CAHIER D’ÉTUDES
WORKING PAPER

N° 37

REGIONAL WAGES AND MARKET POTENTIAL IN THE ENLARGED EU: AN EMPIRICAL INVESTIGATION

Thomas Y. MATHÄ
Allison SHWACHMAN

APRIL 2009

BANQUE CENTRALE DU LUXEMBOURG
EUROSYSTÈME

Regional wages and market potential in the enlarged EU: An empirical investigation[♦]

Thomas Y. Mathä[♣]
Central Bank of Luxembourg

Allison Shwachman[♠]
Clark University

Abstract:

This paper empirically analyses the link between market potential and regional wages in the enlarged EU. We extend previous studies of EU regions in several ways. 1) we analyze the link between market potential and wages for the EU27, 2) correct for spatial autocorrelation present in the data, showing that by neglecting spatial autocorrelation the strength of the relationship between market potential and wages may be underestimated, 3) decompose total market potential into several geographical components and analyze their respective contributions to explaining the geographical wage structure, and 4) explore which regions have gained the most from European integration by calculating counterfactual market potential if they could only trade with other regions within the same country.

Keywords: Market potential, market access, regional wages, distance, European Union

JEL-Codes: F12, F15, R11, R12.

♦ We would like to thank Carsten Schürmann for providing us with the dataset on bilateral distances between EU regions, Michael Pflüger and our colleagues at the BCL for their comments on earlier drafts. The views expressed in this paper are personal and may not necessarily reflect those of the respective institutions the authors are affiliated with.

♣ Economics and Research Department, e-mail: thomas.mathae@bcl.lu.

♠ Economics Department, e-mail: ashwachman@clarku.edu. This paper was written while Allison was visiting the Banque centrale du Luxembourg. She would like to thank *The Henry J. Leir Charitable Foundation* for making her research at the Banque centrale du Luxembourg possible.

Nicht-technische Zusammenfassung

Die „Neue Ökonomische Geographie“ hat zum zentralen Anliegen, die räumliche Wirtschaftsstruktur und die Ballung der wirtschaftlichen Aktivität zu erklären. Sie sieht die räumliche Ballung als einen durch zirkuläre und kumulative Kräfte bedingten endogenen Prozeß. Der entscheidende theoretische Ansatzpunkt ist dabei die Fokussierung auf Marktgrößeneffekte, die in mikroökonomisch fundierten Modellen des allgemeinen Gleichgewichts mit unvollkommenem Wettbewerb und Transportkosten auf dem Gütermarkt dargestellt werden.

Eine empirisch verifizierbare Aussage ist, dass die geographische Einkommensstruktur eng mit dem Marktzutritt oder Marktpotential eines Wirtschaftstandortes zusammenhängt. Unter letzterem versteht man im Allgemeinen den Zutritt eines Wirtschaftsstandortes zu den Märkten anderer Wirtschaftsstandorte. Es wird im Allgemeinen als entfernungsgewichtete Summe der wirtschaftlichen Aktivität aller Wirtschaftstandorte berechnet. Der grundlegende Ansatzpunkt der Neuen Ökonomischen Geographie ist somit offenkundig. In einer Welt der Marktgrößeneffekte, in der Firmen sich auf die Produktion bestimmter Güter spezialisieren, sind von Absatzmärkten weit entfernt liegende Wirtschaftstandorte benachteiligt, da im Durchschnitt höhere Transportkosten bezahlt werden müssen und folglich geringere Umsätze erzielt werden, was letztendlich zu geringeren Lohneinkommen der Beschäftigen führt.

Das vorliegende Arbeitspapier untersucht anhand von Regionaldaten für die erweiterte Europäische Union inwieweit dieser Sachverhalt zutreffend ist. Die Schätzungen belegen, dass ein signifikanter Zusammenhang zwischen dem Marktpotential eines Wirtschaftstandortes und dem Lohneinkommen der Beschäftigten besteht. Steigt der Marktzutritt einer Region um das Doppelte, so steigt das durchschnittliche Lohneinkommen um 6–15 Prozent in den alten EU Wirtschaftsstandorten und 37–59 Prozent in den Wirtschaftsstandorten der neuen EU Mitgliedstaaten. Zudem erzielt die Verbesserung des Ausbildungsniveaus eine beträchtliche volkswirtschaftliche Rendite. Steigt der Anteil der aktiven Bevölkerung mit Hochschulausbildung um 1 Prozentpunkt, so steigt das durchschnittliche Lohneinkommen um ungefähr 1 Prozent; dieses gilt sowohl für die Regionen der alten EU15 Staaten als auch der neuen Mitgliedsstaaten.

Unterteilt man das Gesamtmarktpotential in verschiedene geographische Einheiten, so ergibt sich, dass das lokale Marktpotential eines Wirtschaftsstandortes, d.h. das Marktpotential, welches ein Wirtschaftsstandort auf Grund der Nachrage nach eigenen Produkten aufweist, das Einkommensniveau der Beschäftigten maßgeblich bestimmt, während unmittelbar angrenzende Wirtschaftsstandorte keinen signifikanten Einfluß ausüben. Der Marktzutritt zu allen anderen Wirtschaftsstandorten in der erweiterten EU ist jedoch von ähnlicher Bedeutung wie das lokale Marktpotential. Für das Lohneinkommen ist von Bedeutung, wo sich ein Wirtschaftstandort geographisch gesehen im Vergleich zu der wirtschaftlichen Aktivität anderer Wirtschaftstandorte in der EU befindet, d.h. wie weit entfernt seine Absatzmärkte im Vergleich derer anderer Wirtschaftsstandorte sind. Diese Gesamtergebnisse auf Luxemburg übertragend, ließe sich die Schlußfolgerung ableiten, dass Luxemburgs Einkommen und Wohlstand vor allem auf seiner eigenen Wirtschaftskraft und dem sehr guten

Marktzutritt zu allen anderen Wirtschaftstandorten in der EU beruhen. Die angrenzenden Wirtschaftsstandorte der Großregion dagegen hätten keine Relevanz.

Der Wirtschaftsstandort Luxemburg hat in der Vergangenheit in ausgesprochenem Maße von der Europäischen Integration profitiert. Die Resultate eines Gedankenexperimentes zeigen, dass der Wirtschaftsstandort Luxemburg mit am härtesten getroffen werden würde, wenn morgen alle Staatsgrenzen in der EU schlössen und der Warenaustausch zum Erliegen käme. Luxemburg hat demnach ein großes Interesse an einer Weiterführung und Vertiefung der Europäischen Integration.

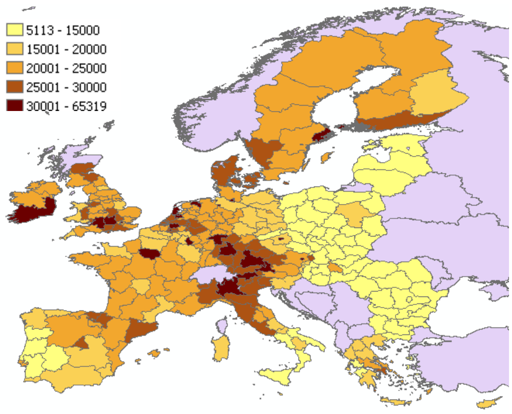
1. Introduction

Regional economic activity and income vary substantially across Europe. With the accession of 10 countries in 2004, socio-economic disparities in the EU have doubled. According to the EU regional policy, one in four regions have a Gross Domestic Product (GDP) per capita less than 75% of the EU27 average and one third of EU citizens, amounting to approximately 170 million people, reside in the poorest regions which receive assistance under the "convergence" objective. Concern over these inequalities has lead the EU to allocate 36% of its annual budget until 2013 to "cohesion," helping less advantaged regions transform their economies to become more competitive. Given the magnitude of this financial investment (46.9 billion euro in 2008 alone) it is important to identify the determinants of income and the causes of these income inequalities so that policies can be most effective. It is also important for high-income regions, such as Luxembourg, to understand why they are so successful so that they can maintain if not improve their economic well-being.

Regional income inequalities may be caused by a combination of factors including differences in the quantity and quality of production factor endowments, technology, public infrastructure, government institutions, etc… This paper focuses on one remarkably simple factor, geography. We seek to answer the question: To what extent can geography explain the structure of regional income? By geography, we mean what is commonly referred to as "second nature" geography, which examines how proximity to consumer markets and input suppliers may explain differences in income.[1] Distance has direct effects on transaction costs, including transportation costs and information costs. Do these costs affect income, and if so, by how much?

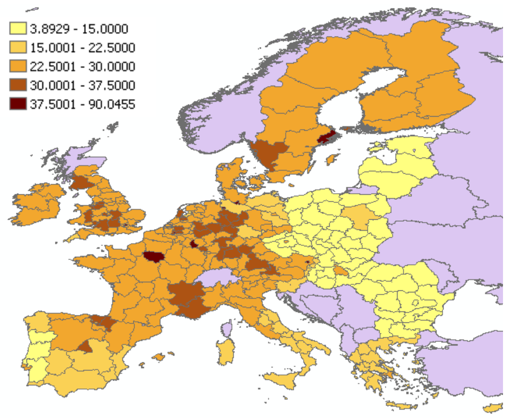
[1] "First nature" geography examines how the physical characteristics, climate and natural resources of an area affect income.

Figure 1: GDP per capita



Note: in 2004, euro pps-adjusted

Figure 2: Labor compensation per employee



Note: in 2004, euro pps-adjusted

We examine the effects of geography on regional income using a New Economic Geography (NEG) framework. Initiated by Krugman (1991), NEG seeks to explain the agglomeration of economic activity in space, with a special attention to the geographical distribution of firms and the geographical variations in prices and costs. The main focus is on the impact of market size, which is incorporated in micro-founded general equilibrium models with imperfect competition, increasing returns and transportation costs. NEG emphasizes the role of proximity or distance, and thus the access to markets or "market potential". In a world with increasing returns and specialization of production, firms in locations farther away from their markets will incur higher transportation costs, earn lower revenues and consequently pay their workers lower wages. Thus, a given location has a market potential that is increasing in economic activity and decreasing in distance. It is a commonly used tool and can be thought of as a measure of a location's access to demand (markets) calculated as a distance-weighted sum of the volume of economic activity (such as GDP or employment) in all locations. The landscape of regional GDP per capita and labor compensation (wage) per employee depicted in Figure 1 and Figure 2 reveals that both measures tend to be higher in central European regions than in peripheral regions; the so called core-periphery pattern. This suggests that geography matters.

In this context, this paper explores empirically how geographical location affects regional income in the EU27. We closely follow the approaches by Hanson (2005), Redding and

Venables (2004), Head and Mayer (2006), Breinlich (2006) and others who have used the NEG framework to estimate structural parameters. Our results corroborate previous empirical findings and extend previous studies in four ways.

First, we analyze the link between market potential and wages for the EU27 based upon Redding and Venables' (2004) NEG framework that relates market potential to the maximum level of wages a firm in each region can afford to pay. Our results support the theoretical predictions of the NEG framework and indicate that market potential is a key variable in explaining the spatial distribution of wages in the EU. We find a substantial difference between EU15 regions and EU new member state (NMS) regions as EU15 regional wages are significantly less sensitive to changes in market potential than wages in new EU member state regions. Wages are also strongly linked to higher education in the EU15. In addition, we find that regional wages respond to unemployment in the expected way.

Second, we correct for spatial autocorrelation detected in the data. The application of spatial econometric techniques extends econometric methods currently used in most studies of regional income inequalities by considering previously ignored spatial effects. Our results suggest that neglecting spatial autocorrelation may result in underestimating the strength of the relationship between market potential and wages.

Third, we decompose market potential into several components and show that a region's domestic market potential (due to the region's demand for its own goods), is a decisive factor determining the region's wage; adjacent regions do not have any significant impact. However, market access to all other EU regions is of similar importance as domestic demand. This is particularly the case for regions in the new member states of the EU.

Fourth, we identify which regions have gained most from European trade integration by calculating counterfactual market potential for the hypothetical case in which regions could only trade with other regions within the same country.

The remainder of the paper is organized as follow: Section 2 presents a brief discussion of the empirical literature of regional income disparities with a focus on NEG studies. Section 3 presents the theoretical framework used to empirically investigate the link between regional income and market potential. Section 4 outlines the empirical framework and describes the data. Section 5 discusses the results of the econometric estimations and section 6 concludes.

