

Sanchis-Guarner, Rosa; Lopez-Bazo, Enrique

Conference Paper

Are Skilled Workers More Attracted to Economic Agglomerations?

46th Congress of the European Regional Science Association: "Enlargement, Southern Europe and the Mediterranean", August 30th - September 3rd, 2006, Volos, Greece

Provided in Cooperation with:

European Regional Science Association (ERSA)

Suggested Citation: Sanchis-Guarner, Rosa; Lopez-Bazo, Enrique (2006) : Are Skilled Workers More Attracted to Economic Agglomerations?, 46th Congress of the European Regional Science Association: "Enlargement, Southern Europe and the Mediterranean", August 30th - September 3rd, 2006, Volos, Greece, European Regional Science Association (ERSA), Louvain-la-Neuve

This Version is available at:

<https://hdl.handle.net/10419/118210>

Standard-Nutzungsbedingungen:

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

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

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

Terms of use:

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

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

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

Are Skilled Workers More Attracted to Economic Agglomerations?

Rosa Sanchis-Guarner^{*†}- Enrique López-Bazo[‡]
AQR Research Group - IREA, Universitat de Barcelona
Preliminary draft-May 2006

Abstract

The New Economic Geography (NEG) models explain the formation and the consolidation of economic agglomerations through a self-sustained process in which a trade-off between centripetal and centrifugal forces takes place. The centripetal forces work through two mechanisms: those known as the backward and the forward linkages. The latter mechanism predicts that workers will be attracted towards economic agglomerations by higher real wages due to smaller price indexes in those regions with greater market potential. Recent contributions have tried to assess this assumption simultaneously estimating some of the structural coefficients of the NEG models. Nevertheless, in all these cases, migrants have been considered to be homogenous. Conversely, the basic hypothesis of this paper suggests that the effect of the market potential on the probability of migrating depends on the human capital level of the workers. The introduction of this assumption is consistent with previous empirical evidence that shows that human capital increases the probability of migration of the workers, accentuating the opportunity cost of not migrating. In agreement with the predictions derived from our theoretical model, the results obtained for the inter-provincial migrations in Spain in the two last decades confirm that human capital has played a remarkable role in determining migration of skilled workers towards economic agglomerations. Specifically, a higher human capital in both home and host regions of the migrants has a positive effect on the predicted aggregated migration flows between these two regions.

1 Introduction

One of the most remarkable features of the Spanish economic performance during the last 50 years has been the exceptional migration movements that took place in both the

^{*}Corresponding author: Departament d'Econometria, Estadística i Economia Espanyola, Facultat de Ciències Econòmiques i Empresariales, Avd. Diagonal, 690, 08034 Barcelona; email: r.sanchis-guarner@ub.edu, elopez@ub.edu; Tel: +34 934021011; +34 934037041; Fax: +34 934021821

[†]R.Sanchis-Guarner acknowledges financial support from the Ministerio de Educación y Ciencia, Secretaría de Estado de Universidades e Investigación, Programa Nacional de Formación de Profesorado Universitario.

[‡]E. López-Bazo acknowledges financial support from the Ministerio de Educación y Ciencia, Plan Nacional de I+D+I, proyecto SEJ2005-00714/ECON.

internal and the external dimensions. The relocation of people, especially if they are workers, has significant consequences on the regions that receive and that expel migrants, and even more if the relocation of workers affect the relative human capital endowments.

In this work we try to assess empirically one of the less treated mechanisms of agglomeration suggested in the New Economic Geography (NEG) literature: the forward linkage. Our contribution is to take into account the human capital endowments of migrants, a factor that had not been considered in the previous contributions that have addressed the movements of workers towards agglomerations in a NEG framework. Our basic hypothesis is the fact that, as it has been stressed in the literature, educated migrants are more prone to migrate in response to regional differences in wage and employment rates.

We follow Crozet (2004) in developing a theoretical framework from which we specify a core equation that relates migration to agglomerations and to human capital endowments. A reduced form of this equation is estimated empirically for the Spanish provinces for the period 1988-2002. Results suggest that human capital endowments of both the home and the host region matter when explaining migration, and consequently point to misspecification of those previous contributions that relate migration to market potential without considering the role played by human capital. As a result, previous estimates might be seriously biased as omitted human capital endowments are expected to be correlated with the measures of agglomeration used in those studies.

The paper is organised as follows. In section 2 we revise the literature on NEG and on migrations, paying special attention to the role that human capital plays in the movement of workers. In addition, the recent evolution of Spanish internal inter-provincial movements is reviewed to motivate the article. In section 3 the theoretical framework is developed in two subsections. In the first one, we review the main characteristics of the theoretical model developed in Crozet (2004). In the second one, we introduce a modification to consider the role that human capital may be playing in the movements of workers towards economic agglomerations. In section 4 we describe the data sources and we perform some exploratory analysis to briefly describe the recent Spanish migration patterns, the direction of the flows and the relationship between some variables of interest. In section 5 we address the empirical validation of the model. Firstly we settle the two estimable counterparts of the reduced forms of the theoretical equations developed in section 3, with and without taking into account human capital of the home and the host regions. These estimable equations are similar to traditional gravity equations. Next we describe some econometric issues and the estimation strategy followed by the analysis of the major results obtained. Finally, section 6 concludes.

2 Migrations, agglomerations and the role of human capital

This work considers whether the human capital of the migrants may have played an important role in the new patterns of migration flows that appeared in Spain since the late 70's. To address this issue we adopt a framework which to our knowledge has never been

used before in the context of the analysis of internal Spanish migrations taking into account human capital. This framework is New Economic Geography (henceforth NEG), the bases of which were formulated in the beginning of the 90's in the seminal papers of Krugman (1991 and 1992) and Venables (1996). Models of general equilibrium with monopolistic competition are used to explain why agglomerations may arise in certain locations and which are the mechanisms that cause their emergence and their reinforcement. Theoretical advances have been abundant (Ottaviano and Puga, 1998; Neary, 2001). However, Fujita et al (1999) have highlighted the importance of assessing empirically the validation or the falsification of the theoretical foundations of the NEG, while some authors (Head and Mayer, 2003) have pointed out that a lot of research in this field remains to be done. In this sense, our paper provides new evidence on one of the mechanisms of the NEG to which less attention has been addressed: the movements of workers towards economic agglomerations (forward linkage).

The NEG models explain the formation and the consolidation of economic agglomerations through a self-sustained process in which a trade-off between centripetal and centrifugal forces takes place. The interaction of scale economies, interregional labour migration, transport costs and a spatially immobile source of demand results in a Hirschman-type (1958) self-sustained cumulative process that explains the formation and the maintenance of economic agglomerations. The centripetal forces work through two mechanisms: those known as the backward and the forward linkages. The backward linkage explains the attraction of firms toward regions with high market potential due to scale economies and the presence of transport costs. A higher access to market assures higher profits for the firm and thus higher nominal wages. The forward linkage explains the attraction of mobile workers towards highly industrialised regions. Workers would migrate attracted by higher real wages in regions with higher market potential as the price indexes in these locations are smaller.

The empirical works in NEG can be classified in four broad categories. On the one hand, the more numerous are those that try to validate some aspects related to the backward linkage. Some works study the "home-market effect", or the fact that location near big markets provides advantages to local firms. In this line we find Davis and Weinstein (1999, 2003) or Feenstra et al (1998). A second stream studies the effect of market potential on factor prices, and more specifically, the spatial correlation of nominal wages and market potential. Some studies estimating wage equations are those of Hanson (1998), Tirado et al (2003), Redding and Venables (2004), Combes and Overman (2004) and Head and Mayer (2005). Other works have focused in the study of influence of market potential in firms' locations, say, if firms tend to locate in regions with high access to demand. For instance, we find the article of Crozet et al (2004), Head and Mayer (2002) or Forslid et al (2002). Finally, few studies focus in the forward linkage, such as that of Crozet (2004).

Crozet (2004) estimates a reduced form derived from a theoretical model with three sectors and R regions within the NEG framework, in the spirit of Hanson's (1998) modification of Krugman's model (1991). Migration choice results from a comparison of the

perceived quality of life in the various locations. A worker will migrate due to regional differences in wages and in employment probabilities, but his probability of movement will be dampened by migration cost, which is assumed to depend only on the distance between the home and the host region. An estimation based in this theoretical model for five European countries over the 80's and the 90's shows strong evidence in favour of his model, confirming that migrants follow market potentials. In the same line, two other articles analyse migration flows following Crozet's (2004) and Crozet's (2002) approaches. They are, respectively Pons et al (2004) for the Spanish provinces in the inter-war years (20's and 30's) and Zuh and Poncet (2003) for workers across China between 1985 and 1995. Nevertheless, in all these cases migrants have been considered to be homogenous in their skills level.

Conversely, the basic hypothesis of this paper suggests that the effect of the market potential on the probability of migrating might be influenced by the human capital level of workers (potential migrants). We thus follow Crozet's approach taking into account that the labour force is heterogeneous in their level of skill.

Relocation of labour that affects to regions and cities through migration movements is one of the most important problems of the regional and spatial economy, not only because this phenomenon gives us the opportunity to verify key hypotheses of the theory of regional growth (does interregional mobility of the factors produce convergence or divergence in the regional income per capita?), but also because it illustrates the advantages of taking into account in the economic analysis the space and the geographical distribution of economic activity across it.

Studies on migrations are usually classified in those that deal with the causes and those that deal with the consequences of labour migrations. The first ones, focused in analysing the forces that originate the movements, are much more abundant. They normally specify a "migration equation" relating a variable embodying the migration phenomenon (in flows or in rates) to variables representing the pecuniary and non-pecuniary conditions of the host and the home regions that might cause the movements. This equation is generally empirically tested in order to assess the main determinants of migrations. However, the analysis of the consequences the relocation of population has, and particularly the relocation of the labour force, is much more important because of the effects it may have on regional growth and on the differences of income per capita or on employment rates across space. Moreover, both the causes and the consequences of the phenomenon are strongly interrelated and in most cases it is very difficult to tell the causes from the consequences, since normally there is an enormous interdependence between the phenomena we are considering (convergence, income, employment, wages and migrations).

The theoretical and empirical debate on regional convergence or divergence is closely related to migration across regional boundaries. Classical equilibrium models state that workers movements following higher incomes in richer regions will end in an equalization of wages and therefore of income per capital across regions. Other models, as the endogenous human capital and the disequilibrium model, are based on the fact that the movements

of the labour force, as they normally occur from the regions with the worse economic endowments towards the regions with the better ones, result in a widening of the differences between them, so the final consequence of migrations is an increase in spatial disparities. Transfers of labour force from peripheral to central regions, especially of skilled workers, will have important consequences in both home regions (brain drain) and host regions (increase in productivity, knowledge spillovers). See Greenwood (1993), Dolado et al (1994), Friedberg and Hunt (1995), Beine et al (2001) and Borjas (1999) for discussion and empirical evidence on the economic consequences of migrations. Empirical evidence for the Spanish case is mixed: some papers prove that workers movements contribute to decrease differences in regional income (Raymond and García-Greciano, 1999; Dolado, González-Páramo and Roldán, 1994) while other authors show the diverging role of labour migration in the process of interregional convergence (Bentolila and Dolado, 1991).

The migratory phenomenon in Spain has been object of numerous studies (for a review see Bover and Velilla, 2005; De la Fuente, 1999 and Ródenas, 1994a). Although most studies are carried out using aggregate data, some works use individual data (Gil and Jimeno, 1993; Antolín and Bover, 1997; Bover and Arellano, 2002; Devillanova and García Fontes, 2004) to identify which regional economic factors may influence migration decisions, taking into account personal characteristics. These works highlight the importance of the interactions between individual characteristics and regional variables. Personal characteristics, such as age, gender or being registered as unemployed, not only have an important direct effect on migration but they also alter the effect of regional economic variables on migration. In this sense, human capital of the potential migrant would have a noteworthy role in the probability of migrate to a region. Several studies point out that, both at the Spanish (Ródenas, 1994b; Devillanova and García-Fontes, 2004; Antolín and Bover, 1997; Serrano, 1998) as at the international level (Molho, 1987; Owen and Green, 1992; Ritsilä and Ovaskainen, 2001; Ritsilä and Haapanen, 2003; Pekkala, 2003; Greenwood, 1997), a greater human capital endowment of the worker increases its propensity to migrate.

This idea agrees with the human capital theory (Sjaastad, 1962) that contemplates migration as an individual decision resulting from a utility maximising process. The migration decision is assumed to be the outcome of personal, household and regional characteristics. Individuals maximise their expected wage according to the human capital they are bearing (Becker, 1975). The potential migrant calculates its expected net benefit as the difference of the individual costs and benefits issuing from a change in location, so he compares his expected benefit net of migration costs in his home region and in the various alternative regions. We should take into account not only the pecuniary factors (wages or employment probability) but also the non-pecuniary ones as, for instance, a more pleasant physical or cultural environment or psychological and cultural costs derived from a change in location. These latter factors, although they can be crucial in the decision, do not enter in the cost-benefit equation, as they do not have a direct translation in monetary units. Nevertheless, some studies control for these non-pecuniary factors, as economic variables have shown to be limited in their explanatory power of migration movements.

After the high gross migration flows during the 50's and the 60's, during the decade of the 70's until the mid 80's the Spanish internal migration flows have been said to be scarce. These low flows are in accordance with the European context of low mobility (Decressin and Fatàs, 1995) and specially if compared with the US. Some reasons given in the literature have been the existence of rigidities in the labour and in the housing markets (mobility barriers) or a generalised increase of the unemployment rate since mid-80's. As asserted in Serrano (1998), low migration flows may be a consequence of an unequal distribution of human capital across Spanish provinces, so the difference in wages may be expressing regional differences in human capital endowments. In their review of inter-regional migration in Spain, Bover and Velilla (2005) show that following a very intensive period of internal migration (1960-1973), during the late 70's and early 80's there was a considerable decline in inter-regional migration in Spain. However, since 1982, after the prolonged fall of the earlier years and despite consistently high aggregate unemployment rates, gross inter-regional migrations started to increase. Indeed, since the early 80's and compared with the 1960-73 period it is not so much the number of inter-regional migrants that has changed but their characteristics and the causes that may be motivating their decisions to change their province of residence.

