head and Mayer, 2006). Unlike grid based methodologies (Dijkstra et al., 2011), the standard geometrical methods ignore that the main city of coastal regions is not generally located at the geographical center of the region.⁶ Additionally, the geometrical arguments do not consider that Market Potential is the summation of IMP and EMP, so the weight of IMP in HMP depend on the sample size considered to evaluate the external component. For the sample analyzed in this paper the median weight of IMP in

⁴ The correlation of the log of real GVApC and the log of nominal remuneration per worker is 0.81, but the data of the latter variable have lower quality. See Breinlich (2006) for further discussion about GVApC.

⁵ The works of Redding and Venables (2004) and Bouhol et al. (2008) are among the exceptions.

⁶ The coastal NUTS 3 regions account for 40% of the population and territory of the 27 members of the European Union (Collet and Engelbert, 2013).

HMP is 9,7% when 1/3 of the radius is taken for internal distances⁷ and only 5,1% when 2/3 is used. There is no available empirical evidence about the link of these figures with the weight of the domestic regional market for the trade relationships of the firms in the median region. The benchmark measure of internal distances used in this paper is based on the approach giving more weight to the internal markets, $d_{ii} = 1/3 \cdot r_i$, which is similar to the 40% of the radius considered by Cambridge Econometrics (2015).

Table 1. Cross-sectional correlations of variables for European regions (logs, year 2008)

	GVApc	EMP	MP	HK	KSpC	GVA	GVA H&R	GVA density	POP	POP density	TFP
GVA per capita (GVApc)	1.000	0.478	0.559	0.539	0.851	0.428	0.129	0.542	0.069	0.340	0.977
External Market Potential (EMP)	0.478	1.000	0.959	0.236	0.399	0.360	0.114	0.667	0.205	0.619	0.519
Market Potential (d_{ii} 1/3) (MP)	0.559	0.959	1.000	0.300	0.461	0.521	0.282	0.825	0.350	0.774	0.610
Human Capital (HK)	0.539	0.236	0.300	1.000	0.384	0.295	0.057	0.318	0.109	0.212	0.516
Capital stock per capita (KSpC)	0.851	0.399	0.461	0.384	1.000	0.294	0.021	0.409	-0.005	0.235	0.720
GVA	0.428	0.360	0.521	0.295	0.294	1.000	0.784	0.603	0.931	0.561	0.470
GVA hotels and restaurants (H&R)	0.129	0.114	0.282	0.057	0.021	0.784	1.000	0.479	0.813	0.501	0.172
GVA density	0.542	0.667	0.825	0.318	0.409	0.603	0.479	1.000	0.447	0.975	0.594
Population (POP)	0.069	0.205	0.350	0.109	-0.005	0.931	0.813	0.447	1.000	0.482	0.141
Population (POP) density	0.340	0.619	0.774	0.212	0.235	0.561	0.501	0.975	0.482	1.000	0.410
Total factor productivity (TFP)	0.977	0.519	0.610	0.516	0.720	0.470	0.172	0.594	0.141	0.410	1.000

Note: See the data appendix. For 206 regions, TFP, in logarithmic form, is built as $GVApc-1/3KSpC$.

Table 1 shows the correlations of the main variables, in logarithmic form, utilized in this paper. The first four variables will be used to estimate the benchmark wage-type equations in the next section. Human capital and the last six variables of the table will be used to build the alternative External ‘Market Potential’ variables in the estimations in Table 3 below. GVA in hotels and restaurants was selected because its low correlation with other variables in the table.⁸ The density variables are defined in terms of geographical areas, as is frequent done in studies inspired by Urban Economics. Human capital and total factor productivity (TFP) are also considered because of their potential to capture spillovers affecting the marginal costs in equation (10).

3.3. The spatial distribution of the data

The NEG is about trade relationships defined in space so the spatial features of the variables used to test the wage-type equation becomes crucial. Figure 1 shows quantile maps

⁷ Using the 1/3 measure for internal distances, the weight of IMP on HMP is higher than 45% for the regions of Berlin, Hamburg, Madrid, Paris, Vienna, Athens and Inner London.

⁸ While in Finland or Denmark the H&R sector was 1% of the 2008 GVA, in Spain was 7% and in Greece 10%. GVA H&R might capture some peculiarities related to geography, such as weather or cultural characteristics, though its logarithmic correlation with regional population is 0.8.

of the logarithms of per capita gross value added and Market Potential in the year 2008. Their values are divided into seven quantiles and darker colors are associated with higher levels of the variables. In spite of the visual limitations of choropleth maps, Figure 1 allows distinguishing a core-periphery pattern of the European regional income per capita, known since the work of Clark et al. (1969). Only a few regions with high GVApc seem to be located out of the so called 'blue banana', particularly those of Nordic countries. The logarithm of a Harris' measure of External Market Potential built with GVA shows an even more concentrated spatial distribution due to the smoothing effects of the summation in equation (13). Therefore, in the context of a wage-type equation, EMP is able to capture in a stylized way the global core-periphery pattern of the depend variable (Bruna et al., 2015). The right map of Figure 1 shows that the inclusion of IMP in the measurement of HMP does not change too much the stylized spatial representation that EMP provides about GVApc.

In Geostatistics a global spatial trend corresponds to a systematic variation of the values of a variable with the geographic space coordinates. Its presence in the data affects the detection of local clustering (Bivand et al., 2008, 260). However, when an explanatory variable presents a global spatial distribution similar to the one of the dependent variable, the probability of being significant in a regression increases, particularly if local differences are captured by a spatial model. Given the importance of this issue, Figure 2 shows a model of the global spatial trends detected in the data of GVA per capita and External Market Potential. The plots show the predictions of these two variables using only the spatial coordinates of the regional centroids. Similar values of each predicted variable are represented by colored lines (technical details are provided in the Appendix). The centroids are represented by points in order to facilitate the interpretation of the plots.