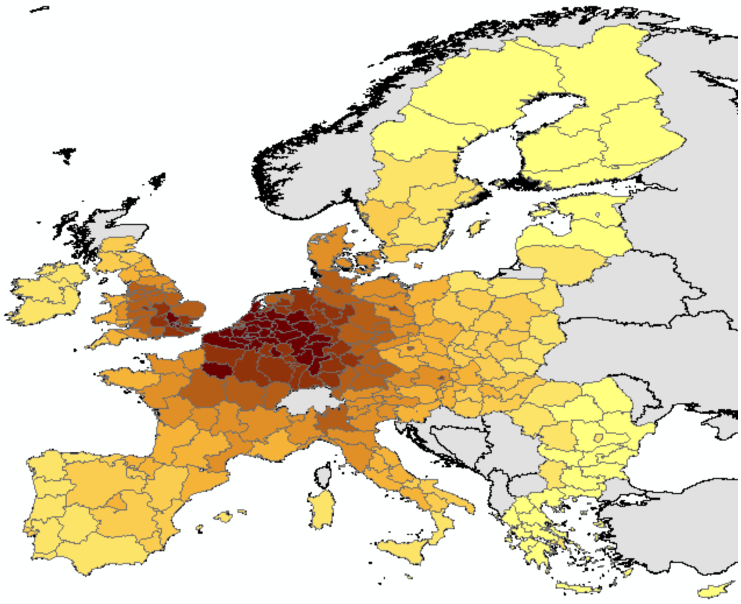
2. Previous empirical literature

The idea of market potential goes back at least as far as Harris (1954) who argued that a region's attractiveness as a production site is dependent upon its access to markets or its "market potential." Harris defined the market potential of region i as a weighted sum of the purchasing power of all regions, with the weights inversely dependent on distance. Mathematically, Harris' market potential may be represented by the following function:

(1) $$MP_i = \sum_j (x_j / d_{ij}),$$

where x_j is a measure of purchasing power or economic activity in region j and d_{ij} is the distance between regions i and j. Figure 3 shows a quantile map of market potential calculated according to Harris' formula for the EU27 regions, where x_j refers to the GDP in region j and distance is calculated as actual travel times to all other regions based on the dataset from Schürmann and Talaat (2002). We see a clear core-periphery pattern, as regions in the central countries in Europe (Luxembourg, Germany, Belgium, etc…) have the highest market potential; whereas, regions in the peripheral countries (Greece, Romania, Finland, etc…) show the lowest market potential. The core-periphery pattern displayed in Figure 3 is similar to that found in Schürmann and Talaat's (2002) European Peripherality Index as well as other previous studies (e.g. Combes and Overman, 2004). Table 1 also reinforces this fact. The 10 most peripheral regions have low market potential. Conversely, the 10 most central regions have high market potential.

Figure 3: Quantile map of market potential according to Harris' formula



Note: Own calculations; regional GDP data for 2004; actual travel time data from Schürmann and Talaat (2002).

Table 1: Most and least central regions in the EU27

	Most Central Regions			Most Peripheral Regions	
	NUTS ID	Region		NUTS ID	Region
1	DEB3	Rheinhessen-Pfalz	1	CY00	Cyprus
2	DE71	Darmstadt	2	GR42	Notio Aigaio
3	DE72	Gießen	3	GR43	Kriti
4	DE12	Karlsruhe	4	GR41	Voreio Aigaio
5	DEC0	Saarland	5	FI1A	Pohjois-Suomi
6	DE25	Mittelfranken	6	FI13	Itä-Suomi
7	DE26	Unterfranken	7	SE33	Övre Norrland
8	LU00	Luxembourg	8	FI19	Länsi-Suomi
9	DEB1	Koblenz	9	FI18	Etelä-Suomi
10	DE11	Stuttgart	10	MT00	Malta

Note: Own calculations based on actual travel time data from Schürmann and Talaat (2002).

While Harris' market potential function is intuitively appealing and produces the core-periphery patterns noticeable in the real world, it is ad-hoc and lacks microeconomic foundations. NEG provides those microfoundations and embeds them in general equilibrium with imperfect competition, increasing returns to scale and transport costs. Importantly, the geographical structure of income and economic activity arises endogenously from circular and cumulative forces. From these models a structural equation can be derived that closely resembles Harris' (1954) market potential function, which has commonly become known as the "wage equation" of NEG models. A number of empirical studies have recently tried to estimate this structural relationship between regional wages and market potential.[2] Those studies can be divided into two different strands.

The first strand examines the effects of geography at a national or regional level using a model with mobile labor and real wage equalization, most often the Helpman (1998) model. Helpman alters Krugman's model by adding housing stock to act as the main driving force for dispersion of economic activity. Studies in this first category directly estimate the wage equation usually using non-linear techniques. Hanson (2005) was one of the first to empirically investigate the predictions of NEG models. He estimates the Helpman model using US panel data and finds that regional variation in earnings is associated with market potential. Using panel data on Italian provinces, Mion (2004) estimates a linear version of the Helpman model and finds that demand linkages influence the spatial distribution of wages. Roos (2001) and Brakman et al. (2004) investigate the existence of a spatial wage structure in Germany, also finding a positive relationship between market potential and wages. Niebuhr (2006) investigates the importance of market potential in explaining regional income disparities and the geographic extent of demand linkages for a cross section of European regions in the EU15, excluding Sweden. Her results indicate a positive relationship between market potential and regional wages in Europe.

The second strand of empirical studies stems from the work of Redding and Venables (2004) who develop a theoretical trade and geography model to examine the relationship between economic geography and income at the international level. They develop a two-step procedure, which avoids having to make the assumptions of real wage equalization and labor mobility. First, they estimate a gravity type relationship for bilateral trade flows between countries. Using the trade equation estimates they derive market access for each country, a measure of a market potential based upon the export demand each country faces given its geographical location relative to that of its trading partners. Finally, they derive a

---

2 For a review of the main ingredients of NEG and the empirics with regard to agglomeration and trade, see Overman et al. (2003) and Head and Mayer (2004a).

wage equation by using the zero profit condition for firms, which defines the maximum level of wages a firm in each country can afford to pay, given its market access. This equation allows for the estimation of the relationship between market access and wages. Using data on 101 countries, their results show that market access is important in explaining cross-country variation in per capita income.

Breinlich (2006) follows the Redding and Venables approach to examine the regional income structure in the EU15. As bilateral trade data is not available for European regions, he assumes that interregional trade flows are determined by similar forces as international trade flows. Using a sample of 193 European regions, his results indicate that market access is a significant determinant of income in the EU15. Head and Mayer (2006) estimate a similar model to examine the relationship between wages and employment and market potential. Using data on 13 manufacturing industries and 57 European regions, their findings show that wage adjustment is the main path to spatial equilibrium. Wages respond to market potential in a positive way. While employment also adjusts, it does not do so in any consistent or significant manner.

This study aims to contribute to the second strand of the empirical NEG literature. There are several reasons why we choose to follow this second approach. First, as noted by Breinlich, this method has the advantage that trade costs are estimated more accurately because they are derived from a gravity equation as opposed to depending solely on bilateral distance. Second, in deriving the wage equation we will use the firms' zero profit condition and assume labor immobility. This assumption is more realistic than that made in the Helpman model, given the low degree of labor mobility in the EU.[3]

Previous empirical studies of European regions provide support for the predictions of theoretical NEG models. However, these studies have often focused on single countries and are limited to the EU15. This study analyzes the significance of market potential in explaining

[3] According to a recent labor mobility survey 75% of Europeans have never lived outside their region of birth (Vandenbrande et al., 2006). The European Commission (2006) judges that only about 1.2% of the working age population in the EU15 countries has moved residence within the country from one NUTS2 region to another since the year before (based on results from the European labor force survey).

income inequalities in the EU27, to our knowledge this is the first study to do so. This study also deviates from the majority of previous work in that we explicitly consider spatial autocorrelation. It is likely that economic variables in one region are correlated with economic variables in neighboring regions. According to Overman and Puga (2002) and the OECD (2005), regions have unemployment outcomes that are closer to neighboring regions than to other more distant regions in the same country. We show that this correlation extends beyond unemployment to other economic variables, including wages. This idea is reinforced from Figure 1, which illustrates that regions with high incomes are often neighbored by regions with high incomes; whereas, regions with low income are surrounded by regions with low incomes. In consideration of this spatial correlation, we employ spatial econometric techniques, which are hitherto rarely implemented in previous NEG studies.

3. Theoretical framework

The theoretical framework closely follows the framework developed by Redding and Venables (2004) and subsequently applied by Head and Mayer (2006) and Breinlich (2006), and others. On the demand side, we assume that the world consists of i = 1, 2, 3, … R regions. We concentrate on the manufacturing sector that operates under increasing returns to scale and imperfect competition in which each firm produces a variety of a differentiated product. Demand in location j for manufacturing goods can be found by maximizing the CES utility function of a representative consumer:

$$(2) \qquad U_j = \left[\sum_{i=1}^{R} n_i x_{ij}^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}} \qquad \text{s.t.} \ \sum_{i=1}^{R} n_i p_{ij} x_{ij} = E_j \qquad \sigma > 1$$

E_j represents region j's expenditure on manufactures. n_i is the number of firms (varieties of goods) in region i and x_{ij} is region j's demand of goods produced in i. σ is the constant elasticity of substitution between any two varieties, which is common across all regions. p_{ij} is the price of varieties produced in i and sold in j. Manufacturing firms sell their varieties in all regions and thus incur transportation costs. These costs take the iceberg form introduced by Samuelson (1954) in which a fraction of any shipped good "melts away" and thus for every unit shipped only $1/ T_{ij}$ units arrives at the destination. Therefore, $p_{ij} = p_i T_{ij}$ where

p_i is the price at the point of production and $T_{ij} > 1$ represents the transportation costs between regions i and j. ($T_{ij} = 1$ represents costless trade). We define G_j as the price index for manufacturing varieties in region j defined over the prices of varieties produced in i and sold in j (p_{ij}).

(3) $$G_j = \left[\sum_{i=1}^{R} n_i p_{ij}^{1-\sigma} \right]^{1/(1-\sigma)}$$

By optimization, it can be shown that region j's demand for each variety produced in i is given by:

(4) $$x_{ij} = p_{ij}^{-\sigma} E_j G_j^{\sigma-1}$$

On the supply side, firms maximize profits, which can be represented by the following function:

(5) $$\pi_i = \sum_{j=1}^{R} p_i x_{ij} - w_i^{\alpha} z_i^{\beta} c_i (F + x_i),$$

where w_i is the wage rate with input cost share α and z_i is the price of the mobile factor of production with input cost share β and where $\alpha + \beta = 1$.[4] α and β are common across all regions. The total production of each firm is x_i. The firm experiences increasing returns to scale represented by a fixed input cost c_iF and a marginal input cost, c_i. Firms set prices as a constant markup $[\sigma/(\sigma-1)]$ over marginal cost such that:

(6) $$p_i = w_i^{\alpha} z_i^{\beta} c_i [\sigma/(\sigma-1)]$$

Assuming free entry and exit, profits will be driven to zero and thus equilibrium output is: $x_i = \bar{x} = F(\sigma-1)$. Inserting this result and the assumption $p_{ij} = p_i T_{ij}$ into the demand function (equation (4)) and summing over j we obtain:

(7) $$p_i^{\sigma} \bar{x} = \sum_{j=1}^{R} T_{ij}^{1-\sigma} E_j G_j^{\sigma-1}$$

Inserting the profit maximizing price of equation (6) into equation (7) we attain:

(8) $$\bar{x}(w_i^{\alpha} z_i^{\beta} c_i (\frac{\sigma}{\sigma-1}))^{\sigma} = \sum_{j=1}^{R} T_{ij}^{1-\sigma} E_j G_j^{\sigma-1}$$

4 Some previous studies also include a composite intermediate input, which allows for the investigation of the relationship between supplier access and wages. As here we are most interested in the relationship between market potential and wages, and due to the extreme multicollinearity problems that arise between supplier access and market potential we do not include this input.

