Department of Economics
Working Paper No. 220

Trade Costs and Income in European Regions

Christoph Hammer
Aurélien Fichet de Clairfontaine

February 2016

WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

Trade Costs and Income in European Regions*

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

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

Vienna University of Economics and Business (WU)
February 2016

Abstract

Using a New Economic Geography (NEG) model, this study estimates the relationship between regional per capita income levels and the proximity of regions to large markets. Market access cannot be observed directly, so it has to be constructed. We follow a two-step-procedure of Redding and Venables (2004) and use results of a spatially-filtered gravity model to infer market access. To this end, we make use of a new dataset of constructed bi-regional trade flows between (and within) 240 European NUTS-2 regions (from 25 European countries excluding Bulgaria, Croatia and Romania) for the year 2010 (Thissen et al. 2014, IPTS). In a second step we test the hypothesis that access to large markets increases factor incomes. We find robust evidence that supports this hypothesis on a regional level. Controlling for a variety of factors that drive income differences, our findings highlight the robustness of the role of market access in explaining the uneven spatial distribution of income.

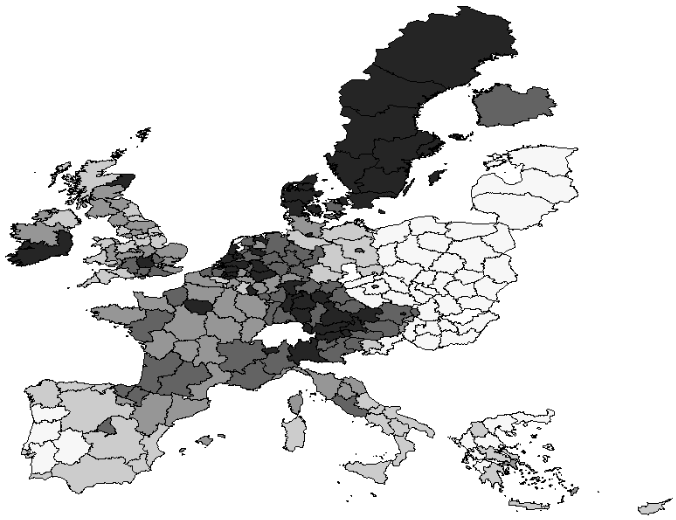
Keywords Wage equation · Gravity · European regions · New Economic Geography
JEL Classification F12 · F14

1 Introduction

Differences in incomes across countries are one of the most studied questions in economics. Models of growth theory usually constitute the main framework studying questions of income differences. However, in growth theory, spatial aspects usually play a very limited role, goods and factors are assumed to be immobile. A look at the data reveals that economic activity is not uniformly distributed across space. There are centers of economic activity and there is also an economic periphery. This is true on the international scale down to the regional and even the city scale. Figure 1 shows the distribution of gross value added (GVA) per capita of European NUTS-2 regions for 2010 [1], darker shades indicate a higher GVA. There seems to be a "core-periphery" structure not only at an European scale but also within countries where the region around the capital city is also the economic core with the highest GVA per capita.
Ever since, the seminal work of Krugman (1991) models of "New Economic Geography" (NEG) have attempted to shed some light on this uneven distribution using a general-equilibrium framework with a focus on geography represented by transport- or trade costs. The change in the spatial structure is a result of a combination of

*Corresponding author: Christoph Hammer (christoph.hammer@wu.ac.at)
[1]http://epp.eurostat.ec.europa.eu/portal/page/portal/nuts_nomenclature/introduction

Figure 1: GVA in 2010, Source: Cambridge Econometrics



Darker shades of grey correspond to higher distribution quantiles of the regional GVA.

forces that promote agglomeration or dispersion respectively. Unlike growth theory (goods and factors immobile) and trade theory (factors immobile), NEG models assume mobile goods and factors. While there are many possible explanations for the formation of agglomerations in space such as physical endowment or the physical geography of a region – what Krugman labelled "first-nature geography" – forces that reflect behaviour of optimizing mobile economic agents are the focus of NEG models. One interesting feature of CES-type NEG models is the so-called "wage-equation"[2]. 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 wage structure in which wages are higher in regions that have better access to markets. In turn, this leads to a relocation[3] of firms and workers to those regions. Depending on the specific model, agglomeration in the high-income "core-region" might present a stable long-run equilibrium. There are, however, also variants following Puga (1999) where this process is eventually reversed and firms relocate to the periphery. In this case, dispersion forces outweigh agglomeration forces. Increasing house prices in the core due to immigration of firms and workers would be one example for such a dispersion force. The relative strength of agglomeration- and dispersion forces in NEG models is driven by the degree of trade integration.

The concept of increased factor remuneration in areas with good access to demand, was first introduced by Harris (1954), who used simple geographic distance to weigh demand from distant regions. Following empirical literature that studies the wage equation is dominated by a two-step approach put forward by Redding and Venables (2004). They first derive a structural trade equation to estimate bilateral trade flows using a gravity model. This is necessary to construct trade costs, which then are used

[2]The name arises from the fact that labour is the only production input in the first NEG model by Krugman (1991). It also holds for other production inputs, since it is about factor remuneration in general.

[3]See Baldwin et al. (2003) for a detailed discussion of different NEG model types. They mainly differ in their assumptions about the factors of production and their assumed (im-)mobility. Head and Mayer (2004a) discuss the early literature from an empirical point of view and provide a list of testable hypotheses of NEG models.

to calculate market- as well as supplier access. In a second step, the wage equation is estimated using the constructed variables. They apply their model to a cross-section of over 100 countries and find positive evidence on the importance of market- and supplier access in explaining income differences across countries.

The wage equation seems to be a very robust[4] relationship. In a similar paper, Head and Mayer (2004b) study the importance of market access for Japanese investors in selected European NUTS-1 regions. Hanson (2005) does not use the two-step approach but estimates the wage equation in a cross-section of US counties including house prices and controlling for heterogeneous workers. Head and Mayer (2006) then incorporate industry, time and intra-national variation into the wage equation. They also consider the possibility that higher demand not only leads to higher wages but higher employment. Their findings suggest that wages respond more to changes in demand, however this response seems to be very industry-specific. Breinlich (2006) applies the two-step approach to a more disaggregated data set of 193 NUTS-2 regions for the 1975-1997 period. He also confirms the role of market access in explaining income differences, although he argues that the trade-cost-saving motive[5] for locating in a core-region might only be a small part of the overall picture. Theoretical NEG models that include intermediate inputs[6] also highlight the importance of supplier access. That is, the attractiveness of a region also depends on the access to sources of inputs, not only output. While Redding and Venables (2004) also control for supplier access with fixed effects, Amiti and Cameron (2007) build supplier access using input-output information on the regional level for Indonesia. Input-output tables provide information about the use of intermediate inputs by a specific industry. Since they cannot match this on the firm level they refer to it as “potential rather than actual suppliers” (Amiti and Cameron, 2007, p.20). They also control for other potential explanations for agglomeration like labour pooling or technological spillovers. Both explanations yield statistically significant results. On the one hand, firms benefit from the proximity to other firms using similar input-mixes, which increases their productivity and thus wages. On the other hand, proximity to firms in the same industry leads to a decrease in wages. This might be due to competition effects outweighing positive externalities. They also show that most effects are highly localized and benefits from good market- and supplier-access at about 100km and 260km respectively. Amiti and Smarzynska Javorcik (2008) perform a very similar study for foreign direct investments (FDI) decisions of foreign firms into China. They transform the wage equation to study how the change in the number of new firms entering a region is dependent on market- and supplier access. An additional feature is the possibility to allow for heterogeneity across industries, which shows that market- and supplier access matter for all industries but differ in magnitude. Hering and Poncet (2009) study the wage equation for 29 Chinese provinces over 1995-2002. In addition to previous studies, they try to identify through which channels market access might influence wages. One example for such an indirect channel comes from increased incentives to human capital formation in agglomerated areas due to positive knowledge externalities. Their results suggest that market access still explains part of the variation in wages even after controlling for potential indirect channels. (Hering and Poncet, 2009, p.12). On a micro level for Chinese workers Hering and Poncet (2010) also find a positive effect of market access on wages. However, the effect is stronger for high-skilled workers and for those working in internationally oriented firms. Their work also confirms one hypothesis of NEG models, namely that a further integration into the world economy might lead to an even stronger wage gradient within China when labour is not mobile enough.

For the estimation of the wage-equation, constructing market access requires a speci-

[4]See Bosker and Garretsen (2010) for a meta-study on the wage equation.

[5]As Redding and Schott (2003) put it:“Because firms in remote locations pay greater trade costs on both their sales to final markets and their purchases of imported intermediate inputs, they have less value added available to remunerate domestic factors of production”. For more details on how market access influences the level of wages see section 2.4.

[6]So called “vertical-linkages” models, see for example Krugman and Venables (1995).

fication of trade costs. 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 (e.g., tariffs, product standards, language, etc.), see for example Fujita et al. (2001), Fujita and Thisse (2002), Fingleton and McCann (2007) or Combes et al. (2008) for a more detailed discussion. Trade costs are not only vital in terms of theory but are still relevant in real world trade. Empirical work has shown that, while pure transport costs have decreased over the last decades, trade still is impeded by distance, borders and other factors mentioned above (for a survey see Head and Mayer (2015)).

Since market access is a trade-cost-weighted measure of demand, we first estimate bilateral trade costs, using a spatial interaction model of the gravity type, by applying a new data set[7] of constructed trade flows between NUTS-2 regions. In constrast to previous research relying on the estimation of the trade equation, the spatial interaction model used in this study accounts for origin- and destination spatial dependences by incorporating spatial filters (see LeSage and Pace (2010)). Second, we construct market access using fitted values of the trade costs in order to estimate the wage equation.

We pay a lot of attention to the robustness of our results. Since our approach consists of two steps, there might be a large variation of outcomes depending on the specifics of both steps (see Bosker et al. (2010)). We will also compare the theoretically derived specifications of market access to more *ad-hoc* versions such as the Harris market potential or a spatial autoregressive model.

The contributions to the literature are many fold: i) having a focus on European regions we are able to control for regional determinants of income, ii) we make use of a spatially-filtered trade equation estimation for the construction of measures of market access, making the process more reliable, iii) we perform a benchmark exercise, comparing the theoretically based measures of market access against *ad-hoc* construction schemes in order to test for model accuracy and iv) the estimation a spatial version of the wage equation (as a robustness check) helps shed some light on possible spatial externalities.