The model based only on geography predicts that the highest values of GVApc are in the Norwegian Sea, instead on the blue banana, because it is capturing a decreasing global trend from North to South. However, Figure 2 also confirms that, on average, EMP is not a bad proxy for the spatial distribution of GVApc in terms of a regression setting. Though this interpretation is consistent with NEG theory, the test below will show that it may be misleading.

Figure 1. Choropleth maps of the logs of GVA per capita and Market Potential (2008)

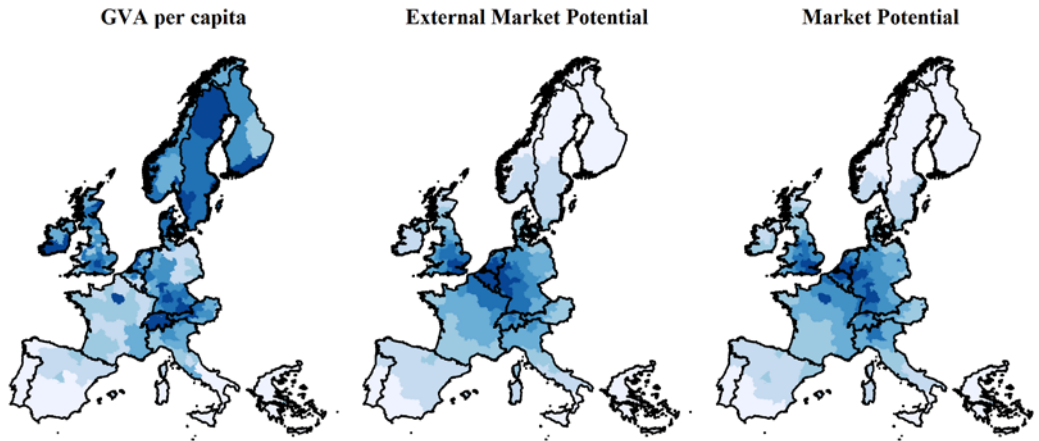
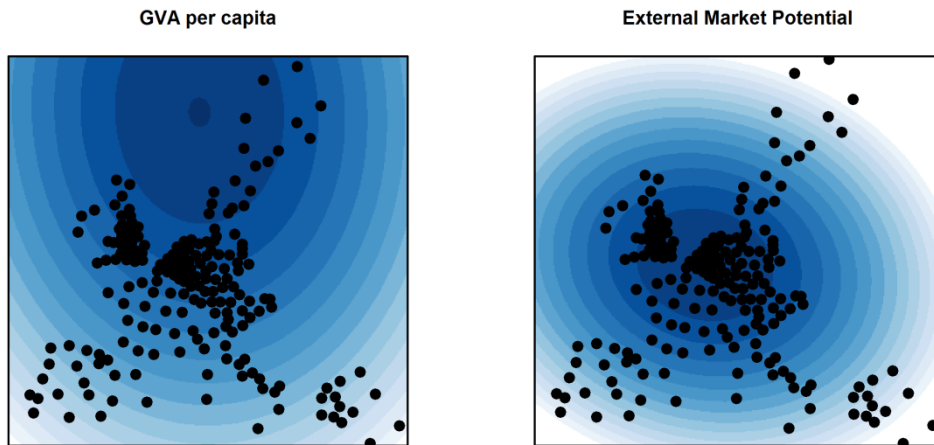


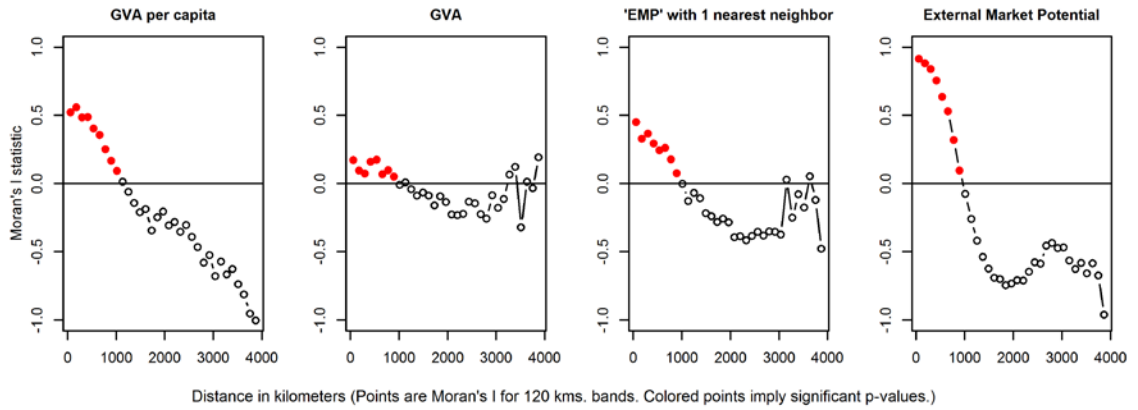
Figure 2. Predictions of log of GVApC and EMP on a grid from a polynomial trend surface of order 2



The previous figures show global spatial patterns. Alternatively, the W matrix used in this paper to test for spatial autocorrelation and estimate spatial econometric models is designed to capture local clustering. With that matrix, Moran's tests reveal that all the variables in Table 1 present positive spatial autocorrelation.⁹ Figure 3 bridges the gap between the global and local approach to the spatial distribution of the variables. It shows Moran's coefficients at intervals of Euclidean distance (correlograms) for four logarithmic variables. For each of them, the similarity of values decreases as the distance between their spatial coordinates increases. Moran's I detect significant spatial autocorrelation (colored points) for observations at distances below about 1,000 kilometers.