Actually, other works have suggested the existence of a change in the migratory patterns of the Spanish interregional flows from the mid-70's. For example, Antolín and Bover (1997) emphasize an increase of the migration flows going from and towards the richest regions (more balanced net flows than before). Indeed, short-distance migration in Spain has steadily increased, beginning from the early 80's as said in Ródenas (1994a). García-Barbancho (1960) and Delgado-Cabeza (1998) point out the increase of the intraregional migrations in detriment of interregional flows and also the increase of migration flows of workers pertaining to sectors that require a higher human capital. Moreover, as showed in Antolín and Bover (1997), higher education not only increases directly the probability of migrating but individuals with higher education tend to be more sensitive to their region's unemployment. The positive effect of education on migration is important and it is an indicator of the different profile of migrants as compared to the 1960-73 period. From the analysis of this work, Bover and Velilla (2005) conclude that reasons that make people migrate to another region in Spain in the 80's and early 90's are probably different from the reasons that made people move in the 60's and 70's. People that move between regions nowadays are people with higher education and they seem to do so in search of cheaper housing, better quality of life and perhaps professional promotion. Unemployment will only increase the probability of migration if the individual has higher education. In addition, Bover and Arellano (2002), using micro-data from the annual residential migration (Estadística de las Variaciones Residenciales - EVR) to analyse the determinants of the increase in intra-regional migrations from the 80's, show that education has a positive and significant effect on the probability of migrating.

3 Theoretical model

This section is divided in two subsections. In the first one, we expose the main features of the theoretical model developed in Crozet (2004). He combines a NEG framework, inspired in Hanson's (1998) modification of Krugman's model, with a model of migration based on Tabuchi and Thisse (2002). His aim is to relate labour migration to the geography of production through real wage differences, *i.e.*, to specify a migration equation relating worker's movements to the regions' market potentials. In a second subsection, we introduce a modification in the theoretical model to take into account the role of human capital on migratory movements.

3.1 Migration equation à la Crozet (2004)

There are R regions in the economy. There are 3 sectors of activity: a traditional sector (Z), producing an homogeneous good under perfect competition, and two sectors producing horizontally differentiated varieties under monopolistic competition. These sectors produce manufactures (X) and non-traded services (Y) respectively.

To produce their goods, the sectors use two factors of production, immobile and mobile labour. Each of these sector-specific factors is fixed in supply. Mobile workers are employed in the services ($L_{i,t}^Y$) and manufactured goods ($L_{i,t}^X$) sectors and immobile workers are employed in the traditional sector. Immobile workers are uniformly distributed across the economy while mobile workers can move to each and every region within the economy.

3.1.1 Migration choice

A mobile worker, k , can choose to locate in its home region, j , or move to any other location i within the economy ($i \in [1, R]$). His choice of location results from a comparison of the perceived quality of life in all the possible alternatives. Thus, the potential migrant compares the expected benefits of migrating to the cost of moving. For simplicity, we assume that the worker only considers pecuniary benefits, specifically, the real expected wage. On the other hand, the worker's movement involves a migration cost, which we have assumed to positively depend only on the distance between the home and the host regions. Workers will take the decision of moving to a specific location maximising their expected real wage net of migration costs.

Formally, if we denote $\omega_{i,t}$ as the real wage at date t , $\rho_{i,t}$ as the probability of finding an employment at date t (so $\omega_{i,t}\rho_{i,t}$ is the expected real wage at date t), d_{ji} as the distance between regions j and i , the migration decision between region j (home) and region i (any alternative host region) is given by the maximisation of the following objective equation:

$$\begin{aligned} \pi_{ji,t}^k &\equiv V_{ji,t}^k + \epsilon_i^k = \ln[\omega_{i,t}\rho_{i,t}] - \ln[d_{ji}(1 + bF_{ji})]^\lambda + \epsilon_i^k \\ &= \ln\left[\omega_{i,t}\rho_{i,t}[d_{ji}(1 + bF_{ji})]^{-\lambda}\right] + \epsilon_i^k, \quad i \in [1, R], \end{aligned} \quad (1)$$

where $[d_{ji} (1 + bF_{ji})]^\lambda$ is the migration cost, λ and b are strictly positive coefficients, and F_{ji} is a dummy variable which takes value 1 if regions j and i do not share a common border (migration cost increases if we move to non-contiguous regions). $V_{ji,t}^k$ is the net satisfaction of living in i at date t for a worker k located originally in region j and ϵ_i^k is an unobservable stochastic component capturing k 's personal perception of the characteristics of region i .

We assume that there exists a gap of time between the moment the worker compares the alternative locations and the moment of the migration choice. As a consequence, a worker k from region i will choose to locate in region i at date t if $V_{ji,t-1}^k > V_{jr,t-1}^k$, $\forall r \neq i$. Nevertheless, we can only determine a probability of migration and not the migration movement itself.

Therefore, in order to find a functional form to this probability we have to set an appropriate hypothesis over the distribution of ϵ_i^k : we assume that this term follows a negative double exponential distribution, which results in a multinomial logit probability (McFadden, 1974). Therefore, the probability of migrating from region j to region i is given by the logit function:

$$P(M_{ji,t}) = \frac{e^{V_{ji,t-1}^k}}{\sum_{r=1}^R e^{V_{jr,t-1}^k}}. \quad (2)$$

The expected migration flow from region j to region i at date t is $L_{j,t}P(M_{ji,t})$, where $L_{j,t}$ is the potential migrant population (mobile workers) in region j at date t , and the total migration outflow leaving region j is $L_{j,t} [1 - P(M_{jj,t})]$. The share of emigrants from region j choosing to go to region i is given by the equation:

$$\frac{migr_{ji,t}}{\sum_{i' \neq j} migr_{ji',t}} = \frac{L_{j,t}P(M_{ji,t})}{L_{j,t} [1 - P(M_{jj,t})]} = \frac{e^{V_{ji,t-1}^k}}{\sum_{r=1}^R e^{V_{jr,t-1}^k} - e^{V_{jj,t-1}^k}}. \quad (3)$$

Applying logarithms we finally obtain:

$$\ln \left[\frac{migr_{ji,t}}{\sum_{i' \neq j} migr_{ji',t}} \right] = \ln \left[\omega_{i,t-1} \rho_{i,t-1} [d_{ji} (1 + bF_{ji})]^{-\lambda} \right] + a_{j,t-1}, \quad (4)$$

where

$$a_{j,t-1} \equiv - \ln \left[\sum_{r=1}^R e^{V_{jr,t-1}^k} - e^{V_{jj,t-1}^k} \right]. \quad (5)$$

3.1.2 Consumption and production

Traditional goods are freely tradable across regions while services are non-tradable and manufactured goods are shipped between regions subject to traditional iceberg transport costs. We assume that a fraction of the good $(\tau_{ij} - 1) / \tau_{ij}$ melts away in transportation so $\tau_{ij} > 1$ units of the good have to be exported from region i to deliver one unit to region j . This transports cost is assumed to be an increasing function of the distance between the

two regions d_{ij} (Hummels, 1998):

$$\tau_{ij} = B d_{ij}^\delta, \quad \forall i \in [1, R], \delta > 1 \text{ and } B > 0. \quad (6)$$

Consumers have identical Cobb-Douglas preferences over goods:

$$U_{i,t} = C_{yi,t}^\phi C_{xi,t}^\mu C_{zi,t}^{1-\mu-\phi}, \quad (7)$$

where ϕ , μ , and $(1 - \mu - \phi)$ are the expenditure shares for manufactured goods, services and the traditional good respectively. Consumers demand all the manufactured varieties produced in any region of the economy but can only consume services produced in the region where they live, as services are non-tradable.

$C_{zi,t}$ is the quantity of traditional goods consumed in region i at date t . $C_{xi,t}$ is the composite of manufactured products varieties given by the CES function:

$$C_{xi,t} = \left(\sum_{m=1}^{n_{x,t}} c(m)_{xi,t}^{(\sigma_x-1)/\sigma_x} \right)^{\sigma_x/(\sigma_x-1)}, \quad i \in [1, R], \quad (8)$$

where σ_x denotes the elasticity of substitution between manufactured varieties, $c(m)_{xi,t}$ is the quantity consumed of variety m in region i at date t and $n_{x,t}$ is the number of available manufactured varieties produced in the economy ($n_{x,t} = \sum_{i=1}^R n_{xi,t}$). The composite of service varieties is also defined by a CES function as:

$$C_{yi,t} = \left(\sum_{m'=1}^{n_{yi,t}} c(m')_{yi,t}^{(\sigma_y-1)/\sigma_y} \right)^{\sigma_y/(\sigma_y-1)}, \quad i \in [1, R], \quad (9)$$

where σ_y denotes the elasticity of substitution between services varieties, $n_{yi,t}$ is the number of available service varieties produced within the region i and $c(m')_{yi,t}$ is the quantity consumed of variety m' in region i at date t .

Manufactures and services sectors are under monopolistic competition, employing mobile workers to produce horizontally differentiated varieties. The production of each variety is subject to economies of scale. The requirement of mobile workers to produce a quantity q_θ ($\theta = x, y$) of any variety is $l^\theta = \varepsilon_\theta + \beta_\theta q_\theta$ ($\theta = x, y$), where ε_θ and β_θ ($\theta = x, y$) are fixed and marginal input requirements. As $n_{xi,t}$ and $n_{yi,t}$ varieties are produced in region i the sectoral employment in the region is then:

$$L_{i,t}^\theta = n_{\theta i,t} (\varepsilon_\theta + \beta_\theta q_\theta), \quad \theta = x, y, \quad i \in [1, R], \quad (10)$$

$L_{i,t}$ is the total number of mobile workers in region i at date t ($L_{i,t} = L_{i,t}^x + L_{i,t}^y$).

Producers maximise their expected benefits and solve the producer's problem. If we denote $w_{i,t}$ as the nominal wage of mobile workers in region i at date t (given to the producer), the producer fixes the variety price as a fixed mark-up over its marginal cost.

The price of a variety produced in region i at date t is:

$$p_{xi,t} = \frac{\sigma_x}{1 - \sigma_x} \beta_x w_{i,t} \quad \text{and} \quad p_{yi,t} = \frac{\sigma_y}{1 - \sigma_y} \beta_y w_{i,t} \quad i \in [1, R]. \quad (11)$$

Equilibrium in the production side (free entry) requires zero profits at a firm level to avoid changes in the locations of the enterprises. Using equations (10), (11) and the zero profit condition we obtain an expression for the number of varieties produced in each region:

$$n_{xi,t} = \frac{L_{i,t}^x}{\varepsilon_x \sigma_x} \quad \text{and} \quad n_{yi,t} = \frac{L_{i,t}^y}{\varepsilon_y \sigma_y} \quad i \in [1, R]. \quad (12)$$

Consumers maximise their utility subject to a budget constraint. Solving the consumer's problem we it is standard to derive the demand of a manufactured variety produced in any region which depends on the CES price index of the aggregate of manufactured goods. The manufactures price index for a region i measures the minimum cost of purchasing a unit of the composite of manufactured goods (8) and correspondingly for the services (9). These indexes can therefore be thought as an expenditure function. For the services the expression of the CES price index is:

$$P_{yi,t} = \left(\sum_{m'=1}^{n_{yi,t}} p_{yi,t}^{1-\sigma_y} \right)^{1/(1-\sigma_y)} = n_{yi,t}^{1/(1-\sigma_y)} p_{yi,t}. \quad (13)$$

Recalling that manufactures are shipped between regions and using (6), the expression of the CES price index for manufactured goods is given by the equation:

$$P_{xi,t} = \left[\sum_{r=1}^R \left(\sum_{m=1}^{n_{xr,t}} (\tau_{ir} p_{xr,t})^{1-\sigma_x} \right) \right]^{1/(1-\sigma_x)} = \left[\sum_{r=1}^R \left(n_{xr,t} \left(B d_{ir}^\delta p_{xr,t} \right)^{1-\sigma_x} \right) \right]^{1/(1-\sigma_x)}. \quad (14)$$

Equation (14) tells us that the price index in region i decreases the higher the share of manufactures is concentrated in the region or in the regions where region i has a good "access to market" because fewer goods paying high transport costs are imported. We can consequently see this price index as the inverse of a market potential¹ function, in the manner that central regions will have lower price indexes due to its centrality. Likewise, a higher number of services produced in the region will lead to a lower services price index, as the index decreases with the number of varieties produced.

¹Usually, in NEG models, *access to markets* is proxied by a measure inspired in Harris' (1954) *market potential functions*.

3.1.3 Relating migrations to market potential

If we consider the traditional good as the numeraire ($p_z = 1$ for all regions), real wage of mobile workers in region i at date t is:

$$\omega_{i,t} = \frac{w_{i,t}}{P_{yi,t}^\phi P_{xi,t}^\mu}. \quad (15)$$

Taking into account the previous comments, if we hold nominal wage constant, workers' income would be higher in regions with a greater access to markets, said, regions producing a large number of varieties (both manufactures and services) and importing manufactured goods from nearby regions. This effect turns central regions attractive locations to mobile workers, as their expected real wage will be higher, and in this way it works as a mechanism of agglomeration.

We can hence connect equation (4) with the previous results. Using (11), (12), (13), (14), and (15), we can rewrite (4) to obtain the expression relating migrations to market potentials:

$$\ln \left[\frac{migr_{ji,t}}{\sum_{i' \neq j} migr_{ji',t}} \right] = \ln \left[\left(\sum_{r=1}^R L_{r,t-1}^x \left(w_{r,t-1} d_{ir}^\delta \right)^{1-\sigma_x} \right)^{\mu/(\sigma_x-1)} \right] + \ln \left[\left(L_{i,t-1}^y \right)^{\phi/(\sigma_y-1)} \right] \quad (16)$$

$$+ \ln \left[w_{i,t-1}^{1-\phi} \rho_{i,t-1} \right] + \ln \left[d_{ji} (1 + bF_{ji}) \right]^{-\lambda} + \tilde{a}_{j,t-1},$$

where²:

$$\tilde{a}_{j,t-1} \equiv a_{j,t-1} + [\mu/(\sigma_x - 1)] \ln [1/(\varepsilon_x \sigma_x)] + [\phi/(\sigma_y - 1)] \ln [1/(\varepsilon_y \sigma_y)] \quad (17)$$

$$- \mu \ln [\sigma_x/(\sigma_x - 1)] - \phi \ln [\sigma_y/(1 - \sigma_y)] - \mu \ln [B] - \mu \ln [\beta_x] - \phi \ln [\beta_y].$$

The LHS of equation (16) is the share of migrants from region j having chosen to locate in region i . On the RHS, the third and the fourth term are straightforward linked to "traditional migration equations" and represent expected nominal wage and cost of migrations respectively. The proportion of workers moving towards region i will increase with nominal wage and the probability of finding a job in the host region. The movements would nevertheless be attenuated by the cost of migration, which can be approximated by the fourth term and which increases with bilateral distances and when migrating to a non-contiguous region. However, what is really essential in this equation is the first term, and to a lesser extent, the second term. These terms denote region i 's *access to markets*. The second one is the price index for services varieties in the host region while the first one is the price index for manufactures in region i . The latter corresponds to a *market potential function* and it is the one that relates labour migration to the location of

²This equation differs from equation (13) in Crozet (2004). We have discussed our results with the author and our expression is analytically correct. Anyway, it does not have any relevant consequences at the empirical level, as all the terms included in $\tilde{a}_{j,t-1}$ are invariable between regions and can be replaced by a time trend or time fixed effects and by home region fixed effects.

manufactured activities. We can try to estimate empirically this relationship in order to validate one of mechanisms of agglomeration predicted by the NEG: the forward linkage. The parameters of interest can be obtained from this expression (elasticity of substitution between varieties, parameters from the transport cost function, shares of expenditure on manufactures and services). Strictly speaking, we can obtain them from the estimation of the indexes of prices of services and manufactures.