Finally, rearranging this equation, we arrive at the wage equation:

(9) $$w_i = A(MP_i)^{1/\alpha\sigma} z_i^{-\beta/\alpha} c_i^{-1/\alpha} \text{ with } A = \bar{x}^{-1/\alpha\sigma}(\frac{\sigma}{\sigma-1})^{-1/\alpha} \text{ and}$$

(10) $$MP_i = \sum_{j=1}^{R} T_{ij}^{1-\sigma} E_j G_j^{\sigma-1}$$

where MP_i is the market potential of region i and A is a constant common across all regions. Equation (9) forms the basis of our empirical analysis. It says that the maximum wage a firm in region i can afford to pay is a function of the price of the mobile factor of production *(z_i)*, technology *(c_i)* and market potential *(MP_i)*. We define market potential of region i as the sum of the transportation-cost-weighted market capacities for all regions. Market capacity is expenditure multiplied by price index, and can be interpreted as a region's real purchasing power. Taking logs on both sides of equation (9) gives the non-linear equation to be estimated:

(11) $$\ln(w_i) = \ln(A) + \frac{1}{\alpha\sigma}\ln(\sum_{j=1}^{R} T_{ij}^{1-\sigma} E_j G_j^{\sigma-1}) + \varepsilon_i$$

where the error term *(ε_i)* encapsulates the mobile factor of production *(z_i)* and technological (c_i) differences between regions. We estimate equation (11) using the strategy introduced by Redding and Venables (2004) that involves a two-step procedure. In the first step, trade data is used to obtain estimates of market capacities and bilateral trade costs. The second step uses the prior obtained estimates to construct a measure of market potential, which is then used as an explanatory variable in the wage equation. We follow the two-step approach because there is no data available for the price index, G, and this approach has the advantage that trade costs are estimated more accurately as they are derived from a gravity equation as opposed to being assumed to depend solely on bilateral distance.

Therefore, the second equation employed in our empirical analysis is the trade equation, which we derive simply by aggregating equation (4).

(12) $$n_i p_i x_{ij} = n_i p_i^{1-\sigma} T_{ij}^{1-\sigma} E_j G_j^{\sigma-1}$$

Equation (12) gives total exports from region i to region j ($n_i p_i x_{ij}$), which is a function of bilateral transportation costs ($T_{ij}^{1-\sigma}$), market capacity of the importing country ($E_j G_j^{\sigma-1}$) and supply capacity of the exporting country ($n_i p_i^{1-\sigma}$). Market capacity is as described above. Supply capacity is the number of firms multiplied by their price.

4. Empirical model

4.1. Trade equation estimation and construction of market potential

In analogy to the Redding and Venables approach, ideally, we should first estimate the trade equation to obtain estimates of bilateral trade costs between regions (φ_{ij})[5], the market capacity (m_i) and supply capacity (s_i) for each region. These estimates can be defined as $\varphi_{ij} = T_{ij}^{1-\sigma}$, $m_i = E_i G_i^{\sigma-1}$ and $s_i = n_i p_i^{1-\sigma}$. In the econometric model, supply and market capacity are captured using regional exporter and importer fixed effects, *EX* and *IM* respectively. According to Redding and Venables (2004, p. 75) this approach "… has the advantage of capturing relevant country characteristics that are not directly observable but are nevertheless revealed through trade performance (for example the degree of openness of the country, and the values of prices and price indices within the country)."[6]

Unfortunately, data on bilateral trade flows between European regions does not exist. Therefore, following Head and Mayer (2006) and Breinlich (2006), we use bilateral trade data between countries in the first step to estimate market capacity and trade costs of individual countries. Then, we employ Head and Mayer's (2006) expenditure allocation rule, which allocates the estimated market capacity of country J to subunits j (regions within J) according to their shares of national GDP. By using this approach we assume that interregional trade flows are determined by forces analogous to those determining international trade flows.[7] For example, we assume that the distance penalty, the amount by which trade is hindered due to distance, is the same for regional trade as for international trade.

5 This parameter is commonly referred to as the "freeness" of trade. See Baldwin et al. (2003)

6 Redding and Venables (2004) also consider an alternative specification in which country dummies are replaced by economic and geographic variables and find that the main results are robust to either specification.

7 Studies on trade within countries have shown that gravity equation estimates yield similar results to distance effect estimates. See Combes et al. (2005).

By taking logs and rewriting the trade equation (12) using the definitions of supply capacity, market capacity and trade costs, we obtain our equation to be estimated:

(13) $\ln(n_i p_i x_{ij}) = \ln(s_i) + \ln(m_j) + \ln(\varphi_{ij})$

The dependent variable is the value of exports from location i to location j. We capture supply capacity ($\ln(s_i)$) of the exporting country, using exporter fixed effects, EX_I and market capacity ($\ln(m_j)$) of the importing country using importer fixed effects, IM_J. For example, the dummy variable EX_{LU} will equal one if Luxembourg is the exporter and zero otherwise. Similarly, IM_{LU} will equal one if Luxembourg is the importer and zero otherwise. Bilateral transport costs ($\ln(\varphi_{ij})$) are a multiplicative function of distance between capital cities (d_{IJ}), border effects (B_{IJ}), adjacency effects (L_{IJ}) and language effects (L_{IJ}). Therefore, letting I and J represent two European countries, equation (13) is estimated by the following:

(14) $ln(X_{IJ}) = \Psi_I EX_I + \Omega_J IM_J + \delta ln(d_{IJ}) + \gamma B_{IJ} + \theta L_{IJ} + \Lambda C_{IJ} + \varepsilon_{IJ}$,

where X_{IJ} is the value of exports from country I to country J. Following previous studies, this estimation includes internal trade (X_{II}), which is equal to production minus total exports. d_{IJ} is the distance in kilometers between capital cities in country I and country J. [8] B_{IJ} is a dummy variable intended to capture the reduction in trade due to crossing a national border and thus equals one when I is not equal to J and zero otherwise. L_{IJ} is a dummy variable that equals one if country I and country J share an official language. C_{IJ} is a dummy variable that equals one if I and J are adjacent and zero otherwise.

We use the parameter estimates from equation (14) to help calculate market potential, as market potential is defined as the sum of the transportation-cost-weighted market capacities for all regions. The coefficient estimates of the importer fixed effects provide estimates of country market capacities as they capture all the determinants (including expenditure and price level) of countries' propensities to demand imports from all partners. From the esti-

[8] A region's distance to itself is calculated as: $d_{ii} = (2/3)\sqrt{area/\pi}$, which represents the average distance from the region's center to all other points in the region assuming the region is circular. A country's distance to itself d_{II} is a weighted sum of all regional internal distances within that country, with the weights dependent on regional shares of national GDP or $d_{II} = \sum_{i,i\in I}(d_{ii} y_i / y_I)$.

mates, we calculate country market capacity as $\hat{m}_J = \exp(\hat{\Omega})$. To compute regional market capacity, we follow Head and Mayer's (2006) expenditure allocation rule, which allocates the estimated market capacity of country J to subunits j (regions within J) according to their shares of national GDP. Assuming homotheticity, expenditure in region j is $E_j=(y_j/y_J)E_J$, where (y_j/y_J) is region j's share of national GDP. We also assume that the price index is the same for each region within a country. Therefore the market capacity of region j is $\hat{m}_J = \exp(\hat{\Omega})(y_i / y_J)$. Using estimated coefficients from equation (15), we calculate bilateral transport costs as:

(15) $$\hat{\phi}_{ij} = d_{ij}^{\hat{\delta}} \exp(\hat{\gamma} B_{ij} + \hat{\theta} L_{ij} + \hat{\Lambda} C_{ij}) \ ,$$

where d_{ij} is the distance between regions. B_{ij} equals zero and L_{ij} equals one if two regions are in the same country. C_{ij} equals one if region i and region j are adjacent. Combining regional market capacities and bilateral transport cost estimates, we construct market potential following equation (10):

(16) $$M\hat{P}_i = \sum_j \hat{\phi}_{ij} (y_j / y_J) \exp(\hat{\Omega}_J)$$

In the second step, we estimate the wage equation using the calculated values of market potential. Taking logs of the wage equation (11) and letting $ln(A) = \alpha_0$ and $(1/\alpha\sigma) = \alpha_1$ we have the following:

(17) $$\ln(w_i) = \alpha_0 + \alpha_1 \ln(M\hat{P}_i) + \varepsilon_i \ ,$$

where the error term is as previously described. This estimation will allow us to establish the extent to which variation in wages may be explained by geographical location and access to markets.

4.2. Data and descriptive statistics

Regional data used in this paper are primarily drawn from Eurostat's REGIO database. Regional coverage is based upon "The Nomenclature of Territorial Units for Statistics" (NUTS), which presents a sub-national disaggregation of regional economies. NUTS is a hierarchical classification system that divides each country into a number of NUTS1 regions, each of which is then divided into a number of NUTS2 regions and likewise into

NUTS3 regions. This paper uses data at the NUTS2 level. It is important to keep in mind when interpreting the results that regional units are often based on existing administrative rather than functional considerations and follow no consistent standards throughout the EU. We obtain spatial data, including a shapefile, from Eurostat's Geographic Information System database (GISCO). Central points for each region are calculated in the spatial software program GEODA as the average of the x and y coordinates of the regions' vertices.[9] For each pair of regions, we then calculate the 'great circle' distance between central points, meaning the distance measured along the surface of the earth. We also provide results using Schürmann and Talaat's (2002) dataset in which interregional distances are represented by estimated road travel time between region capitals. While slightly outdated, this dataset takes into account road types, speed limits, border delays and other important considerations for travel time estimations. Both measures provide similar results; the results based on great circle distance between central points is used in most of the results presented, while the results based on the Schürmann and Talaat (2002) dataset are presented in Section 5.5. Bilateral trade data, at the national or NUTS0 level are taken from Eurostat's COMEXT database. The final dataset comprises 254 NUTS2 EU27 regions and our analysis will focus on a cross section for the year 2004.[10]

Table 2: Descriptive Statistics

Variable	Obs	Mean	Median	Std. Dev.	Min	Max
Wage per employee	254	23,608	24,544	10,008	3,893	90,046
GDP per capita	254	21,014	28,810	8,205	5,113	65,319
Education	254	0.24	0.24	0.08	0.08	0.50
Agriculture share	254	0.04	0.03	0.04	0.00	0.20
Industry share	254	0.28	0.28	0.08	0.08	0.52
Services share	254	0.68	0.68	0.09	0.39	0.92
Total employment	254	805	627	633	13	4,701
Area in km2	254	16,733	10,684	20,155	161	165,296

Note: Data refer to 2004. Wage per employee and GDP per capita in PPS. Employment in thousands.

In our data sample, average annual wages are approximately 23,600 euro and the average GDP per capita is approximately 21,000 euro. However, the disparities between regions are evident as labor compensation (wages) per employee and GDP per capita in the

9 For more information see www.geoda.uiuc.edu.
10 Please see Appendix A for list of regions and variable definitions.

wealthiest region are over 22 and 12 times greater respectively than their level in the poorest region. In the average region, 24% of the economically active population has pursued tertiary education (ISCED levels 5 & 6). This share ranges from 8% to 50% within our sample. The economic composition in the average regions is 4% agriculture, 28% industry and 67% service. Finally, the descriptive statistics on employment and area size reveal large variations across regions that must be taken into account when analyzing results.

Before proceeding to the formal econometric analysis we examine what spatial differences among NUTS2 regions are visible in the data and might potentially affect regional wages. Do certain locations or regional categories exhibit consistently lower or higher wages?

Table 3: Wage differences across regional categories

Regions	Nobs	Mean	Std. Err.	T-test on means equality	Wilcox rank sum test
EU15 vs.	200	26.92	0.59	Pr(|T| > |t|) = 0.00	Prob > |z| = 0.00
New member states	54	11.33	0.68	t = -17.3539	z = -10.56
Capital vs.	27	30.36	3.76	Pr(|T| > |t|) = 0.06	Prob > |z| = 0.13
Non-Capital	227	22.81	0.52	t = -1.9876	z = -1.52
Coast vs.	111	23.70	0.55	Pr(|T| > |t|) = 0.89	Prob > |z| = 0.88
Non-Coast	143	23.54	1.03	t = -0.1401	z = 0.15

Note: Data for 2004.

