Location effects and wages: evidence from Italian panel data on employees*

Giordano Mion[†] and Paolo Naticchioni[‡]

March 2004

Abstract

There is an important literature, both theoretical and empirical, concerning the impact of spatial variables on economic structure and performance. In this paper we attempt to identify and separate the different kind of agglomeration forces on the Italian wage distribution. Using an Italian administrative panel database (INPS), and adding some other geographical economic variables, we argue that the standard urbanization and specialization externalities entail a significant and positive effect on wages in the Italian labour market. Furthermore, we claim that provincial wage distribution is also positively affected by the potential market demand, coming from adjacent locations, and from export. These results confirm the theoretical predictions of new economic geography.

Keywords: Spatial Economics, Wage distribution

JEL Codes: .

*We thank for helpful comments and suggestions. We are also grateful to for providing us with local data on distances and economic indicators.

†Giordano Mion: CORE (Université Catholique de Louvain-la-Neuve, Belgium), Address: CORE-UCL, Voie du Roman Pays 34, 1348 LLN, Belgium. Email: mion@core.ucl.ac.be.

‡Paolo Naticchioni, Università di Roma "La Sapienza", Universitè Catholique de Louvain (IRES). Email: paolo.naticchioni@uniroma1.it

1 Introduction

Italy is probably one of the best example of how economic activities can distribute in a very unequal way across space. So far, many explanation have been put forward to explain the North-South divide. Give some examples here. However, the relatively recent new economic geography literature (NEG) has provided a collection of general equilibrium models explicitly dealing with space, and capable to account for many salient features of the economic landscape[1]. Agents choose their location on the basis of market-price incentives. Then, the combination of increasing returns at firm level with market power (usually in the form of monopolistic competition) and transportation costs, give rise to an endogenous agglomeration, provided that centripetal forces are sufficiently strong.

One common feature of these model, that is particularly important in our paper, is that equilibrium nominal wages are higher in agglomerated areas. More precisely, the concentration of firms and consumers create a positive externality (of a pecuniary nature) that increases factors remuneration in those locations that offer a better access to final demand (usually referred as market potential). In order to test for this prediction, we will use a panel data on individual workers for Italy by controlling for both individual characteristics as well as for localized externalities (like Urbanization and Marshall-Arrow-Romer ones) that are likely to affect wages.

Market access affects are supposed to be small in Italy, because in most sector there is a lower bound on wages that is set at national level. However, although quantitatively small, we do find evidence of them. Introduce some results here.

The rest of the paper is organized as follows.

2 The model

The NEG literature offers the possibility to treat agglomeration in a flexible and rigorous way by means of increasing returns (IRS), imperfect competition, and product differentiation. In this Section, we are particularly concerned with Helpman's (1998) model, which will be the theoretical ground on which we will construct the econometric analysis.

Imagine an economy consisting of Φ locations, two sectors (the manufacturing sector M and the housing sector H), and one production factor (labor). The M-sector produces a continuum of

[1] See Fujita, Krugman and Venables (1999), and Fujita and Thisse (2002) for a review of the literature.

varieties of a horizontally differentiated product under increasing returns to scale, using labor as the only input. Each variety of this differentiated good can be traded among locations incurring in iceberg-type transportation costs.[2] Referring to two generic locations as j and k ($j, k = 1, 2, ..., \Phi$), we thus have that for each unit of good shipped from j to k, just a fraction $v_{j,k} = f(d_{j,k})$ of it, where $d_{j,k}$ is distance between the two locations and $f()$ is a decreasing function. This means that, indicating with $p_{m,j}$ the mill price of a variety produced in location j, the corresponding delivered price for the consumer living in k would be $p_{m,j}/v_{j,k}$. Firms receive mill prices while consumers pay delivered. The H-sector provides instead a homogeneous good, housing, that cannot be traded and whose amount in each location (H_j) is supposed to be exogenously fixed. Its price $P_{H,j}$ can therefore differ from one place to another and is determined by the equilibrium between local supply and demand.

