Span Econ Rev (2006) 8:83–112
DOI 10.1007/s10108-006-9007-x

REGULAR ARTICLE

Estimating Krugman's economic geography model for the Spanish regions

Armando José Garcia Pires

Published online: 16 May 2006

Abstract This paper estimates Krugman's (J Polit Econ 99:413–499, 1991) economic geography model using data from the Spanish NUTS 3 regions. The econometric formalization endogenously determines wages in a region as a function of income and wages in other regions. The specification adopted also allows us to study the relation between the agglomeration of economic activity, increasing returns and market access. The first result obtained is that the Spanish economy exhibits a spatial wage structure: wages in a region are positively determined by income and wages in neighboring regions. In second place it is found support for the structural relations of the underlying theoretical model, indicating the importance of scale economies and transport costs in shaping the Spanish economic geography.

Keywords "New" economic geography · Increasing returns · Market structure · Transport costs · Market potential

JEL Classification F12, R12

Introduction

The starting point of this paper is the so-called "new" economic geography (NEG) of Krugman (1991), Venables (1996) and Fujita et al. (1999), just to mention a few.[1] The NEG studies the location in space of economic activity and the main features of this literature can be summarized as follow: monopolistic

A. J. Garcia Pires (✉)
Norwegian School of Economics and Business Administration,
Department of Economics, Helleveien 30, 5045 Bergen, Norway
e-mail: armando.pires@snf.no

[1] For a review see Fujita et al. (1999); Neary (2001); Fujita and Thisse (2002) and Baldwin et al. (2003).

competition, iceberg trade costs and "home market" and "agglomeration" effects. Monopolistic competition follows directly from the very influential paper by Dixit and Stiglitz (1977), while iceberg trade costs from Samuelson (1954). The "home market" effect is a trademark of the "new" trade theory (see for example Krugman 1980). The new feature of the NEG (besides the combination of the aspects mention above) is the existence of "agglomeration" effects that arise endogenously for certain parameter values.

The NEG highlights the importance of increasing returns, trade costs and demand patterns on the location of economic activity. Accordingly, agglomeration is promoted when scale economies are strong, trade costs are low and demand patterns are biased for manufactured goods. The rational is that economies of scale encourage firms to concentrate production in few locations in order to exploit scale gains; low trade costs allow firms to serve distant markets from central locations while demand patterns biased for industrial goods supports larger agglomerations of firms.

The large bulk of the NEG is theoretical, however, in recent years a growing attention has also been given to empirics.[2] For example Knarvik et al. (2002a) and Redding and Venables (2004) study, respectively, the economic geography of the EU and the world; Davis and Weinstein (1999) test for "home market" effects; Quah (1996) and Puga and Overman (2002) use non-parametric estimation to analyze economic spatial interactions; others like Haaland et al. (2002) and Garcia Pires (2005) apply general equilibrium analysis to "new" economic geography models.

The objective of this paper is also to contribute to the NEG empirical literature. However, the empirical implementation of the NEG theory is not straightforward since the latter is characterized by multiple spatial equilibria. This poses problems in terms of interpreting empirical results. Nevertheless, one special feature of the NEG is that it predicts a spatial wage structure: wages tend to be higher in regions closer to larger markets. It is then developed an econometric formalization based on Krugman (1991) that endogenously determines wages in a region as a function of income and wages in other regions. The specification adopted allows us to study the relation between agglomeration, increasing returns and market access. Specifically, this paper tests for the spatial distribution of economic activity in Spain by estimating the structural parameters of the Krugman (1991) model: elasticity of substitution, trade costs and share of income spent on industrial goods.

This type of structural estimation for NEG models was first developed by Hanson (1998, 2005) for the US. Hanson's (1998, 2005) empirical methodology was also implemented in other country studies. For example, Roos (2001) and Brakman et al. (2004) for Germany, De Bruyne (2003) for Belgium, Mion (2004) for Italy and Kiso (2005) for Japan.[3] However, contrary to this paper, in

[2] For a review see Head and Mayer (2004), Overman et al. (2003) and Combes and Overman (2004).

[3] Paluzie et al. (2005) start like us from Krugman (1991) but only estimates a market potential function similar to Harris (1954) for the Spanish NUTS 3 regions.

all these studies the econometric formalization is based on the Helpman (1998) variant of the Krugman (1991) economic geography model.

The main differences between the two models are that while in Krugman (1991) the constant returns/perfect competitive sector produces a homogeneous good (agriculture) that is freely traded; in Helpman (1998), instead, the homogeneous good is non-tradable (housing). The implication of this assumption is the introduction of an extra centrifugal force, given that prices of housing are higher in the region with more population. As a result, the two models differ in terms of the impact of a reduction in trade costs: in Krugman (1991) reducing trade costs promotes agglomeration; while in Helpman (1998) it promotes dispersion. Note that these comparative-static results do not implicitly mean that the Krugman (1991) and the Helpman (1998) models give different previsions on the spatial evolution of industry. As shown by Puga (1999) they just represent two sides of the bell-shaped curve that relates transport costs with regional inequality: a reduction from high trade costs means an increase in agglomeration, like in Krugman (1991); while a reduction from low trade costs means more dispersion, like in Helpman (1998).[4]

As a result, given that the Helpman (1998) model predicts less extreme spatial patterns than Krugman (1991), it is usually defended that the former is more suitable for empirical implementation (see Combes and Overman 2004; Hanson 1998, 2005; Brakman et al. 2004; Mion 2004). In fact, in the two-regions case the Krugman (1991) model can have full concentration of economic activity with one region attracting all industry (black-hole location). Clearly, such spatial patterns are not easily observed in the real world. However, these authors seem to forget that this extreme configuration of space in the Krugman model only arises in its two-regions version (i.e., Krugman 1991). If Krugman (1991) is generalized to many regions (as in Krugman 1993), the spatial economic patterns derived do not encompass anymore black-hole locations, but more dispersed outcomes where all regions can host some industry. Given this, it is then argued by us that the multi region version of the Krugman (1993) model is also suitable for empirical estimation.

Also, the housing sector in Helpman (1998) is somewhat strange to be modeled as non-tradable in a regional setting, as it is the case in the empirical implementation by Hanson (1998, 2005), Roos (2001), Brakman et al. (2004), De Bruyne (2003), Mion (2004) and Kiso (2005). This implies for example that a consumer in Madrid cannot contract a housing firm from Barcelona to build a house. If that can happen in an international context, such is hardly the case at the regional level.

In addition, and as we have already called the attention before, the Krugman (1991) model focus more in agglomeration forces while the Helpman (1998) model in dispersion forces, i.e., the former attaches more importance to a reduction from high trade costs, while the latter from low trade costs. This is in our view important because one of the central concerns of the "new" economic

[4] Alonso-Villar (2005) can also help to reconcile Krugman (1991) and Helpman (1998) by showing the different role that transport costs have for upstream and downstream producers.

geography literature is the study of agglomeration forces (much more than the dispersion forces). In this sense the empirical test of Krugman (1991, 1993) allows us to look more directly to "agglomeration" effects.

Moreover, as defended by Puga (1999), Krugman (1991) is by nature a regional economy model. Then since Spain has a very strong regionalist culture it can be a particularly interesting case study for the test of the Krugman model. With this we hope to contribute to a better understanding of the Spanish economic geography.

The empirical strategy to test the Krugman model is then divided in two steps. In the first step, it is made a very simple description of the Spanish economic geography in terms of regional wages and GDP. In the second step, the Krugman (1991) model (for short KM) and a reduced form market potential function (MPF) in the line of Harris (1954) are tested econometrically. The first result obtained is that the Spanish economy exhibits a spatial wage structure: wages in a region are positively determined by income and wages in neighboring regions. In second place it is found support for the structural relations of the underlying theoretical model, indicating the importance of scale economies and transport costs in shaping the Spanish economic geography.

Besides this section this paper has more five sections. The next section introduces Krugman's (1991) economic geography model. After we derive the empirical model and we discuss some of the econometric issues involved. Next we present the statistical sources and the data used in this study. Then we show and comment the econometric results obtained in this paper. We conclude by discussing the main findings.

Economic geography model

This section presents very briefly the multi region version of the Krugman (1991) model (see Krugman 1993). This model considers J regions ($j = 1, 2 \ldots, J$); two productive sectors: the traditional sector and the industrial sector; two goods: the homogeneous good (H) and a composite differentiated good (D); and two factors of production: workers (that are perfectly mobile) and farmers (immobile). The H-good is produced under constant returns to scale by the traditional sector (this good is the *numeraire*); and the D-good is produced under increasing returns to scale by the industrial sector. The industrial sector is footloose (in the sense that can move production between locations) and the traditional sector is immobile. The factors of production are sector-specific: workers can only be employed in the industrial sector and farmers on the traditional sector.

The model has a two-stage budgeting decision with upper tier symmetrical preferences of the Cobb-Douglas type:

$$U = H^{1-\mu} D^{\mu}, \tag{1}$$

where H and D are the consumption, respectively, of the homogeneous good and of the differentiated composite good; and μ is the share of income spent on differentiated goods, with $0 < \mu < 1$.

The consumption of the composite differentiated good is a symmetric CES utility function *à la* Dixit and Stiglitz (1977):

$$D = \left[\sum_{i=1}^{r} \delta_i^{\frac{(\sigma-1)}{\sigma}} \right]^{(\sigma/(\sigma-1))}, \tag{2}$$

where σ represents the elasticity of substitution between varieties (with $\sigma > 1$, i.e., products are homogenous if σ tends to infinite and varieties are very differentiated if σ is close to one), δ_i is the demand for variety i of the composite good and r is the number of varieties in the economy (i.e., $i = 1, 2, \ldots, r$).

The technology of the increasing returns sector is given by the usual linear cost function: $T_{Dij} = f + cQ_{Dij}$, where T_{Dij} are the workers used in the production of variety i in the region j, f are the fixed costs, c is the unit variable cost and Q_{Dij} is the quantity of variety i produced in region j.