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 gives details to the data sets, while estimation results are presented and analysed in section 5. Finally, section 6 concludes.

2 Theoretical Model

While space is not explicitly modelled in growth theory, it is often accounted for by augmenting growth equations with a spatial autocorrelation component. However, spatial linkages are not explained economically. They are taken into account by modelling a spatial structure that is purely defined by some form of geographical neighbourhood (contiguity, k-5 nearest neighbours, etc.). While it is necessary to account for spatial correlation from an econometric point of view, it is not satisfactory from an economic point of view[8]. The role for spatial interaction remains inside a "black box". In models of economic geography, on the other hand, spatial structures are the outcome. We will use such models that explain why and how income is related to spatial aspects, which is the focus of our question. Moreover, Head and Mayer (2011) show that the "wage-equation" is even more general and can be derived from different trade[9] models. It is exactly this relationship that we exploit in our analysis.

[7]Details on the construction of the trade matrix can be found in Thissen et al. (2013a) and Thissen et al. (2013b)

[8]We will, however, also contrast our results to a spatial formulation.

[9]According to Head and Mayer, a gravity-type trade equation forms the basis of derivation. This gravity equation is consistent with models of product differentiation, comparative advantage or firm heterogeneity.

There is a big variety[10] of NEG models. They differ in their use of functional forms of utility and costs 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, Ottaviano (2007), Ottaviano and Robert-Nicoud (2006) and Robert-Nicoud (2005) 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 behaviour of the model, the "short-run"[11] equilibrium properties are nearly identical. We will use the implications of the short-run equilibrium to explain differences in income across regions.
The first generation[12] of NEG models is not solvable analytically but relies on numerical solutions. We chose to use the so-called "Footloose entrepreneur"(FE) model by Forslid and Ottaviano (2003) which is solvable analytically and fits our question best.

2.1 Demand

Assume that there are two sectors of production and two types of labour (skilled and unskilled) that spend their income locally. While the constant-returns-to-scale sector (A) acts as a numeraire, the sector of interest is often dubbed "manufacturing" (M) and it features increasing-returns-to-scale internal to the firm. Firms in that sector produce different varieties under monopolistic competition. This means that every firm produces one variety, so there are as many firms as varieties. This model is based on the popular framework developed by Dixit and Stiglitz (1977).
Utility of a representative consumer in this model economy comes from consumption of goods from both sectors A and M.

$$U = A^{1-\mu} M^{\mu} \tag{1}$$

where μ represents the share of expenditure that is spent on the bundle (varieties) of manufactured goods M. Consumers face a budget constraint $E = p_A A + PM$ where E denotes total expenditure, p_A is the price of one unit of good A and P is a price index for all varieties of the M-good.[13] Under this constraint demand for both types of goods takes the familiar form

$$A = \frac{(1-\mu)E}{p_A} \qquad M = \frac{\mu E}{P} \tag{2}$$

Consumers, however, not only face the decision of consuming A and M, but also the decision of which or how much varieties of M to consume. For the M-good, consumers' sub-utility takes the CES form

$$M \equiv \left[\sum_i^N x_i^{(\sigma-1)/\sigma} \right]^{\sigma/(\sigma-1)} \tag{3}$$

Utility increases with the amount of consumption x_i. Further, one can show that this functional form implies the love-for-variety-effect. This means, utility increases with the numbers of varieties available to the consumer – they love variety for varieties' sake. The parameter σ represents the constant elasticity of substitution between any two varieties. Different varieties are imperfect substitutes, so every producer

[10] See Baldwin et al. (2003) for a survey of theoretical models of NEG.

[11] 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" equilibrium.

[12] See Krugman (1991) and Krugman and Venables (1995).

[13] There is no saving in this model, hence income equals expenditure $Y = E$ on the aggregate level.

exerts some form of monopoly power. Considering budget constraints, this yields the following demand function for a variety j with σ elasticity

$$x_j = p_j^{-\sigma} \mu E P^{\sigma-1} \tag{4}$$

$$P = \left[\sum_i^N p_i^{1-\sigma} \right]^{1/(1-\sigma)} \tag{5}$$

μE denotes expenditure on manufacturing goods, P the price index of manufacturing goods and p_j the free-on-board (FOB or mill-) price of variety j.

2.2 Supply

The demand side is virtually identical in all models with a CES sub-utility function. Most of the differences between NEG models come from the supply side and the assumed cost function. The present model assumes that there are two types of labour input, skilled (H) and unskilled (L). The constant-return sector only uses unskilled labour which leads to marginal cost pricing $p_A = w_L$. The term "skilled labour" is synonymous to human capital or entrepreneurs. Entrepreneurs are assumed to be mobile between regions while unskilled labour is assumed to be immobile between regions. Further, it is assumed that in order to set up a firm it requires fixed costs of α units of entrepreneurs H. Unskilled labour is only used in the marginal cost part, i.e.,

$$C(x_i) = w_H \alpha + w_L \beta x_i \tag{6}$$

where x_i is the quantity produced, α and β the fixed and marginal input requirements respectively, w_L the price (wage) of unskilled labour and w_H the price (wage) of the entrepreneur. This might be a quite reasonable assumption about production. As Combes et al. (2008) put it: "In fact, in a growing number of industries, production is divided into several activities, starting with the product design and ending with its marketing and distribution, which all require skilled workers, while the actual production can often be performed by unskilled workers".
The first order condition leads to

$$\pi_i = p_i x_i - w_H \alpha - w_L \beta x_i \tag{7}$$

$$p_i = w_L \beta \frac{\sigma}{\sigma - 1} \tag{8}$$

(8) shows that the price for any variety is a constant mark-up over variable costs, in this case, the wage of unskilled labour. (8) is also a result of the assumption that individual firms can ignore their effect on the overall price index, i.e., $\partial P / \partial p_i = 0$. There is no strategic interaction among firms in this model.

2.3 Trade Costs

So far we focused on a representative consumer and a typical firm in this model economy, irrespective of their location. Now, assume there are $r = 1, ..., R$ identical[14] regions. The index for varieties is dropped, since they enter utility symmetrically.
The importance of space in NEG models is given by the assumption that shipping (manufacturing-) goods to other regions is costly. Virtually all NEG models assume iceberg-type trade costs[15]: in order for one unit to arrive at the destination, τ_{rs} units

[14] "Identical" refers to the endowment with factors, resources, technologies used by firms, etc. Forslid and Ottaviano (2003) show, that the equilibrium properties also hold for the asymmetrical case.

[15] See Fingleton and McCann (2007) for a discussion about the theoretical and empirical implications of the iceberg assumption.

have to be shipped. This results in a higher price for exporting and a slightly different price index[16]

$$p_{rs} = p_r \tau_{rs} \tag{9}$$

$$P_r = \left[\sum_s^R n_s p_{rs}^{1-\sigma} \right]^{1/(1-\sigma)} \tag{10}$$

where n_s represents the number of firms, i.e., the number of varieties produced in region s. This means that mill-pricing is still the optimal strategy for firms. Firms charge the same price irrespective of their location. The price of selling to other regions only differs by the trade costs τ_{rs}. The demand from region s for goods of region r becomes

$$x_{rs} = p_r^{-\sigma} \mu E_s P_s^{\sigma-1} \tau_{rs}^{1-\sigma} \tag{11}$$

Summing over all (export-) markets, a typical firm in region r now faces demand

$$x_r = p_r^{-\sigma} \sum_s \mu E_s P_s^{\sigma-1} \tau_{rs}^{1-\sigma} \tag{12}$$

where μE_s denotes expenditure on manufacturing goods in region s. Further, trade costs have to be accounted for in the price and the amount being shipped. Note that trade is free within a region $\tau_{rr} = 1$, this guarantees that we can interpret $\tau_{rs}^{1-\sigma}$ as the "freeness of trade" ϕ (see Baldwin et al., 2003). When trade is completely free, ϕ is 1, if trade costs are prohibitively high, ϕ approaches 0. (12) also shows that demand faced by a firm in region r depends on the competition[17] in other regions P_s.

2.4 Short-run equilibrium

The short-run equilibrium is now defined as a situation where good- and labour markets are cleared simultaneously. The difference to the long-run equilibrium comes from the mobility of the entrepreneurs. Only in the long-run equilibrium, they have no more incentive to move across regions to the region with the highest profitability. Goods that are produced on the competitive market (A) are assumed to be shipped freely across regions, which equalizes prices as well as wages of unskilled labour $w_{L,r} = w_{L,s}$. Thus, good A is chosen as the numeraire, giving $p_{A,r} = w_{L,r} = 1$. The clearing condition for the good market of A will be dropped due to Walras' law. Turning to the manufacturing sector, the fixed cost requirement α means that in equilibrium the number of firms n_r is proportional to the number of entrepreneurs. The labour market clearing condition for entrepreneurs H is thus given by

$$n_r = \frac{H_r}{\alpha} \tag{13}$$

Another short-run equilibrium condition is free entry and exit of firms. Firms enter until every firm generates zero operating profits, or equivalently, profits only cover fixed costs.

$$(p_r - \beta) x_r = \frac{x_r p_r}{\sigma} = \alpha w_{H,r} \tag{14}$$

which then determines the break-even amount of output per firm

$$x_r = \bar{x} = (\sigma - 1) \alpha w_{H,r} \tag{15}$$

According to (12) the good market equilibrium is defined as

$$\bar{x} = p_r^{-\sigma} \sum_s \mu E_s P_s^{\sigma-1} \tau_{rs}^{1-\sigma} \tag{16}$$

[16] The index rs denotes origin region r and destination region s. $\tau \in [1, +\infty)$

[17] Competition is higher in regions where many varieties are produced. Less varieties have to be imported subject to trade costs which lowers the price index in the respective region.

Using (15), we can rewrite

$$(\sigma - 1)\alpha w_{H,r} = p_r^{-\sigma} \sum_s \mu E_s P_s^{\sigma-1} \tau_{rs}^{1-\sigma} \tag{17}$$

In order to estimate this equation and for an easier interpretation, we transform it further[18]

$$w_{H,r} = \kappa \sum_s E_s P_s^{\sigma-1} \tau_{rs}^{1-\sigma} \tag{18}$$

which Fujita et al. (2001) and Redding and Venables (2004) call the "wage equation". It restates the market clearing condition for manufacturing goods in terms of factor remuneration. Given the current location of firms (i.e., given the market size) and given they make zero operating profits, (18) shows the maximum a firm can pay their entrepreneurs as a function of market access (dubbed MA henceforth). MA is defined as the trade-cost-weighted sum of market capacities $E_s P_s^{\sigma-1}$.
Another way of interpreting this equilibrium condition is that there is a bidding process (at equilibrium market prices) of firms for entrepreneurs, which only ends when all remaining operating profits are zero and totally cover the wage bill of skilled labour (see also (14)). Remember that any type of worker/entrepreneur spends his or her income locally. Therefore, the location of H determines the relative market sizes. For the empirical part, we take this distribution as given to see if it can help explain differences in income across regions according to (18), ie. we estimate a short-run equilibrium.