⁹ The p-values of the Moran's tests under the randomization assumption is zero. The Moran's I of the variables ranks from 0.17 for population to 0.92 for External Market Potential.

Figure 3. Correlograms for several variables in log form: Moran's I on distance intervals of 120 kms.



The correlogram of GVA and GVAp_c are quite different, probably due to a lack of correction for heterogeneity in the case of GVA. The correlogram of External ‘Market Potential’ built for the nearest neighbor ($\sum_{j \neq i}^1 d_{ij}^{-1} GVA_j$) is more similar to the one of the dependent variable studied in this paper. When ‘EMP’ is built for the two nearest neighbors, for the three ones, and so on (not shown), the point representing Moran’s I for the first interval moves upwards towards its value in the correlogram of External Market Potential ($\sum_{j \neq i}^{219} d_{ij}^{-1} GVA_j$). When ‘EMP’ is built for the five nearest neighbors, the plot already presents the snakelike shape of the right plot in Figure 3. However, the correlograms of all these variants of EMP are only approximate representations of the correlogram of GDPpc, which will be relevant for the results in section 4.3.

4. Testing the wage-type equation for European regions

4.1. Benchmark cross-sectional estimations

Table 2 shows alternative empirical wage-type equations that will be the reference for the sensibility analysis in sections 4.2 and 4.3. The OLS estimations in columns (1) to (4) and (6)-(7) present spatially autocorrelated residuals (zero p-value of Moran’s I), so SEM models are estimated in columns (5), (8) and (9). Columns (1) and (2) confirm that increasing the role of internal market potential (d_{ii} equal to 1/3 of the radius instead of 2/3) improves the linear fit but does not have a relevant effect on the estimate of Market Potential. Columns (1) to (5) show that the estimated elasticity of GVAp_c to Market Potential is in the range 0.3-0.4. This elasticity is in the range 0.2-0.3 for External Market Potential (columns 6 to 9), though the inclusion of national or local effects, in columns (7)

and (8), reduces its statistical significance. Both type of effects has to be jointly considered in column (9) to get a 1% significance level for EMP, with a t-Student of only 3.5. However, comparing columns (6) and (9), the inclusion of country dummies and a SEM parameter does not substantially change the estimate of EMP.

Table 2. Cross-sectional estimations of a wage-type equation for 220 European regions (year 2008)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
(Intercept)	6.450*** (0.489)	6.516*** (0.514)	7.883*** (0.465)	7.068*** (0.843)	6.820*** (0.518)	8.242*** (0.459)	8.214*** (0.931)	9.355*** (0.682)	9.115*** (0.638)
Market Potential (d_{ii} 1/3)	0.365*** (0.050)		0.285*** (0.043)	0.390*** (0.080)	0.416*** (0.047)				
Market Potential (d_{ii} 2/3)		0.360*** (0.053)							
Human Capital			0.411*** (0.058)	0.365*** (0.078)	0.398*** (0.055)	0.455*** (0.059)	0.524*** (0.091)	0.570*** (0.053)	0.595*** (0.057)
External Market Potential						0.259*** (0.044)	0.309** (0.098)	0.161* (0.071)	0.224*** (0.064)
λ					0.526*** (0.075)			0.736*** (0.050)	0.465*** (0.081)
Country dummies	No	No	No	Yes	Yes	No	Yes	No	Yes
R-squared	0.313	0.283	0.464	0.797		0.421	0.748		
Adj. R-squared	0.310	0.280	0.459	0.779		0.415	0.725		
AIC	73.72	83.09	21.06	-160.46	-187.99	38.15	-112.80	-82.96	-130.48
p-value Moran's I	0.000	0.000	0.000	0.000		0.000	0.000		
Moran's I residuals	0.587	0.550	0.614	0.227		0.555	0.174		
Sum squared errors	17.52	18.29	13.67	5.18	4.29	14.77	6.43	7.45	5.65

Note: Variables are in log form (see Table 1). The dependent variable is GVApc. Standard errors are in parentheses. Columns (1) to (4) and (6)-(7) are estimated by OLS and include heteroskedasticity robust standard errors. Columns (5), (8) and (9) show estimations of a SEM by maximum likelihood. * Significant at 10% level; ** at 5% level; *** at 1% level.

4.2. Versions of External 'Market Potential' built with alternative variables

An empirical test about the NEG interaction channels would have to prove that the variable of Market Potential defined in terms of market size has significantly higher explanatory power than alternative measures based on other variables. The following exercise is a first attempt to address this issue.

Columns (1) and (8) of Table 3 show a simple estimation of the wage-type equation including only External Market Potential, built with GVA. The other columns show the results when External 'Market Potential' is built with some of the variables in Table 1. The intercepts captures scale differences and all the estimates of 'EMP' are within a range of 0.2-0.4. It is worth noting the relatively high adjusted coefficient of determination when 'EMP' is built with human capital (column 7) or total factor productivity (column 9). This two variables, as well as the similarity of the results in Table 3, reveal that EMP might not be capturing trade accessibility to the markets. The inclusion of measures

of Internal ‘Market Potential’ in Table 3, using the same variables, only improves the coefficient of determination, because of using data from the same region whose GVAp_c is estimated. Adding control variables does not change the conclusions.

