Trade Costs and Income in European Regions: Evidence from a regional bilateral trade dataset

Hammer, Christoph; Fichet de Clairfontaine, Aurélien

DOI:
10.57938/eef3671c-7785-4d1c-8462-9c210eee4451

Published: 01/01/2016

Document Version:
Publisher's PDF, also known as Version of record

Trade Costs and Income in European Regions: Evidence from a regional bilateral trade dataset *

Christoph Hammer[1] and Aurélien Fichet de Clairfontaine[2]

[1]Institute for International Economics and Development, E-mail: chhammer@wu.ac.at.
[2]Institute for International Economics , E-mail: afichet@wu.ac.at.

Vienna University of Economics and Business (WU)
October 2016

Abstract

Using a New Economic Geography (NEG) model, this study estimates the relationship between regional per capita income and the market accessibility of regions. This accessibility cannot be observed directly, so it has to be constructed. We follow a two-step-procedure as suggested by Redding and Venables (2004) and use results of a gravity-type model to infer "real market potential". To this end, we make use of a novel dataset of bi-regional trade flows between (and within) 254 European NUTS-2 regions (for 26 European countries excluding Bulgaria and Romania) for the year 2010. In a second step we test the hypothesis that access to domestic as well as to large foreign markets increases factor incomes. We find evidence that supports this hypothesis on a regional level. This also holds when we control for other potential income determinants. In order for the estimates to be unbiased, we additionally take the spatial structure of the data into account. Our findings indicate that, although the specification derived from theory should be able to capture some spatial spillovers, additionally controlling for spatial autocorrelation in the residuals is necessary to fit the European data.

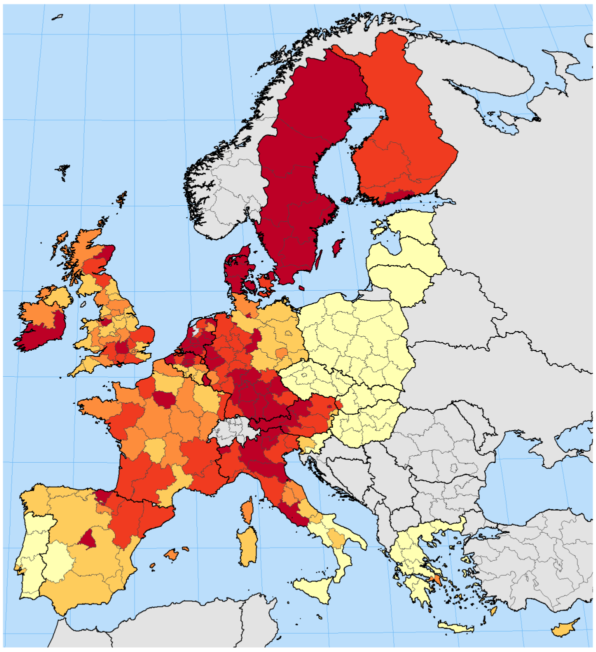
Keywords Wage equation · Trade Equation · European regions · New Economic Geography
JEL Classification F12 · F14 · R12

1 Introduction

Economic activity is not evenly distributed across space. This is true for comparisons between cities, regions as well as countries. One implication of that are varying living standards across space, which can be seen for the case of European regions in figure 1. It shows gross value added (GVA) per capita, as a measure of income, in 2010 for 254 NUTS-2 regions. Darker shades correspond to higher distribution quantiles of the regional GVA per capita.[1] The figure shows what is often said in spatial economics: "space matters". High-income regions are often neighbors to other high-income regions and vice versa[2]. There seems to be a high-income geographic core (including Nordic countries and in parts of the United Kingdom (UK)) and a lower-income geographic periphery, especially looking at the new member states and most Mediterranean regions. These differences in income and living standards are of interest for residents as well as politicians. Lagging regions within countries as well as on the European scale are often the target of regional policy that promotes income convergence. While there is a vast literature on possible explanations for differences in income (-growth) we want to focus on a specific factor, namely the relative location of firms and their access

*Corresponding author: Christoph Hammer (chhammer@wu.ac.at)
[1]`http://epp.eurostat.ec.europa.eu/portal/page/portal/nuts_nomenclature/introduction`
[2]Figures for gross domestic product (GDP) per capita or wages per capita look very similar.

Figure 1: Gross value added per capita, Eurostat



to markets. Ever since the seminal work of Krugman (1991), models of "New Economic Geography" (NEG) have attempted to shed some light on the uneven economic distribution using a general-equilibrium framework with a focus on geography represented by transport- or trade costs. They aim to explain agglomeration patterns by allowing mobile production factors to move across space to regions with the highest rewards. NEG models assume that goods, as well as some factors, are mobile, unlike growth theory (goods and factors immobile) and trade theory (factors immobile). While historic and institutional factors or the physical geography of a region – what Krugman labeled "first-nature geography" – can help explain agglomeration, NEG models focus on forces that reflect behavior of optimizing (mobile) economic agents.
One central element of NEG models that is derived from optimizing agents is the so-called "wage equation"[3]. It states that the "maximum wage that each firm in a specific region can afford to pay is a function of trade-cost-weighted market- and supply capacities" (Redding and Venables, 2004, p.58). This means, NEG models imply a spatial structure in which remunerations of factors are higher in regions that have better access to markets. We will use this implication for our purpose to see if access to markets plays a role in shaping the geographic income structure of European regions.
The concept of increased factor remuneration in areas with good access to demand, was first introduced by Harris (1954), who used geographic distance to weight income from distant regions. The NEG wage equation has some potential advantages over the basic concept of Harris. It is derived from microeconomic theory, it considers competition among firms in export markets and trade costs are usually defined more broadly than as a function of distance only.[4]

The wage equation seems to be a very robust empirical relationship with Redding and Venables (2004) as one of the pioneering contributions.[5] They first derive a structural trade equation to estimate bilateral trade flows using a gravity model. Market access (also sometimes called "real market potential") is then constructed with fitted values of the trade

[3]The name arises from the fact that labor is the only production input in the first NEG model by Krugman (1991).
[4]Theory that is not based on a perfectly competitive goods market.
[5]See Bosker and Garretsen, 2010 for a meta-study on the wage equation.

costs in order to estimate the wage equation. We adopt this approach but in contrast to earlier studies, our analysis is based on a novel cross-regional trade dataset for 254 European NUTS-2 regions in 2010, which was provided by the JRC, European Commission.[6] One the one hand, this permits to infer trade costs rather than defining trade costs as mere inverted distance. Trade costs represent every form of friction in shipping goods, services, people or ideas over space. In the most general case one can think of the sum of transport costs, information and time costs, institutional and cultural barriers (such as tariffs, product standards and language).[7] On the second hand, having a focus on European regions lets us control for regional determinants of income (e.g. Breinlich (2006)).

We show that regions' access to markets is a significant determinant of regional income. When disaggregated into a regionally domestic part and a foreign part (trade-cost-weighted sum of foreign demand), the positive relationship with income of the latter is robust to the inclusion of control variables. However, some more recent contributions mention the role of unobserved spatially correlated factors (Barde, 2010, Kosfeld and Eckey, 2010 or Bruna et al., 2015). Ignoring possible spatial spillovers (apart from the mechanisms modeled in NEG) will potentially bias the results. Despite the striking similarity between the wage-type equation and a spatial autoregression, our results reveal the presence of unexplained spatial autocorrelation in the residuals. This issue is addressed by adding a spatial error component to the model, which does not significantly affect the parameter estimates.

The paper is structured as follows: in section 2 we derive the main equations of interest from a theoretical NEG model. Section 3 describes the estimation procedure or how we get from theory to empirics. Section 4 provides details on the datasets, while estimation results are presented and analyzed in section 5. Finally, section 6 concludes.