Labor is supposed to be freely mobile, and its (exogenous) total amount in the economy is equal to L. The equilibrium spatial distribution of our workers-consumers is thus determined by both wages (w_j), and prices prevailing in each location. We will denote L_j, with $\sum_{j=1}^{\Phi} L_j = L$, as labor in location j, and $\lambda_j = L_j/L$ as the corresponding share of total workers. Preferences and technology do not directly depend upon the location where consumption and production take place, but only indirectly through prices. Therefore it is notationally convenient to describe them, as well as firms' behavior, without explicitly referring to any particular location.

Preferences are identical across all workers. As usual in NEG models, they are described by the standard Cobb-Douglas utility function with CES type sub-utility for the differentiated product, i.e.:

$$U = (C_M)^{\mu} (C_H)^{1-\mu} \qquad 0 < \mu < 1 \tag{1}$$

where C_M stands for an index of the consumption of the M-sector varieties, while C_H is housing consumption. We assume that the modern sector provides a continuum of varieties of (endogenous) size N, the consumption index C_M is thus given by[3]:

[2] The term transportation costs does not simply refers to shipment costs but in general to all costs and impediments of doing business in different markets, like information costs, language differences, etc.

[3] In the original Helpman [1998] formulation, as well as in Krugman [1991] and Krugman [1992], N is not a mass but instead the finite number of varieties provided by the market. However, as pointed out by Fujita and Thisse [2001], this approach is conceptually misleading for the monopolistic competition framework. In fact, in order to be consistent with the requirement that firms are negligible with respect to the market, we should consider a continuum of them. If we do not and use instead an integer number of firms, strategic interactions actually dominates (d'Aspremont, Dos Santos Ferreira and Gerard-Varet [1996]). However, the way N is actually treated

$$C_M = \left[\int_0^N c_m(s)^\rho ds \right]^{1/\rho} \qquad 0 < \rho < 1 \tag{2}$$

where $c_m(s)$ represents the consumption of variety $s \in [0, N]$. Hence, each consumer has a love for variety and the parameter $\sigma \equiv 1/(1-\rho)$, varying from 1 to ∞, represents the (constant) elasticity of substitution between any two varieties. The bigger is σ the more varieties are substitutes: when σ is close to 1 the desire to spread consumption over all varieties increases. If Y denotes the consumer income, then the demand function for a variety s coming from utility maximization is:

$$c_m(s) = p_m(s)^{-\sigma} \, \mu Y (P_M)^{\sigma - 1} \qquad s \in [0, N] \tag{3}$$

where $p_m(s)$ is here the consumer-price (or delivered price) of our generic variety and P_M is the price-index of the differentiated product given by:

$$P_M \equiv \left[\int_0^N p_m(s)^{-(\sigma-1)} ds \right]^{-1/(\sigma-1)} \tag{4}$$

Technology is the same across locations. Each variant of the differentiated product needs labor to be produced. The relation between the amount of labor used ($l(s)$) and the quantity of variant s produced ($c(s)$) is given by:

$$l(s) = fix + \beta c(s) \tag{5}$$

where fix and β are, respectively, the fixed and the marginal labor requirements. The presence of the fixed cost fix clearly imply increasing returns. Without loss of generality we choose the unit for labor such that $\beta = 1$. Since preferences exhibits a symmetric love for diversity and since there are increasing returns to scale but no scope economies, each variety is produced by a single firm. Moreover, as soon as each firm is supposed to be small relative to the market, firms eventually producing more than 1 (up to a set of zero measure) variety would act as if they were actually different. In turn, this implies an identity between the mass of firms and the mass of varieties with the output of each firm equating the demand for the corresponding variety, the latter coming from consumers spread all over the Φ locations.

by Helpman, is such that final results are virtually unchanged. Nevertheless, we prefer to use here the continuum formulation.