Table 3. Cross-sectional OLS estimations for alternative variables of External ‘Market Potential’

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
(Intercept)	6.804*** (0.524)	8.329*** (0.424)	6.580*** (0.450)	6.809*** (0.570)	7.881*** (0.415)	8.870*** (0.197)	11.009*** (0.139)	6.587*** (0.541)	8.059*** (0.271)
EMP (GVA)	0.333*** (0.054)							0.353*** (0.056)	
‘EMP’-GVA H&R		0.283*** (0.071)							
‘EMP’-GVA per capita			0.382*** (0.050)						
‘EMP’-GVA density				0.207*** (0.037)					
‘EMP’-Population					0.331*** (0.064)				
‘EMP’-Populat. density						0.228*** (0.039)			
‘EMP’-Human capital							0.380*** (0.053)		
‘EMP’-TFP									0.377*** (0.052)
R-squared	0.228	0.121	0.285	0.193	0.180	0.196	0.260	0.265	0.290
Adj. R-squared	0.225	0.117	0.282	0.189	0.176	0.193	0.257	0.261	0.287
Sum squared errors	19.68	22.42	18.23	20.58	20.91	20.49	18.86	16.58	16.01
N	220	220	220	220	220	220	220	206	206

Note: Variables are in log form. The dependent variable is GVAp_c. The alternatives External ‘Market Potential’ are logs of Harris’s (1954) indexes built with some of the variables in Table 1 in levels, though the latter table shows their correlations in log form for comparability among tables. Heteroskedasticity robust standard errors are in parentheses (see Table 2). The 14 regions of Norway and Switzerland are omitted in column (9), so column (8) is presented for comparability of samples.

The reason for the similar results in Table 3 is the following. In spite of the low correlations showed in Table 1 for some of the variables, the lowest correlation between the alternative ‘EMP’ variables in Table 3 (not shown) is 0.88. Therefore, what is relevant is the way in which EMP is built and not so much the variable used to build it or a possible consideration of a proxy for S_j^M in equation (12). All the variables built as a sum over all the regions in the sample weighed by inverse distances present a smooth spatial distribution of values, centered around the blue banana (maps not shown) and, therefore, are able to capture the core-periphery spatial pattern in GVAp_c described in section 3.3. The regions with the lowest distances to all the other regions are those in the geographical center of the sample so the inverse distance weighing scheme locates there the center of the ‘EMP’ alternatives. The smooth spatial core-periphery pattern of EMP (see Figure 1) is replicated in all the ‘EMP’ variables because of the summation of spatially autocorrelated

variables, as those in Table 1, for all the regions in the sample. The values of these variables are similar among neighboring regions,¹⁰ while more distant regions tend to present values with higher degree of variation (see Figure 3). Given that peripheral regions tend to have bigger distances to all the other regions, the values receiving less weight are those with less similar information.

4.3. Versions of External ‘Market Potential’ built by number of nearest neighbors

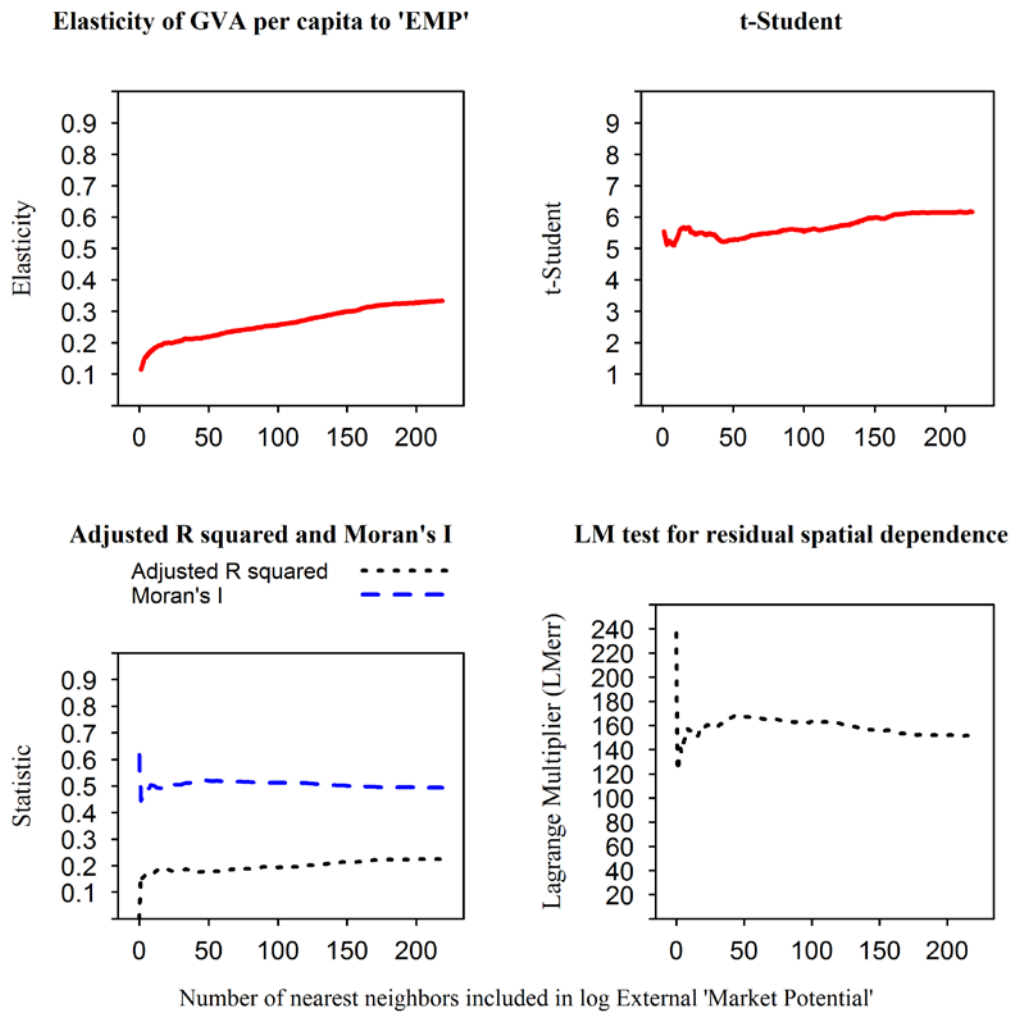
The following sensibility analysis studies the empirical effects of re-defining EMP built with GVA in different variables depending on how many nearest neighbors are considered: first for the nearest neighbor, then for the two nearest neighbors, and so on.¹¹ 219 variables were created and used to estimated 220 regressions of alternative wage-type equations. For zero neighbors the regression omits the ‘EMP’ variable and for 219 neighbors the variable is the same than External Market Potential.