As expected, EU15 regions have higher average wages than new EU member state regions. Cartographic evidence of this fact is shown in Figure 2 and also confirmed by a t-test and Wilcoxon rank sum test, the latter being robust to non-normality. Both tests indicate statistically significant differences in the mean, as shown in Table 3. At the regional level, evidence of higher average wages in capital cities is mixed as the t-test is marginally significant and the Wilcoxon rank test is not statistically significant. However, in Figure 2, many of the regions containing capital cities stand out as having high wages. Our results provide no evidence of systematic differences in average wages between coastal and non-coastal regions.

5. Results

5.1. First stage: Gravity estimation

Table 4 gives the results of the OLS estimation of equation (15). As expected, distance has a negative effect on trade and is statistically significant at the 1% level. The coefficient estimate for distance (-1.4) is similar to that found by Redding and Venables (2004) and Head and Mayer (2004b).[11] The estimates suggest that after controlling for bilateral distance, border effects do not have a statistically significant influence on trade flows. Despite the fact that there are open borders between EU countries, this finding is surprising and unexpected. Language and adjacency have the expected positive signs and are statistically significant at the 5% and 10% levels respectively. The OLS model explains approximately 92% of the cross-section variation in bilateral trade flows.

Silva and Tenreyro (2006) argue that it is inappropriate to use OLS to estimate a log-linearized gravity equation in the presence of heteroskedasticity. They show using Jensen's inequality that under heteroskedasticity OLS estimation can lead to biased estimates.[12] The authors propose using Poisson pseudo-maximum likelihood estimation. A Poisson regression assumes that the distribution mean is equal to the variance. This assumption is not valid for our data since it is characterized by overdispersion: the variance is larger than the mean. Therefore, we estimate a negative binomial model, which allows for over-dispersed data, in addition to the OLS model. The coefficient estimates for the distance and importer and exporter dummy variables are similar to the OLS estimates. The coefficient estimate for border is now statistically significant, indicating that crossing a national border reduces trade. Language is significant at the 10% level, suggesting that sharing a language increases trade. According to our estimates, trade between countries that share an official language is 34% (calculated as $e^{0.292}-1$) greater than trade between countries without a common language. Head and Mayer (2004b) find that trade increases 50% with a common language. Our smaller result may be due to the fact that we are using more current data

11 Redding and Venables' coefficient estimates for distance range from -1.7 to -1.4; Head and Mayer's distance effect averages -1.4.

12 Jensen's inequality implies that $E(\ln y) \neq \ln E(y)$. In classical regression, heteroskedasticity only implies inefficient estimates. In our case, the bias comes from the fact that once transformed, heteroskedastic errors may be correlated with the covariates.

than Head in Mayer (2004), who used data from 1988–1995 or because we are using data for the EU27. The declining influence of sharing an official language may be both a reflection of increased internationalization of economic activity as well as an increase in the number of people learning and speaking European languages beyond their native language in the past decade. Finally, the coefficient estimate of adjacency is statistically significant at the 1% level and implies that adjacency increase trade by approximately 50%.

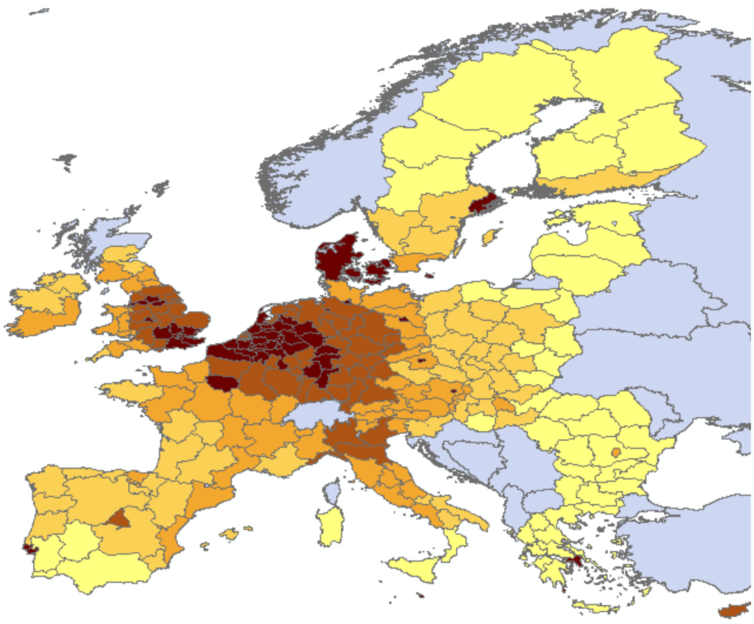
Table 4: Results of the trade equation

Dep Var ln(exports)	OLS	Neg. Binomial
ln(distance)	-1.371***	-1.406***
	(0.090)	(0.066)
Border	-0.281	-0.811***
	(0.443)	(0.243)
Language	0.354*	0.292*
	(0.213)	(0.153)
Adjacency	0.339**	0.405***
	(0.135)	(0.119)
Importer fixed eff.	Yes	Yes
Exporter fixed eff.	Yes	Yes
Observations	729	729
R-squared	0.919	

Note: OLS estimated with robust standard errors. Std. err. in parentheses. p<0.10, ** p<0.05, *** p<0.01

We use the trade equation coefficient estimates from the negative binomial regression to calculate market potential according to equation (16). This will be our main variable of interest and includes both domestic and external market access. Figure 4 shows a cartographic view of market potential by region. There is a clear core-periphery pattern with regions in the economic center of Europe (Belgium, Luxembourg, the Netherlands and Western Germany) displaying the highest market potential and regions in the periphery (Bulgaria, Romania, Northern Finland…) displaying the lowest market potential. London and other capital cities stand out as having high market potential relative to their surrounding regions.

Figure 4: Calculated market potential for EU27 regions



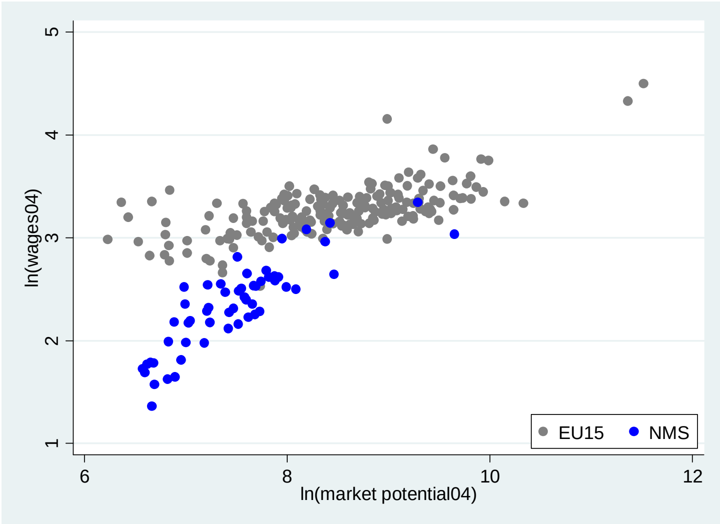
Note: Data and calculations for 2004.

5.2. Second stage: Wage equation

We now estimate the wage equation, using the calculated market potential variable. The scatter plot in Figure 5 displays a positive relationship between wages and market potential. This relationship is robust and not caused by the influence of few regions. Two distinct groups clearly emerge from the scatter plot, the EU15 regions and the NMS regions. These two groups are associated with different slopes and intercepts. Figure 5 illustrates that EU15 regions have higher wages and are not as sensitive to market potential as NMS regions. To control for these differences, we estimate separate regressions for EU15 and NMS regions.[13]

[13] An alternative method would be to include dummy variables for the EU15 and interaction terms.

Figure 5: Scatter plot between market potential and wages



Note: Data and calculations for 2004.

Equation (17) represents a limited explanation of regional wage structure as there are certainly other factors besides market potential that contribute to the spatial distribution of wages. Therefore, we include a number of control variables in the regression model in an effort to check the robustness of the relationship between market potential and wages. Wages may be significantly affected by worker heterogeneity; therefore, we include an education variable in the wage equation. We control for the economic composition of regional economies by including variables for shares of agriculture and industry in each region's GDP. These can be thought of crude proxies for factor endowment differences. Additionally, we include a control for regional long-term unemployment rates, where long-term unemployment is defined as the share of persons in unemployment for one year or longer in the total number of persons in the labor market. This variable is less sensitive than the traditional unemployment rate to short-run transitory shocks that may be problematic in a cross-section regression. Thus, the final equation to be estimated is:

$$(18) \qquad \ln(w_i) = \alpha_0 + \alpha_1 \ln(\hat{MP_i}) + \alpha_2 Edu_i + \alpha_3 IndShare_i + \alpha_4 Agishare_i + \alpha_5 LTunemp_i + \varepsilon_i$$

As the wage regression includes the calculated market potential variable derived from a previous regression, we cannot rely on OLS standard errors as the stochastic error in the wage equation would include the residuals from the trade equation, violating OLS assumptions. As in Redding and Venables (2004) and Breinlich (2006) we use bootstrap methods with 200 replications to explicitly consider the presence of the calculated market potential variable.

Table 5: Wage equation: baseline results

Dep Var ln(wages)	Regions in EU15	Regions in new member states
	OLS	OLS
	(1)	(2)
ln(MP)	0.062**	0.419***
	(0.022)	(0.073)
	[0.030]	[0.073]
Education	0.900***	0.497
	(0.179)	(0.495)
	[0.202]	[0.531]
Agishare	-3.502***	-3.265***
	(0.625)	(0.563)
	[0.703]	[0.663]
Indshare	-0.147	-0.289
	(0.194)	(0.407)
	[0.221]	[0.415]
LTunemp	-0.001	-0.006**
	(0.001)	(0.003)
	[0.001]	[0.003]
Constant	2.722***	0.058
	(0.206)	(0.658)
	[0.284]	[0.667]
Moran's I	0.197	0.074
P-value	0.00	0.13
R squared	0.53	0.86
# observations	191	53

Note: OLS standard errors in (), boot-strapped standard errors in [].
* p<0.10, ** p<0.05, *** p<0.01

Column 1 of Table 5 displays the OLS regression results of the wage equation for the EU15.[15] The coefficient estimate for total market potential is 0.062 and statistically significant at the 1% level. This result is slightly lower than those found in previous studies; Head and Mayer (2006) find an average coefficient of 0.12 after controlling for human

[15] For comparative reasons with Section 5.4, island regions (Åland, Sicily, …) are not included in the regressions of Table 5 as queen contiguity spatial weight matrices cannot be calculated for island regions with no contiguous neighbors.

capital and Breinlich finds a coefficient of 0.083 after controlling for human and physical capital formation. Our results indicate that on average, doubling market potential for an EU15 region would result in a 6.2% increase in wages. As expected, there is a strong and statistically significant relationship between education and regional wages in the EU15. A 10 percentage point increase in the number of people pursuing tertiary education would result in an almost 9% increase in regional wages, ceteris paribus. Intuitively, a high percentage of agriculture has a negative impact on wages; whereas, wages are not sensitive to differences between industry share and service share, which is our base group. As expected, there is a negative relationship between long term unemployment rates and wages; however, this coefficient is not statistically significant. This OLS regression explains approximately 53% of the cross variation in regional wages.

Column 2 of Table 5 displays the OLS regression results of the wage equation for the NMS regions. As evident from Figure 5, NMS regions' wages are more sensitive to market potential than EU15 regions' wages. Total market potential is statistically significant at the 1% level and our results indicate that doubling market potential for NMS regions would result in an increase in wages of 41.9%. This coefficient is substantially larger than the market potential coefficient for the EU15 regions, implying that wages in NMS regions are approximately 7 times more sensitive to changes in market potential than wages in EU15 regions. The coefficient estimate for education has the expected positive sign, but surprisingly is not statistically significant. This result may in part be due to the relatively small variation of the share of the economically active population that has pursued tertiary education. In addition, a high percentage of production in agriculture results in lower regional wages. As expected, there is a negative and statistically significant relationship between long-term unemployment and wages. The OLS model explains 86% of the cross sectional variation in NMS regions' wages.

Recall from the structural derivation of the wage equation that the coefficient estimate of market potential is equal to $1/(\alpha\sigma)$, where α is the labor input share and σ is the elasticity of substitution between any two varieties or the degree of production differentiation. Thus if we assume that labor is the only input ($\alpha=1$), then from the EU15 regression we can calculate σ = 16.1. This finding for the EU15 regions is relatively large compared to those reported in the previous literature where estimates of σ usually lie between 5 and 10.[16] At σ = 2.4, for new member state regions, the opposite holds true relative to the previously reported literature (which typically excludes those regions).