3 Estimation strategy

This section provides an overview of the proposed empirical implementation. First we derive the estimation specification of the wage equation. Secondly we show how to obtain the necessary measures, which come from a gravity equation.
Redding and Venables (2004) estimate a version of (18) which differs in the assumed production factors. Their dependent variable is a composite of costs for immobile labour, some mobile production factor as well as some c_r, the marginal input requirement or "technological difference". In most studies the dependent variable is proxied by GVA. Breinlich (2006) estimates (17) where the technological differences are captured by the error term. To arrive at an estimable expression we take the natural log of (18) and add a disturbance term

$$\ln(w_r) = \gamma + \varphi \ln(MA_r) + \varepsilon_r \tag{19}$$

where $\varepsilon_r \sim N(0, \sigma^2)$.
Since MA is not directly observable, we have to construct it. Head and Mayer (2004a), Combes et al. (2008) and Bosker and Garretsen (2010) summarize important dimensions along which the literature differs when it comes to implementing (19).
First, the wage equation can be estimated directly or in a two-step approach. In this case, directly means that trade costs are already incorporated in the wage equation and estimated together with other parameters of the model, see for example Hanson (2005). For that it is necessary to use non-linear least squares methods which often are highly sensitive to the starting values. On the other hand, it is possible to estimate parameters like σ which are important for simulation studies[19]. An alternative

[18] κ captures all constants.

[19] Even if the structural parameters can be extracted, there is still a problem of interpretation. There might be a problem of interpreting structural parameters derived from a two-region model, estimated with a multi-region data set. Bosker et al. (2010) discuss this issue at length and propose an empirical strategy that first estimates the parameters of interest econometrically. They then use the results to calibrate a more sophisticated multi-regional NEG model to simulate the long-run distribution of firms. This approach is beyond the scope of the present paper but should be kept in mind for interpreting the results.

method follows Redding and Venables (2004) who first estimate a trade equation of the gravity-type in order to get estimates for trade costs between regions. In a second step, obtained estimates are then used to construct measures of MA. This approach, however, does not allow to identify the theoretical parameters separately.

3.1 The trade equation

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

$$n_r p_r x_{rs} = n_r p_r^{1-\sigma} \mu E_s P_s^{\sigma-1} \tau_{rs}^{1-\sigma} \quad (20)$$

This states that exports from r to s are a function of supply capacity $sc = n_r p_r^{1-\sigma}$, market capacity $mc = E_s P_s^{\sigma-1}$, as well as bilateral trade costs $\tau_{rs}^{1-\sigma}$. The right-hand-side of (20) shows the similarity to the definition of MA. It is exactly this similarity that is used to construct MA with the results of the estimated trade equation. We define the total value of goods exported from region r to s as exports $\tilde{x}_{rs}$ and take the natural logarithm to arrive at a specification that can be estimated

$$\ln(\tilde{x}_{rs}) = c + sc + mc + (1-\sigma)\ln(\tau_{rs}) + \varepsilon_{rs} \quad (21)$$

where c^{20} is a constant term controlling for the mean measurement error in the $\tilde{x}_{rs}$'s and $\varepsilon_{rs} \sim N(0, \sigma^2)$.
Supply and market capacity measures (i.e., sc and mc) are not observable and are therefore proxied by exporters and importers fixed effects ξ_s and ξ_r.
Regarding the functional form, multiple approaches on how to proxy both capacity measures for the trade equation estimation, i.e., sc and mc in (21), have been used over time. In early studies, logged GDPs, expenditures as well as remoteness terms were highly popular although they implicitly impose severe restrictions ($mc = \mu E_s$ in the case of the remoteness terms). Modern practice in estimating trade equation has advocated the use of importer and exporter fixed effects as proxy of sc and mc, yielding consistent estimates of $\tau_{rs}^{1-\sigma}$. The exporter ξ_r and importer fixed effects ξ_s can be written as

$$\begin{aligned} \xi_r &= I_n \otimes \iota_n \\ \xi_s &= \iota_n \otimes I_n \end{aligned} \quad (22)$$

where, for n regions, I_n is a n–by–n identity matrix and ι_n a 1–by–n vector of ones. In our econometric setting, we follow this latter definition of multilateral resistance terms using fixed effects.
$\tau_{rs}^{1-\sigma}$ is approximated by a distance deterrence function D_{rs} construed to include spatial, institutional and cultural separation factors that are defined in a further section. Inserting the deterrence function D_{rs} in (21) we obtain

$$\ln(\tilde{x}_{rs}) \quad = \quad \kappa + \delta_1 \xi_r + \delta_2 \xi_s + \ln D_{rs} + \varepsilon_{rs} \quad (23)$$

where

$$\begin{aligned} sc = n_r p_r^{1-\sigma} &\equiv \delta_1 \xi_r && (24) \\ mc = \mu E_s P_s^{\sigma-1} &\equiv \delta_2 \xi_s && (25) \\ \tau_{rs}^{1-\sigma} &\equiv D_{rs} && (26) \\ & && (27) \end{aligned}$$

In turn, this specification allows us to obtain estimates for the empirical counterparts of mc and $\tau_{rs}^{1-\sigma}$ which we will combine later to obtain the correct theory-consistent

[20] It captures the constant $\frac{1}{E}$.

measure of MA.
Taking into account the spatial nature of bilateral trade flows (see LeSage and Pace 2010), eigenvector spatial filters were added to the gravity model specification in (21) [21]. The eigenvectors identified are used as additional explanatory variables in (21) to filter out any remaining unexplained spatial dependence in the residuals. Spatial filtering relies on a spectral decomposition of the transformed spatial weight matrix MWM, where W is an N-by-N spatial weight matrix

$$W = W_n \otimes W_n \tag{28}$$

that captures spatial dependence between origin-destination flows from regions neighbouring both the origins and destinations, labelled origin-to-destination dependence by LeSage and Pace (2010). W_n is a row-stochastic n-by-n spatial weight matrix that describes spatial neighbourhood relationships between the n European regions. This matrix has – by convention – zeros in the main diagonal, and non-negative elements in the off-diagonal cells. Specifically, the (r,s)-th element of W_n is greater than zero if r and s are neighbouring regions.[22] $\otimes$ denotes the Kronecker product, and M is the N-by-N projection matrix $M = I_N - \iota_N \iota_N^{'} \frac{1}{N}$ where I_N is the N-by-N identity matrix, and ι_N the N-by-1 vector of ones.
The orthogonality properties of eigenvectors make the spectral decomposition useful for lower rank approximations to MWM (see Pace et al. 2013). The usual approach is to keep all the eigenvectors associated with the largest magnitude eigenvalues and discard the rest. This involves partitioning the eigenvalues and vectors into two sets, a set of eigenvectors associated with the largest Q eigenvalues and a set of eigenvectors associated with the smallest $N - Q$ eigenvalues of MWM. We follow Tiefelsdorf and Griffith (2007) to identify and optimise the subset of Q eigenvectors by stepwise integration of the eigenvectors. Including Q eigenvectors, (23) is rewritten as:

$$\ln\left(\tilde{x}_{rs}\right) = \kappa + \delta_1 \xi_r + \delta_2 \xi_s + \ln(D_{rs}) + \sum_{q=1}^{Q} \psi_q E_q + \varepsilon_{rs} \tag{29}$$

where E_q is the $q \in Q$ eigenvector and ψ_q is respective coefficient. The OLS residuals of the non-filtered trade equation defined in (21) yield a significant Moran's I of 0.355 on average while the same measure obtained from the filtered version of the model is only equal to 0.035.

3.2 The construction of market access and the wage equation

In estimating the trade equation (29) we find empirical counterparts for $\mu E P_s^{1-\sigma}$ as well as $\tau_{rs}^{1-\sigma}$, which are needed to construct market capacity $\mu E P_s^{1\sigma} \tau_{rs}^{1-\sigma}$ of a region s. Following Redding and Venables (2004), we use those counterparts to obtain the market access measure

$$\widehat{MA}_r = \sum_s \hat{\xi}_{s,s} (D_{rs})^{\hat{\gamma}} = \hat{\xi}_{r,r} (D_{rr})^{\hat{\gamma}} + \sum_{s \neq r} \hat{\xi}_{s,s} (D_{rs})^{\hat{\gamma}} \tag{30}$$

where $\hat{\xi}_{s,r}$ and $\hat{\xi}_{s,s}$ are the estimated importer fixed effects of region r and s respectively and $\hat{\boldsymbol{\gamma}}$ the parameters vector of the components of the deterrence function D_{rs}. The first term of the right-hand side of (30) corresponds to the domestic MA, the second term to the foreign MA. This again shows that MA is a trade-cost-weighted sum of demand from all potential markets. (31) is rewritten by including the terms of (30) which results in

$$\ln(w_r) = \gamma + \varphi \ln(\widehat{MA}_r) + \varepsilon_r \tag{31}$$

[21]see Chun 2008, Fischer and Griffith 2008, Chun and Griffith 2011, Griffith and Fischer 2013.

[22]Neighbours may be defined using contiguity or measures of spatial proximity such as cardinal distance (for example, in terms of the great circle distance) or ordinal distance (for example, in terms of k-nearest neighbours). In this application, we use the concept of k-nearest neighbours with $k = 5$ to define W.

and can be estimated using standard OLS.