Firms know consumers' demand and choose prices in order to maximize their profits given by:

$$\pi(s) = p_m(s)q(s) - w[fix + q(s)] \tag{6}$$

where w is wage paid by our generic firm and $q(s)$ is its output. However, when they look at demand structure, i.e. equation (3), it is likely that they consider Y and P_M as given. Since each of them has a negligible influence on the market, it may accurately neglect the impact of a price change over both consumers' income and the price index. Consequently, (3) implies that each firm faces an isoelastic downward sloping demand with elasticity given by our parameter σ. Solving first order conditions yields the common equilibrium relation between the optimal price, elasticity of demand, and marginal cost:

$$p_m(s) = \frac{w}{1 - (1/\sigma)} \tag{7}$$

Under free entry, profits are zero. This implies, together with equation (7), that the equilibrium output is a constant given by:

$$q(s) = q = (\sigma - 1)fix \tag{8}$$

Note that this relation is true wherever our firm is located. As a result, in equilibrium a firm's labor requirement is also unrelated to firms' distribution:

$$l(s) = l = \sigma\ fix \tag{9}$$

so that the total mass of firms in the manufacturing sector (N) is constant and equal to $L/(\sigma\ fix)$.

Firms and consumers have an address in space and must choose a location. We can now summarize the long-run spatial equilibrium of our economy by means of five equations introducing space indexes on preferences and technology. The first equilibrium requirement comes from utility maximization. Our Cobb-Douglas utility function is in fact such that the (optimal) share of expenditure on each product is constant and equal to the corresponding exponent. If $E_{H,j}$ denotes consumers' expenditure on houses in location j, Y_j the corresponding income, and $C_{H,j}$ total housing consumption in that region, then $\forall j = 1, 2, ..., \Phi$ we have:

$$E_{H,j} \equiv p_{H,j}C_{H,j} = p_{H,j}H_j = (1 - \mu)Y_j \tag{10}$$

where the second equality comes from the equilibrium between local supply and demand of houses ($C_{H,j} = H_j$).

Since there is free entry and exit and, therefore, zero profit in equilibrium the value of the manufacturing production in each region equals factor earnings ($w_j \lambda_j L$). If we now suppose that each individual owns an equal share of the total housing stock, then income in location j is given by[4]:

$$Y_j = \left[\lambda_j \frac{1-\mu}{\mu} \sum_{k=1}^{\Phi} \lambda_k w_k L \right] + \lambda_j w_j L \tag{11}$$

Moreover, for a spatial distribution of workers to be an equilibrium, there should be no incentive to move. As they are perfectly mobile, this implies an equalization of real wages in the long run:

$$\frac{w_j}{(P_{M,j})^{\mu} (P_{H,j})^{1-\mu}} = \frac{w_k}{(P_{M,k})^{\mu} (P_{H,k})^{1-\mu}} \qquad \forall j, k = 1, 2, ..., \Phi \tag{12}$$

Finally, as shown rigorously in Fujita and Thisse [2001], the last two equilibrium relations are:

$$P_{M,j} = \kappa_1 \left[\sum_{k=1}^{\Phi} \lambda_k (w_k \, v_{j,k})^{1-\sigma} \right]^{1/(1-\sigma)} \tag{13}$$

and

$$w_i = \kappa_2 \left[\sum_{k=1}^{\Phi} Y_k (P_{M,k} \, v_{i,k})^{\sigma-1} \right]^{1/\sigma} \tag{14}$$

with $\kappa_1 \equiv \rho^{-1} (H/\sigma f)^{1/(1-\sigma)}$and $\kappa_2 \equiv \rho \, [\mu/(\sigma-1)f]^{1/\sigma}$. Equation (13) comes from optimal pricing rule (7) and zero profit condition (8). Condition (14) express the equilibrium between supply and demand of labor in each location and comes from firm equilibrium labor requirement (9) and consumers' demand (3).

Depending on the level of transportation costs ($v_{j,k}$), elasticity of substitution among varieties (σ) and the share of traded goods in consumers' expenditure (μ), manufacturing activities will be dispersed or agglomerated. In the second case firms will be disproportionately distributed with

[4]From equation (10) total housing expenditure in our Φ locations is given by $E_H = \sum_{k=1}^{\Phi} E_{H,k} = (1 - \mu) \sum_{k=1}^{\Phi} Y_k$. Moreover, $\forall \, k$ we have $\mu Y_k = w_k \lambda_k L$ and taking the sum we get $\sum_{k=1}^{\Phi} Y_k = 1/\mu \left[\sum_{k=1}^{\Phi} w_k \lambda_k L \right]$. Combining these two relations we finally get equation (11).

respect to a region size. In particular, indicating with (H_j) the stock of housing of region j, those locations with an above (below) average endowment will have a more (less) than proportional share of manufacturing in equilibrium. Both a higher share of tradable goods (μ) or a lower elasticity (σ) induce more agglomeration. In the case of μ, this is due to the fact that concentration of firms and consumers in the same place allow to avoid transportation costs thus increasing real consumption. The greater is the share of these goods in the consumption of migrating workers, the stronger is this centripetal force. The role of σ is instead to counterbalance the usual centrifugal force that works against concentration: price competition. A lower elasticity of substitution σ makes in fact varieties more differentiated, relaxing local competition among sellers.