5.3. *Decomposing market potential*

We break down market potential into several subcomponents to see which regions contribute the most to a given region's market potential and to what extent different subcomponents of total market potential help to explain wages. Rather than calculating the market potential of region i by summing the transportation-cost-weighted market capacities of all regions j, we restrict what constitutes j. We separate total market potential into three subcomponents. First, we calculate domestic market potential (*MP_D*) in which j is restricted to i. *MP_D* can be thought of as market potential defined only by a region's internal demand. In a world without any inter-regional trade, a region's domestic market potential would equal its total market potential (*MP*). The second subcomponent of market potential is spatial lag of order 1 market potential (*MP_L(1)*) in which j is restricted to regions which are contiguous with i. Thus *MP_L(1)* is market potential calculated by only including a region's immediate, surrounding, contiguous neighbors. The final component of market potential is the 'rest of Europe' market potential (*MP_ROE*) in which only regions outside spatial lag 1 are included in the calculation. Mathematically, the measures may be represented by the following equation

(19) $$MP_i = MP\_D_i + MP\_L(1)_i + MP\_ROE_i \ ,$$

where $MP\_D_i = MP_{ii}$, $MP\_L(1)_i = \sum_{j \neq i,\, j \& i \ contiguous}^{R} MP_{ij}$ and $MP\_ROE_i = \sum_{j \neq i,\, j \notin L(1)}^{R} MP_{ij}$.

Table 6 shows the shares of market potential for all regions, EU15 and new EU member state regions, capital and noncapital regions, and a select number of individual regions. We see that 'rest of Europe' market potential appears to dominate total market potential, reflecting the fact that EU regional economies are well integrated. Capital regions represent

[16] See Head and Ries (2001), Head and Mayer (2006)

an exception since their domestic market potential is on average the largest share of their MP, due to large home markets. The share of domestic MP for the new EU member state regions is smaller than for the EU15 regions, most likely due to a smaller home market.

Luxembourg provides a special case when decomposing market potential. As illustrated in Figure 4, Luxembourg has a high total market potential relative to other European regions. This finding suggests that Luxembourg is benefiting from a geographical advantage and is an attractive location for firms due to its relatively high market access. However, Luxembourg's domestic market potential as a share of its total market potential is small relative to other EU15 regions and to capital regions. This suggests that the country's high total market potential is primarily due to its central location and interaction with European regions that are beyond spatial lag 1 and not due to its internal demand or its interaction with its immediate neighboring regions. Thus Luxembourg's relationships with other European regions are crucial in determining its market potential, and thus attractiveness.

The contrast is striking if one compares Luxembourg to Brussels, whose domestic market potential accounts for 75% of its total market potential. Both regions have relatively high overall total market potentials; however, the shares of their respective components are drastically different. Brussels can attribute much of its high total market potential to its internal demand and home market, while Luxembourg can attribute its high market potential to its interactions with other regions. However, one should note that the entire country of Luxembourg is one region; whereas, Brussels and several other capital regions are often restricted to metropolitan areas. Thus Luxembourg's domestic market potential faces a higher distance penalty and a smaller weight simply because it is a larger region in terms of area.[17]

[17] In fact, Brussels represents the smallest region in terms of area at only 161 square km. Luxembourg's area is 2,586 km^2.

Table 6: Decomposing market potential

Market Potential Shares			
	MP_D	MP_L(1)	MP_ROE
All regions	27.9%	25.5%	46.6%
EU15 regions	29.4%	27.2%	43.4%
NMS regions	22.6%	19.3%	58.1%
Capital regions	63.6%	8.1%	28.3%
Non-capital regions	23.9%	27.4%	48.7%
Luxembourg	27.0%	6.5%	66.5%
Brussels	75.4%	15.1%	9.5%

Note: Data and calculations for 2004.

Table 7 examines the extent to which each subcomponent of market potential can explain wages. Looking first at the EU15 results displayed in column 1, our results indicate that on average a doubling of *MP_D* would result in a more than a 5.5% increase in regional wages, ceteris paribus. Surprisingly, neither *MP_L(1)* nor *MP_ROE* has a significant relationship with regional wages in the EU15 according to our results. This may be due to the presence of spatial autocorrelation. The Moran's I statistics for the residuals from the OLS regressions indicate that there is spatial correlation in the residuals and consequently that OLS assumptions are violated. Given the extent of European integration, the presence of spatial correlation is not surprising, as shocks to one region are often transmitted to neighboring regions. We examine this issue further in Section 5.4, where we use a spatial error model to correct for spatial linkages between regions. In column 2, the OLS results indicate that for NMS regions *MP_ROE* has approximately twice the impact on NMS wages as *MP_D*. A doubling of domestic market potential would result in a 19.1% increase in NMS wages; while, a doubling in *MP_ROE* would result in a 40.9% increase in NMS wages.

Table 7: Wage equation: separation of market potential into components

	Regions in	
Dep Var ln(wages)	EU15	new member states
	OLS	OLS
	(1)	(2)
ln(MP_D)	0.056***	0.191***
	(0.017)	(0.033)
	[0.021]	[0.044]
ln(MP_L(1))	-0.005	0.009
	(0.015)	(0.037)
	[0.016]	[0.040]
ln(MP_ROE)	0.023	0.409***
	(0.027)	(0.109)
	[0.029]	[0.117]
Education	0.835***	0.581
	(0.182)	(0.583)
	[0.219]	[0.776]
Agishare	-2.921***	-2.179***
	(0.702)	(0.716)
	[0.716]	[0.732]
Indshare	-0.079	-0.378
	(0.197)	(0.377)
	[0.204]	[0.444]
LTunemp	-0.001*	-0.006**
	(0.0009)	(0.003)
	[.0009]	[0.003]
Constant	2.712***	-0.762
	(0.209)	(0.857)
	[0.231]	[0.979]
Moran's I	0.242	0.137
P-value	0	0.01
R squared	0.52	0.89
# observations	191	53

Note: OLS standard errors in (), bootstrapped standard errors in []. * p<0.10, ** p<0.05, *** p<0.01

5.4. Spatial autocorrelation

One shortcoming of previous applications of Redding and Venables' methodology is that they rarely implement spatial econometric analysis. They assume each region to be an isolated, independent entity in which income is influenced by market potential and several region-specific controls. Tobler's (1970) first law of geography states: "Everything is related to everything else, but near things are more related than distant things." In other words, it seems probable that regions are spatially dependent due to integration, knowledge and technology spillovers, and other factors. Spatial dependence implies that economic variables in one location are correlated with those in neighboring locations.

Spatial dependence can take two forms: i) under spatial error dependence unobserved explanatory variables or shocks are correlated across space and ii) under spatial lag dependence the dependent variable in one location is directly affected by the dependent variable in a neighboring location. The empirical NEG literature usually assumes no correlation in the error terms. However, this assumption is violated if there is spatial dependence and consequently spatial correlation in the residuals. Ignoring spatial dependence and spatial correlation can lead to misspecification and biased results. Therefore, it is important to consider alternative models in our analysis.

We examine Moran's I statistics from the OLS regression presented in Table 5 and Table 7 to see if spatial correlation is present in the error terms. Moran's I is a weighted correlation statistic used to measure deviations from spatial randomness.[19] A significant Moran's I indicates the presence of spatial correlation. If spatial autocorrelation is present, Lagrange Multiplier tests can be used to determine the appropriate econometric model.

We employ the spatial error model in which a spatial autoregressive error term is included as an independent variable in addition to the other explanatory variables.[20] The model can be estimated by maximum likelihood and is represented by the equation $Y = \beta X + \varepsilon$, where $\varepsilon = \lambda W\varepsilon + u_i$. W is a symmetric spatial weight matrix that provides the structure of assumed spatial relationships. W contains a "w_{ij}" term for every combination of regions within the dataset. The two most commonly used weighting matrices are based upon contiguity or distance between regions. Arora and Brown (1977) and Hordijk (1979) suggest using a binary contiguity weighting matrix when dealing with spatial models for error terms. Therefore, we employ a "queen" contiguity weighting matrix in which w_{ij} is equal

---

19 $\text{Moran's I} = \frac{N}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2}$ where N is the number of regions, x is the variable of interest, x-bar is the mean, and w is the spatial weight matrix.

20 Results of the Lagrange Multiplier tests show that a spatial error model is appropriate and thus we do not estimate the spatial lag model.

to one if regions i and j are contiguous and zero otherwise. In this model, u is assumed to be independently and identically distributed; whereas, ε depends on the weighted average of the errors of neighboring regions, with λ measuring the strength of the relationship. This model corrects for spatial correlation by allowing shocks or unobservable variables in one region to spillover onto neighboring regions.

Few NEG studies have implemented spatial econometrics techniques. A notable exception is Niebuhr (2006) who finds spatial autocorrelation in EU regional data and subsequently runs a spatial error model to address the problem. She finds that taking spatial autocorrelation into account does not alter the significance of market potential in explaining income. Hering and Poncet (2007), who also follow the Redding and Venables' methodology, implement a spatial lag model in a study on market access and income in Chinese cities; they find that spatial relationships between Chinese cities matter significantly.

In Table 8, we run Moran univariate analyses to test for the existence of spatial correlation in the dependent variable and each of the explanatory variables, except for market potential and its subcomponents. The market potential variables are not tested because of their innately spatial nature. Clearly, if one region has a high market potential and is close to large markets, then its neighbor will also have a high market potential and be close to large markets. All the variables display positive and statistically significant spatial correlation. The results indicate that regions are likely to be situated close to regions with similar economic compositions, education levels, unemployment rates and wages, and as such may be seen as evidence of European integration.

Table 8: The presence of spatial autocorrelation

	EU15 regions		NMS regions	
	Moran's I	p-value	Moran's I	p-value
Agriculture (share in GVA)	0.5054	0.001	0.5951	0.001
Industry (share in GVA)	0.2786	0.001	0.3094	0.001
Education (share ISCED 5&6)	0.5941	0.001	0.2083	0.014
Long term unemployment rate	0.7974	0.001	0.3488	0.001
ln(wages)	0.3264	0.001	0.4935	0.001

Note: Data and calculations for 2004.

Table 9: Wage equation: addressing spatial autocorrelation

	Regions in		
Dep Var ln(wages)	EU15	EU15	new member states
	Spatial error model		
	(1)	(2)	(3)
ln(MP)	0.105***		
	(0.028)		
ln(MP_D)		0.104***	0.189***
		(0.017)	(0.034)
ln(MP_L(1))		-0.009	-0.002
		(0.014)	(0.035)
ln(MP_ROE)		0.077**	0.517***
		(0.034)	(0.110)
Education	1.130***	1.037***	1.098**
	(0.223)	(0.234)	(0.551)
Agishare	-2.486***	-0.309	-1.506***
	(0.641)	(0.666)	(0.741)
Indshare	-0.206	-0.006	-0.088
	(0.195)	(0.192)	(0.344)
LTunemp	-0.0005	-0.001	-0.004*
	(0.001)	(0.001)	(0.003)
Constant	2.288***	1.887***	-1.684**
	(0.241)	(0.264)	(0.852)
Lambda	0.421***	0.564***	0.445***
	(0.077)	(0.065)	(0.147)
Moran's I	-0.027	-0.035	-0.001
P-value	0.35	0.31	0.56
R squared	0.59	0.65	0.90
# observations	191	191	53

Note: Standard errors in (), bootstrapped standard errors in []. * p<0.10, ** p<0.05, *** p<0.01

Table 9 displays the spatial error model results of the wage equation for the EU15 and NMS regions. First, we note that the spatial error models improve the overall fit compared to the OLS estimations. These now explain 65% and 90% of the wage variation across EU15 and new member state regions. In column 1, the coefficient estimate for total market potential is statistically significant and larger in magnitude than the OLS estimate, indicating that on average, doubling market potential for an EU15 region would result in an increase in wages of 10.5%. Thus if we assume that labor is the only input ($\alpha=1$), then we can theoretically interpret this coefficient estimate as σ = 9.5, which is now much closer to estimate ranges of between 5 and 10 previously reported in the literature. The estimate of the λ coefficient is statistically significant and equal to 0.42. The Moran's I statistic for the residuals is close to zero; it is no longer statistically significant suggesting that inclusion of the spatial lag in the error term has eliminated the problem of spatial correlation in the residuals.