4 Data, Variables and Specifications

Our trade data covers bilateral trade in goods and services for six broad NACE 1.1. sectors among 240 NUTS-2 European regions[23] from 25 European countries for the year 2010. This data set is a version of Thissen et al. (2013a,b) tailored for the use of the *RHOMOLO* model, which is developed by the JRC-IPTS, European Commission, see Brandsma et al. (2015). The IPTS kindly provided the data set for the present paper.
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 A to a region B are also imports of a region B coming from a region A). An additional feature that was taken into account are re-exports, see Lankhuizen and Thissen (2014).
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 data set thus comprises $240^2 = 57,600$ observations of intra- or interregional trade flows among European regions, after filtering out Romania, Bulgaria (trade flows data for both countries were obtained using gravity equations, their regions are therefore filtered out in order to avoid replication) and non-continental territories such as the Portuguese, French and Spanish islands.[24]
The variables are defined as follows: geographical distance between European regions is measured in terms of great circle distance between region's economic centres. Additionally, two alternatives of measuring geographical distance are tested against the baseline great circle distance measure ($d_{rs}^{(gcd)}$), namely the population-weighted geographical distance ($d_{rs}^{(pop)}$) as defined in Head and Mayer (2002) and the geographical distance approximated by travel time ($d_{rs}^{(tt)}$). The population-weighted geographic distance is defined in Head and Mayer (2002) 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:

$$d_{rs}^{(pop)} = \left(\sum_{k \in r} (pop_k/pop_r) \sum_{l \in s} (pop_l/pop_s) d_{kl}^{(gcd)} \right) \tag{32}$$

The travel time measure is obtained using the Google Maps API[25], computing the average travel time by car between two regions' economic centers.
The estimation of the wage equation relies on a second dataset, containing the previously constructed MA measures, data for control variables as well as for instruments of MA.
Data on regional GVA (as left-hand side of the (31) are taken from the Cambridge Econometrics Regional Database (2015). Further control variables in the wage equation control for human capital absorption capacity of a region, often referred to as a major determinant of income differentials (Krueger and Lindahl, 2001) as well as for labour market characteristics (such as unemployment rate, net replacement rate). Additionally, an index for product market regulation (PMR) tries to capture potential productivity-enhancing effects of less regulated markets (see Conway et al., 2005). At country level, the following variables are available: Share of tertiary education in

[23] The complete list of NUTS-2 European regions used in this study is provided in the Appendix.
[24] The full detailed list of NUTS-2 regions is available in the Appendix.
[25] `https://developers.google.com/maps/`

the population (Ter.Educ.), Number of issued patents in 2010 (Patents), Unemployment rate (Unemp.). Data for these variables are taken from the ESPON Database[26] (2013). At the country level, we use Product market regulation index and Net replacement rate (obtained from the Eurostat database, EMCO, 2014 and OECD Labour Market indicators respectively).

5 Results

5.1 Trade equation

The first stage of our analysis consists of the estimation of the parameter vector $\boldsymbol{\gamma}$ of the deterrence function and the corresponding market capacity proxy's parameter δ_2 of the trade equation (29) in order to construct MA measures as defined in (30). 240^2 regional bilateral trade flows are regressed on two vectors of 240 importer and 240 exporter dummies and on up to eight variables of trade barriers (i.e., geographical distance, country border, contiguity, language barrier, etc.). In order to stay as close as possible to the theoretical model, only trade flows on NACE 1.1 sector CDE (manufacturing) are used.
Facing a large number of candidate models, we tested several specifications of the deterrence function and ranked them according to their estimated log-likelihood and as well as to their information criterion. The deterrence function is construed to include a separation distance variable, i.e., geographical distance, and spatial separation factors depicting bilateral barriers conditional on regions' institutional, spatial or cultural characteristics. Comparing specifications, we chose a model that includes four spatial separation factors, namely geographical distance d_{rs}, a country border variable b_{rs}, country contiguity c_{rs} and language barrier l_{rs}. Complete results for all specifications tests are shown in Table 7 in the Appendix. Table 1 reports the estimates of the deterrence function components as well as the measure of residuals' autocorrelation (Moran's I) for the preferred model.
With the exception of geographical distance, all spatial (pairwise) separation factors are defined as binary variables taking the value of one or zero. Country border c_{rs} is equal to one if two regions r and s are separated by a country border. Contiguity measure takes a value of one if regions r and s share a common border (contiguous), zero otherwise. Finally language similarity l_{rs} is indicates whether regions' languages are similar (therefore equal to one), it is important to note that regional languages are taken into consideration.
The table includes results for the three alternative proxies of geographical distance. Column i) depicts estimates obtained with great circle distance (in km.) as proxy, column ii) population-weighted great circle distance (in km.) and column iii) travel time (in seconds). The significance of the results only slightly differs between the three specifications. Population-weighted great circle distance is seen as a larger barrier than its two counterparts (-0.814 compared to -0.796 and -0.781), country contiguity, however, seems to represent the most constraining burden to trade (ranging from -0.970 to -0.977). The border effect is about half of the geographical distance effect (ranging from -0.381 to -0.386). Overall, the trade equation seems to fit the data well with a reported average adjusted-R^2 of 0.850. Furthermore, the spatial autocorrelation of the residuals, as seen by the reported Moran's I, is drastically reduced with the addition of the spatial filters (average Moran's I is 0.032).

5.2 Wage equation

Out of the three proxies of geographical distance (great circle distance, population-weighted population distance and travel time) three MA measures are constructed according to (30).

[26] http://database.espon.eu/db2/

Table 1: OLS estimation of the trade equation for i) logged great circle distance, ii) logged population-weighted great circle distance, iii) logged travel time. Observations $n = 240 \times 240 = 57,600$

Distance proxy	i) *gcd*	ii) *pop*	iii) *tt*
Dependent Variable: Exports from r to s			
Great Circle Distance $d_{rs}^{(gcd)}$	-0.796^{***}		
	(0.011)		
Population Weighted $d_{rs}^{(pop)}$		-0.814^{***}	
		(0.011)	
Travel Time $d_{rs}^{(tt)}$			-0.781^{***}
			(0.011)
Border b_{rs}	-0.386^{***}	-0.385^{***}	-0.381^{***}
	(0.036)	(0.036)	(0.036)
Contiguity c_{rs}	-0.977^{***}	-0.974^{***}	-0.970^{***}
	(0.015)	(0.015)	(0.015)
Language l_{rs}	-0.157^{***}	-0.160^{***}	-0.151^{***}
	(0.028)	(0.028)	(0.028)
Adjusted R^2	0.865	0.866	0.865
logLik.	-75504.34	-75433.17	-75558.42
AIC	152534.7	152392.3	152642.8
Moran's I (filtered)	0.032	0.031	0.033
Moran's I (non filtered)	0.366	0.349	0.349

Notes All models include exporter (240) and importer (240) fixed effects. Model specification: (1) logged great circle distance, (2) logged population-weighted great circle distance, (3) logged travel time. $d_{rs}^{(gcd)}$: logged geographical distance (in km.), $d_{rs}^{(pop)}$ logged population-weighted geographical distance (in km.), $d_{rs}^{(tt)}$: logged distance travel time (in seconds), b_{rs}: 1 if separated by a country border, c_{rs}: 1 if share a common border, l_{rs}: 1 if different spoken languages. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

As a benchmark for further calculations and to see how well the theoretically derived MA performs, we also use the *ad-hoc* Harris market potential (HMP), as introduced in Harris (1954). The potential of region r is defined as the sum of gross value added of all its export destinations weighted by their remoteness relative to r:

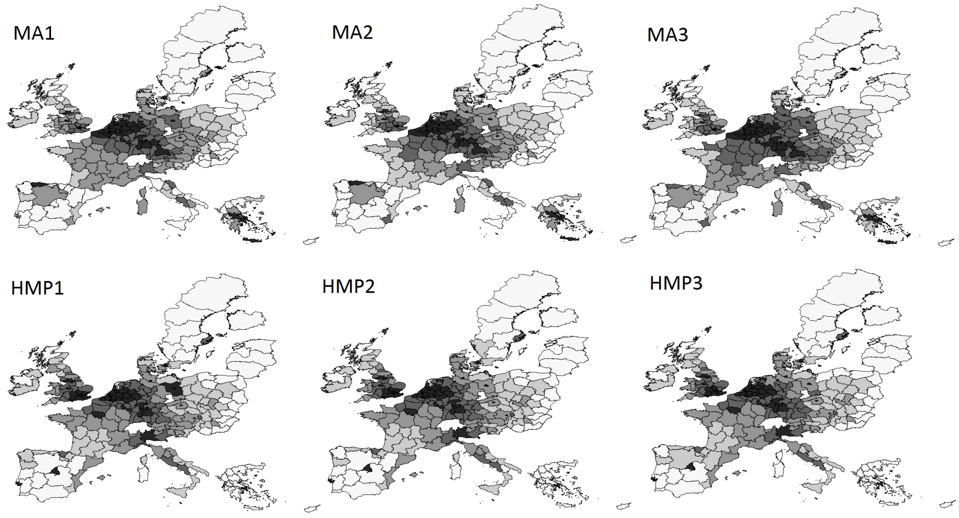
$$HMP_r = \sum_s W_{rs}.GVA_s \tag{33}$$

where W_{rs} is an element of a $n \times n$ spatial weight matrix $\boldsymbol{W}$ that can take different forms for modelling remoteness such as inverse geographical distance, Queen contiguity, k-nearest neighbours, etc.

Two major distinctions between MA (as defined in (30)) and HMP are: First, the neighbourhood structure in HMP is defined in an *ad-hoc* way, while it depends on estimated trade costs in the case of MA. Second, HMP imposes the restriction that market capacity is only composed by μE_s, thus setting $P_s^{\sigma-1}$ equal to one. In other words, there is no price competition assumed. The literature sometimes distinguishes "nominal market potential" when the price is equal to one and "real market potential" when the price index is taken into account.