3.2 Is there any role for human capital?

Our contribution is fundamentally focused in the inclusion of human capital in the model. As stated in the previous section, we expect a greater human capital endowment to increase the probability of migrating. Following *human capital theory* of migrations, individuals maximise their expected wage according to the HK they are bearing, and not according to average wage levels (Becker, 1975). Skilled workers face a higher opportunity cost of unemployment in their home region, as their expected wage is larger (so they lose more if they stay in a region with lower wages) and the probability of finding a job in the destination regions is higher. This fact increases their probability of migrating to a region with better economic conditions. Furthermore, they face less uncertain when encountering a change of location, due to smaller costs of information (for instance, of the economic conditions in the destination region) and because they hope to have better employment opportunities in the host region than non-skilled. We can also consider that they have less psychological costs due to cultural and educational factors, so it is likely it will be easier for them to adapt to the new location.

Following this reasoning, we modify the previous framework in two respects: considering that higher human capital reduces migration cost and considering that higher human capital increases migration benefits. This modifications are introduced in the equation the potential migrant maximises in order to decide a change of location (1), and would be reflected in our core equation relating migrations to market potentials (16).

3.2.1 Modification in the migration cost function

The migration's cost for a worker of region j moving to region i at date t and endowed with a level of human capital $h_{j,t}$ is now defined by equation:

$$C_{ji,t}^{hk} = \left[d_{ji} (1 + bF_{ji}) e^{-\gamma h_{j,t}} \right]^\lambda, \quad (18)$$

where $\gamma > 0$ denotes the sensitivity of migration cost to the endowment of HK. Following Schwartz (1973), we expect the negative effect of distance on the probability of migration to decrease with the educational level of the potential migrant. We have accordingly added a new element $e^{-\gamma h_{j,t}}$, that accounts for the fact that skilled workers are sensitive to distance in a less extent than non-skilled workers. Thus, migration cost decreases with the human capital endowment as a consequence of the interaction of two parameters: λ ,

distance elasticity of migration cost and $-\gamma$, human capital elasticity to migration cost.

3.2.2 Modification in the migration expected benefits function

We assume that the relevant probability of employment is now $\rho_{i,t}^*$, and that it increases with human capital in the host region:

$$\rho_{i,t}^* = \rho_{i,t} e^{\psi h_{i,t}}, \quad (19)$$

So migration's expected benefit is given by equation:

$$EB_{ji,t}^{hk} = \ln [\omega_{i,t} \rho_{i,t}^*] = \ln [\omega_{i,t} \rho_{i,t} e^{\psi h_{i,t}}], \quad (20)$$

where $\psi > 0$ denotes the effect worker's human capital endowment in their migration expected benefit. That is, the benefits a worker can draw of a change in location are higher if the migrant is skilled than if he is not. This is collected by the term $e^{\psi h_{i,t}}$.

3.2.3 Modification of final equation

The objective function that the potential migrant maximises is now:

$$\pi_{ji,t}^{*k} = V_{ji,t}^{*k} + \varepsilon_i^k = \ln [\omega_{i,t} \rho_{i,t} e^{\psi h_{i,t}} [d_{ji} (1 + bF_{ji})]^{-\lambda} e^{\lambda \gamma h_{j,t}}] + \varepsilon_i^k, \quad i \in [1, R]. \quad (21)$$

Following the same strategy than in the previous subsection and introducing the modifications performed in equations (18) and (20), we can derive a new migration equation:

$$\begin{aligned} \ln \left[\frac{migr_{ji,t}}{\sum_{i' \neq j} migr_{ji',t}} \right] &= \ln \left[\left(\sum_{r=1}^R L_{r,t-1}^x \left(w_{r,t-1} d_{ir}^\delta \right)^{1-\sigma_x} \right)^{\mu/(\sigma_x-1)} \right] + \ln \left[\left(L_{i,t-1}^y \right)^{\phi/(\sigma_y-1)} \right] \\ &+ \ln \left[w_{i,t-1}^{1-\phi} \rho_{i,t-1} \right] + \ln [d_{ji} (1 + bF_{ji})]^{-\lambda} + \psi (h_{i,t-1}) + \gamma \lambda (h_{j,t-1}) + \tilde{a}_{j,t-1}, \end{aligned} \quad (22)$$

where $\tilde{a}_{j,t-1}$ is the same as in (5).

As in equation (16), the most important element to be tested is the first one. We expect to validate empirically the *forward linkage* through the structural estimation of this equation and to find values to the NEG parameters we are interested in. Furthermore, we expect these estimated parameters will be robust to the theoretical and empirical arguments that point out to a higher propensity of migration for high educated workers. Additionally, through the estimation of the empirical counterpart of (22) we can also assess the effect of human capital on the migration decisions towards economic agglomerations.

4 Data and descriptive analysis

4.1 Data sources

There are several alternative sources providing data on migratory movements for studying internal migrations in Spain, each of them having their methodology and therefore providing different data³. We have the population census (Censos de la población de España and Padrón municipal de habitantes), which have been used, among others, by Olano (1990) and by Ródenas (1994a); the Migration Survey⁴ (*Encuesta de Migraciones* - EM) which is obtained from the Spanish Labour Force Survey (*Encuesta de Población Activa* - EPA), which has been used by Antolín and Bover (1997) or the records of Spanish Social Security (*Cuentas de Cotización de la Seguridad Social, Fichero Técnico de Cuentas de Cotización* - SSR) used by Devillanova and García-Fontes (2004). Finally we have the Statistics on Residential Changes⁵ (*Estadística de las Variaciones Residenciales* - EVR), provided by the National Institute of Statistic (*Instituto Nacional de Estadística* - INE), and which has been used by, for instance, Arellano and Bover (2002), Bentolila and Dolado (1991) and Ródenas (1994b).

This latter source is constructed from the recording in the register of the inhabitants (Padrón municipal de habitantes) when arriving or leaving a municipality. We have decided to use this source because of the arguments stated in Ródenas and Martí (1997). A noteworthy characteristic of this data is that it measures migrations (movements) but not migrants. From the series we have constructed inter-provincial migration flows matrices for the period 1988-2004. We have excluded non-peninsular territories (Balears, Canarias, Ceuta and Melilla) because of the particularities migration to these provinces may have and for simplicity reasons. We finally have 47x47 matrices from which we have eliminated intra-provincial flows data because, even though they are very important in number during the period, they are out of our interest. In these matrices we have flows of migrations from the province of origin (home region) to the province of destination (host region). These matrices will be the basis to construct our dependent variable “*share of migrants from region j to region i* ”.

Inter-provincial bilateral distances ($d_{ji} \forall i \neq j$) are proxied by “*distances in kilometres by road*”⁶ between the capitals of each province. We have used this proxy because for our empirical purpose it is more convenient, as distances are associated to freight transport costs and to migration costs so it is important to take into consideration geographical elements. Besides inter-provincial distances ($d_{ji} \forall i \neq j$), we need a proxy for the internal distances since the market potential term of both equations (16) and (22) includes, as

³For some discussion see Ródenas (1994a), Ródenas (1994b), Ródenas and Martí (1997), Ródenas and Martí (2002), Martí and Ródenas (2004). In these works they analyse why these sources differ and what each of them do actually gather.

⁴An advantage of this survey is that it provides information about the relationship of the migrant to the labour market, a characteristic that makes this survey a useful source in some cases.

⁵Before 1999 this data series was known as “Migrations” (*Migraciones*) and from that year the data series name becomes “Statistics on Residential Changes” (*Estadística de las Variaciones Residenciales*).

⁶Source: <http://www.softguides.com/spain/regiones/distancias.html>

for goods imported abroad, transport costs charged on manufactured goods produced and consumed locally. Following Redding and Venables (2004), the internal distance is proxied by $dji = (2/3)\sqrt{S_i/\pi}$. S_i denotes the surface of the region which we have obtained from the INE.

We have profusely explored the Valencian Institute of Economic Research's (*Instituto Valenciano de Investigaciones Económicas* - IVIE) database on "Human Capital in Spain and its provincial distribution". This database is basically constructed from information of the Spanish Labor Force Survey (*Encuesta de Población Activa* - EPA) and contains abundant information on the evolution of the educational levels of the Spanish population for a long period (1964-2001). It contains, among other, information about active and employed population by level of education and employed population by economic sector and level of education. We have hence constructed sectoral employment series and employment rate (calculated as one minus the unemployment rate) for the set of Spanish provinces from this dataset for the period 1986-2001. We have also employed this source to construct other variables used in the empirical application as unemployment rates, employment growth and manufactures and tradable services employment proportion in home and host regions.

We also employed data from another dataset coming from IVIE and the BBVA Foundation called "Capital Stock in Spain and its territorial distribution 1964-2002". It provides annual series on public and private capital stock and investment at a provincial level for the period 1964-2000. We have constructed two variables which will be used in the empirical application that approximate the productive and the social public infrastructure endowments of the Spanish provinces for the period 1987-2001. Productive infrastructure gathers public capital on roads, motorways, railways, airports and urban structures in millions of euros. Social infrastructure gathers public capital on education and health services.

Following Serrano (1997) we have constructed a *proxy* for the human capital of each province: *the average years of education*. We use the previous dataset to calculate the proportion of employed population of each level of education and then we impute 0 years for illiterate employees, 3.5 years for employees without education or with primary education, 11 years employees with secondary education, 16 years for those with higher education (first cycle), and 17 for those with higher education (second cycle). This variable is constructed for total employees for the period 1986-2001.

A *proxy* for the nominal wages has been constructed combining two sources: Regional Spanish Accounts (*Contabilidad Regional de España* - CRE, from the INE) and the IVIE employment dataset. From the former we have obtained a current aggregate series for worker's remuneration which we have divided by employees to obtain a *proxy* of the nominal worker's wage by provinces for the period 1986-2001 in current euros.

Finally, following Arellano and Bover (2002), we use data⁷ coming from the Valuation

⁷As stated in the legal conditions for using the Valuation Society Ltd. dataset in housing prices, these data only gather the values collected in the society's valuations, not the entire population.

Society Ltd. (*Sociedad de Tasación S.A*⁸) for the housing prices. These data refer to average prices in nominal euros for square meter of a representative new dwelling in the provinces capitals.

4.2 Descriptive analysis

As already mentioned in section 2, several authors have highlighted the fact that, from the beginning of the 80's there has been a remarkable increase in the inter-provincial migratory movements (Ródenas and Martí, 2002 and Bover and Velilla, 2005). Figure 1 shows immigration and emigration rates (migrants over population for the 1st January) and migration intensity rate (total flows of migrations over population for the 1st January) per 1000 inhabitants. As we can see, for the period 1988-2002 the evolution of the series is increasing, except for particular years (1990-1991 and 1995-1996) where the three rates decrease. In the first part of the period, the increase in the rates has been moderate (between 14-17%), and if we observe the values in 1988 and 1995 they are almost equivalent. It is during the second part of the period when all three rates experienced a sharp increase with growth rates over 40% between 1995 and 2002. Therefore, annual rates have growth more than 62% for the overall period, especially the emigration rate which has growth more than 66% in 15 years.

The analysis of the evolution of these rates only inform us of the magnitude of the movements, but it says nothing about their origin and destination. To shed some light on this issue we have constructed tables 1 and 2. Table 1 shows the five provinces with higher and lower net immigration rates⁹ in 1988, 1995 and 2001 and table 2 does so for migration intensity rates. The relative position each province has in relation to the rest of provinces intuitively indicates which provinces have received more migrants, table 1, and which provinces have been more active in their migratory behaviour, table 1. The provinces that have received more migrants are traditionally situated in the Mediterranean coast (Girona, Tarragona, Castellón, Alicante, and Málaga) or next to Madrid (Guadalajara and Toledo). This reveals a tendency to move towards locations near “traditional” attraction poles (Madrid, Barcelona, Valencia, and Sevilla) and towards “pleasant” locations. On the other hand, the provinces which have “expulsed” more migrants are also traditional migrants receivers of the 50's and the 60's (Barcelona and provinces of the Bask Country), and provinces that have traditionally expelled migrants (Teruel, Jaén, Palencia or Cáceres). Undoubtedly, behind a great deal of this behaviour are “return migration flows”, that is, migrants who moved to the economic poles in the 50's and 60's and that, after retirement, return to their original place of birth. Besides this, migratory intensity rates reveal an active migratory behaviour in the same provinces as before, *i.e.*, those in the Mediterranean coast and near Madrid and Barcelona. Among the five more active provinces we find Teruel, which may appear here because it expulses an important part of its population.

⁸<http://web.st-tasacion.es/html/menu6.php>

⁹Net immigration rate is calculated as net migrants (emigrants minus immigrants) over population in January the 1st of each year.

Among the most static provinces we find some that are situated in the North of the peninsula (Galicia and Asturias). As said in the literature, the reasons that could explain these patterns are varied: a generalised increase in the welfare and economic situation in the country as a whole, a generalised increase in the unemployment rate, rigidities in the housing and labour markets, social benefits as the unemployment benefit may have lessen the traditional economic reasons that moved the migrants two decades ago , so nowadays workers may change their location motivated by different reasons.