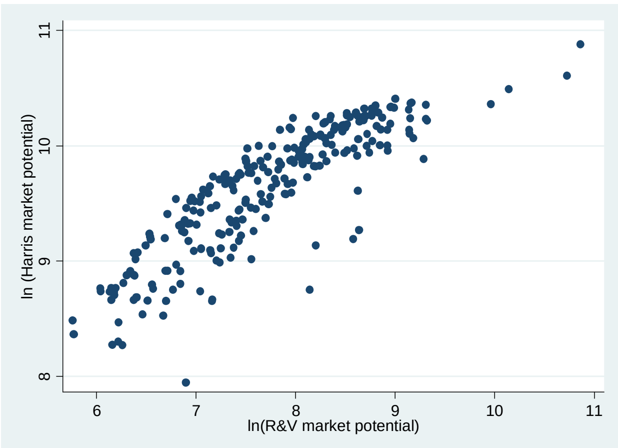
Column 2 and 3 of Table 9 displays the results of the spatial error model when market potential is broken down into its subcomponents for the EU15 and NMS regions respectively. In column 2, the coefficient estimate for *MP_ROE* is now statistically significant, indicating that a doubling of MP_ROE would result in a 7.7% increase in EU15 regional wages. The coefficient estimates for *MP_D* and *MP_ROE* are statistically significant and larger in magnitude than the coefficient estimates from the OLS results in Table 5. This was also the case with total market potential and suggests that not controlling for spatial error correlation may have introduced a downward bias in the coefficient estimates on market potential. This suggests that previous work not considering spatial correlation may have underestimated the impact of market potential on wages in the EU15. The λ coefficient is positive and statistically significant and the Moran's I statistic for the residual indicates that the spatial autocorrelation has been appropriately treated.

In Column 2 of Table 5, the Moran's I statistic for the residuals in the OLS estimation for new member state regions indicated no evidence of spatial autocorrelation. However, the residuals of OLS regression in Column 2 of Table 7 were spatially autocorrelated. This finding suggests that, for NMS regions, the total market potential does a better job capturing spatial effects than its subcomponents. In the spatial error model for the NMS regions, the λ coefficient is positive and statistically significant and the Moran's I statistic for the residual is close to zero and not statistically significant. The coefficient estimate for domestic MP is similar to that from the OLS model and the estimate for *MP_ROE* is larger, indicating that doubling *MP_ROE* would result in a more than 51% increase in regional wages. Also, education is statistically significant at the 5% level, which was not the case in the OLS estimation. The estimates also indicate that a 10 percentage point increase in the number of people pursuing tertiary education would result in a 10% increase in wages in new EU member state regions; a result which is similar to that found for the EU15.

5.5. Alternative estimations and endogeneity

By calculating market potential based on the trade equation, which uses national level data, we assumed that interregional trade flows are subject to the same determinants as international trade flows and that the price index in each country is the same across regions. In an effort to test the robustness of this measure of market potential, we also calculate an ad-hoc measure using Harris' (1954) formula given in equation (1). Figure 6 shows a scatter plot of our previously calculated measure of market potential following the Redding and Venables (2004) methodology and Harris' measure of market potential. The estimates using Harris' measure are reported in Table 10 and remarkably similar to those reported in Table 5, Table 7 and Table 9.

Figure 6: Market potential according to Redding & Venables (2004) and Harris (1954)



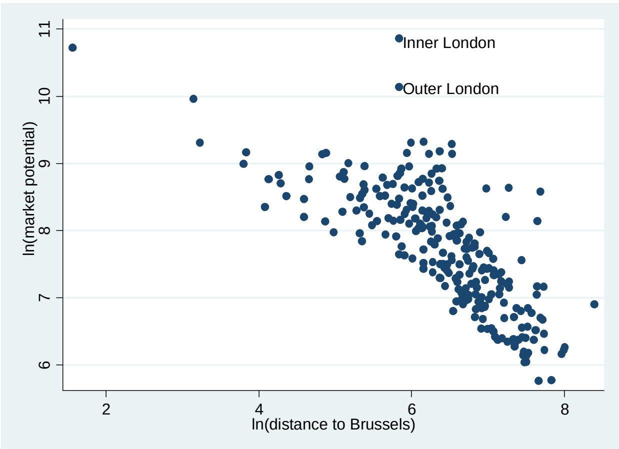
Note: Data and calculations for 2004.

One potential problem in our wage equation is endogeneity. Market potential is a transportation-cost-weighted sum of regional purchasing power. Purchasing power depends on wages and thus there is possible reverse causality. For example, an increase in wages in region i will (ceteris paribus) cause purchasing power in region i to increase and consequently the market potential of region i to increase.

One solution to this endogeneity problem is to exclude domestic market potential from total market potential, i.e. calculate the market potential of region i summing over all regions j not equal to i; we call this measure foreign market potential (*MP_F*). However, as noted by Knaap (2006) and Head and Mayer (2006), this solution is not perfect. As previously discussed, capital regions are often dominated by *MP_D*. Therefore, when we exclude *MP_D* and calculate *MP_F* there are counter-intuitive inversions for the peripheral regions surrounding capital and other economically large regions. For example, the region of Outer London has a higher *MP_F* than the region of Inner London.

Another solution to the endogeneity problem would be to use an instrumental variable approach. An ideal instrument would be correlated with market potential, but not correlated with the error term in the wage equation. Previous studies have found that geographic variables are a promising choice. Redding and Venables (2004) use distance from the nearest major economic center (New York, Brussels or Tokyo); Head and Mayer (2006) use distance from Brussels, and Breinlich (2006) uses distance from Luxembourg as instruments. We follow this approach and take the distance to Brussels, representing the approximate economic center of the EU, as an instrument for total market potential. Figure 7 shows a scatter plot of market potential and distance to Brussels. There is a clear relationship between the variables, suggesting that distance to Brussels should be a good instrument.

Figure 7: Market potential and distance to Brussels



Note: Data and calculations for 2004.

Table 10: Alternative wage equation estimations and robustness checks

Dep Var ln(wages)	Regions in EU15				Regions in new member states			
	OLS	OLS	OLS	IV	OLS	OLS	OLS	IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ln(MP)	0.061***			0.099***	0.419***			0.587***
	(0.019)			(0.032)	(0.073)			(0.155)
	[0.029]				[0.074]			
ln(MP_Harris)		0.115***				0.637***		
		(0.0303)				(0.126)		
ln(MP_F)			0.046***				0.309***	
			(0.016)				(0.088)	
			[0.014]				[0.095]	
Education	0.939***	1.040***	1.027***	0.837***	0.497	1.424**	1.042	0.776
	(0.174)	(0.212)	(0.169)	(0.186)	(0.488)	(0.620)	(0.696)	(0.566)
	[0.191]		[0.212]		[0.531]		[0.690]	
Agishare	-3.481***	-3.140***	-3.743***	-2.809***	-3.265***	-2.735***	-3.596***	-2.237
	(0.594)	(0.554)	(0.560)	(0.738)	(0.578)	(0.688)	(0.853)	(1.061)
	[0.722]		[0.533]		[0.663]		[0.854]	
Indshare	-0.145	-0.227	-0.276	-0.094	-0.29	-0.253	-0.404	-0.213
	(0.173)	(0.218)	(0.173)	(0.176)	(0.385)	(0.436)	(0.480)	(0.409)
	[0.204]		[0.234]		[0.415]		[0.485]	
LTunemp	-0.001	-0.001	-0.001	-0.001	-0.006**	-0.006**	-0.007**	-0.004
	(0.0008)	(0.0009)	(0.0008)	(0.0008)	(0.003)	(0.0028)	(0.003)	(0.003)
	[0.0009]		[0.001]		[0.003]		[0.003]	
Constant	2.715***	2.078***	2.905***	2.409***	0.058	-3.242**	1.021	-1.352
	(0.175)	(0.305)	(0.129)	(0.266)	(0.664)	(1.318)	(0.790)	(1.326)
	[0.250]		[0.123]		[0.667]		[0.822]	
CY & MT Dummies					Yes	Yes	Yes	Yes
R squared	0.56	0.56	0.54	0.54	0.87	0.86	0.83	0.86
# observations	199	199	199	199	55	55	55	55

Note: Standard errors in (), bootstrapped standard errors in []. * p<0.10, ** p<0.05, *** p<0.01

Column 1 of Table 10 shows the results of the OLS wage equation estimation for all EU15 regions, including 8 island regions. The results are similar to those in column 1 of Table 5. Column 2 displays the results of the wage equation when market potential is calculated according to the Harris measure of market potential. The coefficient estimate for the Harris measure is larger in magnitude. In column 3, total market potential is replaced by foreign market potential in the wage equation. As expected, the coefficient estimate is smaller in magnitude, but still statistically significant at the 1% level. Also, the R^2 is slightly lower, suggesting that only including foreign market potential as opposed to total market potential results in a loss of explanatory power. Finally, the instrumental variable regression results are displayed in column 4. The coefficient estimates are similar in terms of magnitude and significance to those in the column 1.

For comparison, column 5 of Table 10 displays the results of the wage equation for all NMS regions. Column 6 displays the results when market potential is calculated according to Harris' formula. The coefficient estimate for the Harris measure of market potential is still significant and positive, but larger in magnitude than our theoretically based market potential. *MP_F*, shown in column 7, is statistically significant and positively related to wages. Again, as foreign market potential contains less information than total market potential, the magnitude of the coefficient and the R^2 are smaller than in previous regressions. The instrumental variable regression results shown in column 8 indicate that on average doubling market potential in a NMS region would result in more than a 50% increase in wages, ceteris paribus.

5.6. Estimates based on actual travel time

As a final robustness check, we use Schürmann and Talaat's (2002) dataset of distances between regions to calculate market potential. In this dataset, interregional distances are represented by estimated road travel times between region capitals. While slightly outdated, the dataset has the advantage of taking into account road types and their quality, speed limits, border delays and other important considerations into the travel time estimations; therefore, they provide estimates based on a better reflection of geographic reality. However, as Bosker and Garretsen (2007) note, travel time is likely to be influenced by infrastructure quality, which in turn is related to income. This may actually exacerbate the previously discussed endogeneity problem. When market potential is calculated using travel times instead of great circle distances, we don't find significant changes in our measurement of market potential (see Figure 4). The regression results for EU15 and NMS regions are displayed in Table 11.

Table 11: Wage equation: estimates based on actual travel time

Dep Var (lnwages)	Regions in						
	EU15				new member states		
	OLS	OLS	Spatial error	Spatial error	OLS	OLS	Spatial error
	(1)	(2)	(3)		(4)	(5)	
ln(MP)	0.096***		0.149***		0.375***		
	(0.022)		(0.026)		(0.071)		
	[0.030]				[0.073]		
ln(MP_D)		0.055**		0.097***		0.168***	0.169***
		(0.016)		(0.016)		(0.031)	(0.032)
		[0.020]				[0.038]	
ln(MP_L(1))		-0.012		-0.011		-0.0202	-0.032
		(0.016)		(0.015)		(0.033)	(0.030)
		[0.017]				[0.034]	
ln(MP_ROE)		0.061**		0.108***		0.397***	0.464***
		(0.026)		(0.033)		(0.084)	(0.086)
		[0.029]				[0.076]	
Education	0.868***	0.830***	1.108***	1.064***	0.478	0.537	0.947*
	(0.174)	(0.179)	(0.220)	(0.228)	(0.547)	(0.559	(0.533)
	[0.193]	[0.210]			[0.530]	[0.718]	
Agishare	-2.720***	-2.407***	-1.557**	-0.189	-2.914***	-2.163***	-1.610***
	(0.643)	(0.706)	(0.649)	(0.707)	(0.746)	(0.714)	(0.739)
	[0.696]	[0.736]			[0.684]	[0.640]	
Indshare	-0.056	-0.039	-0.083	0.002	-0.246	-0.381	-0.128
	(0.189)	(0.197)	(0.190)	(0.192)	(0.428)	(0.375)	(0.347)
	[0.207]	[0.199]			[0.438]	[0.446]	
LTunemp	-0.002*	-0.002*	-0.0006	-0.001	-0.007**	-0.006**	-0.005*
	(0.0008)	(0.0009)	(0.001)	(0.001)	(0.003)	(0.003)	(0.003)
	[0.009]	[.0009]			[0.003]	[0.002]	
Constant	2.376***	2.417***	1.805***	1.587***	0.127	-0.816	-1.485**
	(0.216)	(0.222)	(0.251)	(0.275)	(0.703)	(0.753)	(0.753)
	[0.296]	[0.254]			[0.729]	[0.715]	
Lambda			0.466***	0.554***			0.397***
			(0.074)	(0.066)			(0.154)
Moran's I	0.213	0.248	-0.028	-0.033	0.049	0.119	0.001
P-value	0.00	0.00	0.35	0.32	0.22	0.03	0.34
R squared	0.557	0.556	0.63	0.656	0.840	0.877	0.902
# observations	191	191	191	191	53	53	53

Note: Standard errors in (), bootstrapped standard errors in []. * p<0.10, ** p<0.05, *** p<0.01

There is a positive and significant relationship between market potential and wages for both EU15 regions and NMS regions. The magnitude of the coefficient estimates for market potential are similar in size to those found when great circle distances were used for the distance measure. The coefficient estimates for the other independent variables are also similar in magnitude and significance to those reported in earlier regressions.