3 Market access vs localized externalities

In order to give a useful interpretation of the kind of investigations we want to deal with, as well as to link them to previous studies, one has to come back to Harris's (1954) market-potential concept. Actually, Harris's (1954) market-potential relates the potential demand for goods and services produced in a location $j = 1, 2, .., \Phi$ to that location's proximity to consumer's markets, or:

$$MP_j = \sum_{k=1}^{\Phi} Y_k g(d_{jk}) \tag{15}$$

where MP_j is the market potential of location j, Y_k is an index of purchasing capacity of location k (usually income), $d_{jk}$is the distance between two generic locations j and k and $g()$ is a decreasing function. The higher is the market potential index of a location, the higher is its attraction power on production activities.

In Helpman's model, a good measure of the attractiveness of location j is given by the equilibrium nominal wages w_j. Although firms makes no profits in equilibrium (no matter where they are located), the wage they can afford expresses their capacity to create value once located in a particular region. In fact, if centripetal forces take over, those locations that attract more firms and consumers will also have higher equilibrium nominal wages, thus leading to a positive correlation between agents' concentration and w_j. Combining equilibrium equations and applying logarithms to simplify things one get the following expression for $\ln(w_j)$:

$$\ln(w_j) = \kappa_3 + \sigma^{-1} \ln \left[\sum_{k=1}^{\Phi} Y_k^{\frac{1-\sigma(1-\mu)}{\mu}} H_k^{\frac{(1-\mu)(\sigma-1)}{\mu}} w_k^{\frac{(\sigma-1)}{\mu}} v_{j,k}^{(\sigma-1)} \right] \tag{16}$$

where κ_3 is a function of behavioral parameters (μ, σ). Equation (16) really looks like a market-potential function. It tells us that as long as agglomeration forces are at work ($\sigma(1-\mu) < 1$), the nominal wage in location j (and so firms' profitability) is an increasing function of the weighted purchasing power coming from surrounding locations (Y_k), with weights inversely related to distances d_{jk} trough the transport technology function $f(.)$ (this is the market access component).

This kind of equation is not a peculiar feature of Helpman (1998) only. Using different NEG models, Fujita, Krugman and Venables (1999), for instance, obtain several reduced-form equilibrium equations that looks like market potential functions. In particular, variable expressing the attractiveness of a location (as profits, or wages) turns out to be a positive function of the level of economic activity in the surrounding regions with weights inversely related to distances. As further shown by Puga (1998), this spatial inequalities are stronger the less mobile are workers. The Italian situation is actually characterized by a scarce labor mobility, and this should foster local imbalances. Nevertheless, in many sectors there is a bottom floor imposed at national level on wages, and this greatly reduces regional wages gaps that are instead mapped into strong differences in unemployment rates. The net effect of these two opposing forces on local wages cannot be known a priori.

In order to capture agglomeration forces, we will use a measure of market potential based on Households' income and distances among Italian provinces. Furthermore, as a proxy for the accessibility to foreign markets, we will also construct a measure of trade openness for each province that controls for its sectoral specialization.

Say that, contrary to localized externalities, pecuniary externalities have a large spatial extent and needs spatially weighted variables instead of simple local indicators.....give the indicators of localized externalities.

4 Data description

In this paper we utilize an administrative database supplied by INPS (the Italian social security institute). More specifically, we work on a panel version of this database, elaborated by ISFOL.

The sample units are salaried full-time workers[5] in all private sectors but agriculture. The panel is an employer-employee database, constructed merging INPS employee information database with the employer information database and covers 14 years from 1985 to 1998. The sample scheme has been set up to follow individuals born on the 10^{th} of March, June, September and December, and therefore the proportion of this sample on the Italian employees population is approximately of 1/90.