A set of HMP measures are obtained by using different definition of spatial weight matrices: inverse geographical distance proxied by great-circle distance, population-weighted great circle distance and travel time (respectively dubbed *gcd*, *pop* and *tt*

Figure 2: MA and HMP measures for 229 regions



(MA1/HMP1): measure of MA (or HMP) obtained by using *gcd* as proxy of geographical distance, (MA2/HMP2): with *pop* as proxy of geographical distance, (MA3/HMP3) with *tt* as proxy of geographical distance.
Darker shades of grey correspond to higher distribution quantiles of the MA or HMP measures.

henceforth).[27] The best performing HMP based on different definition of the spatial weight matrix is chosen based on information criterion and on the log-likelihood obtained from the estimation of the wage equation. Results are, however, available for all possible HMP candidates in Table 8 in the Appendix.
Constructed MA measures are illustrated in Fig. 2 on a European regional map. The three top panels $MA1$, $MA2$ and $MA3$ plot constructed MA on the basis of estimated partner fixed effect and deterrence function with *gcd* as geographical distance proxy, *pop* as proxy and *tt* as proxy respectively. The three lower panels plot HMP. A clear core can be seen in south England, Benelux and through large parts of Germany, a semi-periphery for French regions as well as for central eastern European Countries and a periphery for extreme south and extreme north regions. When comparing the six measures, neither MAs nor HMPs show large discrepancies, with perhaps the exception of Greek regions being advantaged in the MA and capital regions such as Ile-de-France, Madrid and Lisbon (as easily seen by the darkest shades) having larger HMP values than MA. The core is also more concentrated in the MAs than in the HMPs. Table 2 additionally gives an overview of the top ten regions with highest MA and underline the relative similarity of the three measures. As seen in the table, only minor changes are observable within and between measures.
Empirically, the wage equation (31) is estimated by proxying wages w_r by regional GVA per capita weighted by (national) average working hours, using the generated measure of MA ($\widehat{MA}_r$) and adding a set $\boldsymbol{X}$ of control variables and/or country fixed effects. Thus, the empirical wage equation is written as:

$$\ln(GVA_r^{(p.c.)}) = \gamma + \varphi \ln(\widehat{MA}_r) + \boldsymbol{X}'\boldsymbol{\theta} + \varepsilon_r \tag{34}$$

An overview of the relationship between the (logged) MA (hence $\overline{MA}$) and (logged) GVA per capita weighted by average working hours is given by the following estima-

[27] Further tests have been done using Queen contiguity and k-nearest neighbours (for $k = 2, \ldots, 10$) spatial weight matrices for the computation of HMP. Information criterion extracted from the estimation of the wage equation indicate that HMPs constructed with inverted geographical distance spatial weight matrices (with *gcd*, *pop* or *tt* as proxy) achieved better fit and more reliable estimates than any other matrices (k-2 to k-10 nearest neighbours, Queen contiguity spatial weight matrices).The results are not shown in this working paper but are available upon request.

Table 2: Top 10 Regions' MA and HMP (*gcd*,*pop*,*tt*)

MA	*gcd*	*pop*	*tt*	HMP	*gcd*	*pop*	*tt*
1	UKI1	UKI1	UKI1	1	UKI1	UKI1	UKI1
2	BE10	BE10	BE10	2	BE10	BE10	BE10
3	UKI2	EL42	EL42	3	AT13	AT13	AT13
4	EL42	UKI2	UKI2	4	UKI2	DE60	DE60
5	DE30	AT13	AT13	5	DE30	DE30	DE30
6	AT13	DE30	DE30	6	DE60	UKI2	UKI2
7	DE60	DE60	DE60	7	CZ01	CZ01	CZ01
8	FI20	FI20	FI20	8	UKG3	UKG3	UKG3
9	DE50	UKG3	DE50	9	DE50	DE50	DE50
10	UKG3	DE50	UKG3	10	FR10	FR10	FR10
Mean Spearman correlation							
	0.917	0.916	0.894		0.913	0.909	0.910

NUTS-2 Classification Names: AT13 Vienna, BE10 Brussels-Capita Region, CZ01 Prague, DE30 Berlin, DE50 Bremen, DE60 Hamburg, EL42 South Aegean, FI20 Helsinki, FR10 Ile-de-France, UKG3 West Midlands, UKI1 Inner London West, UKI2 Inner London East

tion:

$$\begin{aligned}\ln(GVA_r^{(p.c.)}) =& 0.645^{\star\star\star} + 0.509^{\star\star\star}\ln(\widehat{MA}_r) + \varepsilon_r\\ & [0.131] \quad\quad [0.074]\\ R^2 =& 0.228\\ MI =& 0.687^{\star\star\star}\end{aligned}$$

where $\widehat{MA}$ is the mean value of the three constructed MA. The values in square brackets are bootstrapped standard errors. The positive effect of an increase in a region's MA is clearly observable, however, the regressor alone does not account for spatial autocorrelation of regional income per capita as seen by the significant high Moran's I measure (MI). The same observations can be made when HMP is used as a regressor, this variable seems to explain regional income per capita variance to a larger extent than MA and fits the data better:

$$\begin{aligned}\ln(GVA_r^{(p.c.)}) =& -4.155^{\star\star\star} + 0.379^{\star\star\star}\ln(\widehat{HMP}_r) + \varepsilon_r\\ & [0.400] \quad\quad [0.039]\\ R^2 =& 0.330\\ MI =& 0.710^{\star\star\star}\end{aligned}$$

We then augment the model with factors controlling for a region's human capital absorption capacity, labour characteristics as well as cross-country variations (in the form of fixed effects or country-level factors). Regional share of tertiary education (Ter.Educ.) and the number of patents issued in 2010 (Patents) control for labour productivity as well as accumulation of human capital that helps expand production and income (Krueger and Lindahl, 2001). The unemployment rate (Unemp.) and Net Replacement Rate (NRR) account for regional labour characteristics that drive incomes down and up respectively. Product market regulation(PMR) controls for the regulatory environment. More specifically, the coefficient estimate for PMR is expected to be negative, as stronger burdens to competition prevent efficient of resources and therefore may harm productivity that in turns affect income negatively (Conway et al., 2005). Country cross-variations are captured either through control variables at the country level (PMR, NRR) or through country fixed effects. The specifications are defined as follows:

W1: Market access + Country fixed effects

W2: Market access + Ter.Educ.+ Patents + Unemp. + PMR + NRR

W3: Market access + Ter.Educ.+ Patents + Unemp. + Country fixed effects

Altogether, for the wage equation a total of $3 * 3 = 9$ models (three proxies for geographical distance and three wage equation specifications) are tested.[28]. However, since results between specifications are very similar, we restrict our discussion to the MA constructed with population-weighted great circle distance (*pop*).
As mentioned in former studies, the wage equation suffers from endogeneity issues which lead to inconsistent estimates and therefore incorrect conclusion about the estimated effect of MA on income. The departure from the assumptions of the classical linear model can be addressed by instrumenting the endogenous regressors (MA and HMP), following Redding and Venables (2004), Head and Mayer (2006), Hering and Poncet (2009). All instruments reflect geographical features of European regions such as their interconnections, distance to the most central European region (Luxembourg, following Redding and Venables (2004) and Breinlich (2006)) and mean inverted distance to importing partners (Bruna et al., 2015). Additionally, taking into account that MA and HMP are constructed regressors, bootstrapped standard errors are reported. However, for models including country fixed effects (specifications W1 and W3) White robust standard errors are instead used.[29]
Table 3 presents the estimation results for the three model specifications of the wage equation by two stages least squares (2SLS), instrumenting generated MA as well as HMP measures. The upper panel shows the results of the second stage, i.e., the instrumented wage equation and the lower panel the first stage, i.e., the instruments are regressed on MA and HMP. Model selection criteria clearly advocate the use of country fixed effects and regional control variables in the wage equation. The model that includes only country fixed effects also performs better than specification W2: variables aiming at controlling for cross-country variations do not seem to suffice.
The first three columns report results with MA a regressor and the last three columns HMP. Both variable are instrumented by mean inverted distance to importers (MIDIP) and distance to Luxembourg (DistLux). Altogether, the sample consists of 240 European regions, but due to missing data in the variables Patents and Unemployment (for 2010) the sample is reduced to 229. Data availability at the regional level is restricted to only the share of tertiary education and issued patents. The Wu-Hausman test of exogeneity rejects the null hypothesis of the instrumented constructed MA being exogenous at the 0.1% level for models including fixed effects. This is, however, not the case for HMP. For models iv)-vi) instrumented variable regressions could not be proven superior to OLS. Instruments for the 2SLS specification are weighted inverted distance to importing partners (+) and distance to Luxembourg (-). Inclusion of the instruments seem satisfactory as seen by the Hansen-J test (all were not able to reject the null hypothesis of valid over-identification restrictions).
The estimate for MA remains significant throughout the three models and is equal to +0.337 on average, which is about 66% higher than the mean HMP coefficient: +0.202. Human capital, proxied here by the regional share of tertiary educated people and the total number of issued patents in 2010, is a major determinant of regional income (both variables could be included since their correlation is relatively small: 0.300). However, only the share of tertiary educated people shows robust link with income per capita. The coefficient for the Patents variable is not statistically different from zero when controlling for country heterogeneity. As expected, unemployment is a strong impediment to high regional wages and NRR is highly correlated with high income per capita. Regarding PMR, competition friendly countries (with small values) benefit from higher per capita income. No large discrepancies are observed

[28]Table 6 in the Appendix summarizes the list of different deterrence function version as well as the wage equation specifications used in our empirical exercise.

[29]In these cases bootstrap replications were not possible due to a too small number of observations.

Table 3: Estimation of the instrumented Wage Equation

Specification	i) W1	ii) W2	iii) W3	iv) W1	v) W2	vi) W3
Second Stage (2SLS)						
Dependent Variable: GVA_{pc} weighted by Average Working Hours						
MA	0.447***	0.226***	0.339***			
	(0.051)	[0.069]	(0.042)			
HMP				0.246***	0.157***	0.203***
				(0.024)	[0.048]	(0.023)
Ter.Educ		0.310***	0.516***		0.237***	0.330***
		[0.075]	(0.063)		[0.081]	(0.066)
PMR		−0.516***			−0.391**	
		[0.108]			[0.126]	
Patents		1.295***	0.290		1.279***	0.291
		[0.276]	(0.168)		[0.240]	(0.153)
Unemp.		−0.957*	−1.582***		−1.175*	−1.789***
		[0.502]	(0.310)		[0.516]	(0.283)
NRR		0.842***			0.885***	
		[0.090]			[0.087]	
Obs.	240	229	229	240	229	229
RMSE	0.217	0.304	0.160	0.184	0.298	0.146
adj R^2	0.839	0.688	0.913	0.885	0.699	0.928
logLik.	23.369	−47.313	89.132	67.775	−36.732	132.366
AIC	−2.816	−2.363	−3.380	−3.186	−2.455	−3.758
Country FE	Yes	No	Yes	Yes	No	Yes
Moran's I	0.114***	0.469***	0.040	0.175***	0.509***	0.044*
(p-value)	0.000	0.000	0.119	0.000	0.000	0.098
Weak inst.	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***
Wu-Hausmann	0.000***	0.502	0.001***	0.658	0.521	0.698
Hansen-J test	0.519	0.982	0.874	0.351	0.937	0.458
First Stage (2SLS)						
Dependent Variable: MA or HMP						
MIDIP		0.526***			0.860***	
		(0.032)			(0.046)	
DistLux		−0.127***			−0.196***	
		(0.026)			(0.037)	
RMSE		0.312			0.446	
adj R^2		0.648			0.703	