Introducing now transport costs; it is assumed that the homogeneous good is freely trade, while the differentiated good is subject to iceberg trade costs. According to the iceberg hypothesis a part of the good shipped "melts" in transit as expressed in the following equation:

$$V_{ijk} = e^{-\tau d_{jk}}, \tag{3}$$

where V_{ijk} is the fraction of the good i shipped from location j that arrives at k, d_{jk} is the distance between j and k and τ is the trade cost (with $\tau > 0$). Condition (2) and iceberg trade costs imply that the elasticity of demand facing any individual firm is σ. Also in equilibrium due to fixed costs and preference for variety implied by (2), each firm produces only one variety (therefore, the number of varieties equals the number of firms). Conversely, it is assumed that is not profitable for a firm to produce more than one variety.

The profit maximizing behavior of a representative firm is then to set a price as a constant *mark-up* over marginal cost:

$$P_{ij} = \left(\frac{\sigma}{\sigma - 1} \right) c w_j, \tag{4}$$

where w_j is the nominal wage rate in location j and P_{ij} is the price of variety i in region j. Then, the Marshall-Lerner price-cost mark-up is $\sigma/(\sigma - 1)$. The higher this ratio, the higher the degree of monopoly power by a firm. As a result, Krugman (1991) understands σ as an inverse measure of scale economies, given that it can be thought as a direct measure of price distortion and as an indirect measure of market distortion due to monopolistic power. Specifically, given that $\sigma/(\sigma - 1)$ is bigger that one, Krugman (1991) sees this as the presence of increasing returns to scale. Conversely, if σ is very high economies of scale are low, and the reverse for low σ.[5]

[5] Note, however, that σ is a parameter of tastes and not technology and as result the interpretation given by Krugman (1991) is not consensual (see Neary 2001).

Another important relation taken from the KM is $\sigma(1-\mu)$, also known as the "no-black hole condition". If this product is bigger than one, then, the economies of scale are weak and/or the expenditure on manufactures is so small such that in equilibrium location of industry depends only in trade costs (and vice-versa).[6]

This model has four equilibrium conditions that are central for the empirical formulation. The first equilibrium condition predicts the equalization of real wages across regions:

$$\frac{w_j}{I_j^{-\mu}} = \frac{w_k}{I_k^{-\mu}}, \tag{5}$$

where I_j is price index of manufactures in region j.

The second equilibrium condition envisages that total income in a region equals labor income in that region:

$$Y_j = (1-\mu)\phi_j - \mu\lambda_j w_j, \tag{6}$$

where Y_j is total income in region j, ϕ_j is the percentage of farmers in region j and λ_j is the percentage of workers in region j. Note that due to immobile farmers Y_j is always positive even if a region does not host any manufacture.

The third equilibrium condition gives the supply of manufacturing goods. According to this, the price index of differentiated good is bigger in regions that have to import a greater percentage of this good:

$$I_j = \left[\sum_k \lambda_k \left(w_k e^{\tau d_{jk}}\right)^{1-\sigma}\right]^{(1/(1-\sigma))}. \tag{7}$$

Finally, the fourth equilibrium condition presents the labor demand function. This tell us that the labor demand is higher in regions with higher final demand:

$$w_j = \left[\sum_k Y_k \left(I_k e^{-\tau d_{jk}}\right)^{\sigma-1}\right]^{(1/\sigma)}. \tag{8}$$

This model has two types of forces affecting the location choices of firms: centripetal forces (that attract firms to center locations), and centrifugal forces (that pull away firms from center locations). In the first case "home market" and

[6] Note that this relation has a different interpretation in Krugman (1991) and Helpman (1998). This is due to the different comparative static results that these two models have in relation to trade costs. While in Krugman (1991), if $\sigma(1-\mu) < 1$, agglomeration occurs independently of trade costs; in Helpman (1998) it means that trade costs play a role in determining the spatial equilibrium. Conversely, if $\sigma(1-\mu) > 1$, in Helpman (1998) dispersion is always stable independently of trade costs; while in Krugman (1991), agglomeration only occurs for low levels of trade costs.

"price index" effects create "demand" and "cost linkages" through the inter-regional mobility of workers. In the second case "competition" effects in factor and product markets cannibalize the profits of firms. The spatial equilibrium or so to say the choice of location by firms depends on the model structural parameters (σ, μ and τ). To be precise, agglomeration of the industrial sector is predicted for low σ, high μ and low τ (and the reverse for dispersion): low σ because it allows the exploitation of scale gains by concentrating production in few locations, high μ since it supports a larger agglomeration of industrial firms and low τ given that it permits firms to supply the different markets from central locations.

Other important thing to note is that Eq. (8) is very similar to Harris (1954) "Market Potential" function (MPF) in the sense that economic activity is bigger in regions that are closer to larger markets. The KM therefore gives microeconomic content to Harris (1954) ad-hoc formulation of market potential. Equation (9) represents a usual formulation for Harris (1954) MPF:

$$\mathrm{MP}_j = \sum_k Y_k g(d_{jk}), \tag{9}$$

where MP_j is the market potential index for location j; and $g()$ is a decreasing function of trade cost.

The empirical part estimates both the MPF, as well as the KM.

Econometric model

In this section the econometric specification for both the MPF and the KM is presented. Following the logic of the NEG, the MPF can be written by making nominal wages as the dependent variable (i.e., $\mathrm{MP}_j = w_j$) and the trade cost function of the iceberg type (i.e., $q\left(d_{jk}\right) = e^{-\beta d_{jk}}$). Then the empirical specification links nominal wages in a location to income in other locations weighted by distance. As result, factoring out a constant (θ) and an error term (ε_j) it is obtained:

$$\log\left(w_j\right) = \theta + \alpha \log\left(\sum_k Y_k e^{-\beta d_{jk}}\right) + \varepsilon_j, \tag{10}$$

where α and β are, respectively, the effect of purchasing power and distance from consumer markets on wages of a region. The remaining variables are as defined in the previous section. According to the MPF hypotheses both α and β must be bigger than zero ($\alpha > 0$ and $\beta > 0$). The empirical test will tell us if these MPF restrictions are verified.

The estimation of the KM presents a problem because there is no available data for the differentiated goods price index (I) at any level of regional aggregation. Therefore, it is not possible to estimate simultaneously all the four equilibrium equations. In alternative we substitute Eq. (5) in Eq. (8) to get:

$$\log\left(w_j\right) = \theta + \sigma^{-1}\log\left(\sum_k Y_k w_k^{\frac{\sigma-1}{\mu}} e^{-\tau(\sigma-1)d_{jk}}\right) + \eta_j, \tag{11}$$

where η_j is an error term. The remaining variables are as defined in the previous section and above for the MPF. According to theory $\sigma > 1, 0 < \mu < 1$, and $\tau > 0$. One of the objectives of the empirical test is to check if these theoretical restrictions of the KM are satisfied.

Besides the structural parameters (σ, τ and μ) some relations taken from the KM will also be estimated, more precisely the ratios $\sigma/(\sigma - 1)$ and $\sigma(1 - \mu)$. Conversely, if parameter estimates are such that $\sigma/(\sigma - 1) > 1$, then, it would be possible to conclude that the Spanish economic geography is subject to increasing returns to scale. Also, if parameter estimates give $\sigma(1 - \mu) < 1$, then scale economies in Spain are sufficiently strong and/or the manufacturing share in consumption is sufficiently high such that economic activity agglomerates for any value of trade costs.

As discussed by Combes and Overman (2004), the estimation of Eqs. (10) and (11) poses some econometric problems. First, since both equations are highly non-linear it is necessary to use non-linear estimation methods. Second, some endogeneity problems arise what can bias the parameter estimates. In fact in both equations income is determined simultaneously together with wages and in Eq. (11) wages enters both on the right and on the left hand side.

To tackle this issue it is followed Hanson (1998, 2005) strategy. Specifically, besides the level specification it is also estimated a time-differenced version of Eqs. (10) and (11) in order to control for unobserved regional factors constant across time (as endowments and comparative advantage)[7] and for temporary shocks that can influence business cycle (as recessions or plant closings). Conversely, the shortcoming of the levels specification is that it does not control for many of the endogeneity problems, while the disadvantage of the differences specification is that information is lost by taking a time differentiated form.

The time differenced specification for the MPF is:

$$\Delta \log\left(w_{jt}\right) = \alpha\left[\log\left(\sum_k Y_{kt} e^{-\beta d_{jk}}\right) - \log\left(\sum_k Y_{kt-1} e^{-\beta d_{jk}}\right)\right] + \Delta\varepsilon_{jt}, \tag{12}$$

where t refers to the observation period and Δ to the lag operator.

[7] It would be more appropriate to use data on endowments and comparative advantage, or instrument for them as in Hanson (2005), however, due to unavailability of data that was not possibly. Hanson (2005) results, however, are not affected when he controls for comparative advantage and endowments. This may be due to the fact, that the Helpman (1998) model (and obviously also Krugman, 1991) does not encompass endowments and comparative advantage as location determinants.

The time differenced specification for KM is:

$$\Delta \log \left(w_{jt}\right) = \frac{1}{\sigma}\left[\log\left(\sum_k Y_{kt} w_{kt}^{\frac{\sigma-1}{\mu}} e^{-\tau(\sigma-1)d_{jk}}\right) - \log\left(\sum_k Y_{kt-1} w_{kt-1}^{\frac{\sigma-1}{\mu}} e^{-\tau(\sigma-1)d_{jk}}\right)\right] + \Delta\eta_{jt}. \quad (13)$$

It is assumed that the random errors $\Delta\varepsilon_{jt}$ and $\Delta\eta_{jt}$ are uncorrelated with the regressors and across regions.