As far as workers information is concerned, the database contains many individual information like age, gender, qualification, place and date of birth, region where the job takes place, date of beginning and end of the current worker contract, the social security contribution, if the worker is either part time or full time, the yearly wage (which does not take into account the number of worked days) and the daily wage.

For the firms this database contains the following information: headquarter region, production region, the average number of employee (or firm size), the sector and the date of start up and shut down (if the firm has shut down in the panel period) of the firm. This means that in this database we have a detailed information of the location of both the company headquarter and the company production plant.

As far as the workplace location is concerned we use the 95 Italian Provinces (Province). This choice represents a good compromise between a detailed classification of the Italian territory and a statistical significance of the analysis. It is worth noting that our database does not represent the universe of all the Italian workers but only a share (1/90) of it.

Moreover, as sectoral decomposition we have chosen the ATECO 91 code system, which divide the Italian economy in 32 sector (using the 'sottosezioni' level in ATECO91). We decide not to use a finer classification in order to have a higher robustness of the results.

In order to investigate the spatial impact of the economic structure on wage distribution we have integrated the INPS database with additional information. First of all, INPS sample is not useful to derive geographical information concerning firm and employment concentration, density, specialization etc., since it is built from the employee database which by definition overestimate the probability to include a big size firm in the sample. For this reason we have to use information from the Italian Census of Industry and Services (ISTAT) to derive the geographical economic

[5] Apprenticeships and part time workers are excluded from the dataset, since the attention is focused on standard labour market contracts (blue collar, white collar and managers).

variables that then we have attached to the worker information of our Panel[6].

More specifically, from this database we have built the employment density and the specialization level of each province, defining these variables as in Combes (2000). The employment density in province j is intuitively defined in the following way[7]:

$$den_{j,t} = \ln \left[\frac{empl_{j,t}}{province\ size_z} \right]$$

while the specialization index for sector s in province j is defined as the ratio of the employment share of sector s in province j divided by the same ratio at the national level.

$$spec_{j,s,t} = \ln \left[\frac{empl_{j,s,t}/empl_{j,t}}{emp_{s,t}/emp_t} \right].$$

It is worth noting that dividing for the same ratio at the national level is useful to compare values for different sectors. Moreover, all geographical variable are taken in logarithms, meaning that parameters represents the elasticity with respect to each variables and in this way they are easily comparable among variables.

We use these two spatial variables because they are proxies for two of the most important externalities in economic geography.

The employment density stands for a proxy of the urbanization externality, defined as intersectoral positive effects deriving from economic agglomeration. The causes behind this externalities might be linked to both information spillovers and market-based forces.

In the former case it is possible to argue that for instance firms have different piece of relevant information that can be exchanged through both the turnover of high skilled workers and informal communication due to close proximity of firms. Actually, even if modern technologies have greatly improved the exchange of information even among location far one another, recent works[8] shows that the importance of local information spillovers is still very relevant.

In the latter case it would be possible to explain the location of a firm close to large input and output markets, in other words in an agglomerated area, because of positive transportation costs,

[6] Note that the Italian Census of Industry and Services is available only in 1991 and 1996, and so we do not have a yearly information concerning density and specialization indexes. Hence, since our estimation period is 1991-1998 we have used the 1991 census values for the sub-period 1991-1995 and the 1996 census values for the period 1996-1998.

[7] Note that the province size in expressed in square km.

[8] See for instance Jaffe et all (1993), Anselin et all (1997).

increasing returns and related externalities deriving from suppliers and consumers behaviours, in a some way general equilibrium approach, as for instance in the New economic geography models.

These two different explanations cannot be easily separated in an empirical work, since they are usually captured by the same variable, in this paper by the employment density.

As far as the specialization index is concerned, it stands as a proxy for the localization economies that firms enjoy when clustering with other firms of the other sector. This concept goes back directly to the marshallian industrial district concept and to the importance of the regional concentration at the industry level.