To complement the analysis we have proceeded next to perform a comparative study of some key variables of our model. It is shown in tables 3, 4 and 5. In these tables we have sorted some provinces with regard to their *relative market potential*. The measure used to represent the market potential of a region is that proposed by Harris (1954) and it is calculated in three moments of time, 1988, 1995 and 2001. This variable characterizes the access to markets of a region and agglomeration of the economic activity, essential phenomena described by the NEG. Market potential *à la Harris* (henceforth MP) measures the demand a region has access to as the average of the economic size of the surrounding regions weighted by distances. For a region i we have:

$$\text{Market Potential}_i = \sum_{r=1}^R (Y_r/d_{ir}). \quad (23)$$

where R is the number of regions in the relevant area, Y_r is the economic size of region r (generally employment or production) and d_{ir} the distance between locations i and r ; $i, r \in [1, R]$. To represent the economic size of the province we have used total employment. Subsequently, relative MP is related to migrations (*average share of migrants*), to some economic variables that may influence migratory movements (*wages, employment, productive specialization*) and to human capital. Through the descriptive analysis of these variables we want to draw some conclusions in order to strengthen the intuitions deriving from the theory and the existing empirical evidence.

Commuting flows may explain why some provinces, even though they appear among the first five in regard to the MP ranking, are below the national average as regards their employment rate or their nominal wage and vice versa. For instance, Toledo has a relative nominal wage below the average in 1995 and 2001, even though its MP is very high. It is a province situated close to Madrid and it can be qualified as a “*dormitory province*”, that is, workers live there not because the high wages or the low unemployment rates but because housing is less expensive there than in Madrid that is where they are employed. This fact would also explain why Toledo and Guadalajara have relative human capital level below the average even though their MP is over the national average. We must keep in mind that the human capital *proxy* has been calculated from the average years of education of the workers, people employed in the province but not necessarily living there. Thus, if workers are employed in Madrid but they live in Toledo or Guadalajara, their human capital will be recorded in Madrid and not in the province where workers

actually live. For the rest of the provinces there exists an almost perfect correspondence between employment rates, nominal wages and education level of the workers, *i.e.* when classified over(below) the national average in regard to their MP, the levels of the other variables are also over(below) the national average. We have also included three columns that add information about the employment. Relative total employment shows us the crucial weight that only three provinces, Madrid, Barcelona and Valencia, have over the whole country. The small weight Toledo and Guadalajara have justify once **again** that they have a high MP due to its geographical location. Conversely, the fact that Cadiz had a level of employment over the average while a MP below is because it is surrounded by peripheral provinces. Two additional columns show relative specialization in manufactures and services, measured as sectoral employees over total employment, which exhibit the same pattern than the total employment.

Finally we have constructed some figures to show the evolution of the inter-provincial disparities in relation to some key variables of our model. They are showed in figure 3. In the top panel of figure 3 we can see that inter-provincial dispersion for both employment rate and nominal wage has considerably decreased, especially for the latter (41% decrease versus 25%). On the other hand, the expected nominal wage¹⁰ dispersion remains almost stable during the period under analysis. The middle panel shows the dispersion for the employment variables (total and the wide sectors). As expected, total employment dispersion is lower than the one for each of the sectors, as provinces may be specialised. In addition, the inter-provincial distribution of employment has remained very stable, as dispersion growth between 1986 and 2001 is practically zero. As for the evolution of the disparities in the case of human capital and market potential, results are quite similar for both variables as can be deduced from the bottom panel in figure 3. Nevertheless, disparities in immigration and migration intensity rates rise considerably during the period revealing an increase in the dispersion of the flows. This agrees with those authors that have claimed the existence of a more balanced but varied bilateral migration flows between Spanish provinces since the mid-80's.

5 Econometric specification and results

5.1 Estimable gravity equations

As in Crozet (2004), we proceed to the estimation of the empirical counterpart of both (16) and (22). These equations are closely related to simple *gravity equations*. The migration flow from region j to region i increases not only with the nominal wage and the employment rate in the host region but also with the economic size of the host region, while it decreases with the bilateral distance between the host and the home region (d_{ji}). These last two variables represent the host region's *Market Potential*. The estimation of the gravity equations will give us some insight on the relationship we are trying to assess. The

¹⁰The expected nominal wage is defined as the product of the nominal wage and the employment rate.

empirical validation of our model will therefore require that workers were attracted towards economic agglomerations, *i.e.*, the forward linkage (16), and, moreover, that there exists a relationship between the share of migrants of any region and to any other region and their endowments of human capital (22).

Our dependent variable in both cases is *the share of migrants moving from region j to region i at date t* ($sh_{ji,t}$). A *proxy* for the probability of finding a job in the host region is the employment rate, $E_{i,t-1}$ (one minus the unemployment rate). As this variable may be correlated with nominal wages¹¹, we consider the *expected nominal wage* as a single variable defined by the product of nominal wage and employment rate (Harris and Todaro, 1970): $probw_{i,t-1} = w_{i,t-1}E_{i,t-1}$. The economic size of the host region will be *proxied* by total employment, $L_{i,t-1}$. Furthermore, variables included in $\tilde{a}_{j,t-1}$ do not depend on destination region i . We replace this elements by home region fixed effects (a_j) and time fixed effects (F_t)¹². We have also have included a dummy variable that indicates if the home and host regions share a common border (V_{ji}). Finally, in order to control for the different size of the provinces, we include the surface of the host region as an explanatory variable (S_i). Thus, the gravity equation for (16) is given by the expression:

$$\ln(sh_{ji,t}) = \ln \left[\frac{migr_{ji,t}}{\sum_{i' \neq j} migr_{ji',t}} \right] = \beta_1 \ln(L_{i,t-1}) + \beta_2 \ln(prob_{i,t-1}) \quad (24)$$

$$+ \beta_3 \ln(d_{ji}) + \beta_4 V_{ji} + \beta_5 \ln(S_i) + a_j + F_t + u_{ji,t}$$

where $u_{ji,t}$ is a well behaved error term.

If we take into account the human capital in the model, equation (24) becomes:

$$\ln(sh_{ji,t}) = \ln \left[\frac{migr_{ji,t}}{\sum_{i' \neq j} migr_{ji',t}} \right] = \beta_1 \ln(L_{i,t-1}) + \beta_2 \ln(prob_{i,t-1}) \quad (25)$$

$$+ \beta_3 \ln(d_{ji}) + \beta_4 V_{ji} + \beta_5 \ln(S_i) + \beta_6 h_{i,t-1} + \beta_7 h_{j,t-1} + a_j + F_t + u_{ji,t}$$

We have introduced two new variables: human capital in the host region ($h_{i,t-1}$) and human capital in the home region ($h_{j,t-1}$). In accordance to our theoretical model, we expect the parameters of both variables to be positive: a higher human capital in the host region increases probability of finding a job and a higher human capital in the home region decreases cost of migration, so migration flows towards region i raise by means of these two mechanisms. All the explanatory variables are lagged one year to avoid problems of simultaneity and to capture the fact that there is some delay between the moment the migration decision is made and the movement.

¹¹In fact, correlation between nominal wages and employment rate is positive.

¹²We have found better results including these time fixed effects instead of a time trend as in Crozet (2004). In addition, we do not include a variable to take into account if host region is an Objective 1 or 2 region, as the majority of Spanish provinces were qualified as objective regions by the EU regional policy over the period under analysis.

5.1.1 The sample selection problem

In our sample we have 32430 observations (crossed inter-provincial flows in the period 1988-2002). Among these observations we observe 498 cases where the flow of migrants between two provinces was zero, resulting in a value of *the share of migrants moving from region j to region i at date t* ($sh_{ji,t}$) of zero in all these cases. As this variable in the empirical specifications of the gravity equations (24) and (25) appears in natural logarithms, we have to drop the observations with value zero from our sample if we want to directly estimate the equation by OLS. In so doing we are incidentally truncating our sample. If we do not take into account the fact that the sample is selected and we simply estimate (24) and (25) by OLS, the estimation of the parameters of our equations might be biased and inconsistent. Some correcting measures must be applied in order to obtain correct estimations of the parameters of interest.

Studies analysing migration have faced the problem of sample selection mainly when estimating the returns to migration. This problem arises because observed patterns of migration are not the outcome of random process but the observed migration behaviour is the outcome of a maximising behaviour on the part of economic agents. It exists abundant literature that examines the magnitude of the impact on earnings that a change in location reports to migrants. In the estimation process, the estimated parameters in the income equations are corrected by a selection equation determining the existence or the absence of previous migration. The most popular example of this kind of studies is that of Nakosteen and Zimmer (1980), which constitutes the first attempt to explicitly acknowledge the problem with self-selectivity in assessing the returns from inter-state migration in the United States¹³.

Our case is somewhat different. Here the dependent variable is a log-function of the share of migrants moving from j to i but we only observe that variable when the flow of migrants moving from j to i is positive. Next we will discuss the specification that will be estimated for equation (25), as we can particularize the other specifications from this general case.

As indicated above, the idea is very simple, as for some pairs of regions it does not exist a positive flow of migrants (46 x 47 possible crossed flows) in a particular moment of time (15 years). As we are dropping the null observations from our sample in order to estimate the parameters of the gravity equations, we must include in the analysis the factors behind the absence of flow between two specific provinces (from j to i) in a specific date (t). Therefore, if we estimate the gravity equations only using the sample with positive values for the endogenous variable and we do not consider the selection problem, we will be omitting the fact that specific values of the explanatory variables have culminated in the absence of flow between some pairs of provinces and, in this way,

¹³Other examples are Robinson & Tomes (1982) for inter-provincial migration in Canada, Hunt & Kau (1985) for repeated migrants in the United States, Borjas et al (1992) for EEUU, Axelsson & Westerlund (1998) for Sweden, Tunali (2000) for Turkey, Pekkala (2002) for Finland, Détang-Dessendre et al (2004) for French youth or Nakosteen & Westerlund (2004) also for Sweden.

in a null value for the share of migrants. As stated above, this might result in biased and inconsistent estimations of the parameters of the gravity equation. This is likely to be the case in the analysis of the set of EU regions in Crozet (2004), although the author did not mention the possible existence of a selection bias.

5.1.2 Estimation strategy

Our equations of interest are the gravity equations (24) and (25), which relate the share of migrants moving from j to i (quantity) to a series of variables approximating agglomeration and human capital. To ease the notation let's write these equations in the general form:

$$y_{ji,t} = x'_{ji,t}\beta + u_{ji,t}. \quad (26)$$

The existence or the absence of a positive flow between two provinces will be gathered in the selection equation that in general terms can be written as¹⁴:

$$z_{ji,t}^* = g'_{ji,t}\eta + \xi_{ji,t}, \quad (27)$$

where $z_{ji,t}^*$ is the "net benefit of migrating from j to i ". This benefit will depend in a set of factors that are gathered in $g_{ji,t}$. Specifically, $z_{ji,t}^*$ is defined as:

$$z_{ji,t}^* = z'_{ji,t} - z'_{j.,t}, \quad (28)$$

where $z'_{ji,t}$ is the "gross benefit of migrating from j to i at date t " and $z'_{j.,t}$ is the "gross benefit of migrating from j to any other region except i including j -not migrating- at date t ".

The sample rule is that $y_{ji,t}$ is only observed when $z_{ji,t}^* > 0$. $z_{ji,t}^*$ determines the existence of a positive or a null flow of migrants from province j to province i at date t (existence of the flow) while $y_{ji,t}$ determines the magnitude of this flow. If $z_{ji,t}^* > 0$, i.e., $z'_{ji,t} > z'_{j.,t}$, we will observe a positive flow, so the value of $sh_{ji,t}$ will be positive and we will be able to calculate the value of $\ln(sh_{ji,t})$. If $z_{ji,t}^* \leq 0$ we will not observe a flow, so the value of $sh_{ji,t}$ will be zero and the value of $\ln(sh_{ji,t})$ cannot be calculated.

However, the variable $z_{ji,t}^*$ is not observed, we only observe its sign, i.e. only the existence or the absence of flow between a pair of provinces ($sh_{ji,t}$ positive or null) at moment of time. We can thus reformulate our model as follows:

$$\begin{aligned} \text{Selection mechanism:} \quad & z_{ji,t}^* = g'_{ji,t}\eta + \xi_{ji,t}, \\ & z_{ji,t} = 1 \text{ if } z_{ji,t}^* > 0 \text{ and } z_{ji,t} = 0 \text{ otherwise-} \\ & \text{Prob}(z_{ji,t} = 1 | g_{ji,t}) = \Phi(g'_{ji,t}\eta) \text{ and} \\ & \text{Prob}(z_{ji,t} = 0 | g_{ji,t}) = 1 - \Phi(g'_{ji,t}\eta). \end{aligned}$$

¹⁴Our reasoning is primarily based on Greene (2003) pp.780. Other useful manuals addressing this issue are Cameron et al (2005) pp.546, Cabrer et al (2001) pp.227 and Wooldridge (2003) pp.585.

Regression model: $y_{ji,t} = x'_{ji,t}\beta + u_{ji,t}$ is observed only when $z_{ji,t} = 1$.

$$(u_{ji,t}, \xi_{ji,t}) \sim \text{bivariate normal } [0, 0, 1, \sigma_u, \rho].$$

Suppose that $z_{ji,t}$ and $g_{ji,t}$ are observed for a random sample of individuals but $y_{ji,t}$ is only observed when $z_{ji,t} = 1$. It can be demonstrated that:

$$E[y_{ji,t} | z_{ji,t} = 1, x_{ji,t}, g'_{ji,t}] = x'_{ji,t}\beta + \rho\sigma_u\lambda(g'_{ji,t}\eta), \quad (29)$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are respectively the density and the cumulative density functions of a normal distribution and $\lambda(\cdot) = \phi(\cdot)/\Phi(\cdot)$ is the inverse of the Mills ratio.