To estimate Eqs. (10), (11), (12) and (13) it was chosen the non-linear least squares (NLS) estimator.[8] However, computer algorithms for non-linear estimations can have some precision problems. The solution proposed by Davidson and MacKinnon (1993) is to run a Gauss Newton Regression (GNR) on the parameters estimated obtained by NLS. Basically the GNR takes the residuals obtained by NLS and checks on the explanatory power of this. If they have little explanatory power (i.e., magnitudes very close to zero, and t-statistics and R^2 very small) it is possible to accept the validity of the estimations.

Other tests on the endogeneity of the independent variables are also performed. First, the four equations (in levels and in differences for the MPF and the KM) are estimated for two samples of regions: all regions and excluding regions with more than 5% of Spain population. The latter specification does not include the regions of Madrid, Barcelona and Valencia. With this we want to control for the possibility that shocks specific to larger regions influence economic activity in the smaller ones. This can happen, because it is especially in more populated regions that shocks can be correlated with the regressors. If the estimated coefficients with the restricted sample are similar to the ones with the whole sample, then it is very likely that the bigger regions do not affect the overall results. Second, two alternative measures of distance (that will be explained bellow) are used to control for possible measurement errors in this variable. Third, it is also run a regression with dummy variables for the most economically active regions of Spain: Madrid, Barcelona and Viscaya. This intends to control for long run regional differences in technology.

Another econometric issue is that this study uses cross-section data and therefore errors can differ between geographical units of analysis. To minimize this, a variance-covariance matrix Heteroskedasticity-consistent is estimated. However, in spite of all the precautions mentioned before to reduce the

[8] The majority of the studies similar to ours also use NLS (Hanson 1998, 2005; Brakman et al. 2004; Roos 2001; De Bruyne 2003 and Mion 2004). Hanson (2005) also uses GMM and Kiso (2005) only GMM. In turn, Mion (2004) besides NLS and IVNLS (instrumental variables non-linear least squares) also proposes a linearization procedure based in Combes and Lafourcade (2001). De Bruyne (2003) also estimates a linear version but by assuming a value for the trade cost parameter. Further, Hanson (1998, 2005) only estimates the model in first differences, while Brakman et al. (2004), Mion (2004); Kiso (2005) only in levels. See Table 1 for a comparative overlook of the different studies.

Table 1 Overview of results from similar studies

Author(s)	Spatial unit	Estimation method	Dependent variable	Estimation results NEG model	Estimation results MPF
Hanson (1998, 2005)	US	NLS and GMM (differences)	Wages	NLS σ: 4.9* to 7.6* μ: 0.91* to 0.98* τ: 1.6* to 3.2* $\sigma(1-\mu)$: 0.08* to 0.65* $\sigma/(\sigma-1)$: 1.2* to 1.3* GMM σ: 1.7* to 2.5* μ: 0.54* to 0.87* τ: 9.9* to 20.3* $\sigma(1-\mu)$: 0.27* to 0.94* $\sigma/(\sigma-1)$: 1.6* to 2.3*	α: 0.24* to 0.43* β: 5.5* to 16.4*
Brakman et al. (2004)	Germany	NLS (levels)	Wages	Real wage equalization σ: 3.1* to 4.9* μ: 0.54* to 12.8 τ: −0.001* to 0.01* $\sigma(1-\mu)$: −36.6 to 2.3 $\sigma/(\sigma-1)$: 1.25 to 1.48* No-real equalization σ: 3.6* τ: 0.003*	α: 0.049* β: 0.092*
Roos (2001)	Western-Germany	NLS (differences and levels)	Wages	σ: 6.2 μ: 0.8* τ: 0.003	Differences: α: −3.4 to 1.03 β: −0.03 to 0.01 Levels: α: 0.03* to 0.08* β: 0.03* to 0.12*
De Bruyne (2003)	Belgium	NLS (differences and levels)	Density of employment	σ: 5.5 μ: 1.62* τ: 0.003*	Differences: α: 0.26* β(Belgium): −1.02 β(EU): 0.65* Levels: α: 0.67* to 0.68* β(Belgium): 0.07 β(EU): 0.1
Kiso (2005)	Japan	GMM (levels)	Wages	σ: 0.98 to 6.14 μ: −0.01 to 0.68 * τ: −26.30 to 1.69 $\sigma(1-\mu)$: 0.97 to 1.99* $\sigma/(\sigma-1)$: −41.57 to 4.41	NA

Table 1 continued

Author(s)	Spatial unit	Estimation method	Dependent variable	Estimation results NEG model	Estimation results MPF
Mion (2004)	Italy	NLS, IVNLS, Linear version dynamic panel (levels)	Wages	NLS σ: 5.9 to 6.7* μ: 0.91* to 0.93 τ: 0.74 to 0.93 $\sigma(1-\mu)$: 0.29 to 0.59 $\sigma/(\sigma-1)$: −1.2* to 1.27* IVNLS σ: 1.9* to 2* μ: 0.16* to 0.18* τ: 0.16 to 0.19 $\sigma(1-\mu)$: 0.22* to 0.24 $\sigma/(\sigma-1)$: 1.98* to 2.1* Panel σ: 3.2*to 3.4* μ: 0.75* to 0.77* τ: 128* to 133* $\sigma(1-\mu)$: 0.77* to 0.78* $\sigma/(\sigma-1)$: 1.41* to 1.44*	NA
Paluzie et al. (2005)	Spain	NLS (levels)	Wages	NA	α: 0.083* to 0.139* β: 0.077* to 0.102*

NA Not applies
*Statistically significant

auto-correlation of errors, this problem can nevertheless arise. We then perform an additional test by looking at the geographical maps of the residuals. If these maps show no geographical pattern it is then reasonable to accept the nature of the errors hypotheses, i.e., that errors are not correlated with regressors and the geographical unity of analysis.

Statistical sources and data

When doing spatial empirical research two problems usually arise. The first one is how to delimitate in the most appropriate way the spatial units of analysis. The second one relates with obtaining credible and comparable data.

For the first problem there are two sometimes-conflicting answers: the normative criteria and the functional criteria. The normative criteria delimitates regions according to some kind of political-administrative division (e.g., in EU and also in Spain the so-called NUTS regions); the functional criteria delimitates regions according to some set of socio-economic characteristics (e.g., urban area).[9] Note that in Spain the functional criteria can be accepted to coincide more or less with the normative criteria (see Biehl 1986). Furthermore, in the econometric model presented above, finer spatial desegregation is preferable since the probability that region specific shocks (embedded in the error terms) influence the independent variable will be smaller. However, too high geographic detail does not permit to capture the type of spatial externalities that the NEG literature highlights. As such regions will be delimitated taking both into account economic and econometric considerations.

The second problem relates with data sources: do they gives us reliable and comparable data? Of course that the second problem most of the times influences the choice in the first one. Anticipating the answer to the second problem it was used data from the EUROSTAT REGIO database, since this contains comparable data for all regions considered in this study (more on this see EUROSTAT 1996).

Given these issues, the spatial unit of analysis chosen was the 47 Continental NUTS 3 regions of Spain (see Table 2).[10] It was excluded the insular regions of Spain (Baleares Islands, Ceuta and Melilla and Canarias) to maintain some continuity along the spatial dimension. Also, apart from the Spanish regions it is not included any other region in Europe or in the world. Therefore this study is conduced in a closed economy framework. This is not totally satisfactory, since Spain is very integrated in the world economy. However, most of the similar studies that we have mentioned above make the same kind of assumption as in here.[11]

[9] More on this see Magrini (1999).

[10] Note that the majority of EU regional studies use the NUTS 2 level (less regional disaggregated). See for example Biehl (1986); Keeble et al. (1988); Neven and Gouyette (1995); Quah (1996); Martin (1999) and Brülhart (1998).

[11] The only exception is De Bruyne (2003) that introduces distance of Belgium regions to other European countries as an explanatory variable.

Table 2 NUTS regions in Spain

NUTS 1	NUTS 2	NUTS 3
Noroeste	Galicia	La Coruña
		Lugo
		Orense
		Pontevedra
	Principado de Asturias	Principado de Asturias
	Cantabria	Cantabria
Noreste	Pais Vasco	Álava
		Guipúzcoa
		Vizcaya
	Comunidad Foral de Navarra	Comunidad Foral de Navarra
	La Rioja	La Rioja
	Aragón	Huesca
		Teruel
		Zaragoza
Comunidad de Madrid	Comunidad de Madrid	Comunidad de Madrid
Centro	Castilla y León	Avila
		Burgos
		León
		Palencia
		Salamanca
		Segovia
		Soria
		Valladolid
		Zamora
	Castilla-la Mancha	Albacete
		Ciudad Real
		Cuenca
		Guadalajara
		Toledo
	Extremadura	Badajoz
		Cáceres
Este	Cataluña	Barcelona
		Gerona
		Lérida
		Tarragona
	Comunidad Valenciana	Alicante
		Castellón de la Plana
		Valencia
Sur	Andalucia	Almería
		Cadiz
		Córdoba
		Granada
		Huelva
		Jaén
		Málaga
		Sevilla
	Murcia	Murcia

The data used is: wages, GDP and distance (see Table 3 for summary statistics). Due to data unavailability, wages are at the NUTS 2 level.[12] All the

[12] Paluzie et al. (2005) were able to obtain wages data at the NUTS 3 level: however, they use a different database from ours (Fundación BBVA).

remaining variables are measured at the NUTS 3 level.[13] In turn distance data was obtained from the CD-Rom "Route 66 Europe". This software contains vast geographical information on the European Continent that includes, amongst other statistics, road distances between all European cities. Two measures of distance are used: simple distance and *hub-and-spoke* (HAS) distance. The simple distance is the shortest road distance between two major NUTS 3 cities (as defined in EUROSTAT 1996). In the HAS distance, instead, goods are first transported to a transportation hub before going to the final destination. It is assumed that the transportation "hub" for a given NUTS 3 region is the major NUTS 2 city where the first is included.[14]

The study covers the average of the years 1981–1982–1983 (referred as 1981), 1988–1989–1990 (referred as 1988) and 1993–1994–1995 (labeled as 1995). The average is taken to avoid influent observations. Note that these three periods refer to some important events for the Spanish economy. The first and the second periods are before and after the accession of Spain to the EU, respectively. The third period is just subsequent to the Single Market program.[15]

Now it is performed a very crude analysis on the spatial distribution of economic activity in Spain. The variables analyzed are wages per worker and GDP. Namely, it is compared the value of a variable in a region to its national average. Taking wages as example, this is as follows:

$$\frac{w_{jt}}{\bar{w}_t}, \tag{14}$$

where w_{jt} are wages per worker in region j in period t, and $\bar{w}_t$ the national average of wages in the same period t.