Besides, we want to investigate the relevance at the empirical level of the impact of potential market demand on the wage distribution. Actually, as shown in the theoretical model in several standard NEG models wages are positively related in equilibrium to the weighted purchasing power coming from surrounding locations. In order to proxy this variable we have used the household income at the province level, weighting the regions with the distance with respect to the considered location. In an analytical way we have defined the potential market demand for location j' in the following way, for a given year t:

$$\widetilde{W}_{j',t} = \ln \left[\sum_{j=1,\ j \neq j'}^{j} \left[1/d(j') \right] W_{j,t} \right],$$

where W_j is the household income at the provincial level and $d(j') \in (1, \infty)$ is the distance weight utilized that is greater than 1 if $j \neq j'$. Of course the farther the location the higher the distance. In this framework $1/d(j')$ represents a proper weight system because it assigns less importance to the potential demand coming from farther locations. We have excluded from the construction of this index the potential demand of the region to which the index is referred (j'), because it is correlated to the employment density already defined.

Furthermore, in order to take into account also the foreign potential demand we have added to the estimations a measure of trade openness for each province that also controls for its sectoral specialization. Hence, we have constructed the following specialization index. First of all we have computed a sectoral pro capite index of exportation given by $\widehat{c}_{s,t} = Exp_{s,t}/L_{s,t}$ where $Exp_{s,t}$ represents the national amount of export in sector s at time t, and $L_{s,t}$ is the national employment of sector s at time t. Using these shares, which are computed at national level, it is possible to derive predicted values of how much a single location should export if these sectoral shares were the same across locations. In other terms we have: $predicted\_Exp_{s,j,t} = L_{s,j,t} * \widehat{c}_{s,t}$. Finally, we can compute our provincial index of export specialization, for a given year t, given by:

$$Exp\_spec_{j,s,t} = \ln\left[\frac{effective\_Exp}{predicted\_Exp_{s,j,t}}\right] = \ln\left[\frac{L_{s,j,t} * c_{s,t}}{L_{s,j,t} * \widehat{c}_{s,t}}\right]$$

where $c_{s,t}$ are the 'effective' sectoral pro capite index of exportation computed at the provincial level. This index tell us how much each province is specialized in the exportation in a specific sector with respect to the national level.

It is worth noting that in this paper we direct our attention only on the two main important forces of agglomeration usually considered in the literature, the density as a proxy of the urbanization externality and the specialization as proxy of the specialization externality. We do not consider instead other forces, like the diversity externality and the competition externality, because their impact is usually less important and because of they might be correlated with other spatial variables. For instance, the diversity variable is usually strongly correlated with the density one.

Since for some of these additional variables we have information only from 1991 on we have decided to use as estimation period the spell 1991-1998. Moreover, in order to have an homogeneous sample we have not considered the atypical contracts (apprenticeship and part time), selecting only standard contracts for blue collar, white collar and managers. Moreover, since we are interested in carrying out panel estimates we have selected only workers that are in the INPS database at least for two years.

Besides, there are some workers who change job more than once in the same year, meaning that there could be more than an observation per worker per year. Indeed, in order to carry out a panel estimation one observation per year per worker is required. For this reason we have considered only the observation related to the job that lasted more for that worker in that year. Furthermore, we have drop out 1% of the workers in the tails of the wage distribution, in both level and yearly variations.

At the end we use an unbalanced database of 92.579 Italian workers for 560.040 observations.

4.1 Descriptive statistics

From the descriptive statistics some basic element concerning the goals of our paper come out quite clearly. First of all, spatial distribution of wage does not seem to be in a uniform way across Italian regions, meaning that location matters. This does not have to be taken for granted, since Italy is a country characterized by a very important centralize wage setting, where each sectoral contract has to respect important constraint, like a sectoral minimum wage. For this reason we were expecting a more uniform distribution of wages. However, it is worth noting that firms are allowed to integrate the national sectoral contract with a company specific contract, in which for example the minimum wage can be increased. Besides, since several standard economic theory suggest that fixing wages above the minimum wage level might represent an efficient solution for the firm (for instance the efficiency wage approach, the insider outsider and/or the wage setting in presence of unions etc.) it is not surprising that wage distribution is affected by economic location.