2 Theoretical Model

There is a big variety of NEG models. They differ in their number of regions, their use of functional forms for utility and costs or production, as well as their assumptions about the mobility of production factors, and therefore the exact mechanisms of agglomeration and dispersion across space. Common to all models is the endogenous market size for the relevant sector. The way in which market size is determined differs across models (migration across regions, migration across sectors, (human) capital accumulation). However, Robert-Nicoud (2005), Ottaviano and Robert-Nicoud (2006) and Ottaviano (2007) show that all two-region-models based on a CES utility function share the same equilibrium properties, they are, according to Robert-Nicoud (2005), "identical twins". While there are differences in the assumptions and in the specific dynamic behavior of the model, the "short-run" equilibrium properties are nearly identical.[8] We follow the notation of Bruna (2015b), who derives a more general NEG model in which most of the mentioned models above are nested, and report only the main equations.[9] For a full derivation of the general model see Bruna (2015b), for a more comprehensive overview over the most common two-region models see Fujita et al. (2001) and Baldwin et al. (2003).

The model world consists of many symmetrical regions $j = 1, \dots, R$, and two sectors of production. One sector produces a homogeneous good under constant returns to scale. The good of this sector usually acts as numeraire and is traded costlessly across regions. The second sector produces a differentiated good under increasing returns. This sector is often labeled as producing "manufacturing goods". Consumers face a budget constraint and

[6] Details on the construction of the trade matrix can be found in Thissen et al. (2013a), Thissen et al. (2013b) and Alvarez-Martinez and Lopez-Lobo (2016)

[7] See for example Fujita et al. (2001), Fujita and Thisse (2002), Fingleton and McCann (2007), Combes et al. (2008) or Head and Mayer (2015) for a more detailed discussion.

[8] A short-run equilibrium satisfies all equilibrium conditions except the migration condition. This condition depends on the specific model and describes why and how the mobile factor moves between regions. When the migration condition(s) are fulfilled, the model is in the "long-run" or spatial equilibrium.

[9] Similar to most contributions to the literature, his model is not fully specified in a NEG-general-equilibrium sense. Migration dynamics are not modeled explicitly since that is not the aim of these papers. The model is specified to be able to derive a "wage-type" equation that is based on goods- and labor-market equilibria, not necessarily on a spatial 2-region-equilibrium.

want to consume both types of goods. Their preferences also display a love for variety of manufactured good. Constrained optimization leads to the demand for a variety g of the manufactured good produced in region j with σ elasticity:

$$x_{gj} = p_{gj}^{-\sigma}\mu E_j P_j^{\sigma-1} \tag{1}$$

$$P_j = \left[\sum_i^N n_i p_i^{1-\sigma}\right]^{1/(1-\sigma)} \tag{2}$$

where x_{gj} denotes the demand of variety g produced in region j and μE_j the expenditure on manufacturing goods in region j. p_i is the mill price which is the same for each good, while n_i is the number of firms or varieties produced in region i, since each manufacturing firm produces one variety. P_j represents an aggregate price index of manufacturing goods and is sometimes called "supplier access" or "competition index". It shows that competition in the export market j affects demand for goods produced in region i.
Goods that are shipped from i to j are assumed to face iceberg trade costs τ_{ij}.[10] Summing over all possible export markets and considering trade costs, the demand for all goods produced in region i becomes

$$x_i \equiv \sum_j^R x_{ij} = p_i^{-\sigma} \underbrace{\sum_j^R \mu E_j \tau_{ij}^{1-\sigma} P_j^{\sigma-1}}_{RMP} \tag{3}$$

where we dropped the variety index. Equation (3) shows that demand is a function of prices and "market access" or "real market potential" (dubbed RMP henceforth). RMP is a trade-cost-weighted sum of market capacities. It differs from the older ad-hoc measure by Harris ("nominal market potential", 1954) in that it considers the competition P_j as well as a potentially more general trade cost function. In order to connect the measure for RMP to income we have to also consider the supply side of the model.
So far, the majority of the NEG models derives similar if not identical demand functions. Features that distinguish most models are found on the supply side. The assumed production function together with the mobility assumptions about the production factors differs between models. Again, for our purposes we follow Redding and Venables (2004), Breinlich (2006) and Bruna (2015b) who propose rather general production functions. The cost of producing x_i takes the form $q_i c_i(x_i + f)$ where q_i denotes the price index of a compound input, c_i is the marginal input requirement (technology) and f represents fixed costs. Defining $q_i c_i \equiv m_i$, firms maximize profits π_i

$$\pi_i = p_i x_i - m_i(x_i + f) \tag{4}$$

Firms in this model take the price of competitors as given, which yields prices that are a mark-up over marginal costs. Inserting $p_i = m_i \frac{\sigma}{\sigma-1}$ and quantities from equation (3) into equation (4) gives

$$\pi_i = \frac{(\sigma-1)^{\sigma-1}}{\sigma^\sigma} m_i^{\sigma-1} \sum_j^R \mu E_j \tau_{ij}^{1-\sigma} P_j^{\sigma-1} - f m_i \tag{5}$$

If one assumes free entry and exit of firms, profits in the long-run are zero, which determines the break-even output $\bar{x}$. Combining the $\bar{x}$ with the demand for manufactured goods in i, gives the mill price that has to be satisfied in order for firms to produce

$$p_i^\sigma = \frac{1}{\bar{x}} \sum_j^R \mu E_j \tau_{ij}^{1-\sigma} P_j^{\sigma-1} \tag{6}$$

Instead of prices, we can use the mark-up rule to write the market-clearing condition in terms of costs, which Redding and Venables (2004) call the "wage-equation", while Bruna

[10]For a detailed discussion on trade cost formulations see Fingleton and McCann (2007).

(2015b) labels it the "generalized wage-type equation".

$$m_i = \frac{\sigma - 1}{\sigma} \left(\frac{1}{\bar{x}} \sum_{j}^{R} \mu E_j \tau_{ij}^{1-\sigma} P_j^{\sigma-1} \right)^{\frac{1}{\sigma}} \tag{7}$$

Dependent on the production factors included, the left hand side consists of an index of factor remunerations. The first theoretical models only assumed labor as production factor, hence the name wage equation. Equation (7) is the basis for answering our question set out in the beginning. It connects the access to markets via trade costs (a function of the relative position in the trade network) to a measure of regional income. Given that firms make zero profits in the long-run, better access to markets lets them pay higher remunerations to their production factors. A region's access to markets increases i) the lower trade costs to their export markets are, ii) the higher the expenditure on manufacturing goods in those markets and, iii) the less competition is present at those markets. As Head and Mayer (2006) point out, a positive demand shock could lead to higher wages (prices) as can be seen in equation (7). Alternatively or additionally, if factor prizes are assumed to equalize, the price index would have to decrease, which is equivalent of an increase in the number of firms and therefore a rise in the supply of necessary production factors. Head and Mayer (2006) find evidence against the latter adjustment for Europe and so in line with Redding and Venables (2004) we assume the response to run through price adjustments only.