From this descriptive analysis some preliminary observations can be made. First, wages are higher in the regions of Madrid, Cataluña and Pais Vasco (see Figs. 1, 2 and 3). Other regions with higher wages are Aragón, Navarra, Asturias and Cantabria. Instead the regions with lower wages are Galicia, Castilla y León, Murcia, Castilla-la Mancha and Extremadura.

Second, GDP (see Figs. 4, 5 and 6) is higher in Madrid, Barcelona (Cataluña) and Viscaya (Pais Vasco) followed by Zaragoza (Aragón), Sevilla (Andalucia), Valencia (Comunidad Valenciana) and La Coruña (Galicia). In turn, GDP is lower in Lugo and Orense (Galicia); Soria, Segovia, Avila, Palencia, Salamanca

[13] Hanson (1998, 2005) computes the independent variables at the county level, but aggregates the dependent variable (i.e., wages) at the state level or across counties within a concentric distance band.

[14] To give an example, consider the NUTS 3 regions of Gerona and Madrid. The "simple distance" is the shortest road distance between the most populated city of the Gerona region (in this case Gerona) and the most populated city of the Madrid region (Madrid). The HAS distance is the distance between Gerona and the *spoke* NUTS 2 region in which Gerona is included (in this case the Catalonia region with the *spoke* located in Barcelona) and from here to Madrid (since the city of Madrid is also the *spoke* of the NUTS 3 region Madrid).

[15] For periods before 1981 data for Spain from the Eurostat REGIO database is very incomplete. That had precluded us from using data previous to 1981.