There is a noticeable difference in the trade equation results when travel times are used instead of great circle distances. The trade equation results for great circle distances are displayed in the appendix. Unlike the results displayed in Table 4, the border coefficient estimate is not statistically significant. However, this is not surprising because border effects (delays) are already incorporated in travel times. Thus distance is drawing some explanatory power away from border.

5.7. *A thought experiment: The effect of abolishing international trade*

We also calculate country market potential using Harris' simple formula in an effort to determine the gains from international trade. As opposed to calculating the market potential of region i by summing the transportation-cost-weighted market capacities of all regions j, we restrict j to only those regions in the same country as i. Thus, country market potential can be thought of as the market potential that a region would have if all national borders were closed tomorrow.[23]

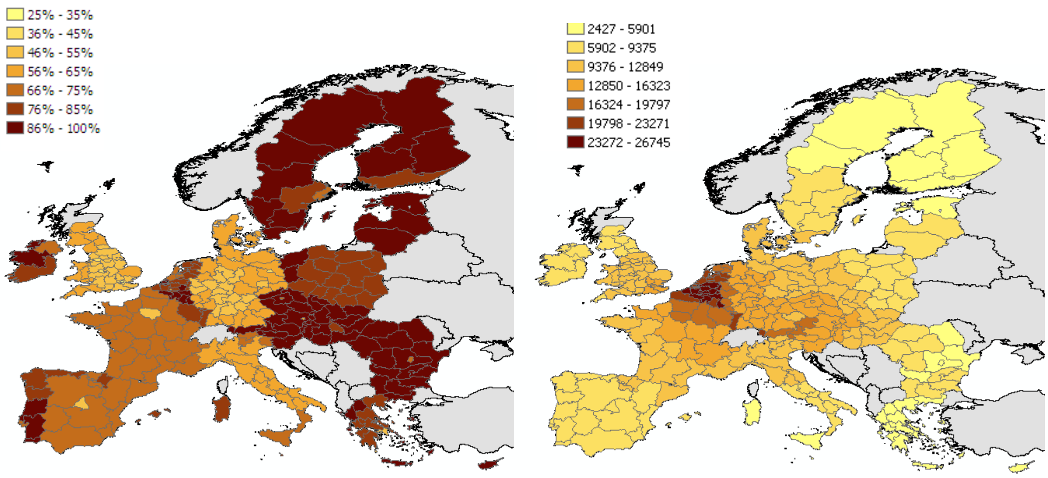
Luxembourg stands out as one of the core regions that benefited the most from European integration and international trade. Its total market potential is high; however, its country market potential is low due to its small size. Figure 8 shows the reduction in market potential in percentage terms (left) and absolute terms (right) from moving to autarky. All regions benefited from trade as their market potential is necessarily increased. In absolute terms, the core regions of Luxembourg, Belgium and the Netherlands see the largest decrease in market potential when trade is confined within national boundaries. This is due to their proximity to large, high income economies such as Germany. However, in relative terms, Luxembourg and the peripheral regions (Finland, Sweden, Latvia, Estonia, Romania, Bulgaria, Portugal, etc..) lose the most from shutting down international trade with a de-

[23] This is similar to the idea of *MP_D*, which is the market potential a region would have in world without international or interregional trade.

crease in market potential of between 85% and 100%. As shown in the empirical analysis, the fall in market potential directly translates into lower wages.

Figure 8: Reduction in market potential when removing international trade

Percent decrease *Absolute decrease*



Note: Calculations based on Harris methodology, data for 2004.

6. Conclusion

Regional wages in Europe are characterized by substantial differences. There are plenty of reasons why this may be the case, including different factor endowments, being located in different countries, different infrastructures, etc… However, one strikingly simple explanation, which has received substantial empirical support, is the fact that regions differ in their market potential, which in turn is by and large determined by the regions' geographical position in Europe. Previous empirical results have shown a strikingly robust relationship between regions' market potential and their wages.

This paper adds to this literature and extends previous studies in several ways. 1) It analyses the link between market potential and wages for the EU27, 2) corrects for spatial autocorrelation present in the data, 3) decomposes total market potential into several geographical components and analyses to which extent each of them contributes to explaining the geographical wage structure, and 4) analyses which regions gain most from European integration, by calculating market potential under the counterfactual assumption that regions only trade with other regions within the same country.

The results not only corroborate previous findings, but also show that by neglecting spatial autocorrelation the strength of the relationship between market potential and wages may be underestimated. We decompose market potential into several components and show that a region's domestic market potential, this is the market potential of a region due to domestic demand for its own goods, is a decisive factor for the region's wage, while adjacent regions do not have any significant impact. However, the market access to all other EU regions is of a similar importance as domestic demand. This is particularly true for regions in the EU new member states. Fourth, we analyze which regions have gained most from European trade by calculating counterfactual market potential for the hypothetical case in which regions could only trade with other regions within the same country. This exercise shows that regions in small central countries, i.e. regions with very good market access, such as Luxembourg, Belgian and Dutch regions benefitted significantly from EU trade integration both in absolute and relative terms. Regions in the outmost EU periphery, i.e. regions with very low market access, also benefit in relative terms but not in absolute terms, owing to the fact that they rank bottom in absolute terms already.

More empirical work on economic geography in Europe is needed to fully understand the relationship between geography and economic well-being. Future research could try to assess how changes in market potential over time have affected the regional wage structure in Europe. The Redding and Venables (2004) two-stage framework is well suited for analyzing the effects of increased trade integration on market potential and regional wage structure.

7. References

Anselin, L. (1988) *Spatial econometrics: methods and models*. Dordrecht: Kluwer Academic Publishers.

Arora, S. and M. Brown (1977). Alternative approaches to spatial autocorrelation: An improvement over current practices. *International Regional Science Review, 2*: 67–78.

Brakman, S., H. Garretsen and M. Schramm (2004). The spatial distribution of wages: estimating the Helpman-Hanson model for Germany. *Journal of Regional Science, 44*: 437–466.

Bosker, M and H. Garretsen (2007). Trade costs, market access and economic geography: why the empirical specification of trade costs matters, CESIfo Working Paper, No. 2071.

Breinlich, H. (2006). The spatial income structure in the European Union – What role for economic geography? *Journal of Economic Geography, 6*: 593–617.

Combes, P.-P., M. Lafourcade and T. Mayer (2005). The trade creating effects of business and social networks: evidence from France. *Journal of International Economics, 66*: 1–29

Combes, P.-P. and H.G. Overman (2004). The spatial distribution of economic activities in the European Union, In J.V. Henderson and J.F. Thisse (Eds.), *Handbook of Regional and Urban Economics 4*: 2845–909. Amsterdam: Elsevier-North-Holland.

European Commission (2006). *Employment in Europe 2006*. Luxembourg: Office for Official Publications of the European Communities.

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

Harris, C. (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. and J. Ries (2001). Increasing returns versus national product differentiation as an explanation for the pattern of U.S.-Canada trade. *American Economic Review*, *91*: 858–875.

Head, K. and T. Mayer (2004a). The empirics of agglomeration and trade, In J.V. Henderson and J.F. Thisse (Eds.), *Handbook of Regional and Urban Economics 4*: 2609–669. Amsterdam: Elsevier-North-Holland.

Head, K. and T. Mayer (2004b). Market potential and the location of Japanese investment in the European Union. *Review of Economics and Statistics*, *86*: 959–972.

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

Helpman E. (1998). The size of regions, in D. Pines, E. Sadka and I. Zilcha (eds). *Topics in Public Economics*, Cambridge: Cambridge University Press.

Hering, L. and S. Poncet (2007). Economic geography, spatial dependence and income inequality in China, CEPII working paper, No. 2007–22.

Hordijk, L. (1979). Problems in estimating econometric relations in space. *Papers, Regional Science Association*, *42*: 99–115.

Knaap, T. (2006). Trade, location and wages in the United States. *Regional Science and Urban Economics*, *36*: 595–612.

Krugman, P. (1991). Increasing returns and economic geography. *Journal of Political Economy, 99*: 483–499.

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

Niebuhr, A. (2006). Market access and regional disparities: New economic geography in Europe. *Annals of Regional Science*, *40*: 313–334.

OECD (2005). *Employment Outlook*. Paris: OECD.

Overman, H., S. Redding and A.J. Venables (2003). The economic geography of trade, production, and income: A survey of empirics, in J. Harrigan and E. Kwan Choi (eds). *Handbook of International Trade, 1*: 353-87. Oxford: Blackwell.

Puga, D. (2002) European regional policy in light of recent location theories. *Journal of Economic Geography, 2*: 372–406.

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

Roos, M. (2001). Wages and market potential in Germany. *Jahrbuch für Regionalwissenchaft*, *21*: 171–195.

Samuelson, P. (1954). The transfer problem and transport costs, II: Analysis of effects of trade impediments. *The Economic Journal, 64*: 264–289.

Schürmann, C. and A. Talaat (2002). The European Peripherality Index, Paper presented at the 42nd Congress of the European Regional Science Association, Dortmund, Germany, 27–31 August 2002.

Santos Silva, J.M.C. and S. Tenreyro (2006). The log of gravity. *Review of Economics and Statistics*, *88*(4): 641–658.

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography*, *46*(2): 234–240.

Vandenbrande, T., L. Coppin, P. van der Hallen, P. Ester, D. Fourage, A. Fasang, S. Geerdes and K. Schomann (2006). *Mobility in Europe: Analysis of the 2005 Eurobarometer survey on geographical and labour market mobility*. European Foundation for the Improvement of Living and Working Conditions. Luxembourg: Office for Official Publications of the European Communities.