3 Estimation Strategy

This section provides an overview of the proposed empirical implementation. Firstly, we derive the estimation specification of the wage equation. Secondly, we show how to construct RMP with variables that are estimated with a trade equation.
As pointed out in Bruna et al. (2015) and Bruna (2015b), the left-hand side of the wage equation is best represented by marginal costs. Which costs should be included? The literature so far usually specified the wage equation in way such that the left-hand side consists of immobile or non-traded factors of production. Redding and Venables (2004) estimate a version of equation (7). Their dependent variable is a Cobb-Douglas composite of costs $m_i = w_i^{\beta} v_i^{\gamma} c_i$.[11] Wages w are earned by labor, which they assume to be immobile, while v_i denotes the remuneration of some mobile production factor. Again, c_i is the marginal input requirement or "technological difference". Breinlich (2006) estimates equation (7) where the technological differences are captured by the error term.
The remuneration of every regionally mobile factor might equalize in the long run. If one thinks that cross-sectional regressions capture long-run relationships, then the long-run must be considered. In this spirit, we interpret the left-hand-side as every factor that is immobile, which is proxied by GVA per employee. On the right-hand-side we have RMP, as well as a variable that captures the technological difference or productivity. We will additionally control for factors that might influence income differences across space.
Taking the natural log of the resulting equation and adding a disturbance term yields

$$\ln(GVA_i/population_i) = \beta_0 + \beta_1 \ln(RMP_i) + \mathbf{X}'\gamma + \varepsilon_i \tag{8}$$

where $\varepsilon_i \sim N(0, \sigma^2)$. RMP is not directly observable, so we have to construct it.
Since we use a novel regional trade dataset, we follow Redding and Venables (2004) who first estimate a trade equation of the gravity-type in order to get estimates for trade costs between regions.
One appealing feature of NEG models with a CES sub-utility function for manufacturing goods is the derivation of a structural gravity equation within the model. The value of total exports from one region to another can be expressed as

$$n_i p_i x_{ij} = n_i p_i^{1-\sigma} \mu E_j P_j^{\sigma-1} \tau_{ij}^{1-\sigma} \tag{9}$$

[11] We suppress the index for the intermediate good.

This states that exports from i to j are a function of supply capacity $sc = n_i p_i^{1-\sigma}$, market capacity $mc = E_j P_j^{\sigma-1}$, as well as bilateral trade costs $\tau_{ij}^{1-\sigma}$. The right-hand-side of equation (9) shows the similarity to the definition of RMP(see equation (3)). We take advantage of this similarity in order to construct RMP with the results of the estimated trade equation. We define the total value of goods exported from region i to j as exports $n_i p_i x_{ij} = \tilde{x}_{ij}$ and take the natural logarithm to arrive at a specification that can be estimated

$$\ln(\tilde{x}_{ij}) = c + sc + mc + (1-\sigma)\ln(\tau_{ij}) + \epsilon_{ij} \tag{10}$$

where c is a constant term controlling for the mean measurement error in the $\tilde{x}_{ij}$'s and $\epsilon_{ij} \sim N(0, \sigma^2)$.[12]

Supply and market capacity measures (i.e., sc and mc) are not observable and are therefore proxied by exporter ξ_i and importer ξ_j fixed effects respectively. $\tau_{ij}^{1-\sigma}$ is approximated by a distance deterrence function D_{ij} construed to include spatial, institutional and cultural separation factors that are defined in a further section. This highlights one of the differences to earlier approaches, where RMP was constructed by weighting GDP by geographical distance instead of a broader measure for trade costs that result from a gravity equation. We insert both sets of fixed effect and the deterrence function D_{ij} in equation (10)

$$\ln(\tilde{x}_{ij}) \quad = \quad \kappa + \xi_i + \xi_j + \ln D_{ij} + \epsilon_{ij} \tag{11}$$

Equation (11) is estimated by Pseudo-Poisson Maximum Likelihood (PPML) as advocated by Silva and Tenreyro (2006) who provide simulation evidence that this estimator is well behaved even when the conditional variance is far from being proportional to the conditional mean.[13]. Origin- and destination spatial dependences induce a bias in the parameters of the deterrence function as well as in the fixed effects. However, the Moran test for spatial autocorrelation in the residuals obtained with PPML shows that these dependencies are significant but negligible (equal to 0.01 on average throughout different specifications).[14]

Estimating equation (11) yields empirical counterparts for $\mu E_j P_j^{1-\sigma}$ as well as $\tau_{ij}^{1-\sigma}$, which are needed to construct market capacity $\mu E_j P_j^{1\sigma} \tau_{ij}^{1-\sigma}$ of a region j. Following Redding and Venables (2004), estimated RMP is obtained with

$$\widehat{RMP}_i = \hat{\xi}_j^i \hat{D}_{ii} + \sum_{j \neq i} \hat{\xi}_j^j \hat{D}_{ij} \tag{12}$$

where $\hat{\xi}_j^j$ is estimated importer fixed effects (subscript j) of region j (superscript j) and $\hat{D}_{ij}$ the estimated deterrence function. The first term of the right-hand side of (12) corresponds to the domestic RMP (dubbed $DRMP$ henceforth), the second term to the foreign RMP (dubbed $FRMP$ henceforth). This again shows that RMP is a trade-cost-weighted sum of demand from all potential markets. Equation (8) is rewritten by including both terms of equation (12) which results in

$$\ln(GVA_i/population_i) = \beta_0 + \beta_1 \ln(\hat{\xi}_j^i \hat{D}_{ii}) + \beta_2 \ln(\sum_{j \neq i} \hat{\xi}_j^j \hat{D}_{ij}) + \mathbf{X}'\gamma + \varepsilon_i \tag{13}$$

where β_1 and β_2 are the coefficients of both $DRMP$ and $FRMP$ respectively. This equation can be estimated by OLS, but a simultaneity bias arising in the construction of the RMP causes this specification to be endogenous. A component of the market potential, i.e., the $DRMP$ defined by $\hat{\xi}_j^i D_{ii}$, contains a proxy for expenditure while the dependent variable is income. Addressing this issue, $DRMP$ is instrumented in a two stages least squares ($2SLS$) by regional components uncorrelated with GVA per capita.

[12] c captures the constant.

[13] In addition, the presence of zeros in the trade data does not affect the performance of the estimator, as it would the standard OLS approach.

[14] In contrast with this result, the estimation of the trade equation with OLS often yield significantly high Moran's I (0.34 on average), therefore requiring the use of spatial filters as a surrogate to obtain unbiased and efficient parameter estimates.

4 Data

Our two-stage analysis relies on two datasets. The first dataset is a novel regional bilateral trade dataset for the estimation of the trade equation and the construction of our variable of interest: RMP. The second is a regional cross-section dataset for the estimation of the wage-type equation.

Our regional trade data covers bilateral trade in goods and services for six broad NACE rev.2 sectors among 254 NUTS-2 European regions from 26 European countries for the year 2010.[15] This data set is tailored for the use of the spatial CGE RHOMOLO model, which is developed by the European Commission, Joint Research Center in Seville, who also made the trade matrix available to us (see López-Cobo M. 2016, forthcoming, which is based on Alvarez-Martinez and Lopez-Lobo, 2016 as well as Thissen et al., 2013a,b).[16]
In essence, national trade flows are broken down with regional data on consumption, investment and production to generate regional make and use tables. These tables are conform national accounts according to the WIOD database (Timmer et al, 2015). The resulting data base is consistent with a series of macro constraints as well as internally consistent (exports from a region i to a region j are also imports of a region j coming from a region i). Importantly, the construction of the interregional and international trade data does not rely on the gravity approach and does not impose any geographical structure on the trade data (Thissen et al. 2013a). Our gravity regression does consequently not just recover the geographical patterns from which the trade data were constructed. Our dataset for the trade equation thus comprises $254^2 = 64,516$ observations of intra- and interregional trade flows among European regions, after we exclude Bulgaria and Romania.[17]
The variables are defined as follows: geographical distance between European regions is measured in terms of great circle distance between region's geographical centers. Additionally, population-weighted geographical distance as defined in Head and Mayer (2002) was used as an alternative in order to account for the fact that often, the economic activity does not take place in the geographical center of a region.[18]. A third proxy for geographical distance is travel time (or commuting time), obtained using the Google Maps API, computing the average travel time by car between two regions' economic centers and therefore incorporating differences in infrastructure.[19] The measurement of intra-regional distance follows the area-based approach of Head and Mayer (2002) for great circle distance and population-weighted distance. In turn, measurement of intra-regional travel time is density-based as introduced by Baradaran and Ramjerdi (2001).
The estimation of the wage-type equation relies on the second dataset, which contains the previously constructed RMP, its disaggregated components $DRMP$ and $FRMP$, data for control variables as well as for instruments of $DRMP$ for 254 NUTS-2 regions. Our dependent variable is GVA per capita (Eurostat). The set of controls includes: 1) share of tertiary education in the population as a proxy of human capital (ShEd, Eurostat), 2) unemployment rate (Unemp., Eurostat), 3) an index of product market regulation (PMR, OECD) index (that tries to capture potential productivity-enhancing effects of less regulated markets, see Conway et al., 2005), 4) net replacement rate (NRR, Eurostat). Additionally, a dummy controls for the presence of a capital city as advocated by Crespo Cuaresma and Feldkircher (2013). Table 6 in the Appendix gives a comprehensive description of both datasets.