Figure 4 shows the estimation results when the regression in column (1) of Table 3 is repeated for all the new ‘EMP’ variables. The estimate of ‘EMP’ increases as the latter variable includes more nearest neighbors. However, once the first nearest neighbor is considered, the improvement of the heteroskedasticity robust t-Student or the adjusted R squared is not relevant. The inclusion of the first nearest neighbor captures locational information, as it can be seen in the decline of the Moran’s I statistic or the LM test for the residuals of each regression. Nevertheless, the residuals continue to be spatially auto-correlated in all the regressions (not show), probably due to the different weighting schemes in the ‘EMP’ variables and the W matrix used to test for residual spatial auto-correlation (Bruna et al., 2015).

¹⁰ The median distance for the fifth nearest neighbor in the regional sample studied here is around 160 kilometers, which implies an inverse distance weight of 0.006.

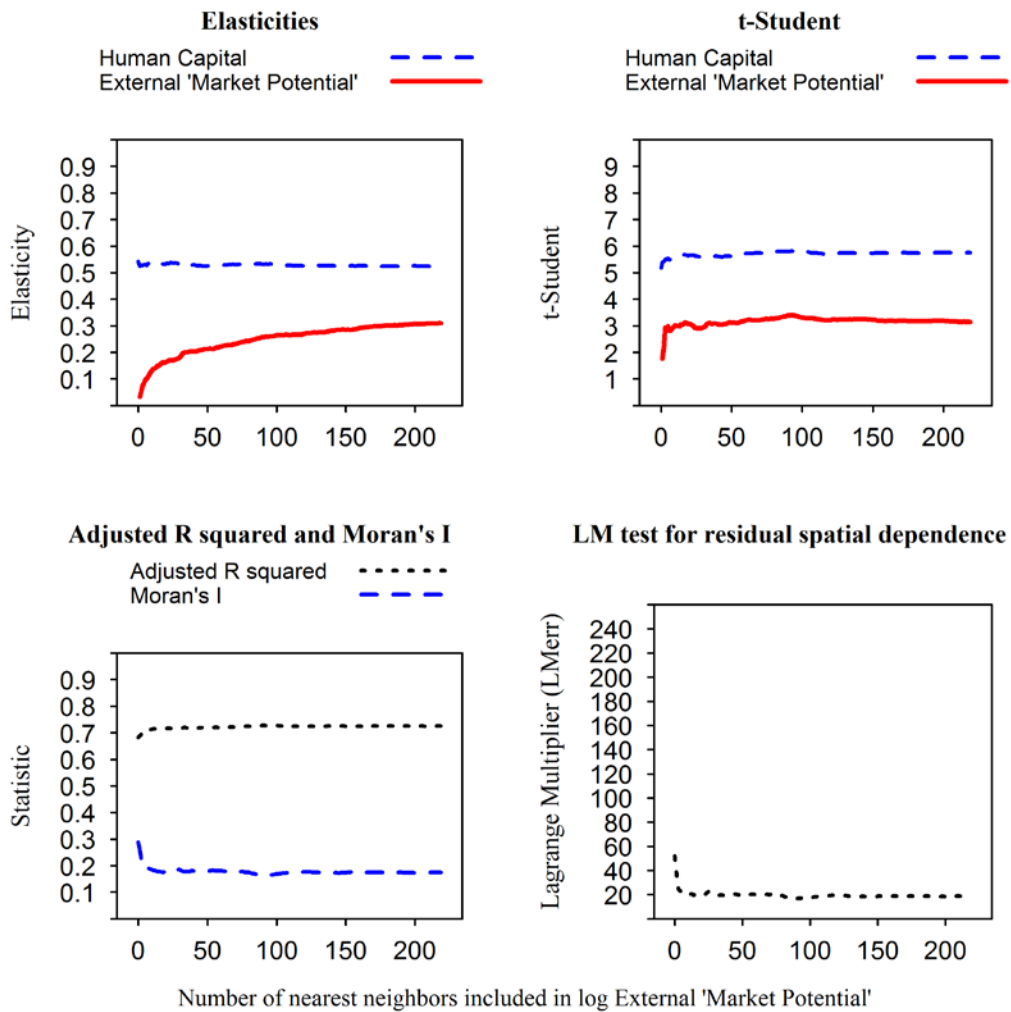
¹¹ The author is grateful to a conversation with James LeSage for his inspiration to do this exercise. See Negreiros (2009) and LeSage and Pace (2014) for a similar discussion in the context of the weights matrix for Spatial Econometrics.

Figure 4. 220 regressions with 'EMP' by number of neighbors: specification in column (1) of Table 3



These results are reinforced when the exercise is repeated for other specifications in Table 2, as it can be seen in Figure 5 for the specification including human capital and country dummies. A particular case is the specification in column (9) of Table 2, including an additional SEM parameter, for which 'EMP' becomes more significant as more nearest neighbors are considered in its definition (not shown). However, the main explanatory power of the regression comes from the human capital variable and from the unexplained part (country effects and the spatial process in the error term). Putting it together with the results in the previous section, that is not a confirmation of the NEG interaction channels.