In the selection equation the dependent variable is $sh_pr_{ji,t}$ which is a binary variable which takes value 1 when $sh_{ji,t}$ is positive and 0 when it is null. The set of factors than may be affecting the probability of migrate from region j to region i at date t can be gathered in several subtypes of factors: "physical and location variables" (d_{ji} , V_{ji} , S_i , ct_i), fixed effects (a_j , F_t), "amenities" ($\inf pd_{i,t-1}$, $\inf soc_{i,t-1}$), "economic structure variables" ($mf_{i,t-1}$, $mf_{j,t-1}$, $sv_{i,t-1}$, $sv_{j,t-1}$) and "economic variables" ($L_{i,t-1}$, $probw_{i,t-1}$, $dif_wn_{ji,t-1}$, $dif_pv_{ji,t-1}$, $dif_hk_{ji,t-1}$, un_i , un_j , eg_i). Besides the variables defined previously we have introduced new variables: $dif_x_{ji,t-1}$, is the difference between the home and the host regions for variable x ; wn states for the nominal wages; pv states for the housing prices; un_i and un_j state for the unemployment rates in the home and the host regions; eg_i states for the annual employment growth rate in the host region; ct_i is a dummy variable that takes value 1 if the host region is a coastal province; $\inf pd_{i,t-1}$ and $\inf soc_{i,t-1}$ are productive and social infrastructure in the host region respectively, and the last four variables represent the sectoral employment proportion over total employment for the manufactures ($mf_{i,t-1}$ and $mf_{j,t-1}$) and for the tradable services ($sv_{i,t-1}$ and $sv_{j,t-1}$) in both home and host regions. As for the gravity equations, regressors in the selection equation are lagged one year. For the regression equations, (24) and (25), we will using the variables and the notation defined in the previous section. To avoid problems of identification we have to remark that several variables that appear in the selection do not appear in the regression equation.

The parameters of the sample selection model can be estimated by maximum likelihood. However, Heckman's (1979) two-step estimation procedure, sometimes called the Heckit estimator, is more usually used¹⁵¹⁶. The Heckman's method is basically performed as follows:

1. Estimate the probit equation (selection mechanism) by maximum likelihood to obtain estimates of η . For each observation of the selected sample ($sh_pr_{ji,t} = 1$) we compute $\hat{\lambda}(g'_{ji,t}\hat{\eta}) = \phi(g'_{ji,t}\hat{\eta})/\Phi(g'_{ji,t}\hat{\eta})$.

¹⁵In fact, MLE estimation is more efficient than Heckman two-steps procedure and it calculates σ_ϵ and ρ directly. On the other hand it has two important disadvantages: it is based on strong distributional assumptions and we can have problems of convergence in the optimization process.

¹⁶Some reasons for the popularity of the Heckit estimator are given in Cameron & Trivedi (2005), page 550.

2. Estimate β and $\beta_\lambda = \rho\sigma_u$ by OLS regression of $y_{ji,t}$ on X and $\hat{\lambda}$.

5.2 Empirical results

We have computed the two-steps Heckman estimator for equations (24) and (25). Results are displayed in tables 6 to 9. We show also the results for OLS estimation in tables 10 and 11, so we can observe the effects that controlling for selection sample has on the estimated parameters. The overall model test statistic is an asymptotically equivalent Wald test that all the parameters in the regression equation being zero (except the constant), and in all cases this test reports that our estimated models are globally significant. Generally the parameters estimated by OLS are biased upwards respect to the estimation by Heckman, except in the host region human capital ($h_{j,t-1}$) parameter in columns (4), (6), (7) and (9), which value is slightly higher when we control by selectivity. In all cases, the individual test of significance of the parameter associated to selectivity, λ , reveals that it is significant, so we can conclude that it is important to control by selection.

5.2.1 Model without human capital

Parameters have generally the expected signs and are significantly different from zero at 1% level in almost all cases. The estimated parameter associated to the bilateral distance between home and host regions variable is always significant at 1% level, having the expected negative value in all cases. Estimated values are around (-1), which are over those estimated for the Spanish case in Crozet (2004). This means that Spanish migrants are averse to distance, and that the aversion increases when migrating to a non-contiguous province, as the estimated parameter of V_{ji} is in all cases high, positive and significant. Migrations are thus mainly pointed towards neighbouring regions.

One should expect the coefficient associated to regional employment to be positive, as we expect migrant to move towards larger regions due to higher wages and lower living costs in bigger local markets. The effect of total employment is also large, positive and significant, suggesting that migrants do prefer large economic regions. In addition, the parameter associated to the surface of the host region is positive and significant in all cases (except in specifications 9 and 15). We can thus infer that our proxy for market potential is significant and positive or, in other words, inter-provincial migrants in Spain prefer not far away provinces with large regional markets (*centripetal dynamics*).

Columns (1), (2) and (3) estimate three different versions of equation (24). The difference between the first two columns is that home region fixed effects are not included in the first estimation while they are so in the second one. The results are essentially the same: as in all cases the parameters have the expected sign and are significant at 1% level. We have performed an F-test that confirms the significance of including regional fixed effects, so the estimation in column (2) is more appropriate than that of column (1). Both estimations are globally significant. In column (3) we have considered the influ-

ence of employment on manufactures, services and traditional sectors separately¹⁷, as we expect the sectoral composition of gross regional product to have different influences on migration choices. In this case, it seems that the influence of economic size is mainly due to services. Local manufacturing employment has a little influence on attracting migrants and this influence is even negative. As for the effect of the expected nominal wage in the host region, our estimates contrasts with the ones obtained for the Spanish NUTS II regions in Crozet (2004). Our estimates for the provinces (NUTS III regions) points to a positive effect of the expected nominal wage on the share of migrants, suggesting that it really acts as an incentive for workers that decide to migrate.

5.2.2 Model with human capital

Next, we have estimated some specifications for the reduced form in equation (25), including human capital in the home and the host region. Columns (4), (5) and (6) in table 6 show the results. The estimated coefficients are jointly significant and the parameters for the basic variables remain significant and show the expected signs. But, the coefficients for the human capital variables are neither significant ($h_{j,t-1}$) nor they show the expected sign ($h_{i,t-1}$). Undoubtedly, this is against our hypothesis on the role of human capital on migration decisions.

Some authors have stated that the unequal spatial distribution of wages in Spain is due to an unequal distribution of human capital, so higher nominal wages are due to higher human capital endowments (Serrano, 1998; De la Fuente et al, 2003 and López-Bazo and Moreno, 2005). Actually, the high correlation between nominal wages and human capital in the host region in our sample is very large (correlation coefficient of 0.912). And, as stressed in some of the above-mentioned studies, the whole set of Spanish regions have experienced a process of continuous growth in the levels of education of their population and labour force. All of this is confirmed in our sample by an R^2 as high as 0.87 when performing a simple regression of years of education on the set of time dummies and nominal wages. As a result, the estimation of the model expanded with the inclusion of human capital variable might be contaminated by strong collinearity if we include in the specification both human capital and expected nominal wage at the same time.

To shed some additional light on the role of human capital we have re-estimated equation (25) but excluding the variable $\ln(\text{prob}w_{i,t-1})$, as we suspect that at least part of the effect of human capital might be captured by the spatial distribution of the expected nominal wage through the Spanish provinces. Results are shown in columns (7), (8) and (9) of table 7. Now, the estimates for the parameters of both human capital variables show the expected signs, although the estimated parameter for the stock of human capital in the home region, $h_{j,t-1}$, is not significant.

When we exclude regional fixed effects from the specification (10 to 15) we find results more according to our theoretical predictions. Results are shown in table 7. Columns (10)

¹⁷Results are robust to the inclusion of different industry classifications.

to (13) state the same specification that columns (4) to (6) but without regional fixed effects, *i.e.* they are modifications of column (1) introducing human capital variables. In column (10) the signs of the parameters associated to the human capital variables are maintained in both columns. However, since the magnitude for the host region human capital parameter in column (10) is approximately the same and its sign is contrary to expected as in column (4), the parameter is no longer significant. On the other hand, compared to the value obtained in column (4), the home region human capital parameter is higher than before (0.129 versus 0.022) and it is now significant. It is worthy to say that if we compare column (10) in table 7 and in table 10, controlling for selection changes the results for the estimated parameters for human capital variables since, although even very small, the parameter associated to host region human capital is positive but not significant if we do not control for selection of the sample, contrary to results in table 7. Furthermore, when we try different specifications introducing human capital variables separately (columns 11 and 12 in table 7), we find different results. Compared to results in columns (5) and (6), we find the some important changes: signs are maintained (host region human capital negative sign and home region human capital positive sign) but now the parameters associated to the human capital endowments in both home and host regions are significant. Therefore, host region human capital parameter in column (11) is significant but still with a different sign to the expected one and, moreover, it is not significant and lower if we do not consider selectivity. Finally, in column (12) we introduce only home region human capital, whose estimated parameter is positive as in (6) but it becomes significant as we remove home region fixed effects. These differences found in the significance and the magnitude (biased upwards when not controlling) of some parameters when correcting or not correcting for selection shows us the importance of taking into account that our sample is selected when we estimate the parameters of interest.

As we have seen, if we introduce at the same time the variables human capital in host region and expected nominal wage in host regions (columns 10 and 12) we are not able to find a significant coefficient with the expected sign for the in the host region human capital. The expected sign and a significant coefficient for $h_{j,t-1}$ is obtained when the regression is run excluding the regional fixed effects and expected nominal wage in host region (columns 13, 14, and 15). The improvement in the result obtained in this case is likely to be due to two causes: the fact that human capital endowment only varies smoothly over time in the period under analysis in the whole set of provinces, causing that regional fixed effects absorbed most of the effects of the level of education in the home region¹⁸ and to the fact that probably expected nominal wage and human capital endowment in the host regions are expressing the same phenomenon. Results show that an increase in one year of education in the host and home regions has a positive effect on migrations. This confirms empirically the intuitions stated in the previous sections and constitutes a stable basis to proceed to the structural estimation of the NEG parameters in equations (16) and (25).

¹⁸The R^2 of a regression between human capital and a set of provincial dummies in our sample is 0.953.

6 Conclusions

This work aims to contribute to NEG empirical assessment in several ways. Firstly, we adopt an uncommon framework to analyse the Spanish migratory phenomenon, that relating workers migration to agglomeration following the NEG framework. Secondly, we provide empirical evidence on the *forward linkage*, one of the two basic mechanisms of agglomeration predicted by the NEG and the one that has attracted less attention so far. Finally, we consider the role that human capital might be playing in migrations towards economic agglomerations, in order to validate empirically the theoretic arguments that point to a higher propensity to migrate of educated workers and to regions with large endowments of human capital.

Our theoretical model is essentially based in Crozet (2004). We obtain a core equation relating migration flows to market potential, and we extend this equation including human capital of both home and host regions as the level of education attained by the potential migrants will affect both the cost and the expected benefit of a change in location.

Previous works that have provided empirical evidence to the forward linkage, as Crozet (2004), Pons et al (2004) and Zuh and Poncet (2003), may have obtained inconsistent estimations of the NEG parameters of interest as they do not take into account the role of human capital in their analysis. In this sense, our contribution is double: we are able to provide a more robust estimation of the parameters of interest and, furthermore, we obtain estimations of the influence of human capital on migration flows in both the home and the host region. If human capital is not taken into account in the analysis the estimation of the parameters in a migration equation *à la Crozet* will be biased. Previous theoretical and empirical works point out an improvement of the model if we consider the effect that the human capital endowment has in the probability of migrate of the workers. If we omit this fact, an omission bias may be affecting our estimations. Moreover, as the omitted variable is related to other variables actually apperaring in the model, as agglomerations and wages (Glaeser & Maré 2001), the bias in the estimation emerges.

Data on Spanish inter-provincial migration flows is used to briefly describe migration-flows magnitudes and directions and their relationship to some economics variables that may explain the relative attractiveness of some regions to workers and especially to the most skilled ones. To provide some empirical evidence to our theoretical model we have estimated two sets of gravity equations (with and without human capital variables) that relate share of migrants to a set of explanatory variables in which we include total employment and bilateral distances as proxies for market potential. We have computed the Heckman two step estimator in order to correct the selection bias.

Nevertheless, the bias in the estimation of the parameters that other works may having as they do not take into account the human capital endowment could be faded. We have highlighted that the effect of the human capital endowment in the host region could be embodied in the nominal wages, so if we include both variables in the estimation at the same time our results are not as forceful as suggested by the model. On the other hand,

the home region fixed effects could be gathering home region characteristics, among which home region human capital endowments. When removing the home region fixed effects and the expected nominal wage in host region, the results seem to validate our theoretical arguments, as the estimated coefficients for stock of human capital in the home and host regions are positive and significant.

Nevertheless, these results are preliminary. The unexpected sign of the parameter estimated for the manufactures employment raises the question of the suitability of only estimate gravity equations to empirically assess our theoretical model. Furthermore, the unexpected signs found for the estimated coefficients associated to human capital variables in the specifications including the expected nominal wage justifies the use of a real market potential function in the spirit of NEG framework. Thus, structural estimations of equations (16) and (25) will be performed in the future. Further extension of this work may include the development of a theoretical model that introduces heterogeneity in the skill level of the workers as a factor of production. Another extension will be the analysis if the impact that human capital endowment has on the parameters of the model, i.e., if there exist a different sensibility of skilled workers as they are attracted towards economic agglomerations.

References

1. Antolín, P. and O. Bover, (1997), "Regional migration in Spain: the effect of personal characteristics and of unemployment, wage and house price differentials using pooled cross-sections", *Oxford Bulletin of Economics and Statistics* 59, pp. 215-35.
2. Axelsson R. and O. Westerlund (1998), "A panel study of migration, self-selection and household real income", *Journal of Population Economics*, volume 11, nb.1, pp.113-126.
3. Becker, G., (1975), "Human capital", The University of Chicago Press.
4. Beine, M., Docquier F. and H. Rapoport, (2001), "Brain Drain and Economic Growth: Theory and Evidence" *Journal of Development Economics*, v. 64, iss. 1, pp. 275-89.
5. Bentolila, S. y J.J. Dolado., (1991), "Mismatch and internal migration in Spain, 1962-86", en F. Padoa-Schioppa (ed.), *Mismatch and labour mobility*, Cambridge University Press, Cambridge.
6. Borjas G.J, Bronars S and J.Trejo, (1992), "Self-selection and internal migration in the United States", NBER Working Papers Series nb.4002.
7. Borjas, George J., (1999), "The Economic Analysis of Immigration", *Handbook of Labour economics*. Volume 3A, pp. 1697-1760, *Handbooks in Economics*, vol. 5. Amsterdam; New York and Oxford: Elsevier Science, North-Holland.