5 Results

In a first step we estimate the trade equation (11). 254^2 regional bilateral trade flows are regressed on two vectors of 254 importer and 254 exporter dummies and on up to five separa-

[15] The complete list of NUTS-2 European regions used in this study is provided in the Appendix.

[16] For more details on the RHOMOLO model see EC JRC REMO (2016)

[17] Trade flows data for both Romania and Bulgaria were obtained using gravity equations, their regions are therefore filtered out in order to avoid replication. The full detailed list of NUTS-2 regions is available in the Appendix.

[18] It is written as the sum of the shares of population of the NUTS-3 composing the origin NUTS-2 times the sum of the shares of population of the NUTS-3 composing the destination NUTS-2

[19] `https://developers.google.com/maps/`

Table 1: PPML estimation of the trade equation (11)

	i) GCDDIST		ii) POPWGCD		iii) TRAVELTIME	
	Est.	S.E.	Est.	S.E.	Est.	S.E.
Dependent variable: Gross regional bilateral exports						
Geo. Dis.	−0.994	0.005	−0.896	0.005	−0.652	0.003
Border	−0.857	0.023	−0.995	0.032	−1.363	0.029
Contig.	0.244	0.016	0.333	0.015	0.513	0.014
Capital	0.556	0.086	0.547	0.064	0.277	0.060
Lang Sim.	0.447	0.029	0.406	0.027	0.322	0.025
Observations	64,516		64,516		63,506	
pseudo R^2	0.983		0.987		0.989	
BIC	2605913		2382155		2065665	
Log.Lik.	–1302444		-1190565		-1032323	

Notes All models include 254 exporter- and 254 importer fixed effects. Model specification: i) logged great circle distance in kilometers, ii) logged population-weighted great circle distance in map unit, iii) travel time in seconds. Geo. Dis.: geographical distance. Border: 1 if separated by a country border. Contiguity: 1 if share a common border. Capital: 1 if a capital is located in one of the two regions. Language Sim.: 1 if common spoken languages. Robust standard errors are reported. pseudo R^2: McFadden R^2.

tion factors composing the deterrence function (i.e., geographical distance, country border, country contiguity, language barrier and capital city). In order to stay as close as possible to the theoretical model, only trade flows on NACE 1.1 sector BCDEF (manufacturing, industry and construction) are used.[20]

Table 1 reports the estimated coefficients from equation (10) (excluding the 508 fixed effects). Results for three different specifications of the deterrence function are shown, using three distinct measures of geographical distance: great circle distance (measured in kilometers), population-weighted great circle distance (measured in kilometers) and travel time (measured in seconds).[21]

With the exception of geographical distance, all spatial (pairwise) separation factors are defined as binary variables that take the value of one or zero. Country border (Border) is equal to one if two regions i and j are separated by a country border. Country contiguity (Contig.) takes on a value of one if regions i and j are located in contiguous countries, zero otherwise. The variable Capital captures whether a capital is located in one of the two regions. And finally, language similarity (Lang. Sim.) indicates whether regions' languages are similar (therefore equal to one). With regard to the latter, regional languages are taken into consideration.

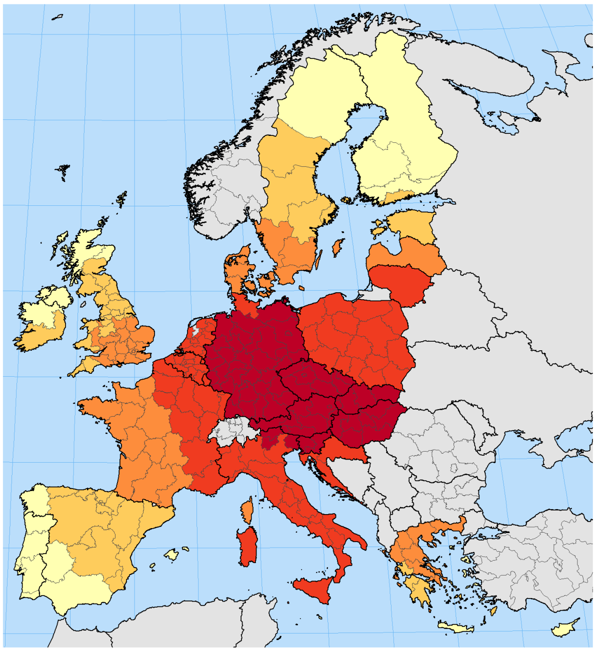
Table (1) lists estimates obtained with three different proxies for distance respectively: great circle distance (in kilometers), population-weighted great circle distance (in kilometers) and travel time (in seconds). While the significance of the results only slightly changes between the three specifications, magnitudes of the coefficients differ. This potentially has an impact for the following calculations as discussed in Bosker and Garretsen (2010).

Geographical distance, proxied by great circle distance, is a larger burden to trade than when proxied by population-weighted great circle distance and travel time (−0.994 compared to −0.896 and −0.652). The opposite is true for the other controls: country border is a stronger impediment to trade in the specification with travel time than for the two others (it is 1.6 times the border effect of the first specification and about 1.4 times the one in the second

[20]The trade data set is divided into five aggregate sectors, one of which is equivalent to BCDEF.

[21]Further specifications, that are not reported in this version of the study, also included the same proxies measured in map unit instead of kilometers and controlled for home bias (whether regions preferences for intraregional trade impact the overall results). Although some parameter estimates differed slightly, they are not significantly different from those reported here.