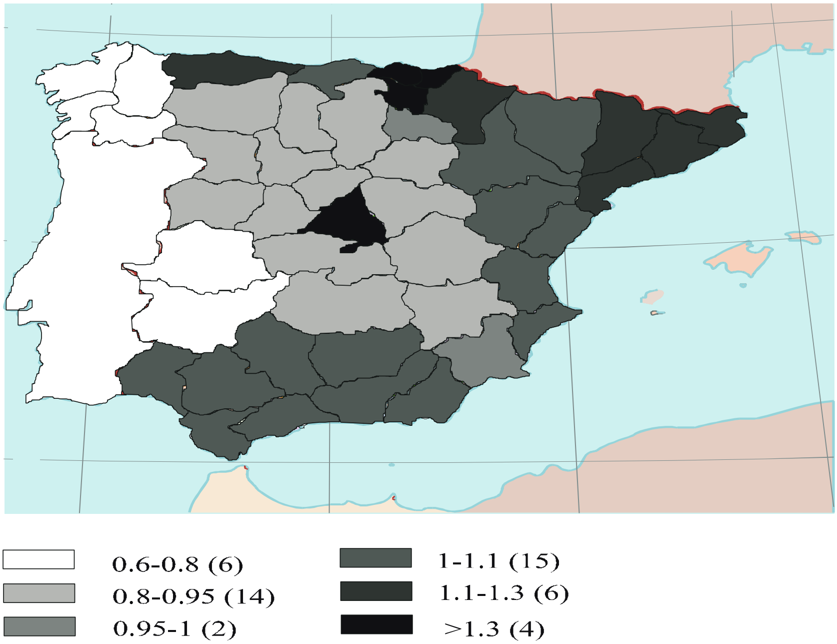

Fig. 1 Spatial distribution of wages 1981

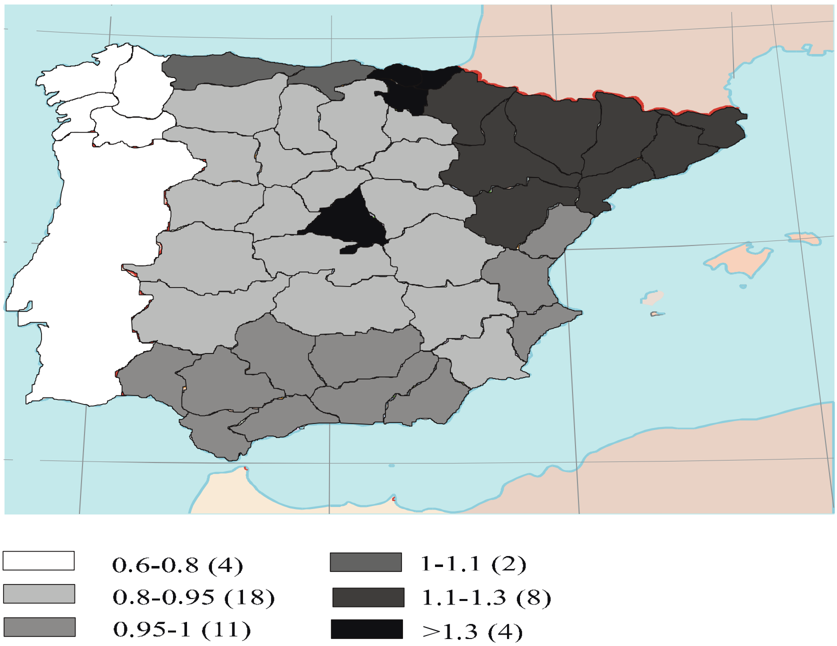

Fig. 2 Spatial distribution of wages 1988

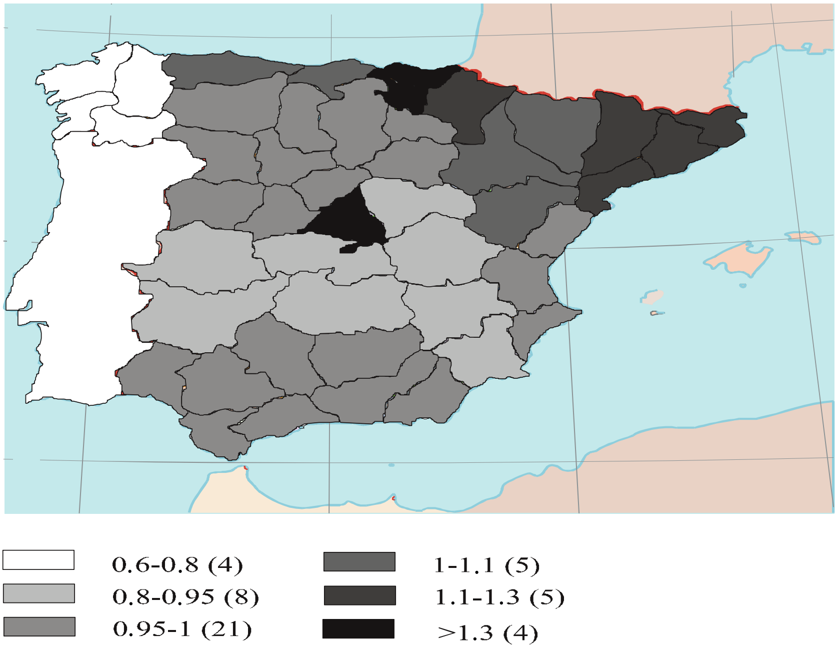

Fig. 3 Spatial distribution of wages 1995

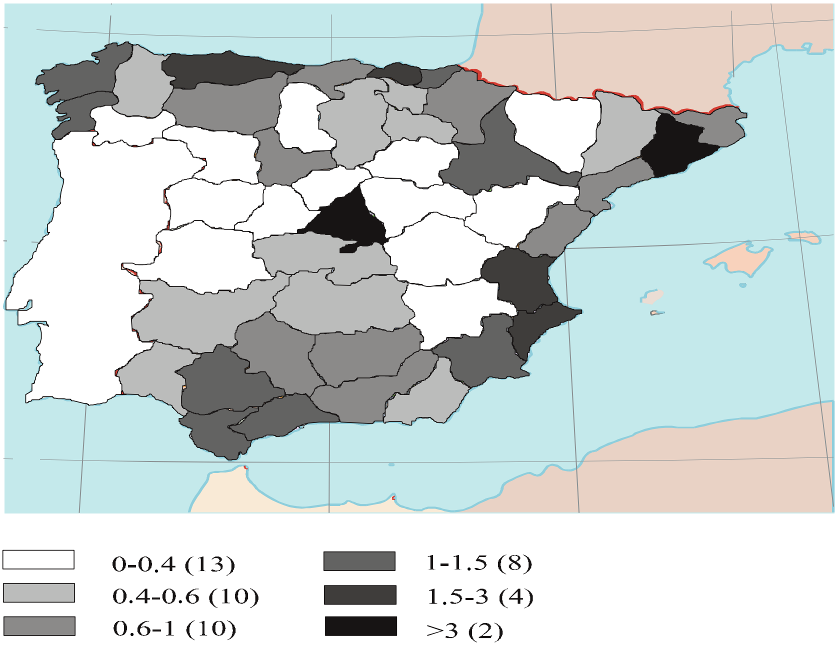

Fig. 4 Spatial distribution of GDP 1981

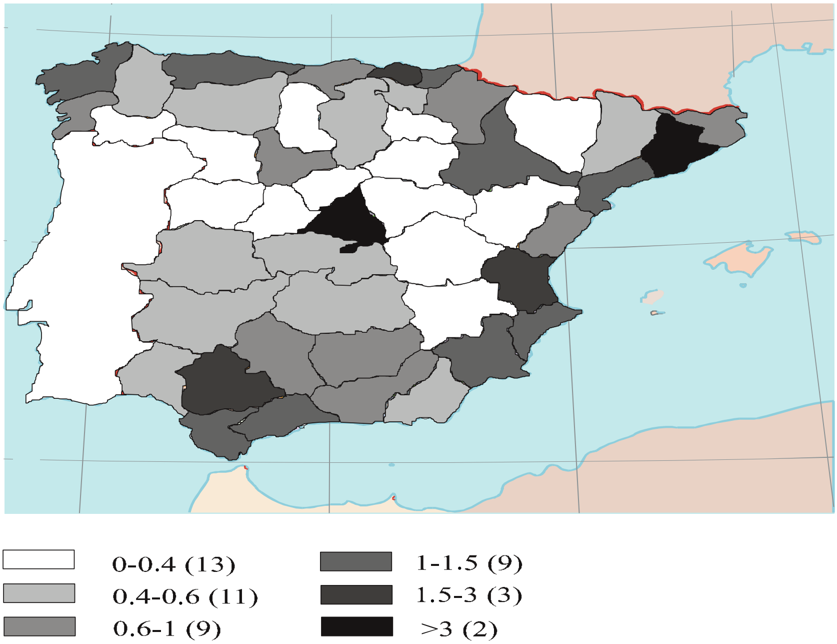

Fig. 5 Spatial distribution of GDP 1988

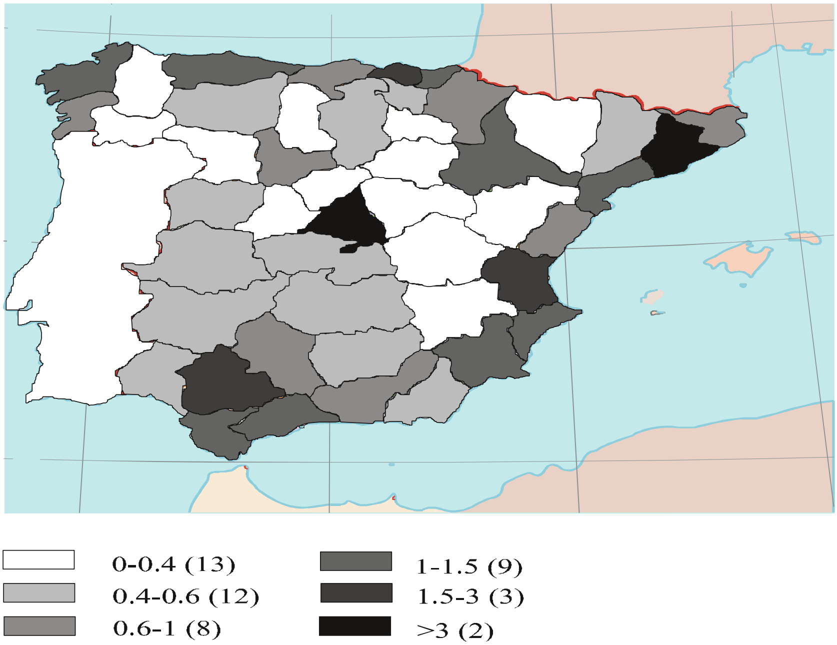

Fig. 6 Spatial distribution of GDP 1995

and Zamora (Castilla y León); Cuenca, Guadalajara and Albacete (Castilla-la Mancha) and Huesca and Teruel (Aragón).

Although not perfect, then, there seems to exist a gradient in wages as predict by the Krugman (1991) model: wages are higher where there is a higher level of economic activity. In turn, the cases that do not match perfectly (La Coruña), this may be due to the fact that wages are measured at the NUTS 2 level and GDP at the NUTS 3 level.

Econometric results

In this section it is presented the econometric results for both the MPF and the KM (in levels and in differences). Note that, as referred before, all regressions were tested against the GNR. All the tests confirmed the validity of our estimates. Also, standard deviations were calculated in two ways: the "usual" standard deviation, and a standard deviation Heteroskedasticity-consistent (only the latter are reported). Since all estimations yield similar standard deviations results, it is then reasonable to believe that errors do not differ very much in the spatial unit of analysis. In addition, the geographical maps of residuals (in levels and in differences for both the MPF and KM) do not show any spatial pattern. Therefore, it is possible to accept the hypothesis of non-autocorrelation of residuals. Lastly, just refer that this study has a strong theoretical component, as such econometric results are interpreted based on the theory.

Market potential function

This subsection shows the econometric results for the MPF. Given that parameter estimations are similar across the different specifications, we only presented in detail results in levels and in differences for all regions and simple distance (Table 4). Tables 5, 6, 7, 8 and 9 show results with the controls mentioned above: excluding high population regions, referred as restricted sample (Table 5); HAS distance with the whole sample (Table 6) and with the restricted sample (Table 7); and dummy variables for the most economic active regions with the full sample (Table 8) and with the restricted sample (Table 9).

In what respects the results in levels, the first thing to note is that the parameter that measures the effects of purchasing power on wages in a region (α) is always positive as predicted by the theory. This means that higher consumer demand raises nominal wages in a location. The point estimates for α are between 0.14 (in 1981) and 0.19 (in 1988). Also corroborating MPF hypothesis, the effect of distance from consumer markets on the economic activity of a region (β) is positive, and ranging from 0.008 (in 1995) to 0.02 (in 1981) per kilometer. This shows that greater distance from consumer markets reduces wages in a region.

Turning now to the results in differences, all coefficients are statistically significant, but not all of them are consistent with theory. The effects of purchasing power on wages are positive as expected, and the point estimates for α are now

Table 3 Summary statistics

Period	Wages (Ecu)	GDP (Ecu)	Direct distance (km)	HAS distance (km)
1995	14 493,84	8 488 787 234		
	(2 086,174)	(12 621 562 342)		
1988	11 294,52	7 010 567 376	517,4404	612,9872
	(1 989,203)	(10 140 535 762)	(260,0536)	(285,8376)
1981	7 119,455	3 547 510 638		
	(1 311,679)	(4 866 928 293)		

The sample is the 47 regions of Continental Spain. Regions definitions are those from EUROSTAT (1996). Standard errors are in parentheses

Table 4 Market potential function: 47 regions, simple distance

	1981	1988	1995	1988–1981	1995–1988
θ	5.75444*	4.84170	5.65975		
	(0.893135)	(1.64675)	(1.35663)		
α	0.140647*	0.189602	0.162944	0.687818*	1.37768*
	(0.039297)	(0.06764)	(0.05483)	(0.009593)	(0.029114)
β	0.02214*	0.010267	0.0087751	−0.005180*	−0.008097*
	(0.007106)	(0.003057)	(0.002572)	(0.001446)	(0.001651)
R^2	0.25751	0.243	0.161	0.161	0.249

All 47 regions are from Continental Spain. See text for definition of simple distance. Heteroskedasticity-consistent standard errors are reported in parentheses
*Statistically significant

Table 5 Market potential function: 44 regions, simple distance

	1981	1988	1995	1988–1981	1995–1988
θ	5.29933*	3.57096	4.00689*		
	(1.32447)	(3.05488)	(2.80537)		
α	0.161462*	0.246272*	0.234698*	0.703615*	1.24182*
	(0.058723)	(0.126325)	(0.114878)	(0.009699)	(0.2459)
β	0.017288*	0.0093884*	0.0081792*	−0.005069*	−0.008194*
	(0.006644)	(0.0035876)	(0.002622)	(0.001522)	0.0014978
R^2	0.212824	0.20273	0.219809	0.173	0.155

The low population sample excludes Madrid Barcelona and Valencia. See text for definition of simple distance. Heteroskedasticity-consistent standard errors are reported in parentheses
*Statistically significant

between 0.68 and 1.38. However, contrary to what is predicted by the MPF theoretical literature, distance has a negative effect on wages. This may be caused by two factors. First, the MPF is a reduced form of the theoretical based Krugman wage equation, and as such it might suffer from omitted variables bias. Second, as one anonymous referee noticed, wages are measured at the NUTS 2 level while the remaining variables at the NUTS 3 level. Wages, in this sense, may be reacting very little in relation to distance.

In terms of time evolution, in levels α increases from the first (1981) to the second period (1988), but decreases a little in the third one (1995). In differences,

Table 6 Market potential function: 47 regions. *hub-and-spoke* distance

	1981	1988	1995	1988–1981	1995–1988
θ	6.18179*	6.2866*	6.88771*		
	(0.697922)	(0.912403)	(0.812111)		
α	0.121114*	0.130848*	0.113787*	0.687853*	1.37764*
	(0.030803)	(0.038111)	0.033238	(0.009591)	(0.029104)
β	0.014609*	0.010202*	0.0086162*	−0.005181*	−0.008080*
	(0.004535)	(0.002873)	(0.0023918)	(0.001454)	(0.001649)
R^2	0.311803	0.281814	0.278080	0.16	0.25

All 47 regions are from Continental Spain. See text for definition of *hub-and-spoke* distance. Heteroskedasticity-consistent standard errors are reported in parentheses
*Statistically significant

Table 7 Market potential function: 44 regions, *hub-and-spoke* distance

	1981	1988	1995	1988–1981	1995–1988
θ	6.58424*	7.18033*	7.66283*		
	(0.776343)	(0.813587)	(0.757798)		
α	0.094132*	0.094504*	0.083083*	0.703629*	1.24186*
	(0.035565)	(0.035687)	(0.032695)	(0.009709)	(0.024595)
β	0.016689*	0.013747*	0.011455*	−0.005082*	−0.008189*
	(0.007055)	(0.006261)	(0.004726)	(0.001540)	(0.001498)
R^2	0.20915	0.153207	0.14733	0.173	0.156

The low population sample excludes Madrid Barcelona and Valencia. See text for definition of *hub-and-spoke* distance. Heteroskedasticity-consistent standard errors are reported in parentheses
*Statistically significant