Second Stage: All models include a constant and country fixed effects. Ter.Educ: regional share of tertiary education, PMR: product market regulation index (country level), Patents: number of patents issued in 2010, Unemp.: regional unemployment rate, NRR: net replacement rate (country level). MA and HMP are instrumented using MIDIP and DistLux.
First Step: All models include a constant. MIDIP: Mean inverted distance from importing partners, DistLux: Distance from Luxembourg. Robust standard errors in parentheses. Bootstrapped standard errors (500 replications) in square brackets.
*** p<0.01

within models with MA and models with HMP.
As robustness checks, estimations of the wage equation were also performed using an even larger set of variables on labour market characteristics, i.e. tax wedge (the difference between before-tax and after-tax wages), government expenditure on labour market and the number of people involved in labour market policies. The results are very similar to those shown in Table 3, the coefficient estimate of MA remains positive (0.331 on average) and data fit (without country fixed effects) is equal to 0.62.
Changes in coefficients along the three specifications are discernible for MA and HMP, both lose in magnitude when country fixed effects are excluded and the largest significant value is obtained from the first specification W1.
An important result is also the spatial autocorrelation measure of the residuals. It indicates that not all income per capita autocorrelation is captured by the regressors. Models ii) and v) yield positive and strongly significant Moran's I that approximates those found in the literature. This indicates that MA (or HMP) alone do not fully control for the spatial pattern of income per capita. However, inclusion of country fixed effects drastically reduces the measure of autocorrelation. More specifically, using regional control variables as well as country fixed effects (W3 for models iii) and vi)) totally accounts for these spatial patterns.
In order to draw a first conclusion, the theoretical measure of regions' proximity to markets (MA) plays a significant role in explaining the uneven distribution of income between European regions. Furthermore, we showed that results are comparable to those obtained with *ad-hoc* HMP measures. What strongly matters, thus supporting the findings of Bosker and Garretsen (2010), is especially the definition of the wage equation and of its components as well as the choice of the deterrence function.[30].
More specifically, as already noted in Bruna et al. (2015), the addition of control variables at the regional level combined with country fixed effects helped partly resolve the issue of spatially autocorrelated residuals. This further strengthens our results as they should be more unbiased.

5.3 A Spatial Wage Equation

Although the estimations of the wage equation have brought satisfying results regarding data fit, there is spatial autocorrelation left when country fixed effect are excluded (only one Moran's I out of six is insignificant). This might raise concerns that MA as well as the controls for regional and country characteristics are not fully able to capture the spatial pattern of income per capita. As a robustness check, we remodel the wage equation by adding a spatial lag component, fully controlling for the underlying spatial autocorrelation of the regressand. A classic spatial autocorrelation regression (dubbed SAR henceforth) is used and is defined as follows in matrix notation:

$$\ln(\boldsymbol{GVA}^{(p.c.)}) = \rho \boldsymbol{W} \cdot \ln(\boldsymbol{GVA}^{(p.c.)}) + \varphi \ln(MA) + \boldsymbol{X'\theta} + \boldsymbol{u} \qquad (35)$$

where $\boldsymbol{W}$ is a 240×240 spatial weight matrix used to control for spatial autocorrelation in the regressand and ρ is its related coefficient. As usual in the literature, the $\boldsymbol{W}$ is defined as a $k = 5$ nearest neighbours spatial weight matrix. Parameters are obtained using Maximum Likelihood estimation (ML Estimation).
A positive and significant $\hat{\varphi}$ even after adding a spatial lag component should provide a hint of robustness for the link between both MA measures and income per capita. The parameter of the SAR model cannot be interpreted directly, so we focus on the direct, indirect and total effects that are reported in Table 4. For all six models, the spatial autocorrelation parameter ρ is positive significant. The point estimate is low when country fixed effects are included (0.260 on average) and averaging 0.600 without. This indicates that even after controlling for regions' proximity to markets

[30] Additionally the form of deterrence function (logged geographical distance) and the choice of estimators (OLS vs. PPML) also have great impacts on the results. Those results are partially shown in Appendix and are also available upon request.

Table 4: ML Estimation of the SAR version of wage equation

Specification		MA W1	MA W2	MA W3	HMP W1	HMP W2	HMP W3
ρ		0.249*	0.589*	0.227*	0.308*	0.612*	0.258*
		(0.057)	(0.057)	(0.058)	(0.053)	(0.053)	(0.053)
MA (*pop*)	Direct	0.307*	0.172*	0.234*			
	Indirect	0.097*	0.211*	0.066*			
	Total	0.404*	0.383*	0.300*			
HMP	Direct				0.253*	0.174*	0.192*
	Indirect				0.106*	0.234*	0.061*
	Total				0.359*	0.408*	0.253*
Ter.Educ.	Direct		0.306*	0.603*		0.219*	0.397*
	Indirect		0.378*	0.170*		0.296*	0.125*
	Total		0.685*	0.773*		0.516*	0.523*
PMR	Direct		−0.116			0.070	
	Indirect		−0.143			0.102	
	Total		−0.259			0.178	
Patents	Direct		0.714*	0.312*		0.569*	0.236$^+$
	Indirect		0.882*	0.088$^+$		0.767*	0.074
	Total		1.596*	0.400$^+$		1.336*	0.311$^+$
Unemp.	Direct		−0.407	−1.092*		−0.565$^+$	−1.257*
	Indirect		−0.502	−0.308*		−0.760	−0.398*
	Total		−0.910	−1.401*		−1.325	−1.656*
NRR	Direct		0.366*			0.378*	
	Indirect		0.452*			0.509*	
	Total		0.819*			0.888*	
Moran's test		0.724	0.000*	0.908	0.074$^+$	0.000*	0.802
AIC		−49.962	16.391	−184.220	−130.910	−7.372	−221.050
Country FE		Yes	No	Yes	Yes	No	Yes

Notes: All models include a constant and country fixed effects. Direct, indirect and total effect obtained out of $MCMC$ 1000 simulations. * p<0.05, $^+$ p<0.1

and for human capital there still remains some unexplained spatial patterns in the dependent variable. The addition of this spatial lag parameter successfully removed spatial autocorrelation in the residuals for the models including country fixed effects (insignificant Moran test), but does not seem to suffice for the second specification W2.

On the side of the regressors, total effects are comparable to results shown in the previous section: MA and HMP still have a positive impact on regional income per capita whereas the latter has relatively smaller coefficients (10% smaller than those of the MA). A high regional share of tertiary education remains an important explanatory factor of regional incomes per capita, in contrast to unemployment. There are no notable differences in the effects of the number of issued patents.

Looking at the direct and indirect effects of the regressors yields interesting new findings. The indirect effect, illustrating neighbours MA (or HMP) effect on one's regional income per capita, is positive and in two cases out of six, larger than the direct effect. Human capital also shows a positive significant indirect effect, indicating that regions with high income are very likely to be located next to other high income regions as well as regions with high share of tertiary education and patents issuance. Domestic unemployment rate shows no effect on the regional distribution of income when net replacement rate is included, however, the spatial correlation becomes significant when controlling for fixed effects instead.

In conclusion, controlling for regressand spatial autocorrelation pattern yields similar results that are comparable with previous results using 2SLS. This further indicates that market access is a robust and important determinant of regions income per capita. Concerning the differences between MA and HMP, the latter still provides the best fit as seen by the log-likelihood and information criterion (AIC), as already shown in the previous section.

6 Concluding remarks

Using a spatially filtered gravity estimation of regional bilateral trade flows and testing for different specifications of Fujita's wage equation, we showed that proximity to large export markets is a robust determinant of a region's per capita income. In other words, better access to consumer markets increases factor rewards. Our theory-consistent measure of market access is comparable to the *ad-hoc* specification of the Harris market potential and even slightly outperforms it when explaining spatial income distribution. Using regional data permitted us to control for various determinants of cross-regional income differences such as knowledge creation (as seen by the number of patents issued in a region), education level and a region's unemployment rate, on a more disaggregated level than in previous studies. Moreover, testing for different measures of market access (relying on distinct proxies for geographical distance and deterrence function specifications) as well as addressing endogeneity issues with instrument variables confirmed once more that specification matters although the overall strength of the link between market access and regional income per capita were not significantly impacted.

This provides a solid basis for extensions and further investigation that is also called for in the literature. First, a test whether the proposed effect of market access on income differs between the six available NACE sectors would be of interest. Second, alternative channels through which market access affects income could be identified. One example is the higher incentives for (human-) capital formation in agglomerated areas. This leads to additional robustness check, namely on the effect of alternative sources of agglomeration on income that might get captured by our measure of market access. The possible contribution of location fundamentals (e.g., physical geography or infrastructure) on income differences has to also be accounted for.

Ultimately the question is of high interest from a regional policy point of view. Further market integration in Europe could lead to higher income as a result of more agglomeration but also to a greater divergence within Europe and also within countries or even regions. Regional Policy aims at income convergence and a catching-up process for the European periphery. More research is needed to understand the possible trade-offs that further integration can have for Europe and its regions.

Acknowledgements

We would like to thank the Institute for Prospective Technologies Studies in Seville for kindly providing us the 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 and at the FIW 2015 conference (Vienna).

References

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

Amiti, M. and Smarzynska Javorcik, B. (2008). Trade costs and location of foreign firms in China. *Journal of Development Economics*, 85(1-2):129–149.

Anderson, J. E. and van Wincoop, E. (2003). Gravity with Gravitas: A Solution to the Border Puzzle. *American Economic Review*, 93(1):170–192.

Anselin, L. (1988). *Spatial econometrics: methods and models*, volume 4. Springer Science & Business Media.

Bachtrögler J., Badinger H., Fichet de Clairfontaine A., Reuter W.H. (2014). Summarizing data using partially ordered set theory: An application to fiscal frameworks in 97 countries. WU Economics Working Paper, No. 181.

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

Barro, R. and Sala-i Martin, X. (1997). Technological diffusion, convergence, and growth. *Journal of Economic Growth*, 2(1):1–26.

Beestermöller, M. and Rauch, F. (2014). A dissection of trading capital: Cultural persistence of trade in the aftermath of the fall of the iron curtain. Technical report, Munich Discussion Paper.