Figure 2: Freeness of trade ($\hat{D}$) for Vienna (AT13)



specification). The contiguity parameter is positive significant but does not compensate for the negative border effect, stating, that on average, intra-national trade is preferred to international trade. This is also the case where trading countries are close or even contiguous. There is, on average, more trade with regions that contain the capital city of a country (0.551 for models i) and ii) and relatively smaller for model iii)). The same is true for trade between regions that share the same language (0.391 on average). Overall, the trade equation seems to fit the data well with a reported average pseudo-R^2 of 0.986.
The choice of the right set of estimates for the construction of the RMP is based on the model-fit and on the Schwartz information criterion (reported in the table). As a result, we base all calculations on geographical distance proxied by travel time (i.e., model iii)).
The five separation parameters build the deterrence function D_{ij}, which Baldwin et al. (2003) calls "freeness of trade", defined on a scale from zero (trade is fully impeded) to one (free trade).[22]
In order to get some intuition, Fig. (5) shows the resulting estimated "freeness of trade" $\hat{D}_{ij} \equiv \hat{\tau}_{ij}^{1-\sigma}$ for the NUTS-2 region of Vienna, Austria (AT13) based on the estimates from model iii). Darker shades indicate a higher freeness of trade, or equivalently, lower trade costs with the respective partner region. From the Viennese perspective, trade with the Czech Republic, Slovakia, Hungary, part of Northern Italy and Germany is seen as freer than with other parts of Europe due to contiguity, proximity, possible language similarities and the size of the market.
According to equation (12), the vector of coefficients (254×1) is exponentiated and multiplied with the estimated deterrence matrix $\hat{D}_{ij}$ (254×254) in order to construct the RMP. The variable is further disaggregated into $DRMP$ (using the diagonal of the deterrence matrix only) and $FRMP$.[23] Table (2) lists the top five regions with the highest RMP, $DRMP$ and $FRMP$. It clearly shows that $DRMP$ represents the largest part of the RMP since four out of five regions found themselves in both lists (also, the average share of domestic market potential approximates 0.62). High $FRMP$, in turn, seems to be found mostly in central European regions (all five regions are found in Benelux or Germany).

[22] It is possible that some values for the freeness of trade exceed one, due to small intra-regional area (and therefore small geographical distance).

[23] The scale of RMP depends on how we treat insignificant fixed effects. We opt to include regions with insignificant partner fixed effects by setting their values equal to zero (thus equal to one when exponentiated).

Table 2: Top 5 regions for RMP, domestic RMP and foreign RMP

RMP		$DRMP$		$FRMP$	
CODE	NAME	CODE	NAME	CODE	NAME
ITC4	Lombardia	ITC4	Lombardia	BE10	Brussels
DE11	Stuttgart	DE11	Stuttgart	DEA3	Munster
FR71	Rhone-Alpes	ES51	Cataluna	DE30	Berlin
ITH3	Veneto	FR71	Rhone-Alpes	DEA5	Arnsberg
DEA1	Dusseldorf	ITH3	Veneto	DEA2	Köln

Notes RMP: Real market potential, $DRMP$ domestic Real market potential, $FRMP$ foreign Real market potential

The wage-type equation (13) is estimated using $2SLS$. $DRMP$ is instrumented by logged regional population, which is validated as instrument by the Weak Instrument F-test throughout the specifications.[24] Table (3) shows the results of our estimations. In the first three columns we test the theoretically derived wage-type equation without including controls other than a set of country dummies. Domestic market potential is positive and significant (column i)) as long as we do not include any other control. In subsequent specifications we add foreign market potential (column ii)), as well as the logged regional share of population in higher education, the logged regional unemployment rate, a dummy for the capital city, an index for product market regulation, as well as the national net replacement rate that is payed out after 13 months of unemployment. With the addition of country dummies (compare columns iv) and v)), the net replacement rate become more important while the regulation index turns insignificant.

Our estimate for foreign market potential $FRMP$ in the full model (column v)) is in a similar range as those in the literature. However, the effect of domestic market potential on regional income is not significantly different from zero. This is also a consequence of controlling for the presence of a capital city in the considered region. $DRMP$ is especially important for high-income regions that are often the region that contain the capital city of a country. When we drop the capital city dummy, domestic potential gains some explanatory power again.

An aspect that is overlooked by the early papers in this literature, is the fact that estimates of the wage-type equation could be biased due to spatial correlation in the residuals. We include two Moran statistics, that test for remaining unexplained spatial correlation with a Queen contiguity and a $k-5$ nearest neighbor spatial weight matrix. The tests for both assumed neighborhood structures show that there is in fact correlation left in the residuals. This points to a weakness of the NEG wage-type equation, which does not successfully capture all spatial correlation across European regions, although RMP is constructed similar to a spatial lag of market capacity. Indeed, equation (13) resembles a spatial autoregressive model (SAR). The deterrence function D is very similar to a trade-weighted geographical distance spatial weight matrix. Furthermore, the regional partner fixed effects ξ_j also obtained from the trade equation are proxies for market capacity mc (see equation (10)). Market capacity is in turn composed by regional expenditure on manufactured goods μE_j as well as the price index $P^{1-\sigma}$. If we assume that regional income approximates expenditure, the fixed effects could represent a proxy for GVA, which is (divided by population) our dependent variable.[25] Thus RMP (the sum of both $DRMP$ and $FRMP$) could be interpreted as the endogenous spatial component of a SAR wage-type equation.

However, the dependent variable of equation (13) is GVA per capita, since it is supposed to proxy an index for regional immobile factor remuneration. The importer fixed effects ξ_j are only weakly linked to GVA per capita (correlation coefficient of about 0.200) and can therefore not be utilized as its substitute. We conclude that our equation cannot interpreted

[24]Moreover, the correlation coefficient of region population and our dependent variable GVA per capita could not proven different from zero by a Pearson Correlation test. It is, however, highly correlated with $DRMP$: 0.663.

[25]The correlation coefficient for GVA and partner fixed effects equals 0.706.

Table 3: 2SLS estimation of the wage equation

	i)	ii)	iii)	vi)	v)
Dependent variable: GVA per capita					
$\ln DRMP$	0.116∗∗		0.051	−0.025	0.013
	(0.052)		(0.047)	[0.060]	(0.036)
$\ln FRMP$		0.676∗∗∗	0.625∗∗∗	0.018	0.144∗
		(0.132)	(0.140)	[0.057]	(0.087)
ShEd				0.318∗∗∗	0.408∗∗∗
				(0.072)	[0.090]
Unemp.				−0.352∗∗∗	−0.268∗∗∗
				[0.053]	(0.043)
NRR				0.597∗∗∗	1.050∗∗∗
				[0.086]	(0.165)
PMR				−0.410∗∗∗	−0.415∗
				[0.110]	(0.218)
Capital				0.356∗∗∗	0.303∗∗∗
				[0.089]	(0.066)
Wu Hausman (p-value)	0.459	-	0.499	0.001∗∗∗	0.027∗∗
Weak Instrument (p-value)	0.000∗∗∗	-	0.000∗∗∗	0.000∗∗∗	0.000∗∗
Moran test ($k-5$ nn, p-value)	0.582∗∗∗	0.103∗∗∗	0.110∗∗∗	0.106∗∗∗	0.106∗∗∗
Moran test (Q, p-value)	0.545∗∗	0.081∗∗	0.080∗∗	0.079∗∗∗	0.006
Observations	250	250	250	232	232
R-squared	0.798	0.825	0.830	0.576	0.901
Country FE	YES	YES	YES	NO	YES

All models include a constant. All models expect iv) include country fixed effects. Wu-Hausman test performs a test of endogeneity of $DRMP$. Weak Instrument tests for the relevance of the instruments used for $DRMP$. Moran test ($k-5$ nn: $k-5$ nearest neighbors, Q: Queen contiguity) tests for the presence of spatial autocorrelation int the residuals. ShEd: regional share of tertiary education, Unemp.: regional unemployment rate, NRR: country net replacement rate, PMR: country product market regulation index, Capital: the region contains a capital city. $DRMP$ is instrumented by logged regional population. Bootstrapped standard errors are reported for the models iv) (square brackets) and White-robust standard errors for the models i), ii), iii) and v). *** p<0.01, ** p<0.05, * p<0.10.

as a SAR.[26]

Although the hypothesis of homoskedastic residual variance could not be rejected by the Pagan-Hall test (Breusch-Pagan test for generalized least squares, the test rejects the null hypothesis only in the case where country fixed effects are not included), we address the issue of remaining unexplained spatial autocorrelation by additionally modeling a spatial error term (i.e., spatial error model SEM).