Table 8 Market potential function: 47 regions, simple distance and regional *dummy*

	1981	1988	1995	1988–1981	1995–1988
θ	−3.63416	−3.38158	−3.04372		
	(7.10498)	(7.1412)	(7.08762)		
α	0.11414*	0.12331*	0.10731*	0.68352*	1.39409*
	(0.02962)	(0.02908)	(0.02885)	(0.010479)	(0.028524)
β	0.032515*	0.028319*	0.02273*	−0.004705*	−0.008346*
	(0.00869)	(0.00665)	(0.00375)	(0.001447)	(0.001497)
Dummy	0.121635	0.593994	0.618024	0.029523*	−0.033063
	(0.352511)	(0.53054)	(0.825679)	(0.010712)	(0.19583)
R^2	0.19782	0.213793	0.208222	0.187367	0.30552

All 47 regions are from Continental Spain. See text for definition of the regional dummy variable and simple distance. Heteroskedasticity-consistent standard errors are reported in parentheses
*Statistically significant

α increases from the first period (1988-1981) to the second (1995–1988). Then, in the 1980s and in the 1990s purchasing power became more important for regional wages. In turn, β in levels continuously decrease from 1981 to 1995 showing that distance has lost importance in the 1980s and in the 1990s in the Spanish economy. This can be due to the improvements on transport and communication infrastructures that Spain has gone through since its adhesion to the

Table 9 Market potential function: 44 regions. simple distance and regional *dummy*

	1981	1988	1995	1988–1981	1995–1988
θ	5.19121*	3.47165	4.04974		
	(1.31093)	(3.15359)	(2.71693)		
α	0.166352*	0.250296	0.233302*	0.701748*	1.24585*
	(0.058264)	(0.130274)	(0.111449)	(0.010135)	(0.025398)
β	0.017803*	0.0091661*	0.008589*	−0.003847*	−0.008176*
	(0.006582)	(0.003542)	(0.002796)	(0.001307)	(0.001435)
Dummy	0.077425	−0.02337	0.077123	0.035698*	−0.012136
	(0.071109)	(0.030093)	(0.052072)	(0.010459)	(0.017269)
R^2	0.225161	0.204289	0.240098	0.187453	0.162741

The low population sample excludes Madrid Barcelona and Valencia. See text for definition of the regional dummy variable and simple distance. Heteroskedasticity-consistent standard errors are reported in parentheses
*Statistically significant

Table 10 Krugman model: 47 regions, simple distance

	1981	1988	1995	1988–1981	1995–1988
θ	4.91956*	4.59820*	5.93014*		
	(1.82701)	(2.27363)	(1.83618)		
σ	4.98730*	4.26683*	5.18728*	1.71391*	2.39952*
	(2.28292)	(1.77684)	(2.07050)	(0.34198)	(0.759267)
μ	−10.0341	−7.31305	−7.13609*	0.85784	0.857713*
	(8.23053)	(4.77127)	(3.52942)	(0.05089)	(0.074665)
τ	0.00378269*	0.00283772*	0.00193187*	0.02338*	0.021295*
	(0.00142353)	(0.00109534)	(0.00072102)	(0.00406)	(0.008109)
$\sigma/(\sigma-1)$	1.25080*	1.30611*	1.23882*	2.40074*	1.71453*
	(0.143593)	(0.166493)	(0.118089)	(0.67104)	(0.387646)
$\sigma(1-\mu)$	55.0302	35.4703	42.2042	0.24705	0.34142*
	(63.4165)	(31.8180)	(31.3989)	(0.09788)	(0.07682)
R^2	0.153341	0.185354	0.178497	0.592	0.536

All 47 regions are from Continental Spain. See text for definition of simple distance. Heteroskedasticity-consistent standard errors are reported in parentheses
*Statistically significant

European Community. In this sense European regional funds helped Spanish regions to become more integrated.

The introduction of other controls confirms the picture from the central case (see Tables 5, 6, 7, 8 and 9). In levels, the trade cost parameter (β) shows very similar magnitudes in all estimations and the same happens with the temporal evolution. Also in levels, the impact of purchasing power on wages (α), changes a little in the alternative specifications. However, the temporal evolution of α in levels is very similar across estimations. In differences, instead, the α parameter shows magnitudes very similar in the alternative formalizations and the same happens with the time evolution. In what respects the regional controls specification, the dummy variable is never statistically significant. As such, regional dynamics of the most advanced regions do not seem to be affecting the Spanish

Table 11 Krugman model: 44 regions, simple distance

	1981	1988	1995	1988–1981	1995–1988
θ	2.45113	2.27596	4.20221*		
	(2.36474)	(2.16490)	(1.60315)		
σ	2.93521*	2.52913*	3.01992*	1.82163*	2.95455*
	(1.05516)	(0.642978)	(0.702487)	(0.09681)	(1.08137)
μ	−3.73113	−2.24888*	−2.40659*	0.86172*	0.880896*
	(2.47616)	(0.939564)	(0.828597)	(0.03932)	(0.065971)
τ	0.00510792*	0.00491522*	0.00338598*	0.01566*	0.0099346*
	(0.0019376)	(0.00146254)	(0.00084698)	(0.00552)	(0.004379)
$\sigma/(\sigma-1)$	1.51674*	1.65397*	1.49507*	2.21710*	1.51163*
	(0.281749)	(0.274982)	(0.172175)	(0.07509)	(0.283061)
$\sigma(1-\mu)$	13.8869	8.21686	10.2876*	0.25189*	0.351898*
	(11.9477)	(4.29542)	(4.69213)	(0.10107)	(0.076264)
R^2	0.191806	0.289050	0.306184	0.462	0.576

The low population sample excludes Madrid Barcelona and Valencia. See text for definition of simple distance. Heteroskedasticity-consistent standard errors are reported in parentheses
*Statistically significant

Table 12 Krugman model: 47 regions, *hub-and-spoke* distance

	1981	1988	1995	1988–1981	1995–1988
θ	5.09561*	5.15818*	6.26141*		
	(1.24675)	(1.71052)	(1.55941)		
σ	5.74790*	5.17031*	6.17103*	1.71136*	2.40186*
	(1.65620)	(1.61419)	(2.02112)	(0.34181)	(0.837591)
μ	−14.1846	−14.2091	−12.2218	0.85041	0.866413*
	(33.4233)	(13.9783)	(8.51957)	(0.05469)	(0.07689)
τ	0.00235941*	0.00197196*	0.00136783*	0.02637*	0.025902*
	(0.00073821)	(0.00058956)	(0.00040003)	(0.00406)	(0.010335)
$\sigma/(\sigma-1)$	1.21062*	1.23979*	1.19339*	2.40575*	1.71334*
	(0.073470)	(0.092815)	(0.075586)	(0.67131)	(0.426208)
$\sigma(1-\mu)$	87.2796	78.6356	81.5920	2.256	0.320857*
	(54.202)	(76.0423)	(57.2048)	(0.09842)	(0.07860)
R^2	0.259542	0.274246	0.256757	0.462	0.47

All 47 regions are from Continental Spain. See text for definition of *hub-and-spoke* distance. Heteroskedasticity-consistent standard errors are reported in parentheses
*Statistically significant

economic geography in the period under analysis. In general, then, results are not very sensitive to the controls performed, what can lead us to have some confidence in the values estimated for the parameters.

It is now compared results in this paper with those from similar studies (see Table 1). Mion (2004) and Kiso (2005) do not test for the MPF. Also, only Roos (2001) and De Bruyne (2003) estimate both the version in levels and differences. In these two papers results differ a little in the two specifications. As such it is only done a direct comparison in between the results in levels and in between the results in differences. First, note that in differences in Roos (2001) and in De Bruyne (2003) the distance parameter has like in here the

Table 13 Krugman model: 44 regions, *hub-and-spoke* distance

	1981	1988	1995	1988-1981	1995-1988
θ	5.45880*	5.38976*	6.28826*		
	(1.06903)	(1.55234)	(1.51513)		
σ	6.28083*	5.08207*	5.63538*	2.36442*	2.70491*
	(1.84497)	(1.47593)	(1.76320)	(0.92564)	(0.916013)
μ	−13.9324	−9.38514	−7.70320	0.85327*	0.863101*
	(34.7493)	(6.04851)	(3.64724)	(0.04337)	(0.070138)
τ	0.00215440*	0.00194709*	0.00138046*	0.00791*	0.0067811
	(0.00074135)	(0.00058332)	(0.00039474)	(0.000593)	(0.003849)
$\sigma/(\sigma-1)$	1.18936*	1.24497*	1.21573*	1.73291*	1.58654*
	(0.066158)	(0.088574)	(0.082060)	(0.08178)	(0.315138)
$\sigma(1-\mu)$	93.7879	52.7780	49.0458	0.34693*	0.370299*
	(59.298)	(36.3824)	(28.0097)	(0.09796)	(0.076336)
R^2	0.219030	0.245557	0.239098	0.454699	0.514

The low population sample excludes Madrid Barcelona and Valencia. See text for definition of *hub-and-spoke* distance. Heteroskedasticity-consistent standard errors are reported in parentheses
*Statistically significant

Table 14 Krugman model: 47 regions, regional *dummy*

	1981	1988	1995	1988–1981	1995–1988
θ	4.85457*	4.81260*	6.06184*		
	(1.94272)	(2.33090)	(1.90418)		
σ	4.90203*	4.52541*	5.45635*	1.72833*	2.5948*
	(2.37836)	(2.03213)	(2.35637)	(0.34947)	(0.879527)
μ	−9.77219	−8.25818	−7.83685	0.85681	0.857656*
	(8.47134)	(5.79067)	(4.19944)	(0.01555)	(0.075364)
τ	0.00384516*	0.00258011*	0.00174762*	0.01307*	0.012915*
	(0.00155598)	(0.00104799)	(0.00070904)	(0.003879)	(0.006128)
Dummy	−0.00728889	0.041671	0.033197	0.023766	0.024231
	(0.079547)	(0.077229)	(0.058838)	(0.01975)	(0.018399)
$\sigma/(\sigma-1)$	1.25628*	1.28366*	1.22440*	2.373*	1.62704*
	(0. 156206)	(0.163506)	(0.118655)	(0.65881)	(0.34581)
$\sigma(1-\mu)$	52.8056	41.8970	48.2170	0.24748	0.369353*
	(64.6893)	(40.4851)	(38.7560)	(0.03979)	(0.077508)
R^2	0. 153441	0.189104	0.182050	0.42183	0.566826