Behrens, K., Ertur, C. and Koch, W. (2012). Dual Gravity: Using Spatial Econometrics to Control for Multilateral Resistance Journal of Applied Econometrics, 27(5):773–794

Bolduc, D., Laferriere, R., and Santarossa, G. (1995). *Spatial autoregressive error components in travel flow models: an application to aggregate mode choice.* Springer.

Bosker, M., Brakman, S., Garretsen, H., and Schramm, M. (2010). Adding geography to the new economic geography: bridging the gap between theory and empirics. *Journal of Economic Geography*, 10(6):793–823.

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

Brandsma, A., Kancs, d'A., Monfort, P. and Rillaers, A. (2015) RHOMOLO: A dynamic spatial general equilibrium model for assessing the impact of cohesion policy *Papers in Regional Science*, 94:197–221.

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., Faía, A., and Lopez-Rodriguez, J. (2014). Market Potential and the curse of distance in European regions.

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

Burger, M., Van Oort, F., and Linders, G.-J. (2009). On the specification of the gravity model of trade: zeros, excess zeros and zero-inflated estimation. *Spatial Economic Analysis*, 4(2):167–190.

Cambridge Econometrics European Regional Database (2015). `http://www.camecon.com/SubNational/SubNationalEurope/RegionalDatabase.aspx`

Chun, Y. (2008). Modeling network autocorrelation within migration flows by eigenvector spatial filtering. *Journal of Geographical Systems*, 10(4):317–344.

Chun, Y. and Griffith, D. A. (2011). Modeling network autocorrelation in space–time migration flow data: an eigenvector spatial filtering approach. *Annals of the Association of American Geographers*, 101(3):523–536.

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.

Crozet, M. (2004). Do migrants follow market potentials? an estimation of a new economic geography model. *Journal of Economic Geography*, 4(4):439–458.

Dixit, A.K. and Stiglitz, J.E. (1977) Monopolistic Competition and Optimum Product Diversity. *The American Economic Review*, 67(3):297–308

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.

Fischer, M. M. and Griffith, D. A. (2008). Modeling spatial autocorrelation in spatial interaction data: An application to patent citation data in the european union. *Journal of Regional Science*, 48(5):969–989.

Fischer, M. M., Scherngell, T., and Jansenberger, E. (2006). The geography of knowledge spillovers between high-technology firms in europe: Evidence from a spatial interaction modeling perspective. *Geographical Analysis*, 38(3):288–309.

Forslid, R. and Ottaviano, G.I.P. (2002) An analytically solvable core-periphery model. *Journal of Economic Geography*, 3(3):229–240

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.*

Griffith, D. A. and Fischer, M. M. (2013). Constrained variants of the gravity model and spatial dependence: model specification and estimation issues. *Journal of Geographical Systems*, 15(3):291–317.

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. (2004a). Chapter 59 The empirics of agglomeration and trade. In J. Vernon Henderson and Jacques-Franois Thisse, editor, *Handbook of Regional and Urban Economics*, volume Volume 4 of *Cities and Geography*, pages 2609–2669. Elsevier.

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. (2011). Gravity, market potential and economic development. *Journal of Economic Geography*, 11(2):281–294.

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.

Helpman, E., Melitz, M., and Rubinstein, Y. (2007). Estimating trade flows: Trading partners and trading volumes. Technical report, National Bureau of Economic Research.

Hering, L. and Poncet, S. (2009). The impact of economic geography on wages: Disentangling the channels of influence. *China Economic Review*, 20(1):1–14.

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

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.

Krugman, P. and Venables, A. J. (1995). Globalization and the Inequality of Nations. *The Quarterly Journal of Economics*, 110(4):857–880.

Lankhuizen, M. and Thissen, M. (2014). *Identifying true trade patterns: correcting bilateral trade flows from re-exports.* ERSA Working Paper.

LeSage, J. and Pace, R. K. (2010). *Introduction to spatial econometrics.* CRC press.

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.I.P. (2002) Agglomeration and Trade Revisited. *International Economic Review*, 43(2):409–435

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.

Pace, R. K., Lesage, J. P., and Zhu, S. (2013). Interpretation and computation of estimates from regression models using spatial filtering. *Spatial Economic Analysis*, 8(3):352–369.

Puga, D. (1999) The rise and fall of regional inequalities. *European Economic Review*, 43(2):303–334.

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

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.

Sellner, R., Fischer, M. M., and Koch, M. (2013). A spatial autoregressive Poisson gravity model. *Geographical Analysis*, 45(2):180–201.

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.

Tiefelsdorf, M. and Griffith, D. A. (2007). Semiparametric filtering of spatial autocorrelation: the eigenvector approach. *Environment and Planning A*, 39(5):1193.

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: 575605.

Westerlund, J. and Wilhelmsson, F. (2011). Estimating the gravity model without gravity using panel data. *Applied Economics*, 43(6):641–649.

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 240 NUTS-2 regions located in a subset of the EU-28 member states (excluding Croatia, Romania and Bulgaria). We exclude the Spanish North African territories of Ceuta y Melilla, the Portuguese non-continental territories Açores and Madeira, and the French Departments d'Outre-Mer Guadeloupe, Martinique, Guyane Française and Réunion. Thus, we include the following NUTS-2 regions:

Austria	Burgenland, Kärnten, Niederösterreich, Oberösterreich, Salzburg, Steiermark, Tirol, Vorarlberg, Wien
Belgium	Prov. Antwerpen, Prov. Brabant-Wallon, Prov. Hainaut, Prov. Limburg (B), Prov. Liège, Prov. Luxembourg (B), Prov. Namur, Prov. Oost-Vlaanderen, Prov. Vlaams-Brabant, Prov. West-Vlaanderen, Région de Bruxelles-Capitale/Brussels Hoofdstedelijk Gewest
Czech Republic	Jihovýchod, Jihozápad, Moravskoslezsko, Praha, Severovýchod, Severozápad, Stredni Morava, Stredni Cechy
Denmark	Hovedstaden, Sjaelland, Syddanmark, Midtjylland, Nordjylland
Estonia	Eesti
Finland	Aland, Etelä-Suomi, Itä-Suomi, Länsi-Suomi, Pohjois-Suomi
France	Alsace, Aquitaine, Auvergne, Basse-Normandie, Bourgogne, Bretagne, Centre, Champagne-Ardenne, Corse, Franche-Comté, Haute-Normandie, Île-de-France, Languedoc-Roussillon, Limousin, Lorraine, Midi-Pyrénées, Nord-Pas-de-Calais, Pays de la Loire, Picardie, Poitou-Charentes, Provence-Alpes-Côte d'Azur, Rhône-Alpes

Germany	Arnsberg, Berlin, Brandenburg, Braunschweig, Bremen, Chemnitz, Darmstadt, Dessau, Detmold, Dresden, Düsseldorf, Freiburg, Giessen, Halle, Hamburg, Hannover, Karlsruhe, Kassel, Koblenz, Köln, Leipzig, Lüneburg, Magdeburg, Mecklenburg-Vorpommern, Mittelfranken, Münster, Niederbayern, Oberbayern, Oberfranken, Oberpfalz, Rheinhessen-Pfalz, Saarland, Schleswig-Holstein, Schwaben, Stuttgart, Thüringen, Trier, Tübingen, Unterfranken, Weser-Ems
Greece	Anatoliki Makedonia, Thraki, Attiki, Ipeiros, Voreio Aigaio, Dytiki Ellada, Dytiki Makedonia, Thessalia, Ionia Nisia, Kentriki Makedonia, Kriti, Notio Aigaio, Peloponnisos, Sterea Ellada
Hungary	Dél-Alföld, Dél-Dunántúl, Észak-Alföld, Észak-Magyarország, Közep-Dunántúl, Közep-Magyarország, Nyugat-Dunántúl
Ireland	Border Midland and Western, Southern and Eastern
Italy	Abruzzo, Basilicata, Calabria, Campania, Emilia-Romagna, Friuli-Venezia Giulia, Lazio, Liguria, Lombardia, Marche, Molise, Piemonte, Puglia, Sardegna, Sicilia, Toscana, Trentino-Alto Adige/Südtirol, Umbria, Valle d'Aosta/Vallée d'Aoste, Veneto
Latvia	Latvija
Lithuania	Lieteva
Luxembourg	Luxembourg (Grand-Duché)
Netherlands	Drenthe, Flevoland, Friesland, Gelderland, Groningen, Limburg (NL), Noord-Brabant, Noord-Holland, Overijssel, Utrecht, Zeeland, Zuid-Holland
Malta	Malta
Poland	Dolnoślaskie, Kujawsko-Pomorskie, Lubelskie, Lubuskie, Lódzkie, Mazowieckie, Malopolskie, Opolskie, Podkarpackie, Podlaskie, Pomorskie, Ślaskie, Świetokrzyskie, Warmińsko-Mazurskie, Wielkopolskie, Zachodniopomorskie
Portugal	Alentejo, Algarve, Centro (P), Lisboa, Norte
Slovakia	Bratislavsky Kraj, Stredné Slovensko, Východné Slovensko, Západné Slovensko
Slovenia	Vzhodna Slovenija, Zahodna Slovenija
Spain	Andalucía, Aragón, Cantabria, Castilla y León, Castilla-La Mancha, Cataluña, Comunidad Foral de Navarra, Comunidad Valenciana, Comunidad de Madrid, Extremadura, Galicia, Islas Baleares, La Rioja, País Vasco, Principado de Asturias, Región de Murcia
Sweden	Mellersta Norrland, Norra Mellansverige, Smaland med Öarna, Stockholm, Sydsverige, Västsverige, Östra Mellansverige, Övre Norrland
United Kingdom	Bedfordshire & Hertfordshire, Berkshire, Buckinghamshire & Oxfordshire, Cheshire, Cornwall & Isles of Scilly, Cumbria, Derbyshire & Nottinghamshire, Devon, Dorset & Somerset, East Anglia, East Riding & North Lincolnshire, East Wales, Eastern Scotland, Essex, Gloucestershire, Wiltshire & North Somerset, Greater Manchester, Hampshire & Isle of Wight, Herefordshire, Worcestershire & Warkwickshire, Highlands and Islands, Inner London, Kent, Lancashire, Leicestershire, Rutland and Northamptonshire, Lincolnshire, Merseyside, North Eastern Scotland, North Yorkshire, Northern Ireland, Northumberland and Tyne and Wear, Outer London, Shropshire & Staffordshire, South Western Scotland, South Yorkshire, Surrey, East & West Sussex, Tees Valley & Durham, West Midlands, West Wales & The Valleys, West Yorkshire