8. Bover, O. and M. Arellano, (2002), "Learning about migration decisions from the migrants: using complementary data sets to model intra-regional migrations in Spain", *Journal of Population Economics* 15, pp. 357-80.
9. Bover, O. and P. Velilla, (2005), "Migration in Spain: historical background and current trends", in: Klaus F. Zimmermann (ed.), *European Migration: What Do We Know?*, CEPR and OUP.
10. Cabrer-Borrás B., Sancho-Pérez A. and, G. Serrano Domingo (2001), "Microeconometría y decisión", Ed. Pirámide, Colección Economía y Empresa, Madrid.
11. Cameron A. C. and K.T. Pravin (2005), "Microeconometrics methods and applications", Cambridge University Press, Cambridge.
12. Combes, PP. y Henry G. Overman, (2004), "The Spatial Distribution of Economic Activities in the European Union", en *Handbook of Urban and Regional Economics*, vol.4, V. Henderson and J.-F. Thisse (eds.), Elsevier-North Holland, Amsterdam.
13. Crozet, M., (2002), "Do migrants believe in market potential?", mimeo.
14. Crozet, M., (2004), "Do migrants follow market potentials? An estimation of a new economic geography model" *Journal of Economic Geography*, Oxford University Press, vol. 4(4), pp. 439-458
15. Crozet, M., Mayer, T. and J-L. Mucchielli, (2004), "How Do Firms Agglomerate? A Study of FDI in France", *Regional Science and Urban Economics*, v. 34, iss. 1, pp. 27-54.
16. Davis, D. and D. Weinstein, (1999), "Economic Geography and Regional Production Structure: An Empirical Investigation", *European Economic Review* 43(2), pp. 379-407.
17. Davis, D. and D. Weinstein, (2003), "Market Access, Economic Geography and Comparative Advantage: An Empirical Assessment", *Journal of International Economics*, 59(1), pp. 1-23.
18. De la Fuente, A., (1999), "La dinámica territorial de la población española: un panorama y algunos resultados provisionales", *Revista de Economía Aplicada*, vol.VII, nº 20. pp. 53-108.
19. De la Fuente, A., Doménech R and J.F. Jimeno, (2003), "Human capital as a factor of growth and employment at the regional level. The case of Spain", IAE WP 610.04.
20. Decressin J. and A. Fatàs, (1995), "Regional Labour Market Dynamics in Europe", *European Economic Review*, v. 39, iss. 9, pp. 1627-55.
21. Delgado-Cabeza, M., (1998), "Las desigualdades territoriales en el Estado Español 1955-1995", *Revista de Estudios Regionales*, 2 (51): pp. 61-90.

22. Détang-Dessendre C., Drapier C. and H. Jayet (2004), "The Impact of Migration on Wages: Empirical Evidence from French Youth", *Journal of Regional Science*, Volume 44 Issue 4, p.661.
23. Devillanova, C. and W. García-Fontes, (2004), "Migration across Spanish Provinces: Evidence from the Social Security Records (1978-1992)", *Investigaciones Económicas*, vol. 28 (3), 461-487.
24. Dolado, J., Goría A and A. Ichino, (1994), "Immigration, Human Capital and Growth in the Host Country: Evidence from Pooled Country Data", *Journal of Population Economics*, v. 7, iss. 2, pp. 193-215.
25. Dolado, J.J, González-Páramo J.M, y J.M. Roldán, (1994), "Convergencia económica entre las provincias españolas: evidencia empírica (1955-1989)", *Moneda y Crédito*, nº 198, pp. 81-118.
26. Feenstra, R, Markusen, J. and A.J. Rose, (1998), "Understanding the Home Market Effect and the Gravity Equation: The Role of Differentiating Goods"; C.E.P.R. Discussion Papers, CEPR Discussion Papers: 2035
27. Forslid, R., J. Haaland and K-H Midelfart-Knarvik, (2002), "A U-shaped Europe? A simulation study of industrial location", *Journal of International Economics* 57(2), pp.273–297.
28. Friedberg, R and J. Hunt, (1995), "The Impact of Immigrants on Host Country Wages, Employment and Growth", *The Journal of Economic Perspectives*, Vol. 9, No. 2, pp. 23-44.
29. Fujita, M., Krugman P.R. y A.J. Venables., (1999), "The spatial economy: Cities, regions and international trade", Cambridge, MA: MIT press.
30. García-Barbancho, A., (1960), "Los movimientos migratorios en España", *Revista de Estudios Agrosociales*, octubre-noviembre, Madrid.
31. Glaeser E.L. and D.C. Maré (2001), "Cities and Skills", *Journal of Labor Economics*, vol. 19, pp.316–342
32. Greene W. H. (2003), "Econometric analysis, 5th edition", Prentice Hall, Upper Saddle River (N.J.),
33. Greene, William H, (1981), "Sample Selection Bias as a Specification Error: Comment", *Econometrica*, Econometric Society, vol. 49(3), pp. 795-98.
34. Greenwood, M. J., (1993), "Migration: A Review", *Regional Studies*, v. 27, iss. 4, pp. 295-96.
35. Greenwood, M. J., (1997), "Internal migration", in Rosenzweig M. R., Stark O. (eds.), *Handbook of Population and Family Economics*, vol. 1B.

36. Hanson, G.H., (1998), "Market Potential, Increasing Returns, and Geographic Concentration," NBER Working Papers 6429, National Bureau of Economic Research, Inc.
37. Harris, J.R and M.P Todaro, (1970), "Migration, Unemployment and Development: A Two-Sector Analysis", American Economic Association, vol. 60(1), pp. 126-42.
38. Head, K and T. Mayer, (2002), "Market potential and the location of Japanese investment in the European Union", CEPR Discussion paper No.3455, Centre for Economic Policy Research, LSE.
39. Head, K and T. Mayer, (2003), "The Empirics of Agglomeration and Trade", CEPR Discussion paper No.3985, Centre for Economic Policy Research, LSE
40. Head, K and T. Mayer, (2005), "Regional Wage and Employment Responses to Market Potential in the EU", C.E.P.R. Discussion Papers, CEPR Discussion Papers: 4908.
41. Heckman (1979), "Sample Selection Bias as a Specification Error", *Econometrica*, Volume 47, Number 1, pp. 153-162.
42. Hirschman, A.O., (1958), "The Strategy of Economic Development", New Haven, CT, Yale University Press.
43. Hummels, D., (1998), "Towards a Geography of Trade Costs", mimeo University of Chicago.
44. Hunt, J.C., and Kau, J. B, (1985) "Migration and Wage Growth. A Human Capital Approach." *Southern Economic Journal* 51, 697-710.
45. Krugman, P.R., (1991), "Increasing returns and economic geography", *Journal of Political Economy*, 99 (3), pp.483-99.
46. Krugman, P.R., (1992) "A dynamic spatial model", NBER Working paper 4219, National Bureau of Economic Research. Cambridge, MA.
47. López-Bazo E. and R. Moreno, (2005), "Regional heterogeneity in the private and social returns to human capital", mimeo.
48. Martí, M. and C. Ródenas Calatayud, (2004), "Migrantes y migraciones: de nuevo la divergencia en las fuentes estadísticas", *Estadística Española*, 46(156), pp. 293-321.
49. McFadden, D.L., (1974), "Conditional Logit Analysis of Qualitative Choice Analysis", in *Frontiers of Econometrics*, ed. P. Zarembka. New York: Academic Press pp. 105-142.
50. Molho, I., (1987), "The migration decision of young men in Great Britain", *Applied Economics*, 19, pp. 221-243.

51. Nakosteen R. and M. Zimmer (1980), "Migration and Income: The Question of Self-Selection", *Southern Economic Journal*, Vol. 46, No. 3, pp. 840-851
52. Nakosteen R. and O. Westerlund (2004), "The effects of regional migration on gross income of labour in Sweden", *Papers in Regional Science*, Volume 83, Number 3, pp. 581 - 595
53. Neary, P., (2001), "Of Hype and Hyperbolas: Introducing the New Economic Geography", *Journal of Economic Literature*, v. 39, iss. 2, pp. 536-61
54. Olano, Rey A., (1990), "Las migraciones interiores en fase de dispersión", *Revista de Economía y Sociología del Trabajo* 8-9, pp. 86-98.
55. Ottaviano, G. and D. Puga, (1998), "Agglomeration in the Global Economy: A Survey of the 'New Economic Geography.'", *World Economy*, v. 21, iss. 6, pp. 707-31.
56. Owen, D. and A. Green, (1992), "Migration patterns and trends", in: T. Champion and T. Fielding (Eds) *Migration Processes and Patterns* (London, Belhaven),
57. Pekkala S., (2002) "Migration and Individual Earnings in Finland: A Regional Perspective" *Regional Studies*, Vol. 36.1, pp. 13-24. (pg 14-16: en contra de Heckman)
58. Pekkala S., (2003), "Migration Flows in Finland: Regional Differences in Migration Determinants and Migrant Types", *International Regional Science Review*, v. 26, iss. 4, pp. 466-82.
59. Pons, J., Silvestre, J., Tirado, D., and E. Paluzie, (2004), "Were Spanish migrants attracted by industrial agglomerations? An analysis for the interwar years in the light of the new economic geography", *Documents de treball de la Facultat de Ciències Econòmiques i Empresariales* n° 121, Universitat de Barcelona.
60. Raymond Bara, J.L. and B. García-Greciano, (1996), "Distribución regional de la renta y movimientos migratorios", *Papeles de Economía Española*, n° 67, pp. 185-200.
61. Redding S. and Venables A., (2004), "Economic Geography and International Inequality", *Journal of International Economics*, January 2004, v. 62, iss. 1, pp. 53-82.
62. Ritsila, J. and M. Haapanen, (2003), "Where Do the Highly Educated Migrate? Micro-level Evidence from Finland", *International Review of Applied Economics*, October 2003, v. 17, iss. 4, pp. 437-48.
63. Ritsila, J. and M. Ovaskainen, (2001), "Migration and Regional Centralization of Human Capital", *Applied Economics*, v. 33, iss. 3, pp. 317-25.

64. Robinson C. and N. Tomes (1982), "Self-selection and interprovincial migration in Canada" *Canadian Journal of Economics* 15, pp. 474-502.
65. Ródenas Calatayud C. and M. Martí, (1997), "¿Son bajos los flujos migratorios en España?", *Revista de Economía Aplicada*, 5 (15), pp.155-171.
66. Ródenas Calatayud C. and M. Martí, (2002), "Migraciones 1990-1999: ¿Qué ha sucedido en la última década?", *Revista Valenciana de Economía y Hacienda*, (5), pp.37-60.
67. Ródenas Calatayud C., (1994a), *Emigración y economía en España*, Universidad de Alicante, Editorial Civitas.
68. Ródenas Calatayud C., (1994b), "Migraciones interregionales en España (1960- 1989): cambios y barreras", *Revista de Economía Aplicada* 2, pp. 5-36.
69. Schwartz A. (1973), "Interpreting the Effect of Distance on Migration", *Journal of Political Economy*, Sept.-Oct. 1973, v. 81, iss. 5, pp. 1153-69.
70. Serrano, L., (1997), "Productividad del trabajo y capital humano en la economía española", *Moneda y Crédito*, (205), pp.79-101.
71. Serrano, L., (1998), "Capital humano y movilidad espacial del trabajo en la economía española", WP-EC98-06. Instituto Valenciano de Investigaciones Económicas (Ivие),
72. Sjaastad L.A., (1962) "The Costs and Returns of Human Migration", *The Journal of Political Economy*, Vol.70, iss.5, pp.80-93.
73. Tirado D., Pons J. and E. Paluzie, (2003), "Industrial agglomerations and wage gradients: the Spanish economy in the interwar period" *ERSA conference papers n° ersa03p113*, European Regional Science Association.
74. Tunali I. (2000), "Rationality of Migration", *International Economic Review*, Volume 41, Number 4, pp. 893-920 (28)
75. Venables, A..J., (1996), "Equilibrium locations of vertically linked industries", *International Economic Review* 37 (2), pp. 341– 359.
76. Wooldridge J. M, (2003), "Introductory econometrics: a modern approach", South-Western Collage, Australia.
77. Zhu N. and S. Poncet, (2003), "Provincial Migration Dynamics in China: Borders, centripetal forces and trade", *ERSA conference papers n° ersa03p148*, European Regional Science Association.

Figures and tables

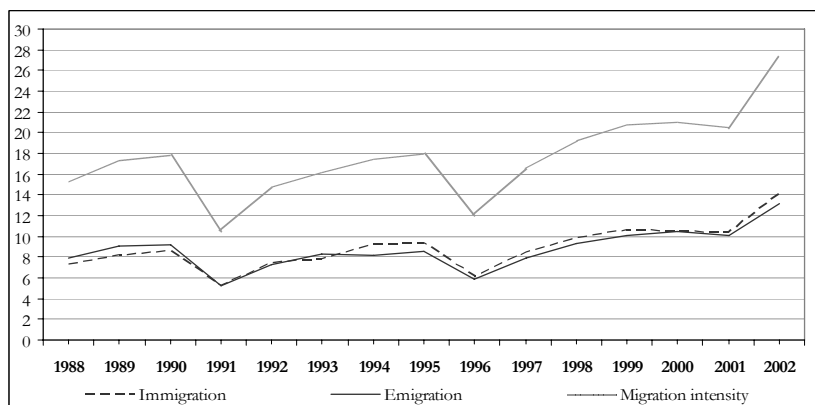


Figure 1: Annual migration rates. Evolution 1988-2002 (for 1000 inhabitants)

NET IMMIGRATION RATE					MIGRATION INTENSITY RATE						
1988		1995		2001	1988		1995		2001		
Girona	8.38	Guadalajara	18.47	Guadalajara	18.05	Guadalajara	28.38	Guadalajara	45.94	Guadalajara	54.40
Tarragona	4.67	Tarragona	7.63	Tarragona	9.47	Girona	24.27	Toledo	27.70	Toledo	32.12
Castellón	4.10	Toledo	6.92	Girona	6.48	Tarragona	22.51	Ávila	26.46	Tarragona	31.93
Alicante	4.04	Girona	4.69	Castellón	6.15	Teruel	21.66	Tarragona	25.29	Girona	28.99
Málaga	3.36	Alicante	4.20	Toledo	6.02	Ávila	21.01	Soria	24.54	Teruel	28.33
Jaén	-4.53	Barcelona	-2.05	Zamora	-3.17	Barcelona	9.16	Sevilla	10.32	Valencia	12.26
Vizcaya	-5.08	Guipúzcoa	-2.19	Cáceres	-3.30	Coruña (A)	8.44	Coruña (A)	10.30	Pontevedra	11.56
Cuenca	-5.45	Jaén	-2.75	León	-3.46	Valencia	8.40	Pontevedra	9.53	Coruña (A)	11.37
Guipúzcoa	-6.31	Palencia	-3.03	Salamanca	-4.34	Pontevedra	7.51	Valencia	9.40	Sevilla	11.29
Teruel	-6.33	Vizcaya	-3.92	Cuenca	-5.52	Asturias	7.07	Asturias	8.75	Asturias	10.63

Tables 1 and 2: Relative position of provinces. Net immigration and migration intensity rates.