Figure 5. 220 regressions with 'EMP' by number of neighbors: specification in column (7) of Table 2

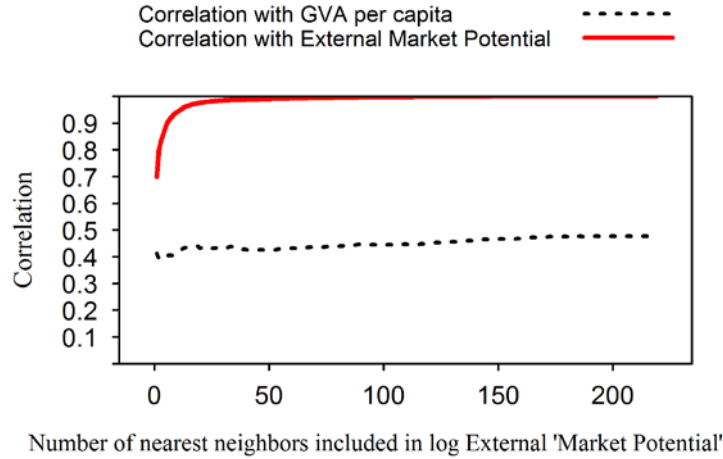


The reason for these results can be observed in Figure 6. The effect of the (weighted) summation makes the correlation of the 'EMP' variables with the full External Market Potential variable to converge very fast to 1 after the inclusion of the first nearest neighbors.¹² However, the inclusion of additional nearest neighbors does not increase the correlation between the 'EMP' variables and the dependent variable of the wage-type equation studied here. The 0.478 correlation between GVAp_c and EMP shown in Table 1 is almost reached when 'EMP' is defined for only the first nearest neighbor. On average, the relative values of the complete EMP and the 'EMP' defined for the nearest neighbor present similar discrepancies when compared with the relative values of GVAp_c. The locational information captured by both variables is different but the estimation error of

¹² This result explains the multicollinearity problems found if an additional 'EMP' variable is included in the regression, built with the regions omitted from the other 'EMP' variable (not shown).

capturing a global spatial trend of the dependent variable is similar to the error made when the focus is shifted to capture local spatial clustering. The result is driven by the statistical properties of the data and consistent with the analysis of the correlograms in Figure 3.

Figure 6. Correlation of the ‘EMP’ variables by number of neighbors with the dependent variable and with External Market Potential built for all the neighbors (logs)



The generalized wage-type equation (10) showed that the effects of Market Potential on marginal costs may come through factor prices or total factor productivity. The nearest neighbor effect shows that the empirical wage-type equation might be capturing local effects. This does not deny NEG theory because the nearest neighbor might be, on average, statistically representative of the accessible market size associated with each location¹³. However, the nearest neighbor effect could also be explained by knowledge spillovers, Urban Economics theories or alternative perspectives about the relationship between agglomeration and trade (Parr et al., 2002). It reveals a lack of identification between local externalities and global spatial trends in the wage-type equation.

The main practical implication of considering all the regions in the sample when building External Market Potential is to reduce the weight of the internal market in a full variable of Market Potential (HMP). The consideration of internal markets have a marginal effect on the estimation of a wage-type equation (see Table 2) because the number of neighbors considered in EMP is unnecessarily high. The inclusion of IMP is more rele-

¹³ The author thanks Giuseppe Arbia for raising this point.

vant when estimating spatial models (Bruna et al., 2015) because those models tend to reduce the significance of the external component, so adding endogenous information of the same region reinforces the significance of HMP.

Finally, Table 4 shows a preliminary robustness analysis when the estimation is controlled for unobserved regional heterogeneity. Panel data ‘wage equations’ have been estimated by Breinlich (2006), Fingleton (2008, 2009), Boulhol and de Serres (2010) or Head and Mayer (2011). Now the novelty is to use a sample of 220 European regions to compare the results of a non-spatial and a SEM specification including Harris’ External Market Potential with those obtained for two artificial ‘EMP’ variables. The first three columns show that the results for EMP are not easily distinguishable from those with variables representing local spillovers. However, once those local effects are captured through a spatial error panel data model, the estimates for the three variables become negative.

Table 4. Fixed effects panel data models with 'EMP' by number of neighbors (1995-2008)

	(1)	(2)	(3)	(4)	(5)	(6)
'EMP' built with the nearest neighbor	0.216*** (0.016)			-0.161*** (0.017)		
'EMP' built with the five nearest neighbors		0.355*** (0.019)			-0.463*** (0.023)	
External Market Potential			1.290*** (0.085)			-0.837*** (0.132)
λ				0.647*** (0.016)	0.789*** (0.011)	0.630*** (0.017)
Adj. R-squared	0.696	0.707	0.699			
Sum squared errors	5.33	5.06	5.24	6.39	8.39	6.42

Note: Variables are in log form. The number of observations is 3,080. Standard errors are in parentheses. All the estimations include time effect. Columns (4) to (6) show estimations of a SEM model by maximum likelihood.

5. Conclusions

This paper posits an alternative approach to the NEG wage-type equation. Firstly, it sets up a theoretical framework emphasizing that the dependent variable of the ‘wage equation’ is marginal costs. This generalized wage-type equation reveals that the phenomenon to be explained is uncertain and depend on factor prices and total factor productivity. Therefore, the interpretation of the empirical results for the right-hand-side of the equation can be done in terms of any of these elements. Then, it proposes two artificial tests about the wage-type equation, that are useful to analyze the key statistical properties of the data conditioning the results. For the first time in the literature, a wage-

type equation is estimated after redefining Market Potential for different variables and different number of nearest neighbors. The results of this sensibility analysis show that the specific way of building Market Potential does not guarantee an unambiguous interpretation of the estimation results.