For this, we augment equation (13) with a endogenous spatial lag structure in the residuals ε:

$$\ln(GVA_i/population_i) = \beta_0 + \beta_1 \ln(\hat{\xi}_j^i \hat{D}_{ii}) + \beta_2 \ln(\sum_j \hat{\xi}_j^j \hat{D}_{ij}) + \mathbf{X}'\gamma + \varepsilon_i \quad (14)$$

$$\varepsilon_i = \lambda \mathbf{W} \varepsilon_i + v_i \quad (15)$$

where $v \sim N(0, \sigma^2)$. The choice of the spatial weight matrix for the endogenous residual spatial lag is based on the Moran's I's reported for the instrument variable regressions of equation (13). The most appropriate seems to be a $k-5$ nearest neighbors spatial weight matrix (W_{k5}). It is worth noting that the variable $DRMP$ is still instrumented by logged

[26]If we assume that regional importer fixed effects ξ_j are exogenous, an alternative interpretation of our equation would be of a spatial lag exogenous (SLX) wage-type equation.

Table 4: SEM estimation of the wage equation (W_{k5})

	i)	ii)	iii)	vi)	v)
Dependent variable: GVA per capita					
$\ln DRMP$	0.116***		0.041	−0.019	0.004
	(0.042)		(0.040)	(0.034)	(0.033)
$\ln FRMP$		0.803***	0.760***	0.050	0.206**
		(0.093)	(0.103)	(0.096)	(0.087)
ShEd.				0.462***	0.416***
				(0.068)	(0.068)
Unemp.				−0.160***	−0.261***
				(0.047)	(0.037)
NRR				0.358***	1.037***
				(0.108)	(0.176)
PMR				−0.499***	−0.370
				(0.139)	(0.226)
Capital				0.345***	0.284***
				(0.051)	(0.051)
λ	0.323***	0.307***	0.313***	0.812***	0.027
	(0.087)	(0.088)	(0.088)	(0.036)	(0.068)
Wu-Hausman (IV, p-value)	0.457	-	0.243	0.000***	0.000***
Hausman test (p-value)	0.030**	0.120	0.135	0.335	0.003
BP test (p-value)	0.628	0.756	0.802	0.000***	0.050*
LR test (p-value)	0.001***	0.002***	0.002***	0.000***	0.741
Observations	249	249	249	244	244
R-squared (Nagelerke)	0.792	0.835	0.836	0.801	0.900
Country FE	YES	YES	YES	NO	YES

All models include a constant. All models expect iv) include country fixed effects. λ is the residual autocorrelation parameter. Wu-Hausman test (IV) performs a test of endogeneity of $DRMP$. Hausman test (res. autoc.) tests for remaining spatial autocorrelation in the residuals. BP test is the Breusch-Pagan homoskedasticity test. LR test is the likelihood ratio test of the SEM against OLS. ShEd: regional share of tertiary education, Unemp.: regional unemployment rate, NRR: country net replacement rate, PMR: country product market regulation index, Capital: the region contains a capital city. $DRMP$ is instrumented by logged regional population. *** p<0.01, ** p<0.05, * p<0.10.

regional population.
The results of the SEM estimation are shown in Table (4). The likelihood ratio tests (testing the model against a simple OLS) and the Hausman tests of residual autocorrelation tend to validate our SEM wage-type equation. Endogeneity tests (Wu-Hausman test) still reject the null hypothesis of exogeneity when no control variables are added to the model (columns i), ii) and iii)). Furthermore, as in the previous estimation, heteroskedasticity is only present in the case where country fixed effects are excluded (model iv)).
The role of domestic market potential does not change and it is only significant in the first specification. The parameter estimate for $FRMP$ becomes larger than in the previous $2SLS$ regression and seems also more precise when residual spatial correlation is controlled for. It takes on a maximum value of 0.803 when only country fixed effect are included (model ii)) and drops to 0.206 when controlling for other potential regional and country income determinants (model v)). Unexplained spatial structure seems to lower the impact of our main variable. However, country fixed effects together with additional controls appear to correct for all unobserved spatial structure in the residuals.

6 Concluding remarks

Models of New Economic Geography claim that there is a positive relationship between a region's access to markets (or real market potential) and its income. This is due to the assumption that firms and/or workers migrate to those regions that are most profitable. Thus, there will be a process of agglomeration as long as there are regions that promise higher rewards for factor remuneration. Market potential is a function of trade costs, since lower trade costs mean an easier access to a specific market. Following this line of reasoning, we utilize a new regional trade data set to estimate bi-regional trade costs, or equivalently, a measure of freeness of trade. We then use these estimates to construct market potential, since it is not observable directly. Estimates from purely theoretically derived specifications show a positive and significant impact market potential can have on income. When we consider differences in institutions and technology, a regions foreign market potential seem to be more relevant than its domestic market potential.

The construction of real market potential is based on a sum of region's income weighted by trade freeness. This is similar to a spatial lag used in the Spatial Econometrics literature. The wage-type equation should therefore, by construction, be able to capture some potential global autocorrelation patterns as Bruna (2015a) points out. However, tests show that this does not suffice to fit the European data. We make use of a spatial error model specification (SEM) in order to correct the estimates by considering remaining unobserved (local) spatial autocorrelation in the residuals. Estimates do not change significantly, as access to foreign markets remains positively associated with high regional income per capita. This is in line with recent research on the topic. One possibility for future analysis is to further open the black box of unexplained spatial correlation in the European data as well as to try to distinguish classic NEG mechanisms from alternative forces of agglomeration.

We conclude that a regions market potential, which is based on trade between regions, can help explain income difference across Europe, even when we consider additional potential factors that drive income, as well as possible correlations across space that are not explained by the equation. This does not necessarily imply a straight-forward policy conclusion. While it seems to be the case that better access to market can increase income to some degree it is not trivial to see how this can be achieved. This might be important for regional policy that promotes convergence and it might not suffice to remove trade impediments in single regions.

Acknowledgements

We would like to thank the European Commission, Joint Research Center in Seville for kindly providing us the trade data, a collaboration that has been made possible thanks to the COST Action IS1104 ("The EU in the new economic complex geography"). The authors are grateful for the numerous helpful feedbacks, critics and suggestions gathered at Vienna University of Economics and Business seminars, the FIW 2015 conference (Vienna), the final GeComplexity conference (Crete, 2016) as well as the ERSA 2016 congress (Vienna).

References

Álvarez-Martínez, M.T. and López-Cobo, M. (2016). Social Accounting Matrices for the EU-27 in 2010. Building a new database for RHOMOLO. *Institute for Prospective Technological Studies, DG-JRC, European Commission*

Amiti, M. and Cameron, L. (2007). Economic Geography and Wages. *Review of Economics and Statistics*, 89(1):15–29.

Baldwin, R., Forslid, R., Martin, P., Ottaviano, G., and Robert-Nicoud, F. (2003). *Economic Geography and Public Policy.* Princeton University Press.

Baradaran, S. and Ramjerdi, F. (2001). Performance of accessibility measures in Europe. *Journal of Transportation and Statistics*, 4(2/3):31–48.

Barde, S. (2010). Increasing Returns and the Spatial Structure of French Wages. *Spatial Economic Analysis*, 5:73–91.

Bosker, M. and Garretsen, H. (2010). Trade costs in empirical New Economic Geography. *Papers in Regional Science*, 89(3):485–511.

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

Bruna, F., Lopez-Rodriguez, J., and Faíña, A. (2015). Market Potential, Spatial Dependences and Spillovers in European Regions *Regional Studies*, 2015:1–13.