1988									
RELATIVE MP À LA HARRIS	AVERAGE SHARE OF MIGRANTS	RELATIVE EMPLOYMENT RATE	RELATIVE NOMINAL WAGE	RELATIVE EXPECTED NOMINAL WAGE	RELATIVE TOTAL EMPLOYMENT	RELATIVE MANUFACTURED SPECIALIZATION	RELATIVE SERVICES SPECIALIZATION	RELATIVE HUMAN CAPITAL	
Madrid	188.91	19.213%	101.53	150.04	169.50	633.05	612.21	821.94	125.16
Barcelona	163.91	9.779%	94.64	129.60	134.90	624.37	1035.47	622.49	114.36
Guadalajara	133.08	0.616%	103.29	106.75	109.04	17.66	19.63	15.67	102.74
Toledo	125.42	1.128%	107.49	83.59	96.46	65.90	79.42	49.52	90.47
Vizcaya	121.41	2.532%	94.82	134.00	131.80	149.93	194.30	168.34	120.62
Cáceres	71.26	1.343%	95.62	77.73	68.89	53.60	19.33	48.32	85.67
Cádiz	67.30	2.029%	73.25	122.54	103.73	104.10	73.50	113.29	94.38
Badajoz	65.04	1.336%	85.27	85.82	68.70	67.24	23.73	70.64	95.41
Almería	62.69	1.430%	101.22	72.93	84.41	54.60	10.71	50.80	95.24
Huelva	60.68	1.059%	81.99	122.10	101.09	42.97	30.55	38.73	91.44

1995									
RELATIVE MP À LA HARRIS	AVERAGE SHARE OF MIGRANTS	RELATIVE EMPLOYMENT RATE	RELATIVE NOMINAL WAGE	RELATIVE EXPECTED NOMINAL WAGE	RELATIVE TOTAL EMPLOYMENT	RELATIVE MANUFACTURED SPECIALIZATION	RELATIVE SERVICES SPECIALIZATION	RELATIVE HUMAN CAPITAL	
Madrid	206.73	17.005%	100.29	134.73	149.86	676.71	546.32	828.65	119.98
Barcelona	187.32	9.336%	99.08	113.58	123.38	665.52	1009.60	664.98	111.89
Guadalajara	145.11	0.832%	108.73	112.48	120.56	17.69	16.53	16.74	98.14
Toledo	135.29	1.591%	102.97	97.58	107.53	59.11	71.96	47.18	89.08
Valencia	127.00	4.333%	95.99	103.89	104.10	280.87	347.07	269.45	106.81
Cádiz	73.48	2.009%	84.89	108.27	78.61	102.10	62.58	115.29	97.06
Ourense	73.04	1.042%	112.34	59.75	65.63	56.84	34.28	35.67	82.32
Badajoz	69.73	1.641%	82.76	89.49	69.31	65.43	31.58	66.14	92.14
Almería	69.39	1.507%	96.73	77.95	85.94	56.64	16.81	55.55	94.66
Huelva	67.43	0.880%	90.46	98.35	89.55	47.13	31.59	47.91	97.47

2001									
	RELATIVE MP À LA HARRIS	AVERAGE SHARE OF MIGRANTS	RELATIVE EMPLOYMENT RATE	RELATIVE NOMINAL WAGE	RELATIVE EXPECTED NOMINAL WAGE	RELATIVE TOTAL EMPLOYMENT	RELATIVE MANUFACTURED SPECIALIZATION	RELATIVE SERVICES SPECIALIZATION	RELATIVE HUMAN CAPITAL
Madrid	214.80	17.965%	103.80	127.18	145.86	722.13	533.35	878.06	114.22
Barcelona	180.04	9.770%	104.23	116.82	133.00	629.85	996.58	603.04	106.09
Guadalajara	149.50	0.828%	104.08	109.92	112.36	18.91	16.96	18.81	104.57
Toledo	139.52	1.653%	101.36	87.99	95.09	65.03	80.97	51.76	93.22
Valencia	129.02	5.132%	100.91	100.26	105.22	294.54	324.48	294.71	107.50
Badajoz	71.06	1.023%	89.13	79.17	65.78	72.59	39.89	69.83	92.96
Lugo	70.99	0.770%	99.16	68.73	65.70	47.43	26.78	35.24	83.34
Almeria	70.46	1.303%	100.12	80.26	91.25	57.71	17.14	57.44	88.05
Huelva	69.26	0.932%	88.31	87.94	77.88	50.32	35.15	39.85	84.63
Ourense	67.71	0.090%	98.43	78.32	75.14	40.28	37.33	35.15	91.79

Tables 3, 4 and 5: Relative positions of provinces. Market potential *à la Harris*.

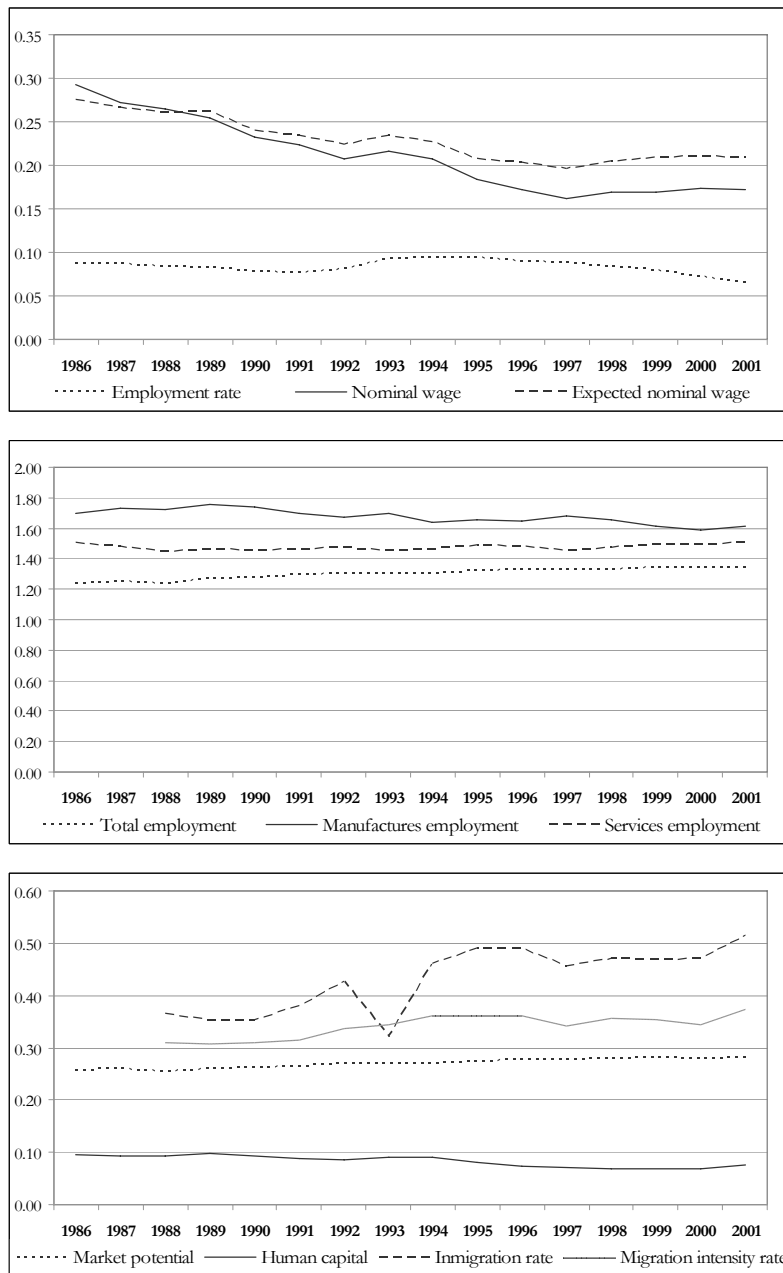


Figure 3: Evolution of coefficients of variation for several variables. Period 1986-2001.

GRAVITY EQUATIONS WITH SAMPLE SELECTION. Heckman. Two-steps estimates									
Dependent variable: share of migrants from region j to region i									
$\ln(\text{sh}_{\{j,t\}})$	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Total employment in host region	0.937***	1.011***		1.015***	1.015***	1.011***	1.008***	1.008***	1.025***
$\ln(L_{\{i,t-1\}})$	0.008	0.006		0.006	1.006	0.005	0.006	0.006	0.006
Manufactured employment in host region			-0.064***						
$\ln(LX_{\{i,t-1\}})$			0.011						
Services employment in host region			1.034***						
$\ln(LY_{\{i,t-1\}})$			0.017						
Traditional employment in host region			-0.016						
$\ln(LZ_{\{i,t-1\}})$			0.017						
Expected nominal wage in host region	0.299***	0.311***	0.221***	0.364***	0.364***	0.311***			
$\ln(\text{probw}_{\{i,t-1\}})$	0.032	0.022	0.031	0.030	0.030	0.021			
Distances between home and host regions	-0.907***	-1.074***	-1.075***	-1.076**	-1.076**	-1.074***	-1.084***	-1.084***	-1.103***
$\ln(d_{\{j\}})$	0.013	0.011	0.010	0.010	0.010	0.010	0.010	0.010	0.010
Common border	1.001***	0.948***	0.957***	0.945***	0.945***	0.948**	0.932***	0.932***	0.910***
$V_{\{j\}}$	0.025	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019
Area in host region	0.061***	0.058***	0.028***	0.056***	0.057***	0.058***	0.032***	0.032***	0.009
$(\text{Area}_{\{i\}})$	0.012	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.008
Human capital in host region				-0.025***	-0.025***		0.062***	0.062***	
$h_{\{i,t-1\}}$				0.010	0.010		0.009	0.008	
Human capital in home region				0.023		0.023	0.025		0.023
$h_{\{j,t-1\}}$				0.016		0.016	0.016		0.016
Regional home region fixed effects									
$a_{\{j\}}$	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time fixed effects (sign)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)
$F_{\{t\}}$	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lambda	-2.465***	-0.775***	-0.563***	-0.780***	-0.780***	-0.776***	-0.832***	-0.832***	-0.898***
	0.144	0.069	0.069	0.069	0.069	0.069	0.071	0.071	0.072
Rho	-2.532	-1.029	-0.769	-1.036	-1.036	-1.030	-1.099	-1.099	-1.179
Sigma	0.973	0.153	0.732	0.753	0.753	0.753	0.757	0.757	0.761
Constant	-7.189***	-6.719***	-4.732***	-7.183***	-7.013***	-6.896***	-4.265***	-3.680***	-3.595***
	0.407	0.251	0.302	0.304	0.278	0.280	0.180	0.166	0.203
No. observations	32430	32430	32430	32430	32430	31932	32430	32430	32430
Censored observation	498	498	498	498	498	498	498	498	498
Uncensored observation	31932	31932	31932	31932	31932	31932	31932	31932	31932
Wald chi2	26900	61335	65580	61299	61298	61336	60273	60271	59289
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

***, **, * significance at 1,5 and 10% level

Table 6: Econometric results. Gravitational equation with sample selection. Heckit estimator (1)

GRAVITY EQUATIONS WITH SAMPLE SELECTION. Heckman. Two-steps estimates						
Dependent variable: share of migrants from region j to region i						
$\ln(\text{sh}_{\{j,t\}})$	(10)	(11)	(12)	(13)	(14)	(15)
Total employment in host region $\ln(L_{\{i,t-1\}})$	0.950*** <i>0.008</i>	0.942*** <i>0.008</i>	947*** <i>0.007</i>	0.942*** <i>0.008</i>	0.934*** <i>0.009</i>	0.958*** <i>0.007</i>
Manufactured employment in host region $\ln(LX_{\{i,t-1\}})$						
Services employment in host region $\ln(LY_{\{i,t-1\}})$						
Traditional employment in host region $\ln(LZ_{\{i,t-1\}})$						
Expected nominal wage in host region $\ln(\text{probw}_{\{i,t-1\}})$	0.375*** <i>0.039</i>	0.359*** <i>0.034</i>	0.334*** <i>0.029</i>			
Distances between home and host regions $\ln(d_{\{j\}})$	-0.890*** <i>0.012</i>	-0.909*** <i>0.013</i>	-0.888*** <i>0.012</i>	-0.897*** <i>0.013</i>	-0.915*** <i>0.014</i>	-0.914*** <i>0.013</i>
Common border $V_{\{j\}}$	1.048*** <i>0.023</i>	1.007*** <i>0.025</i>	1.050*** <i>0.023</i>	1.035*** <i>0.023</i>	0.995*** <i>0.025</i>	1.011*** <i>0.024</i>
Area in host region $(\text{Area}_{\{i\}})$	0.064*** <i>0.011</i>	0.060*** <i>0.012</i>	0.065*** <i>0.011</i>	0.038*** <i>0.011</i>	0.035*** <i>0.012</i>	0.013 <i>0.010</i>
Human capital in host region $h_{\{i,t-1\}}$	-0.019 <i>0.013</i>	-0.029*** <i>0.014</i>		0.070*** <i>0.010</i>	0.057*** <i>0.011</i>	
Human capital in home region $h_{\{j,t-1\}}$	0.129*** <i>0.008</i>		0.130*** <i>0.008</i>	0.127*** <i>0.008</i>		0.122*** <i>0.008</i>
Regional home region fixed effects $a_{\{j\}}$	No	No	No	No	No	No
Time fixed effects (sign) $F_{\{t\}}$	(-) Yes	(-) Yes	(-) Yes	(-) Yes	(-) Yes	(-) Yes
Lambda	-2.045*** <i>0.129</i>	-2.480*** <i>0.145</i>	-2.032*** <i>0.128</i>	-2.121*** <i>0.013</i>	-2.548*** <i>0.148</i>	-2.733*** <i>0.178</i>
Rho	-2.235	-2.543	-2.224	-2.291	-2.836	-2.655
Sigma	0.915	0.975	0.913	0.926	0.986	1.029
Constant	-8.818** <i>0.373</i>	-7.520*** <i>0.442</i>	-8.599*** <i>0.373</i>	-5.782*** <i>0.201</i>	-4.629*** <i>0.181</i>	-5.005*** <i>0.153</i>
No. observations	32430	32430	32430	32430	32430	31932
Censored observation	498	498	498	498	498	498
Uncensored observation	31932	31932	31932	31932	31932	31932
Wald chi2	31701 <i>0.000</i>	26757 <i>0.000</i>	31837 <i>0.000</i>	30619 <i>0.000</i>	25912 <i>0.000</i>	28960 <i>0.000</i>