Why do empirical tests tend to accept the NEG (at least under the scope limitations of this paper)? The way in which Market Potential is built, as a sum for all the observations in the sample produces a spatially smoothed distribution of values. When the weights in the summation are distances with an exponent close to -1, the spatial distribution of the resulting values will be able to proxy in a stylized way the core-periphery pattern of the European regional income per capita. However, many variables measured at the regional level in Europe are spatially autocorrelated and produce similar empirical wage-type equations when Market Potential is constructed with them. Additionally, in terms of relative values, the average discrepancy of Market Potential with the dependent variable is almost the same if Market Potential is defined for the whole sample or for the first nearest neighbor. Or in other words, the estimation error of capturing a global spatial trend of the dependent variable is similar to the error made when the focus is shifted to capture spatial clustering. The main role of using the whole sample to build the variable, as predicted by NEG, is to reduce the weight of an *ad hoc* and endogenous measure of internal market size on the full variable of Market Potential. This result cast doubt on some of the previous empirical evidence supporting NEG's interaction channels for the European regions, due to a lack of identification between local spillovers and global trends in a wage-type equation.

The paper opens several lines of future research. The proposed tests can be studied in other geographical samples. They can be repeated using other proxy variables for Market Potential or using different estimation techniques. The discussion can be extended through the estimation of more sophisticated spatial models or the inclusion of additional control variables. However, the results presented here point to a high degree of uncertainty when interpreting the results of an empirical wage-equation for European regions.

References

Ahlfeldt, G., Feddersen, A. (2008) *Determinants of Spatial Weights in Spatial Wage Equations: A Sensitivity Analysis within a Consistent Framework*. Working Paper

22. Chair for Economic Policy, University of Hamburg. Available at: <http://ideas.repec.org/p/hce/wpaper/022.html>.
- Baldwin, R. E., Forslid, R., Martin, P., Ottaviano, G., Robert-Nicoud, F. (2003) *Economic Geography and Public Policy*. Princeton, N.J.: Princeton University Press.
- Behrens, K., Gaigné, C., Ottaviano, G. I. P., Thisse, J.-F. (2006) Is remoteness a locational disadvantage?. *Journal of Economic Geography*, 6: 347–368.
- Bivand, R., Pebesma, E. J., Rubio, V. G. (2008) *Applied Spatial Data Analysis with R*. Berlin: Springer.
- Boulhol, H., de Serres, A. (2010) Have developed countries escaped the curse of distance?. *Journal of Economic Geography*, 10: 113–139.
- Boulhol, H., de Serres, A., Molnár, M. (2008) The contribution of economic geography to GDP per capita. *OECD Journal: Economic Studies*, 2008: 1–37.
- Brakman, S., Garretsen, H., van Marrewijk, C. (2009a) Economic Geography within and between European nations: the role of Market Potential and density across space and time. *Journal of Regional Science*, 49: 777–800.
- Brakman, S., Garretsen, H., van Marrewijk, C. (2009b) *The New Introduction to Geographical Economics*. Cambridge University Press.
- Breinlich, H. (2006) The spatial income structure in the European Union—what role for Economic Geography?. *Journal of Economic Geography*, 6: 593–617.
- Bruna, F. (2015a) A generalized NEG wage-type equation. In C. Díaz-Roldán, J. Perote (eds) *Advances on International Economics*, pp. 63–82. Newcastle: Cambridge Scholars Publishing.
- Bruna, F. (2015b) What is the dependent variable of the NEG ‘wage’ equation? *Mimeo*, Brunas, F., Lopez-Rodriguez, J., Faiña, A. (2015) Market Potential, Spatial Dependences and Spillovers in European Regions. *Regional Studies*, Cambridge Econometrics (2015) *The European Regional Economic Model (EUREGM)*. Available at: <http://www.camecon.com/SubNational/SubNationalEurope>.
- Clark, C., Wilson, F., Bradley, J. (1969) Industrial location and economic potential in Western Europe. *Regional Studies*, 3: 197–212.
- Collet, I., Engelbert, A. (2013) *Coastal regions: people living along the coastline, integration of NUTS 2010 and latest population grid*. Statistics in focus 30/2013. Eurostat. Available at: http://ec.europa.eu/eurostat/statistics-explained/index.php/Coastal_regions_-_population_statistics.
- Combes, P.-P., Mayer, T., Thisse, J.-F. (2008) *Economic geography: the integration of regions and nations*. Princeton, N.J.: Princeton University Press.
- Dijkstra, L., Annoni, P., Kozovska, K. (2011) *A new regional competitiveness index: Theory, Methods and Findings*. European Union Regional Policy Working Papers 02/2011. Directorate-General for Regional Policy. Available at: http://ec.europa.eu/regional_policy/information/focus/index_en.cfm.
- Doran, J., Fingleton, B. (2015) Resilience from the micro perspective. *Cambridge Journal of Regions, Economy and Society*, 8: 205–223.
- Duranton, G., Puga, D. (2004) Micro-foundations of urban agglomeration economies. In J.V. Henderson, J.-F. Thisse (eds) *Handbook of Regional and Urban Economics* 4, pp. 2063–2117. Amsterdam: North Holland.
- Fingleton, B. (2006) The new economic geography versus urban economics: an evaluation using local wage rates in Great Britain. *Oxford Economic Papers*, 58: 501–530.