All 47 regions are from Continental Spain. See text for definition of the regional dummy variable and simple distance. Heteroskedasticity-consistent standard errors are reported in parentheses
*Statistically significant

wrong sign. Second, all parameters estimates in differences in Roos (2001) are not statistically significant. The same also happens for the distance parameter in levels in De Bruyne (2003). In terms of magnitudes, in differences, ours estimates for α are higher than those for US (Hanson 1998, 2005) and Belgium (De Bruyne 2003). In levels that is also the case in relation to Belgium and Germany (Brakman et al. 2004), but lower than those for Western Germany (Roos 2001). Hence, the effects of purchasing power on wages seem to be bigger

Table 15 Krugman model: 44 regions, regional *dummy*

	1981	1988	1995	1988-1981	1995-1988
θ	1.92928	2.22947	4.19035*		
	(2.51109)	(2.21527)	(1.64267)		
σ	2.72408*	2.51597*	3.01495*	1.76945*	2.91092*
	(0.965437	(0.652405)	(0.720998)	(0.7673)	(1.07378)
μ	−3.36226	−2.23676*	−2.40301*	0.86124*	0.876778*
	(2.24770)	(0.948816)	(0.842855)	(0.041695)	(0.06925)
τ	0.00561697*	0.00494844*	0.00339231*	0.01553*	0.010065*
	(0.00229792)	(0.00152213)	(0.00087794)	(0.000939)	(0.004482)
Dummy	−0.069194	−0.00794319	−0.00215603	0.01525	−0.004813
	(0.038631)	(0.024094)	(0.017145)	(0.012428)	(0.009174)
$\sigma/(\sigma-1)$	1.58002*	1.65964*	1.49629*	2.29963*	1.52331*
	(0.324796)	(0.283880)	(0.177584)	(0.07512)	(0.294056)
$\sigma(1-\mu)$	11.8831	8.14359	10.2599*	0.24553*	0.35869*
	(10.1021)	(4.33415)	(4.79726)	(0.11042)	(0.079921)
R^2	0.201168	0.289190	0.306199	0.463642	0.5781

The low population sample excludes Madrid Barcelona and Valencia. See text for definition of the regional dummy variable and simple distance. Heteroskedasticity-consistent standard errors are reported in parentheses
*Statistically significant

in Spain than in the US, Belgium and Germany, but lower than that for Western Germany.

In turn, as shown in Table 1, our results resemble very closely those by Paluzie et al. (2005). This is especially important since we use a different database from them. The only difference is the time evolution of β: while in Paluzie et al. (2005) the importance of distance increases from 1955 to 1995, here it decreases. This maybe due to the fact that we consider a more recent period (1981–1995) where Spain went through more consistent improvements in transport infrastructure.

Summing up, in Spain nominal wages are positively correlated with the distance-weighted sum of personal income in surrounding regions, confirming therefore Harris (1954) market potential hypothesis. In addition in spite of some differences, our results are also in line with those from similar studies.

Krugman model

This subsection shows results for the KM. Since these are fairly similar across estimations it is only discussed in detail results in levels and differences for all regions and simple distance (Table 10). Econometric results under the other specifications are showed in Tables 11, 12, 13, 14 and 15.

Starting with the levels specification; as predicted by the theoretical model the parameter elasticity of substitution (σ) is bigger than one (ranging from 4.27 in 1988 to 5.19 in 1995). This indicates that markets in Spain are imperfect competitive. As a consequence the price-cost margin ($\sigma/(\sigma-1)$) is also bigger than one and comprehended between 1.23 (in 1995) and 1.3 (in 1988).

This shows that industrial production in Spain is subject to increasing returns to scale. The transport costs parameter (τ) as expected is bigger than zero, meaning that greater distance from consumer markets reduces wages in a region. The point estimates for τ have its maximum in 1981 (0.003 per kilometer) and its minimum in 1995 (0.001 per kilometer).

As it can be seen, most parameters are statistically significant and consistent with the model hypothesis. The exception is the parameter μ and, in consequence, the product $\sigma(1-\mu)$. In fact, estimations for μ are extremely high not satisfying the theoretical restriction, $0 < \mu < 1$. In our view, the main reason for this to happen is that the levels specification does not control for many of the endogeneity problems posed by the estimation of the Krugman model (and also the Helpman 1998 version). In fact Hanson (1998, 2005) abstains from showing results in levels for the US because they are implausible.

Consider now the results in differences. Contrary to the specification in levels, now all parameters are not only statistically significant but also satisfying the theoretical restrictions. The elasticity of substitution is bigger than one indicating again imperfect competitive markets. Also, since $\sigma/(\sigma-1) > 1$, industrial production in Spain is subject to increasing returns to scale and the price-cost margin estimates are between 1.7 (in 1988–1981) and 2.4 (in 1995–1988). The share of expenditure spent on industrial goods (μ) is now statistically significant and most importantly in accordance with the theory. To be precise, differentiated goods have a share in consumption around 0.85. The trade cost parameter is also statistically significant and consistent with the theoretical constraint. The impact of distance in this period is about 0.02 per kilometer. In turn, since $\sigma(1-\mu) < 1$, industrial production in Spain has a tendency to agglomerate for any value of trade costs.

Regarding the time evolution, in levels σ decreases from 1981 to 1988 but increases in 1995. In differences, σ increases from the first period (1988–1981) to the second (1995–1988). In this sense markets have become more competitive over time in Spain. This can be the result of the Single Market program by the European Community that opened the Spanish market to foreign competition. In turn, confirming the MPF findings, trade costs decreased in importance in the period under analysis (for both the specification in levels and in differences). Also, in differences it can be seen that the share of manufactured goods in consumption has remained fairly stable in the 1980s and 1990s.

The introduction of other controls (Tables 11, 12, 13, 14 and 15) gives parameter estimates similar to the ones previously reported. Furthermore the time evolution of the different parameters does not change in both the specification in levels and in differences. In terms of magnitudes there are however some variation. The parameter σ in levels is lower with the restricted sample plus simple distance (Table 11) or plus dummy variables (Table 15). Though, the same does not happen in the restricted sample with HAS distance (Table 13). This might indicate that this measure of distance is more appropriate for the Spanish economic geography. In differences however the value of σ is fairly constant across the different specifications. The same occurs with μ in differences and τ in levels and in differences. Though, note that the impact of distance seems to be slightly bigger in differences than in levels. Related with the regional controls specification (Tables 14, 15), as for the MPF, the regional dummy is never statistically significant. Once again we fail to capture regional technological dynamics. Results are then generally consistent across specifications, what means that they are not sensitive to different controls.

Given that in differences it is obtained a better fit and parameter estimates are more consistent with theory, this specification is (for the sample and the spatial unit of analysis) preferred to the one in levels. The same happens in Hanson (1998, 2005) as it was already mentioned above. However, this is not the case in Brakman et al. (2004), Roos (2001), De Bruyne (2003), Kiso (2005) and Mion (2004) where the levels specification is favored. We think that when the levels specification is preferred, the intercept term in Eqs. (10) and (11) contains a lot of relevant information, i.e., the geography of wages is explained for other reasons besides those given by the MPF or the KM. If the contrary happens (i.e., if the differences specification is preferred), then, the MPF and the KM alone explain much of the spatial distribution of wages. This may also elucidate on the discrepancy between our results with the MPF and the KM: in the former the levels specification is favored while in the latter the contrary happens. Conversely, the MPF in levels performs better than the MPF in differences because in the first the intercept term captures what this reduced form equation is unable to cover up in the second. On the contrary, since the augmented KM equation already encompasses many spatial determinants of regional wages, the differences specification is able to do better than the one in levels because the former controls for some of the econometric problems discussed above.

It is now analyze our results in the light of those from similar studies (see Table 1). In what relates to the parameter μ, this is also awkward in the papers that use the Helpman (1998) model (Hanson 1998, 2005; Roos 2001; Brakman et al. 2004; De Bruyne 2003; Mion 2004; Kiso 2005). As discussed in the introduction, this possibly results from the housing sector being modeled as non-tradable in a regional setting. For example in Hanson (1998, 2005), μ is very close to one meaning that only a negligible part of income is spent in housing (what is not at all true). The same is the case in Roos (2001) and Mion (2004).[16] Also Brakman et al. (2004) estimate values for μ (implausible) close to ours in levels. In addition, in Kiso (2005) and De Bruyne (2003) μ also does not follow the theoretical restriction. It is then questionable the appropriateness of the either the KM or the Hanson–Helpman version to estimate this parameter.

In what respects trade cost, comparisons are more difficult since different measures of distance as well as alternative specifications for the distance function are used across studies. Nonetheless, τ is many times not statistically significant (for example in Roos 2001; Mion 2004; Kiso 2005) or has the opposite sign to what is predicted by the theory (Brakman et al. 2004) or it implies a very

[16] This only happens in Mion (2004) for the NLS estimation. In the IVNLS estimation, μ is between 0.16 and 0.18, which is too low. However in the linearized version, μ ranges from 0.77 to 0.78, which is more in accordance with the share of income spent in housing in Italy (about 20%).

limited impact of distance (Hanson 1998, 2005). On the other hand, ours estimates for τ are in accordance to what is standard in the trade-gravity literature (see the comprehensive review by Anderson and van Wincoop 2004).