Appendix:

A1: Variable Definitions

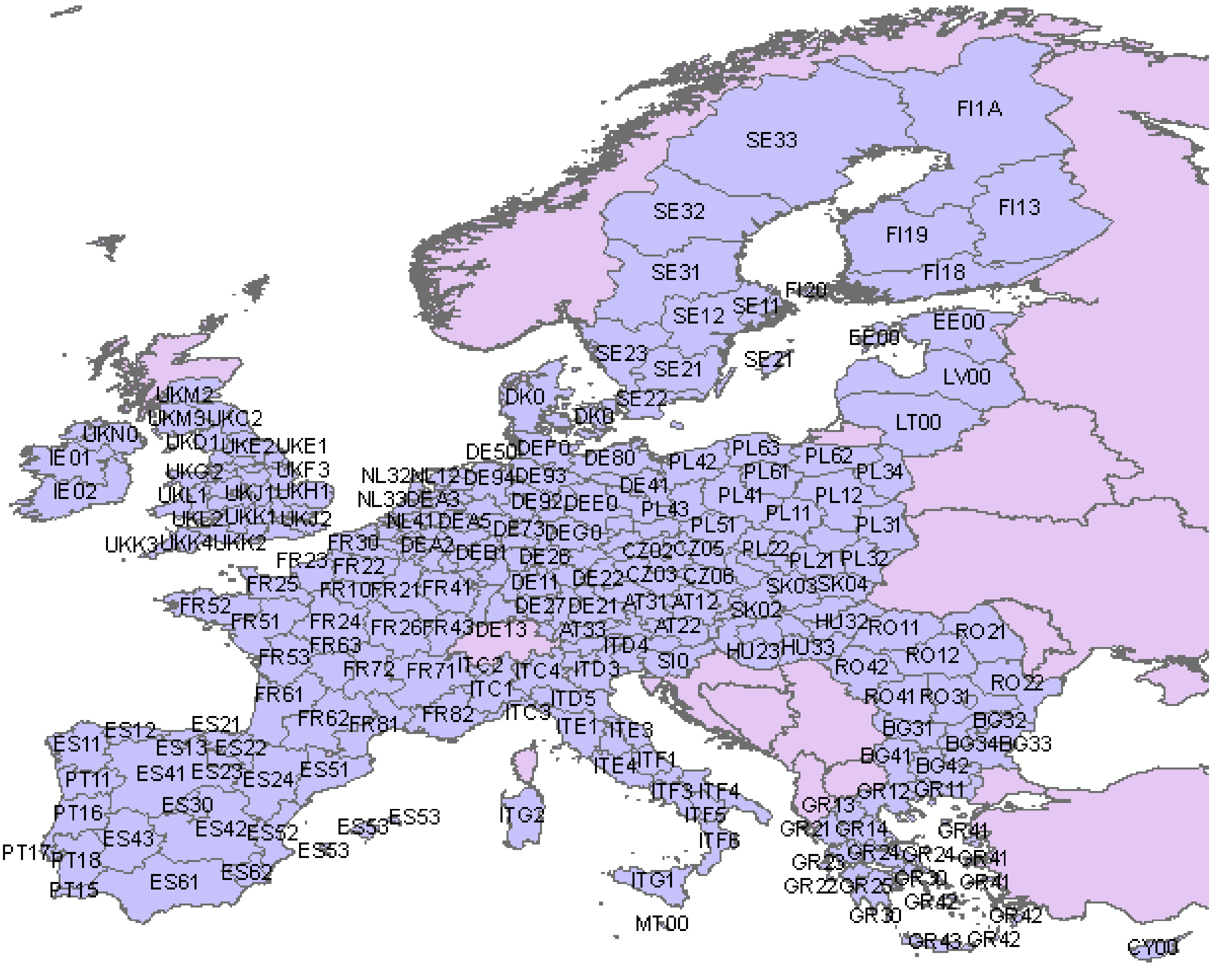
Trade Equation	
Dependent Variable	
ln(*exports*)	log of the value of exports from origin region to destination country in euro
Independent Variables	
ln(*distance*)	log of the distance in kilometers from the origin region to the destination region
border	dummy variable: 1 if origin and destination are not in the same country, 0 otherwise
adjacency	dummy variable: 1 if origin and destination are contiguous, 0 otherwise
language	dummy variable: 1 if origin and destination share an official language, 0 otherwise
Wage Equation	
Dependent Variable	
ln(*wages*)	log of annual compensation per employee in thousand euro
Independent Variables	
ln(*MP*)	log of total market potential where market potential is calculated according to Redding and Venables (2004) methodology
ln(*MP_D*)	log of domestic market potential where market potential is calculated according to Redding and Venables (2004) methodology
ln(*MP_L(1)*)	log of L(1) market potential where market potential is calculated according to Redding and Venables (2004) methodology
ln(*MP_ROE*)	log of rest of Europe market potential where market potential is calculated according to Redding and Venables (2004) methodology
ln(*MP_Harris*)	log of total market potential where market potential is calculated using Harris' (1954) formula
ln(*MP_F*)	log of foreign market potential where market potential is calculated according to Redding and Venables (2004) methodology
Education	the share of the economically active population that has pursued tertiary education (ISCED levels 5 & 6)
Agishare	share of agriculture in regional GDP
Indshare	share of industry in regional GDP
LTunemp	regional long term unemployment rate
CY&MT	Correction dummy variables for Cyprus and Malta which are outliers

Note: All data refer to 2004 if not otherwise reported.

A2: List of Nuts2 regions included in the dataset:

at11 Burgenland (A), at12 Niederösterreich, at13 Wien, at21 Kärnten, at22 Steiermark, at31 Oberösterreich, at32 Salzburg, at33 Tirol, at34 Vorarlberg be10 Région de Bruxelles, be21 Prov. Antwerpen, be22 Prov. Limburg (B), be23 Prov. Oost-Vlaanderen, be24 Prov. Vlaams Brabant, be25 Prov. West-Vlaanderen, be31 Prov. Brabant Wallon, be32 Prov. Hainaut, be33 Prov. Liège, be34 Prov. Luxembourg (B), be35 Prov. Namur, bg31 Severozapaden, bg32 Severen tsentralen, bg33 Severoiztochen, bg34 Yugoiztochen, bg41 Yugozapaden, bg42 Yuzhen tsentralen, cy00 Cyprus, cz01 Praha, cz02 Strední Cechy, cz03 Jihozápad, cz04 Severozápad, cz05 Severovýchod, cz06 Jihovýchod, cz07 Strední Morava, cz08 Moravskoslezsko, de11 Stuttgart, de12 Karlsruhe, de13 Freiburg, de14 Tübingen, de21 Oberbayern, de22 Niederbayern, de23 Oberpfalz, de24 Oberfranken, de25 Mittelfranken, de26 Unterfranken, de27 Schwaben, de30 Berlin, de41 Brandenburg – Nordost, de42 Brandenburg – Südwest, de50 Bremen, de60 Hamburg, de71 Darmstadt, de72 Gießen, de73 Kassel, de80 Mecklenburg-Vorpommern, de91 Braunschweig, de92 Hannover, de93 Lüneburg, de94 Weser-Ems, dea1 Düsseldorf, dea2 Köln, dea3 Münster, dea4 Detmold, dea5 Arnsberg, deb1 Koblenz, deb2 Trier, deb3 Rheinhessen-Pfalz, dec0 Saarland, ded1 Chemnitz, ded2 Dresden, ded3 Leipzig, dee0 Sachsen-Anhalt, def0 Schleswig-Holstein, deg0 Thüringen, DK0 Denmark, ee00 Estonia, es11 Galicia, es12 Principado de Asturias, es13 Cantabria, es21 Pais Vasco, es22 Comunidad Foral de Navarra, es23 La Rioja, es24 Aragón, es30 Comunidad de Madrid, es41 Castilla y León, es42 Castilla-la Mancha, es43 Extremadura, es51 Cataluña, es52 Comunidad Valenciana, es53 Illes Balears, es61 Andalucia, es62 Región de Murcia, fi13 Itä-Suomi, fi18 Etelä-Suomi, fi19 Länsi-Suomi, fi1a Pohjois-Suomi, fi20 Åland, fr10 Île de France, fr21 Champagne-Ardenne, fr22 Picardie, fr23 Haute-Normandie, fr24 Centre, fr25 Basse-Normandie, fr26 Bourgogne, fr30 Nord - Pas-de-Calais, fr41 Lorraine, fr42 Alsace, fr43 Franche-Comté, fr51 Pays de la Loire, fr52 Bretagne, fr53 Poitou-Charentes, fr61 Aquitaine, fr62 Midi-Pyrénées, fr63 Limousin, fr71 Rhône-Alpes, fr72 Auvergne, fr81 Languedoc-Roussillon, fr82 Provence-Alpes-Côte d'Azur, gr11 Anatoliki Makedonia, Thraki, gr12 Kentriki Makedonia, gr13 Dytiki Makedonia, gr14 Thessalia, gr21 Ipeiros, gr22 Ionia Nisia, gr23 Dytiki Ellada, gr24 Sterea Ellada, gr25 Peloponnisos, gr30 Attiki, gr41 Voreio Aigaio, gr42 Notio Aigaio, gr43 Kriti, hu10 Közép-Magyarország, hu21 Közép-Dunántúl, hu22 Nyugat-Dunántúl, hu23 Dél-Dunántúl, hu31 Észak-Magyarország, hu32 Észak-Alföld, hu33 Dél-Alföld, ie01 Border, Midlands and Western, ie02 Southern and Eastern, itc1 Piemonte, itc2 Valle d'Aosta/Vallée d'Aoste, itc3 Liguria, itc4 Lombardia, itd1 Provincia Autonoma Bolzano-Bozen, itd2 Provincia Autonoma Trento, itd3 Veneto, itd4 Friuli-Venezia Giulia, itd5 Emilia-Romagna, ite1 Toscana, ite2 Umbria, ite3 Marche, ite4 Lazio, itf1 Abruzzo, itf2 Molise, itf3 Campania, itf4 Puglia, itf5 Basilicata, itf6 Calabria, itg1 Sicilia, itg2 Sardegna, lt00 Lithuania, lu00 Luxembourg (Grand-Duché), lv00 Latvia, mt00 Malta, nl11 Groningen, nl12 Friesland (NL), nl13 Drenthe, nl21 Overijssel, nl22 Gelderland, nl23 Flevoland, nl31 Utrecht, nl32 Noord-Holland, nl33 Zuid-Holland, nl34 Zeeland, nl41 Noord-Brabant, nl42 Limburg (NL), pl11 Lódzkie, pl12 Mazowieckie, pl21 Malopolskie, pl22 Slaskie, pl31 Lubelskie, pl32 Podkarpackie, pl33 Swietokrzyskie, pl34 Podlaskie, pl41 Wielkopolskie, pl42 Zachodniopomorskie, pl43 Lubuskie, pl51 Dolnoslaskie, pl52 Opolskie, pl61 Kujawsko-Pomorskie, pl62 Warminsko-Mazurskie, pl63 Pomorskie, pt11 Norte, pt15 Algarve, pt16 Centro (PT), pt17 Lisboa, pt18 Alentejo, ro11 Nord-Vest, ro12 Centru, ro21 Nord-Est, ro22 Sud-Est, ro31 Sud – Muntenia, ro32 Bucuresti – Ilfov, ro41 Sud-Vest Oltenia, ro42 Vest, se11 Stockholm, se12 Östra Mellansverige, se21 Småland med öarna, se22 Sydsverige, se23 Västsverige, se31 Norra Mellansverige, se32 Mellersta Norrland, se33 Övre Norrland, SI0 Slovenija, sk01 Bratislavský kraj, sk02 Západné Slovensko, sk03 Stredné Slovensko, sk04 Východné Slovensko, ukc1 Tees Valley and Durham, ukc2 Northumberland, Tyne and Wear, ukd1 Cumbria, ukd2 Cheshire, ukd3 Greater Manchester, ukd4 Lancashire, ukd5 Merseyside, uke1 East Yorkshire and Northern Lincolnshire, uke2 North Yorkshire, uke3 South Yorkshire, uke4 West Yorkshire, ukf1 Derbyshire and Nottinghamshire, ukf2 Leicestershire, Rutland and Northants, ukf3 Lincolnshire ukg1 Herefordshire, Worcestershire and Warks, ukg2 Shropshire and Staffordshire, ukg3 West Midlands, ukh1 East Anglia, ukh2 Bedfordshire, Hertfordshire, ukh3 Essex, uki1 Inner London, uki2 Outer London, ukj1 Berkshire, Bucks and Oxfordshire, ukj2 Surrey, East and West Sussex, ukj3 Hampshire and Isle of Wight, ukj4 Kent, ukk1 Gloucestershire,Wiltshire and Bristol/Bath, ukk2 Dorset and Somerset, ukk3 Cornwall and Isles of Scilly, ukk4 Devon, ukl1 West Wales and The Valleys, ukl2 East Wales, ukm2 Eastern Scotland, ukm3 South Western Scotland , ukn0 Northern Ireland

A3: Map of Nuts2 Regions included in dataset



A4: Additional results: Gravity equation based on actual travel times

Trade Equation Estimation Dep. Var ln(exports)	OLS	Neg. Binomial
ln(distance)	-1.602***	-1.596***
	(0.098)	(0.079)
border	0.529	0.043
	(0.473)	(0.267)
Language	0.298	0.259*
	(0.214)	(0.154)
Adjacency	0.184	0.310**
	(0.138)	(0.123)
Importer fixed effects	Yes	Yes
Exporter fixed effects	Yes	Yes
Observations	729	729
R-squared	0.919	

Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.01
Note: Based on the Schürmann and Talaat (2002) dataset.

BANQUE CENTRALE DU LUXEMBOURG

EUROSYSTÈME

2, boulevard Royal
L-2983 Luxembourg

Tél.: +352 4774-1
Fax: +352 4774 910

www.bcl.lu • info@bcl.lu