Bruna, F. (2015a). Why do empirical tests tend to accept the NEG? An alternative approach to the 'wage equation' in European regions. European Regional Science Association Conference Paper

Bruna, F. (2015b). A generalized NEG wage-type equation. In Díaz-Roldán, Perote, editors, *Advances on International Economics*, pages 63–82. Cambridge Scholars Publishing, Newcastle upon Tyne.

Combes, P.-P., Mayer, T., and Thisse, J.-F. (2008). *Economic Geography: The Integration of Regions and Nations*. Princeton University Press.

Conway, P., Janod, V., and Nicoletti, G. (2005). *Product Market Regulation in OECD Countries: 1998 to 2003*. OECD Economics Department Working Papers, No. 419, OECD Publishing, Paris.

Crespo Cuaresma, J., and Feldkircher, M. (2013). *Spatial filtering, model uncertainty and the speed of income convergence in Europe*. Journal of Applied Econometrics, 28(1):720–741.

EMCO Labour Market Report (2014). Council of The European Union, Brussels.

ESPON European Regional Database (2013). `http://www.espon.eu/main/Menu_Projects/Menu_ScientificPlatform/espondatabase2013.html`

Fingleton, B. and McCann, P. (2007). Sinking the iceberg? On the treatment of transport costs in new economic geography. In Fingleton, B., editor, *New Directions in Economic Geography*, pages 168–203. Edward Elgar Publishing, Cheltenham.

Fujita, M., Krugman, P., and Venables, A. J. (2001). The Spatial Economy: Cities, Regions, and International Trade. MIT Press Books, The MIT Press.

Fujita, M. and Thisse, J.-F. (2002). Economics of agglomeration: Cities, industrial locations, and regional growth. *Cambridge: CambridgeUniversityPress.*

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

Harris, C. D. (1954). The market as a factor in the localization of industry in the united states. *Annals of the association of American geographers*, 44(4):315–348.

Head, K. and Mayer, T. (2002). Illusory Border Effects: Distance mismeasurement inflates estimates of home bias in trade . *CEPII Working Paper Series*, Paris: CEPII.

Head, K. and Mayer, T. (2004b). Market Potential and the Location of Japanese Investment in the European Union. *Review of Economics and Statistics*, 86(4):959–972.

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

Head, K. and Mayer, T. (2015). Chapter 3 - Gravity Equations: Workhorse,Toolkit, and Cookbook. In Elhanan Helpman, K. R. a. G. G., editor, *Handbook of International Economics*, volume Volume 4 of *Handbook of International Economics*, pages 131–195. Elsevier.

Hering, L. and Poncet, S. (2010). Market Access and Individual Wages: Evidence from China, Review of Economics and Statistics, 92(1):145–159

European Commission, DG Joint Research Center (2016). RHOMOLO Model Description: A dynamic spatial general equilibrium model for EU regions and sectors., http://rhomolo.jrc.ec.europa.eu/

Kosfeld, R. and Eckey, H.-F. (2010). Market access, regional price level and wage disparities: the German case., Jahrbuch fuer Regionalwissenschaft, 30:105–128

Krueger, A. B. and Lindahl (2001). Education for Growth: Why and for Whom?, Journal of Economic Literature, 39(4):1101–1136.

Krugman, P. (1991). Increasing Returns and Economic Geography. *Journal of Political Economy*, 99(3):483–499.

López-Rodríguez, J., Faíña, J. A., and López-Rodríguez, J. (2007). Human Capital Accumulation and Geography: Empirical Evidence from the European Union. *Regional Studies*, 41(2):217–234.

Ottaviano, G. (2007). Models of 'new economic geography': factor mobility vs. vertical linkages. In Fingleton, B., editor, *New Directions in Economic Geography*, pages 53–69. Edward Elgar, Cheltenham.

Ottaviano, G. I. P. and Robert-Nicoud, F. (2006). The 'genome' of NEG models with vertical linkages: a positive and normative synthesis. *Journal of Economic Geography*, 6(2):113–139.

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

Robert-Nicoud, F. (2005). The structure of simple 'New Economic Geography' models (or, On identical twins). *Journal of Economic Geography*, 5(2):201–234.

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

Thissen, M., Diodato, D., and Van Oort, F. (2013a). Integrated regional europe: European regional trade flows in 2000. *PBL Netherlands Environmental Assessment Agency publication*, (1035).

Thissen, M., Van Oort, F., and Diodato, D. (2013b). Integration and convergence in regional europe: European regional trade flows from 2000 to 2010. In *ERSA conference papers*, number ersa13p1116. European Regional Science Association.

Tiefelsdorf, M. (2003). Misspecifications in interaction model distance decay relations: A spatial structure effect. *Journal of Geographical Systems*, 5(1):25–50.

Timmer, M.P., Dietzenbacher, E., Los, B., Stehrer, R. and de Vries, G.J. (2015). An Illustrated User Guide to the World InputOutput Database: the Case of Global Automotive Production *Review of International Economics*, 23: 575–605.

Appendix

NUTS is an acronym of the French for the 'nomenclature of territorial units for statistics', which is a hierarchical system of regions used by the statistical office of the European Community for the production of regional statistics. At the top of the hierarchy are NUTS-0 regions (countries) below which are NUTS-1 regions and then NUTS-2 regions. This study disaggregates Europe's territory into 254 NUTS-2 regions located in a subset of the EU-28 member states (excluding Romania and Bulgaria). We exclude the French Departments d'Outre-Mer Guadeloupe, Martinique, Guyane Française and Réunion. Thus, we include the following NUTS-2 regions:

NUTS-2	NAME	NUTS-2	NAME
AT11	Burgenland (AT)	FR71	Rhone-Alpes
AT12	Niederosterreich	FR72	Auvergne
AT13	Wien	FR81	Languedoc-Roussillon
AT21	Karnten	FR82	Provence-Alpes-Cote d'Azur
AT22	Steiermark	FR83	Corse
AT31	Oberosterreich	HR03	Jadranska Hrvatska
AT32	Salzburg	HR04	Kontinentalna Hrvatska
AT33	Tirol	HU10	Kozep-Magyarorszag
AT34	Vorarlberg	HU21	Kozep-Dunantul
BE10	Brussels	HU22	Nyugat-Dunantul
BE21	Prov. Antwerpen	HU23	Del-Dunantul
BE22	Prov. Limburg (BE)	HU31	Eszak-Magyarorszag
BE23	Prov. Oost-Vlaanderen	HU32	Eszak-Alfold
BE24	Prov. Vlaams-Brabant	HU33	Del-Alfold
BE25	Prov. West-Vlaanderen	IE01	Border, Midland and Western
BE31	Prov. Brabant Wallon	IE02	Southern and Eastern
BE32	Prov. Hainaut	ITC1	Piemonte
BE33	Prov. Liege	ITC2	Valle d'Aosta/Vallee d'Aoste
BE34	Prov. Luxembourg (BE)	ITC3	Liguria
BE35	Prov. Namur	ITC4	Lombardia
CY00	Kipros / Kibris	ITF1	Abruzzo
CZ01	Praha	ITF2	Molise
CZ02	Stredni Cechy	ITF3	Campania
CZ03	Jihozapad	ITF4	Puglia
CZ04	Severozapad	ITF5	Basilicata
CZ05	Severovychod	ITF6	Calabria
CZ06	Jihovychod	ITG1	Sicilia
CZ07	Stredni Morava	ITG2	Sardegna
CZ08	Moravskoslezsko	ITH1	Provincia Autonoma di Bolzano/Bozen
DE11	Stuttgart	ITH2	Provincia Autonoma di Trento
DE12	Karlsruhe	ITH3	Veneto
DE13	Freiburg	ITH4	Friuli-Venezia Giulia
DE14	Tubingen	ITH5	Emilia-Romagna
DE21	Oberbayern	ITI1	Toscana
DE22	Niederbayern	ITI2	Umbria
DE23	Oberpfalz	ITI3	Marche
DE24	Oberfranken	ITI4	Lazio
DE25	Mittelfranken	LT00	Lietuva
DE26	Unterfranken	LU00	Luxembourg
DE27	Schwaben	LV00	Latvija
DE30	Berlin	MT00	Malta
DE40	Brandenburg	NL11	Groningen
DE50	Bremen	NL12	Friesland (NL)
DE60	Hamburg	NL13	Drenthe
DE71	Darmstadt	NL21	Overijssel
DE72	Giessen	NL22	Gelderland
DE73	Kassel	NL23	Flevoland
DE80	Mecklenburg-Vorpommern	NL31	Utrecht
DE91	Braunschweig	NL32	Noord-Holland
DE92	Hannover	NL33	Zuid-Holland
DE93	Luneburg	NL34	Zeeland
DE94	Weser-Ems	NL41	Noord-Brabant
DEA1	Dusseldorf	NL42	Limburg (NL)
DEA2	Koln	PL11	Lodzkie
DEA3	Munster	PL12	Mazowieckie
DEA4	Detmold	PL21	Malopolskie
DEA5	Arnsberg	PL22	Slaskie
DEB1	Koblenz	PL31	Lubelskie
DEB2	Trier	PL32	Podkarpackie
DEB3	Rheinhessen-Pfalz	PL33	Swietokrzyskie
DEC0	Saarland	PL34	Podlaskie
DED2	Dresden	PL41	Wielkopolskie
DED4	Chemnitz	PL42	Zachodniopomorskie
DED5	Leipzig	PL43	Lubuskie
DEE0	Sachsen-Anhalt	PL51	Dolnoslaskie
DEF0	Schleswig-Holstein	PL52	Opolskie