***, **, * significance at 1,5 and 10% level

Table 7: Econometric results. Gravitational equation with sample selection. Heckit estimator (2)

SELECTION EQUATION. Probit estimated by ML									
Dependent variable: share of migrants from region j to region i (binary)									
sh_pr_{i,t}	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(8)
Total employment in host region	0.500***	0.852***		0.852***	0.852***	0.852***	0.763***	0.763***	0.763***
ln(L_{i,t-1})	0.104	0.118		0.118	0.118	0.118	0.105	0.105	0.105
Manufactured employment in host region			-0.493						
ln(LX_{i,t-1})			0.263						
Services employment in host region			1.158***						
ln(LY_{i,t-1})			0.322						
Traditional employment in host region			0.273						
ln(LZ_{i,t-1})			0.300						
Expected nominal wage in host region	-0.116	0.577	0.363	0.577	0.577	0.577			
ln(prob_{i,t-1})	0.257	0.362	0.355	0.362	0.362	0.362			
Distances between home and host regions	-0.634***	-1.335***	-1.329***	-1.335***	-1.335***	-1.335***	-1.326***	-1.326***	-1.326***
ln(d_{ij})	0.064	0.104	0.105	0.104	0.104	0.104	0.103	0.103	0.103
Common border	5.354	6.094	6.120	6.094	6.094	6.094	6.090	6.090	6.090
V_{ij}	-	-	-	-	-	-	-	-	-
Area in host region	0.232***	0.229***	0.258***	0.229***	0.229***	0.229***	0.255***	0.255***	0.255***
(Area_{ij})	0.066	0.075	0.080	0.075	0.075	0.075	0.073	0.073	0.073
Regional home region fixed effects									
a_{ij}	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time fixed effects (sign)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)
F_{it}	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Nominal wage differences between home and host regions	-0.000***	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000***	-0.000***	-0.000***
(dif_wn_{ij,t-1})	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Housing prices differences between home and host regions	0.000***	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(dif_pv_{ij,t-1})	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Human capital differences between home and host regions	-0.111***	-0.0147	-0.003	-0.0147	-0.0147	-0.0147	-0.019	-0.019	-0.019
(dif_hk_{ij,t-1})	0.043	0.055	0.059	0.055	0.055	0.055	0.055	0.055	0.055
Unemployment rate in host region	-0.950*	-0.178	-0.952	-0.178	-0.178	-0.178	-0.668	-0.668	-0.668
(Un_{i,t-1})	0.524	0.601	0.657	0.601	0.601	0.601	0.517	0.517	0.517
Unemployment rate in home region	1.862***	0.25	0.347	0.25	0.25	0.25	0.471	0.471	0.471
(Un_{j,t-1})	0.038	1.325	1.327	1.325	1.325	1.325	1.316	1.316	1.316
Employment growth in host region	1.079*	1.405***	1.262***	1.405***	1.405***	1.405***	1.423***	1.423***	1.423***
(Eg_{i,t-1})	0.576	0.618	0.621	0.618	0.618	0.618	0.617	0.617	0.617
Host region coastal province	0.101	0.260***	0.282***	0.260***	0.260***	0.260***	0.286***	0.286***	0.286***
Ct_{ij}	0.075	0.083	0.087	0.083	0.083	0.083	0.082	0.082	0.082
Productive infrastructure in host region	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(infpd_{i,t-1})	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Social infrastructure in host region	0.000***	0.000	0.000	0.000	0.000	0.000	0.000*	0.000*	0.000*
(infsoc_{i,t-1})	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Manufactures employment proportion in host region	0.882	0.205	4.243***	0.205	0.205	0.205	0.686	0.686	0.686
(mf_{i,t-1})	0.616	0.677	2.387	0.677	0.677	0.677	0.607	0.607	0.607
Manufactures employment proportion in home region	2.686***	3.408***	3.386***	3.408***	3.408***	3.408***	3.138***	3.138***	3.138***
(mf_{j,t-1})	0.433	1.445	1.448	1.445	1.445	1.445	1.433	1.433	1.433
Tradable services employment proportion in host region	3.311***	2.372***	0.795	2.372***	2.372***	2.372***	2.890***	2.890***	2.890***
(sv_{i,t-1})	0.731	0.802	1.573	0.802	0.802	0.802	0.734	0.734	0.734
Tradable services employment proportion in home region	7.323***	-0.666	-0.647	-0.666	-0.666	-0.666	-0.612	-0.612	-0.612
(sv_{j,t-1})	0.545	1.216	1.221	1.216	1.216	1.216	1.216	1.216	1.216
Constant	-1.913	-2.744	-1.232	-2.744	-2.744	-2.744	2.437***	2.437***	2.437***
	2.258	3.477	3.350	3.477	3.477	3.477	1.233	1.233	1.233

***, **, * significance at 1,5 and 10% level

Table 8: Econometric results. Gravitational equation with sample selection. Selection equation (1)

SELECTION EQUATION. Probit estimated by ML						
Dependent variable: share of migrants from region j to region i (binary)						
sh_pr_{i,t}	(10)	(11)	(12)	(13)	(14)	(15)
Total employment in host region	0.500***	0.500***	0.500***	0.524***	0.524***	0.524***
ln(L_{i,t-1})	0.104	0.104	0.104	0.090	0.090	0.090
Manufactured employment in host region						
ln(LX_{i,t-1})						
Services employment in host region						
ln(LY_{i,t-1})						
Traditional employment in host region						
ln(LZ_{i,t-1})						
Expected nominal wage in host region	-0.116	-0.116	-0.116			
ln(prob_w_{i,t-1})	0.257	0.257	0.257			
Distances between home and host regions	-0.634***	-0.634***	-0.634***	-0.633***	-0.633***	-0.633***
ln(d_{ij})	0.064	0.064	0.064	0.064	0.064	0.064
Common border	5.355	5.355	5.355	5.356	5.356	5.356
V_{ij}	-	-	-	-	-	-
Area in host region	0.232***	0.232***	0.232***	0.225***	0.225***	0.225***
(Area_{ij})	0.066	0.066	0.066	0.065	0.065	0.065
Regional home region fixed effects						
a_{ij}	No	No	No	No	No	No
Time fixed effects (sign)	(-)	(-)	(-)	(-)	(-)	(-)
F_{jt}	Yes	Yes	Yes	Yes	Yes	Yes
Nominal wage differences between home and host regions	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***
(dif_wn_{i,t-1})	0.000	0.000	0.000	0.000	0.000	0.000
Housing prices differences between home and host regions	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***
(dif_pv_{i,t-1})	0.000	0.000	0.000	0.000	0.000	0.000
Human capital differences between home and host regions	-0.111*	-0.111*	-0.111*	-0.113***	-0.113***	-0.113***
(dif_hk_{i,t-1})	0.043	0.043	0.043	0.043	0.043	0.043
Unemployment rate in host region	-0.950*	-0.950*	-0.950*	-0.866*	-0.866*	-0.866*
(Un_{i,t-1})	0.524	0.524	0.524	0.490	0.490	0.490
Unemployment rate in home region	1.861***	1.861***	1.861***	1.858***	1.858***	1.858***
(Un_{j,t-1})	0.380	0.380	0.380	0.380	0.380	0.380
Employment growth in host region	1.079*	1.079*	1.079*	1.063*	1.063*	1.063*
(Eg_{i,t-1})	0.576	0.576	0.576	0.575	0.575	0.575
Host region coastal province	0.100	0.100	0.100	0.094	0.094	0.094
Ct_{jt}	0.075	0.075	0.075	0.073	0.073	0.073
Productive infrastructure in host region	0.000	0.000	0.000	0.000	0.000	0.000
(infpd_{i,t-1})	0.000	0.000	0.000	0.000	0.000	0.000
Social infrastructure in host region	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***
(infsoc_{i,t-1})	0.000	0.000	0.000	0.000	0.000	0.000
Manufactures employment proportion in host region	0.881	0.881	0.881	0.706	0.706	0.706
(mf_{i,t-1})	0.616	0.616	0.616	0.480	0.480	0.480
Manufactures employment proportion in home region	2.686***	2.686***	2.686***	2.624***	2.624***	2.624***
(mf_{j,t-1})	0.433	0.433	0.433	0.410	0.410	0.410
Tradable services employment proportion in host region	3.310***	3.310***	3.310***	3.143***	3.143***	3.143***
(sv_{i,t-1})	0.731	0.731	0.731	0.631	0.631	0.631
Tradable services employment proportion in home region	7.323***	7.323***	7.323***	7.280***	7.280***	7.280***
(sv_{j,t-1})	0.545	0.545	0.545	0.536	0.536	0.536
Constant	-1.913	-1.913	-1.913	-2.862***	-2.862***	-2.862***
	2.258	2.258	2.258	0.848	0.848	0.848

***, **, * significance at 1,5 and 10% level

Table 9: Econometric results. Gravitational equation with sample selection. Selection equation (2)

GRAVITY EQUATIONS. Linear regression by OLS									
Dependent variable: share of migrants from region j to region i									
$\ln(\text{sh}_{\{j,t\}})$	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Total employment in host region	1.006***	1.032***		1.036***	1.036***	1.032***	1.029***	1.029***	1.050***
$\ln(L_{\{i,t-1\}})$	0.005	0.005		0.005	0.005	0.005	0.005	0.005	0.005
Manufactured employment in host region			-0.062***						
$\ln(LX_{\{i,t-1\}})$			0.011						
Services employment in host region			1.052***						
$\ln(LY_{\{i,t-1\}})$			0.015						
Traditional employment in host region			-0.0219						
$\ln(LZ_{\{i,t-1\}})$			0.016						
Expected nominal wage in host region	0.400***	0.338***	0.231***	0.385***	0.386***	0.338***			
$\ln(\text{probw}_{\{i,t-1\}})$	0.024	0.022	0.031	0.031	0.031	0.022			
Distances between home and host regions	-0.935***	-1.088***	-1.085***	-1.090***	-1.090***	-1.088***	-1.100***	-1.100***	-1.123***
$\ln(d_{\{j\}})$	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010
Common border	1.064***	0.959***	0.965***	0.957***	0.957***	0.956***	0.944***	0.944***	0.919***
$V_{\{j\}}$	0.017	0.017	0.017	0.017	0.017	0.018	0.017	0.017	0.017
Area in host region	0.074***	0.060***	0.031***	0.059***	0.059***	0.060***	0.033***	0.033***	0.007
$(\text{Area}_{\{i\}})$	0.009	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008
Human capital in host region				-0.022**	-0.023**		0.709***	0.070***	
$h_{\{i,t-1\}}$				0.010	0.010		0.007	0.007	
Human capital in home region				0.022		0.022	0.024		0.022
$h_{\{j,t-1\}}$				0.015		0.015	0.015		0.015
Regional home region fixed effects									
$a_{\{j\}}$	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time fixed effects (sign)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)
$F_{\{t\}}$	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	-8.459***	-7.033***	-4.861***	-7.461***	-7.295***	-7.205***	-4.382***	-4.200***	-3.620***
	0.272	0.254	0.308	0.306	0.283	0.280	0.175	0.128	0.158
No. observations	31932	31932	31932	31932	31932	31932	31932	31932	31932
Rbar-squared	0.657	0.707	0.717	0.707	0.707	0.707	0.705	0.705	0.704
Root MSE	0.796	0.737	0.724	0.736	0.736	0.737	0.738	0.738	0.739
F-stat	4428	1484	1478	1441	1463	1461	1456	1478	1473
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

***, **, * significance at 1,5 and 10% level; robust estimation of standard errors

Table 10: Econometric results. Gravitational equation without sample selection. OLS estimator (1)

GRAVITY EQUATIONS. Linear regression by OLS						
Dependent variable: share of migrants from region j to region i						
$\ln(\text{sh}_{\{j,t\}})$	(10)	(11)	(12)	(13)	(14)	(15)
Total employment in host region	1.003***	1.007***	1.004***	0.996***	0.999***	1.026***
$\ln(L_{\{i,t-1\}})$	0.006	0.006	0.005	0.005	0.005	0.005
Manufactured employment in host region						
$\ln(LX_{\{i,t-1\}})$						
Services employment in host region						
$\ln(LY_{\{i,t-1\}})$						
Traditional employment in host region						
$\ln(LZ_{\{i,t-1\}})$						
Expected nominal wage in host region	0.422***	0.414***	0.423***			
$\ln(\text{probw}_{\{i,t-1\}})$	0.033	0.033	0.024			
Distances between home and host regions	-0.904***	-0.936***	-0.904***	-0.912***	-0.944***	-0.940***
$\ln(d_{\{ji\}})$	0.010	0.010	0.010	0.010	0.010	0.010
Common border	1.104***	1.063***	1.104***	1.092***	1.051***	1.062***
$V_{\{ji\}}$	0.017	0.017	0.017	0.017	0.017	0.017
Area in host region	0.075***	0.073***	0.075***	0.047***	0.045***	0.009
$(\text{Area}_{\{i\}})$	0.009	0.009	0.009	0.009	0.009	0.009
Human capital in host region	0.0005	-0.006		0.103***	0.094***	
$h_{\{i,t-1\}}$	0.010	0.011		0.008	0.008	
Human capital in home region	0.171***		0.171***	0.170***		0.166***
$h_{\{j,t-1\}}$	0.006		0.006	0.006		0.006
Regional home region fixed effects						
$a_{\{j\}}$	No	No	No	No	No	No
Time fixed effects (sign)	(-)	(-)	(-)	(-)	(-)	(-)
$F_{\{t\}}$	Yes	Yes	Yes	Yes	Yes	Yes
Constant	-10.014***	-8.537***	-10.020***	-6.663***	-5.224***	-5.538***
	0.306	0.300	0.280	0.147	0.133	0.121
No. observations	31932	31932	31932	31932	31932	31932
Rbar-squared	0.664	0.657	0.664	0.662	0.655	0.661
Root MSE	0.788	0.797	0.779	0.790	0.798	0.793
F-stat	3959	4213	4152	4136	4418	4389
	0.000	0.000	0.000	0.000	0.000	0.000

***, **, * significance at 1,5 and 10% level; robust estimation of standard errors

Table 11: Econometric results. Gravitational equation without sample selection. OLS estimator (2)