- Fingleton, B. (2008) Competing models of global dynamics: evidence from panel models with spatially correlated error components. *Economic Modelling*, 25: 542–558.
- Fingleton, B. (2009) Testing the NEG model: Further evidence from panel data. *Region et Developpement*, 30, 141–158.
- Fingleton, B., Fischer, M. M. (2010) Neoclassical theory versus new economic geography: competing explanations of cross-regional variation in economic development. *The Annals of Regional Science*, 44: 467–491.
- Fujita, M., Krugman, P., Venables, A. J. (1999) *The spatial economy: cities, regions and international trade*. Cambridge: The MIT Press.
- 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: 315–348.
- Head, K., Mayer, T. (2000) Non-Europe: the magnitude and causes of market fragmentation in the EU. *Review of World Economics*, 136: 284–314.
- Head, K., Mayer, T. (2004) The empirics of agglomeration and trade. In J.V. Henderson, J.-F. Thisse (eds) *Handbook of Regional and Urban Economics 4*, pp. 2609–2669. Amsterdam: North Holland.
- Head, K., Mayer, T. (2006) Regional wage and employment responses to market potential in the EU. *Regional Science and Urban Economics*, 36: 573–594.
- Head, K., Mayer, T. (2011) Gravity, market potential and economic development. *Journal of Economic Geography*, 11: 281–294.
- Head, K., Mayer, T. (2014) Gravity equations: workhorse, toolkit, and cookbook. In G. Gopinath, E. Helpman, K. Rogoff (eds) *Handbook of International Economics 4*, pp. 131–195. Amsterdam: Elsevier.
- Keeble, D., Owens, P. L., Thompson, C. (1982) Regional accessibility and economic potential in the European Community. *Regional Studies*, 16: 419–432.
- Kordi, M., Kaiser, C., Fotheringham, A. S. (2012) A possible solution for the centroid-to-centroid and intra-zonal trip length problems. In J. Gensel, D. Josselin, D. Vandenbroucke (eds) *Proceedings of the AGILE International Conference on Geographic Information Science*, pp. 147–152. Avignon, France: AGILE.
- Kosfeld, R., Eckey, H.-F. (2010) Market access, regional price level and wage disparities: the German case. *Jahrbuch für Regionalwissenschaft*, 30: 105–128.
- Krugman, P. (1991) Increasing returns and economic geography. *Journal of Political Economy*, 99: 483–499.
- Krugman, P. (1992) *A Dynamic Spatial Model*. NBER Working Paper 4219. Cambridge, Massachusetts: National Bureau of Economic Research. Available at: <http://ideas.repec.org/p/nbr/nberwo/4219.html>.
- Leamer, E. E., Levinsohn, J. (1995) International trade theory: The evidence. In G.M. Grossman, K. Rogoff (eds) *Handbook of International Economics 3*, pp. 1339–1394. Elsevier.
- LeSage, J. P. (2014) What Regional Scientists Need to Know about Spatial Econometrics. *The Review of Regional Studies*, 44: 13–32.
- LeSage, J. P., Pace, R. K. (2014) The Biggest Myth in Spatial Econometrics. *Econometrics*, 2: 217–249.
- Negreiros, J. (2009) Spatial Neighborhood Matrix Computation: Inverse Distance Weighted versus Binary Contiguity. In A. Car, G. Griesebner, J. Strobl (eds) *Geospatial Crossroads @ GI_Forum' 09: Proceedings of the Geoinformatics Forum Salzburg*, pp. 154–157. Salzburg, Austria: Wichmann Herbert.

- Niebuhr, A. (2006) Market access and regional disparities. *The Annals of Regional Science*, 40: 313–334.
- Overman, H. G. (2004) Can we learn anything from economic geography proper?. *Journal of Economic Geography*, 4: 501–516.
- Parr, J. B., Hewings, G. J. D., Sohn, J., Nazara, S. (2002) Agglomeration and Trade: Some Additional Perspectives. *Regional Studies*, 36: 675–684.
- Popper, K. R. (1959) *The Logic of Scientific Discovery* (2002 edition). London: Routledge.
- Puga, D. (2010) The magnitude and causes of agglomeration economies. *Journal of Regional Science*, 50: 203–219.
- Redding, S. J. (2011) Economic Geography: a Review of the Theoretical and Empirical Literature. In D. Bernhofen, R. Falvey, D. Greenaway, U. Kreickemeier (eds) *Palgrave Handbook of International Trade*, pp. 497–531. London: Palgrave Macmillan.
- Redding, S. J., Venables, A. J. (2004) Economic geography and international inequality. *Journal of International Economics*, 62: 53–82.
- Rosenthal, S. S., Strange, W. C. (2004) Evidence on the nature and sources of agglomeration economies. In J.V. Henderson, J.-F. Thisse (eds) *Handbook of Regional and Urban Economics 4*, pp. 2119–2171. Amsterdam: North Holland.
- Stewart, J. Q. (1947) Empirical mathematical rules concerning the distribution and equilibrium of population. *Geographical Review*, 37: 461–485.

Appendix

A.1 Sample

The regional units studied in this paper are based on Eurostat’s nomenclature of territorial units for statistics (NUTS), 2006 version, at the aggregation level NUTS 2. The sample includes 220 regions from 17 countries: Austria, Belgium, Spain, Finland, France, Greece, Ireland, Italy, Luxembourg, Netherlands, Norway, Portugal, Sweden Switzerland and United Kingdom. The calculations with capital stock data exclude the 14 regions of Norway and Switzerland.

A.2 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. Human capital stock is proxied by Eurostat’s share of population who have successfully completed education at the third level in science and technology fields of study.

Breinlich (2006) and Ahlfeldt and Feddersen (2008) find that building Market Potential with travel times instead of geographical distances does not alter significantly the

results. Great circle distances (d_{ij}) among regional centroids are used here, calculated from GISCO's shape files (© EuroGeographics for the administrative boundaries. Regional areas and the Euclidean distances for Figure 3 are calculated from these files after an EPSG 3035 projection.

A.3 Trend surface of GVApc and EMP

Figure 2 shows level plots of the predictions of the logs of GVApc and EMP on a polynomial of degree 2 of the standardized coordinates of the regional centroids after being projected on the plane. If E and N are these coordinates, the model for the dependent variable studied in this paper is the following: $\ln \text{GVApc} = C + \beta_1 E + \beta_2 N + \beta_3 E^2 + \beta_4 N^2 + \beta_5 EN$. The estimated model is used to predict the variable on an artificial grid of spatial coordinates and those predictions are colored by intervals of values. The white color corresponds to out of range predictions.