Table 6: HMPs, Deterrence function and Wage equation specifications

First Step – Construction of MA measures
Deterrence function specification
D1: $D_{rs} = d_{rs}^{\gamma_1} \exp(\sum_{p=1}^{P-1} b_{rs}^{p} * \gamma_2^{p} + c_{rs} * \gamma_3 + l_{rs} * \gamma_4)$ D2: $D_{rs} = d_{rs}^{\gamma_1} \exp(b_{rs} * \gamma_2 + c_{rs} * \gamma_3 + l_{rs} * \gamma_4)$ D3: $D_{rs} = d_{rs}^{\gamma} \exp(b_{rs} * \gamma_2 + c_{rs} * \gamma_3 + l_{rs} * \gamma_4 + cap * \gamma_5 + s * \gamma_6 + r * \gamma_7 + e * \gamma_8 + i * \gamma_9)$
Second Step – Estimation of the Wage Equation
Wage Equation Specification
W1: $\ln(GVA_{pc}) \sim \ln(MA) + CFE$ W2: $\ln(GVA_{pc}) \sim \ln(MA) + Ter.Educ. + PMR + Patents + Unemp. + NRR$ W3: $\ln(GVA_{pc}) \sim \ln(MA) + Ter.Educ. + Patents + Unemp. + CFE$

Notes: d geographical distance, b country border, c country contiguity, l language barrier, cap capital region, s access to the sea, r region contiguity, e East border, i island, CFE country fixed effects, MA market access, $Ter.Educ.$ Share of tertiary education, PMR Product Market Regulation, $Patents$ Number of Patents issued in 2010, $Unemp$ Unemployment rate, NRR Net Replacement Rate

A - Results of three specifications of the deterrence function D_{rs}

Table 7 reports the parameter estimates from the trade equation obtained using i) three deterrence function specification (listed below), as well as three proxies of geographical distance (*gcd*, *pop* and *tt*). The three specifications are:

D1: Geographical distance + 25 Country border(s)

D2: Geographical distance + Country border + Country contiguity + Language barrier

D3: Geographical distance + Country border + Country contiguity + Language barrier + Capital region + Access to the sea + Region contiguity + Eastern countries border + Island

The first specification D1, listed above, includes a country border variable for each country (therefore 25 country border variables), equal to one if a specific country border is crossed and zero otherwise. Only one country border variable, equally defined for all the countries, is included in the two remaining specifications D2 and D3. The variable is equal to one if two regions are separated by a country border. Contiguity measure is proxied by a region dummy variable taking a value of one if the regions share a common border, zero otherwise. Analog to the contiguity measure, language similarity is a dummy variable that takes a value of one if the regions are not located in the same language area (also taking into account regional specificities and language minorities). Capital region accounts for the presence of the country's capital in at least one of the two regions. Access to the sea is likewise equal to one if at least one region has access to the sea. Region contiguity is defined as equal to one if both regions share a common border. Eastern countries border takes the value of one if one of the two regions is located in an Eastern European country: Czech Republic, Slovakia, Slovenia, Poland, Lithuania, Latvia and Hungary. Finally, Islands is a binary variable equal to one if at least one of the two regions is located on a island (only the case for regions of the UK, Malta and Ireland).

The most constraining barrier seems to be country contiguity, with on average -0.964. It is followed by geographical distance with on average -0.827 for great circle distance, -0.843 for population-weighted great circle distance and -0.805 for travel time. The value of the coefficient estimate of geographical distance is strongly significant for three specification and independent of the proxies used, however their differences is only statistically significant at the 5% level when comparing D1 to D2 and/or D3. The least constraining barrier according to the OLS estimates is the landlocked characteristic of a region (the other case of the variable access to the sea s_{rs}) with a mean negative coefficient equal to 0.087.

The trade equation fits the data relatively well as seen by the adjusted R^2 that is not inferior than 0.86. Furthermore, controlling for origin- and destination dependencies in the flows has brought some positive results since the Moran's I of the residuals (measure the extent to which the models' residuals are spatially autocorrelated) have been drastically reduced (10 times smaller than without filters).

The most striking results from the performed estimations are that i) choice of proxies of geographical distance does not seem to matter since coefficient estimates are not statistically different (at the 5% and especially at the 10% level) and ii) the specification of the trade deterrence only matters for the negative effect of geographical distance on trade, being a large impediment to exports when only country borders are taken into account and a less of a barrier when other spatial factors such as language similarity or country contiguity are added to the model.

Table 7: OLS estimates of the trade equation for three deterrence function's specifications and three proxies of geographical distance

		gcd			*pop*			*tt*	
Spec.:	D1	D2	D3	D1	D2	D3	D1	D2	D3
log(*greatCircDist*)	−0.865*	−0.797*	−0.820*						
	(0.012)	(0.012)	(0.012)						
log(*pop.dist*)				−0.884*	−0.815*	−0.832*			
				(0.012)	(0.012)	(0.012)			
log(*traveltime*)							−0.846*	−0.781*	−0.788*
							(0.012)	(0.012)	(0.012)
Country border b		−0.386*	−0.402*		−0.386*	−0.402*		−0.382*	−0.398*
		(0.037)	(0.037)		(0.037)	(0.037)		(0.037)	(0.037)
Country contig. c		−0.977*	−0.956*		−0.975*	−0.955*		−0.970*	−0.955*
		(0.016)	(0.016)		(0.016)	(0.016)		(0.016)	(0.016)
Language barrier l		−0.158*	−0.112*		−0.160*	−0.113*		−0.151*	−0.119*
		(0.029)	(0.029)		(0.029)	(0.029)		(0.029)	(0.029)
Capital region *cap*			−0.126*			−0.132*			−0.138*
			(0.042)			(0.042)			(0.042)
Access to the sea s			0.086*			0.083*			0.094*
			(0.017)			(0.017)			(0.017)
Region contiguity r			−0.406*			−0.357*			−0.292*
			(0.033)			(0.033)			(0.033)
Eastern countries e			−0.113*			−0.112*			−0.095*
			(0.020)			(0.020)			(0.020)
Islands i			−0.559*			−0.565*			−0.377*
			(0.076)			(0.076)			(0.076)
RMSE	0.921	0.905	0.903	0.920	0.904	0.902	0.922	0.906	0.904
adj R^2	0.861	0.866	0.866	0.861	0.866	0.867	0.861	0.865	0.866
Moran's I	0.048	0.032	0.031	0.048	0.031	0.031	0.050	0.033	0.033

Notes All models include exporter (240) and importer (240) fixed effects. Model specification: (1) logged great circle distance, (2) logged population-weighted great circle distance, (3) logged travel time. $d_{rs}^{GCD)}$: logged geographical distance (in km.), $d_{rs}^{(POPGCD)}$ logged population-weighted geographical distance (in km.), $d_{rs}^{(TT)}$: logged distance travel time (in seconds), b_{rs}: 1 if separated by a country border, c_{rs}: 1 if share a common border, l_{rs}: 1 if different spoken languages. Robust standard errors in parentheses. * p<0.05

B - Market Access effects

Table 8 lists a number of 27 MA coefficients obtained from wage equations with three specifications (W1, W2 and W3), and MA measures computed with three geographical proxies (*gcd*, *pop* and *tt*) and three deterrence function's specification (D1, D2 and D3). The coefficients are reported along to their standard errors, as well as the measures of fit (R^2) and extracted log-likelihood of the estimated wage equation. Regarding the fit of the data, significant differences are perceived only between the wage equation specifications. As mentioned in Bosker and Garretsen (2010), control variables and country fixed effect as components of the wage equation explain a large part of the variance of the MA effects. Excluding the country fixed effects of the model results in a loss of fit, the best fit is however seen for the W3 specification that includes those effects as well as control variables at the regional level. The smallest mean MA effects are observed for the W2/D1 specification being equal to +0.191 on average. Largest coefficients are obtained for the W1 specification, only controlling for country fixed effects. As a conclusion, MA effect is not truly sensible to the choice of geographical distance proxy, since comparing the MA effect reveals a high similarity. Concerning the effect of the composition of the deterrence function D_{rs} on the MA effect, disparities are perceivable especially when comparing D2 with D1 and/or D3, coefficients being relatively larger than their counterparts.

In other words: the specification of the trade equation and wage equation matters. This is one of the main findings of Bosker and Garretsen (2010) who perform a meta-study on the wage equation. They conclude that the various aspects of the trade equation (estimation technique, controls, construction of (internal-) distance, etc.) usually influence the results of the wage equation quite heavily.

Based on two criteria, namely the log-likelihood measure and the model fit, the preferred model combines a MA measure obtained with the D2 deterrence function and W3 specification of the wage equation, thus including country fixed effects and control variables available at the regional level accounting for human capital absorption and labour market characteristics.

Table 8: Estimated MA effect for 27 specifications (W1, W2, W3, D1, D2, D3, gcd, pop, tt)

		W1			W2			W3		
		gcd	pop	tt	gcd	pop	tt	gcd	pop	tt
D1	MA effect	0.324*	0.315*	0.355*	0.192*	0.191*	0.191*	0.228*	0.221*	0.245*
	Std. Error.	(0.033)	(0.035)	(0.035)	(0.035)	(0.034)	(0.037)	(0.029)	(0.028)	(0.032)
	R	0.871	0.873	0.873	0.703	0.705	0.698	0.922	0.922	0.921
	logLik	53.119	54.561	54.641	−45.524	−44.859	−47.375	107.782	108.506	106.704
D2	MA effect	0.326*	0.319*	0.355*	0.205*	0.203*	0.200*	0.253*	0.246*	0.265*
	Std. Error.	(0.037)	(0.036)	(0.039)	(0.043)	(0.042)	(0.045)	(0.029)	(0.028)	(0.032)
	R	0.862	0.864	0.864	0.694	0.698	0.690	0.926	0.926	0.925
	logLik	44.787	46.367	46.876	−48.923	−47.593	−50.431	113.517	114.321	112.127
D3	MA effect	0.288*	0.284*	0.326*	0.197*	0.194*	0.201*	0.226*	0.222*	0.249*
	Std. Error.	(0.034)	(0.033)	(0.038)	(0.039)	(0.038)	(0.043)	(0.027)	(0.026)	(0.030)
	R	0.859	0.860	0.860	0.697	0.698	0.693	0.925	0.925	0.924
	logLik	42.551	43.487	43.346	−47.844	−47.593	−49.503	112.653	112.980	110.848
Country FE		YES			NO			YES		
Control Variables		0			5			3		