Turning now to σ, estimates for this parameter are consistent with those from previous studies. However, while in here σ is always statistically significant, the same is not the case for example in Roos (2001), De Bruyne (2003) and Kiso (2005). In terms of magnitudes, σ is somewhat smaller in Spain than in US. This indicates that Spanish consumers value variety more than US consumers, implying higher mark-ups in Spain. This is also the case for the other EU country studies (Brakman et al. 2004; Roos 2001; De Bruyne 2003; Mion 2004). Also, our estimates for σ are in agreement with those found in the gravity-type of models (see again Anderson and van Wincoop 2004).

In what respects the price-cost margin, most studies indicate the presence of economies of scale at work. However, this ratio is not statistically significant in some of them (for example Brakman et al. 2004; Kiso 2005) and in others the presence of economies of scale is not always confirmed (for example Brakman et al. 2004; Kiso 2005 and in the NLS estimation of Mion 2004).

Additionally, accordingly to ours results Spanish economic geography has a tendency to agglomerate independently of trade costs (since $\sigma(1-\mu) < 1$). The contrary is the case for Germany and Japan, where the parameter estimates for $\sigma(1-\mu)$ indicate that dispersion is always a spatial equilibrium independently of trade costs. In turn for US and Italy, trade costs are important in determining the economic landscape of these two countries. Then, while agglomeration is a very stable feature of the Spanish economy, this is not exactly the same in the US, Italy, Germany and Japan.

In summary: the spatial wage structure in Spain seems to be explained by economies of scale (that makes it important to be closer to larger markets), the expenditure share of manufactures in consumption (that supports larger agglomerations) and distance from consumer markets (that is reduced when in central locations). This confirms the NEG theoretical findings and the empirical results from similar studies.

Discussion

This paper estimates Krugman (1991) economic geography model to the NUTS 3 regions of Spain. Accordingly to estimates obtained, industrial production in Spain is subject to increasing returns to scale and markets are of an imperfect competitive nature. Furthermore, the local level of economic activity and distance from central locations can explain regional wages differences, i.e., larger markets have potentially higher wages.

As such this paper confirms results from other similar studies that use instead the Helpman (1998) variant of the Krugman (1991) model as theoretical basis for the empirical implementation (like Hanson 1998, 2005; Roos 2001; Brakman et al. 2004; De Bruyne 2003; Mion 2004; Kiso 2005).

Nonetheless some questions were overlooked in this study. First this paper focus in a limited set of location determinants. Consequently, it does not test against competitive location theories, but it only looks at some of the hypotheses of the NEG. Others, like Davis and Weinstein 1999, have included additional spatial economic explanations as endowments and comparative advantage. Second, trade costs are assumed to be homogeneous and constant across regions and industries. However, as shown by Limão and Venables 2001 trade costs are heterogeneous across these dimensions. Third, issues related with industrial structure and the economic characteristics of regions were putted aside, but these are certainly important in balancing regional outcomes. Knarvik et al. 2002a present evidence that such is the case for Europe. Fourth, international trade and multinational investment considerations were putted aside. Though, it is acknowledge (see for example Knarvik et al. 2002b; Head et al. 1995) that both are important catalysts of regional development. Fifth, it was ignored the role of "big agents" (i.e., the state) in shaping the economic landscape (see Holmes 1998). This is certainly important for Spain given the weight that regional and central governments have in the economy. On true, no study can cover all these aspects; nevertheless empirical literature on the field should be able to take them into consideration.

In any case, the objective of this paper was to make a structural estimation of the Krugman (1991) model, which focus only in a limited set of spatial explanations, i.e., mainly increasing returns. To introduce other factors (like the ones mentioned above) would oblige us to do this in ad-hoc non-theoretical based way. We have, then decided to keep our empirical model the closest possible to theory in order to test for the NEG previsions. Given our results, we believe that this type of structural models, in spite of being simple, can be an important tool to study empirically the effects of space on economic activity.

Acknowledgements The author is in debt to Renato Flôres, Paula Fontoura, German Calfat, José Pedro Pontes, Margarida Proença, João Santos Silva, Peter Neary and Dermot Leahy for helpful discussion and comments during the preparation of this work. This version benefited greatly from the remarks and suggestions of two anonymous referees. Part of this research was conducted at the University College Dublin as a Marie Curie Fellow. My activity was supported by a grant from *Fundação para a Ciência e a Tecnologia* (Praxis XXI/BM/17786/98). The usual disclaimer applies.

References

Alonso-Villar O (2005) The effects of transport costs revisited. J Econ Geogr 5:589–604
Anderson J, van Wincoop E (2004) Trade costs. J Econ Lit 42:691–751
Baldwin R, Forslid R, Martin P, Ottaviano G, Robert-Nicoud F (2003) Public policies and economic geography. Princeton University Press, Princeton
Biehl D (1986) L'Impact de L'Infrastructure Sur le Development Regional. Commission des Communautés Européennes, Luxembourg
Brakman S, Garretsen H, Schramm M (2004) The spatial distribution of wages: estimating the Helpman-Hanson model for Germany. J Reg Sci 44:437–466
Brülhart M (1998) Trading places: industrial specialization in the European union. J Common Market Stud 36:320–345
Combes P-P, Lafourcade M (2001) Transport costs decline and regional inequalities: evidence from France. CEPR Discussion Paper 2894

Combes P-P, Overman H (2004) The spatial distribution of economic activities in the European union. In: Henderson J, Thisse J-F (eds) Handbook of regional and urban economics, vol 4. North Holland, New York
Davidson R, MacKinnon J (1993) Estimation and inference in econometrics. Oxford University Press, Oxford
Davis D, Weinstein D (1999) Economic geography and regional production structure: an empirical investigation. Eur Econ Rev 43:379–407
De Bruyne K (2003) The location of economic activity: is there a spatial employment structure in Belgium? Mimeo Katholiek University of Leuven
Dixit A, Stiglitz J (1977) Monopolistic competition and optimum product diversity. Am Econ Rev 67:297–308
EUROSTAT 1996 REGIO: Banque de Données Régionales, Description du Contenu. Commission des Communautés Européennes, Luxembourg
Fujita M, Thisse J-F (2002) Economics of agglomeration: cities, industrial location and regional growth. Cambridge University Press, Cambridge
Fujita M, Krugman P, Venables A (1999) The spatial economy: cities, regions, and international trade. MIT Press, Cambridge
Garcia Pires AJ (2005) Market potential and welfare: evidence from the Iberian Peninsula. Port Econ J 4:107–127
Haaland J, Knarvik K-H, Forslid R (2002) A U-shaped Europe? A simulation study of industrial location. J Int Econ 57:272–297
Hanson G (1998) Market potential, increasing returns, and geographic concentration. NBER Working Paper 6429
Hanson G (2005) Market potential, increasing returns, and geographic concentration. J Int Econ 67:1–24
Harris C (1954) The Market as a factor in the localization of production. Ann Assoc Am Geogr 44:315–348
Head K, Mayer T (2004) The empirics of agglomeration and trade. In: Henderson J, Thisse J-F (eds) Handbook of regional and urban economics, vol 4. North Holland, New York
Head K, Ries J, Swenson D (1995) Agglomeration benefits and location choice. J Int Econ 38:223–248
Helpman E (1998) The size of regions. In: Pines D, Sadka E, Zilcha J (eds) Topics in public economics: theoretical and applied analysis. Cambridge University Press, Cambridge
Holmes T (1998) The effects of state policies on the location of manufacturing industry: evidence from state borders. J Polit Econ 106:667–705
Keeble D, Offord J, Walker S (1988) Peripheral regions in a community of twelve member states. Commission of the European Communities, Luxembourg
Kiso T (2005) Does new economic geography explain the spatial distribution of wages in Japan? Mimeo University of Tokyo
Knarvik K-H, Overman H, Redding S, Venables A (2002a) The location of european industry. Euro Econ 2:216–273
Knarvik K-H, Overman H, Redding S, Venables A (2002b) Integration and industrial specialization in the European union. Rev Econ 53:469–481
Krugman P (1980) Scale economies, product differentiation, and the pattern of trade. Am Econ Rev 70:950–959
Krugman P (1991) Increasing returns and economic geography. J Polit Econ 99:413–499
Krugman P (1993) On the number and location of cities. Eur Econ Rev 37:293–298
Limão N, Venables A (2001) Infrastructure, geographical disadvantage and transport costs. World Bank Econ Rev 15:451–479
Magrini S (1999) The evolution of income disparities among the regions of the European union. Reg Sci Urban Econ 29:257–281
Martin P (1999) Public policies, regional inequalities and growth. J Public Policies 72: 85–105
Mion G (2004) Spatial externalities and empirical analysis: the case of Italy. J Urban Econ 56:97–118
Neary P (2001) Of hypes and hyperbolas: introducing the new economic geography. J Econ Lit 39:536–561
Neven D, Gouyette C (1995) Regional convergence in the European community. J Common Market Stud 33:47–65

Overman H, Redding S, Venables A (2003) The economic geography of trade, production and income: a survey of empirics. In: Harrigan J, Choi K (eds) Handbook of international trade, vol 4. Basil Blackwell, London
Paluzie E, Pons J, Tirado D (2005) Potencial de Mercado y Estructura Espacial de los Salarios en las Regiones Españolas, 1955–1995. Cuad Aragon Econ 15:1–25
Puga D (1999) The rise and fall of regional inequalities. Eur Econ Rev 43:303–334
Puga D, Overman H (2002) Unemployment clusters across European regions and countries. Econ Policy 31:115–143
Quah D (1996) Regional convergence clusters across Europe. Eur Econc Rev 40:951–958
Redding S, Venables A (2000) Economic geography and international inequality. J Int Econ 34: 380–400
Roos M (2001) Wages and market potential in Germany. Jahrb Regionalwissensch 21:171–195
Samuelson P (1954) The transfer problem and transport costs. Econ J 64:264–289
Venables A (1996) Equilibrium locations of vertically linked industries. Int Econ Rev 37:341–359