DEG0	Thuringen
DK01	Hovedstaden
DK02	Sjaelland
DK03	Syddanmark
DK04	Midtjylland
DK05	Nordjylland
EE00	Eesti
EL11	Anatoliki Makedonia, THraki
EL12	Kentriki Makedonia
EL13	Ditiki Makedonia
EL14	THessalia
EL21	Ipeiros
EL22	Ionia Nisia
EL23	Ditiki Ellada
EL24	Sterea Ellada
EL25	Peloponnisos
EL30	Attiki
EL41	Boreio Aigaio
EL42	Notio Aigaio
EL43	Kriti
ES11	Galicia
ES12	Principado de Asturias
ES13	Cantabria
ES21	Pais Vasco
ES22	Comunidad Foral de Navarra
ES23	La Rioja
ES24	Aragon
ES30	Comunidad de Madrid
ES41	Castilla y Leon
ES42	Castilla-La Mancha
ES43	Extremadura
ES51	Cataluna
ES52	Comunidad Valenciana
ES53	Illes Balears
ES61	Andalucia
ES62	Region de Murcia
ES63	Ciudad Autonoma de Ceuta
ES64	Ciudad Autonoma de Melilla
ES70	Canarias
FI19	Lansi-Suomi
FI1B	Helsinki-Uusimaa
FI1C	Etela-Suomi
FI1D	Pohjois- ja Ita-Suomi
FI20	Aland
FR10	Ile 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-Comte
FR51	Pays de la Loire
FR52	Bretagne
FR53	Poitou-Charentes
FR61	Aquitaine
FR62	Midi-Pyrenees
FR63	Limousin
PL61	Kujawsko-Pomorskie
PL62	Warminsko-Mazurskie
PL63	Pomorskie
PT11	Norte
PT15	Algarve
PT16	Centro (PT)
PT17	Lisboa
PT18	Alentejo
PT20	Regiao Autonoma dos Acores
PT30	Regiao Autonoma da Madeira
SE11	Stockholm
SE12	Ostra Mellansverige
SE21	Smaland med oarna
SE22	Sydsverige
SE23	Vastsverige
SE31	Norra Mellansverige
SE32	Mellersta Norrland
SE33	Ovre Norrland
SI01	Vzhodna Slovenija
SI02	Zahodna Slovenija
SK01	Bratislavsky kraj
SK02	Zapadne Slovensko
SK03	Stredne Slovensko
SK04	Vychodne Slovensko
UKC1	Tees Valley and Durham
UKC2	Northumberland and Tyne and Wear
UKD1	Cumbria
UKD3	Greater Manchester
UKD4	Lancashire
UKD6	Cheshire
UKD7	Merseyside
UKE1	East Yorkshire and Northern Lincolnshire
UKE2	North Yorkshire
UKE3	South Yorkshire
UKE4	West Yorkshire
UKF1	Derbyshire and Nottinghamshire
UKF2	Leicestershire, Rutland and Northamptonshire
UKF3	Lincolnshire
UKG1	Herefordshire, Worcestershire and Warwickshire
UKG2	Shropshire and Staffordshire
UKG3	West Midlands
UKH1	East Anglia
UKH2	Bedfordshire and Hertfordshire
UKH3	Essex
UKI1	Inner London
UKI2	Outer London
UKJ1	Berkshire, Buckinghamshire and Oxfordshire
UKJ2	Surrey, East and West Sussex
UKJ3	Hampshire and Isle of Wight
UKJ4	Kent
UKK1	Gloucestershire, Wiltshire and Bristol/Bath area
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
UKM5	North Eastern Scotland
UKM6	Highlands and Islands
UKN0	Northern Ireland

Table 6: Data and variables description

Variable	data description	Level of aggregation	Obs.	Mean	Min	Max	Source
Exports $\tilde{x}$	Regional exports, gross value, for sectors B-F	NUTS-2	64,516	152.63	0	266,683.40	JRC, EC
GVA p.c.	Gross value added per capita	NUTS-2	252	23,018.41	4,963.42	74,869.86	Eurostat
RMP	Real market potential	NUTS-2	251	3.52	0.57	10.51	own calculations
$DRMP$	Domestic market potential	NUTS-2	250	2.26	0.20	9.33	own calculations
$FRMP$	Foreign market potential	NUTS-2	250	1.26	0.31	2.70	own calculations
ShEd.	Share of 25-64 olds that completed tertiary education (ISCED 2011, level 5-8)	NUTS-2	268	25.47	9.00	54.50	Eurostat
Unemp.	unemployment rate	NUTS-2	267	9.21	2.70	32	Eurostat
NRR	It is defined as the net income of an unemployed person receiving unemployment and possibly other benefits, expressed as a share of the income earned previously in the job before becoming unemployed.	Country	266	0.29	0	0.75	EC
PMR	Index that summarises a wide array of different regulatory provisions across countries.	Country	249	1.50	0.96	2.21	OECD
Capital	Indicator, equals 1 if a region contains the capital city of the country.	NUTS-2	252				Dummy, own calculations
Geo Dist.	Distance in travel time (seconds) between economic centers of two regions.	NUTS-2	63504	57193	0.17	1860196	Google API
Border	Indicator, equals 1 if the trade flow has to pass a country border.	NUTS-2	64516				Dummy, own calculations
Contig.	Indicator, equals 1 if trading regions are contiguos and separated only by a country border.	NUTS-2	64516				Dummy, own calculations
Lang Sim.	Indicator, equals 1 if trading regions share a common language.	NUTS-2	64516				Dummy, own calculations

Notes EC: European Commission, JRC: Joint Research Center, OECD: Organisation for Economic Co-operation and Development.