First of all, it is interesting to remark that the ratio between the province with highest average wages and the lowest one is 1.52 and this ratio is increasing overtime (1.46 in 1991 and 1.56 in 1998). This result still holds even if the different qualifications are taken into account. For instance, the same rate in 1998 was equal to 1.40 for blue collar workers, 1.53 for while collar and 2.82 for managers.

Even considering a less extreme indicator than the max/min as the P90/P10 ratio we still derive interesting results. In 1998 this ratio for all workers is 1.24, 1.22 for blue collar, 1.17 for white collar and 1.33 for managers, confirming that there is a substantial wage spatial distribution, similar to other European countries[9].

Moreover, also considering the relation among provincial average wages and provincial population (in 1998) it is possible to derive a clear positive correlation between these two variables (0.45) while when considering density instead of population (0.35) the correlation between is slightly lower. These correlation rates may probably be spuries, in the sense that they could be affected by sectoral and qualification composition of the economic structure. It is worth noting, for instance, that the two provinces with the highest average wages in 1998 are the ones also characterized by the highest population (Rome and Milan).

Another interesting figure concerns the relation between average provincial wages and po-

[9] See for instance Combes (2003).

tential market demand, proxied by the variable defined in the previous section. Also using a rough simple correlation (0.53) it is possible to claim that these variables are strongly correlated, suggesting that spatial potential market demand might be a driving force for wage distribution.

Using more appropriate econometric techniques we will show that these rough results derived from descriptive statistics are robust.

5 Econometric Specification

The main equation of our paper is a standard wage equation for panel data, *i.e.* regressing the logarithm of the yearly wage on the covariates[10]. It is worth noting that some of the covariates are in level, like age, gender, qualification dummies, while the ones describing spatial phenomena are reported in logarithms. We use the following notation:

$$w_{i,t} = I\_Char_{i,t} + L\_Ext_{m,j,t} + \widetilde{W}_{j',t} + Exp\_spec_{j,s,t} + \delta_t + u_i + \varepsilon_{i,t} \qquad (17)$$

where u_i is an individual fixed effect (that we will suppose to be correlated with regressors), δ_t is a time effect, the term $I\_Char_{i,t} = gender_i + Age_{i,t} + Age^2_{i,t} + \ln[Firm\_Size_{i,t}] + Blue\_collar_{i,t} + White\_collar_{i,t}$ is a battery of individual characteristics, $Loc\_Ext_{j,t} = Spec_{j,m,t} + Den_{j,t}$ controls (respectively) for marshallian and urbanization externalities in sector s and $\widetilde{W}_{j',t}$ ($Exp\_spec_{j,s,t}$) stands for potential market access (provincial export specialization). The spatial units of analysis (j) are the 95 Italian provinces.

The panel dimension in our analysis is essential to control for endogeneity problems, linked to the fact that the location choice for skilled workers cannot be considered as exogenous. In other terms, a skilled worker will be located, with a higher probability, in a agglomerated province because of a potential better exploitation of his/her higher capacity. In this framework, the higher wages in a agglomerated province might simply derive from a composition effect in the labour force, i.e. higher incidence of skilled labour workers. The panel dimension of our database allow us to compute an individual fixed effects, which is able to capture the unobservable skill differences among workers. In this way the spatial variable coefficients are no longer biased

[10] As wage variable we use the yearly wage paid by the firm to the employee, divided by the number of worked weeks, and then reporting the week wage at the monthly level. We did not use the information of the worked days because Ginzburg (1998) claimed that this variable in the south could be underestimated, leading to higher daily wages in this region, which is indeed supposed to be the poorest Italian area.

because the correlation among individual effects and agglomerated regions is explicitly taken into account.

5.1 Estimation results

In order to test our main hypotheses we carry out OLS, random effects and within estimates. From table 1 it is possible to note that all the controls variables behave in the expected way, men wages are higher than women ones, the age impact is positive and concave, and the firm size and the worker qualification affect positively the wage dynamics.

Moreover, the standard spatial variable coefficients confirm the theoretical predictions and previous empirical results for other countries. The higher are the specialization and urbanization of a location the higher the wage levels. It is important to note that the spatial variable coefficients in both the OLS and random effect estimates are higher than in the fixed effect (within). This clearly confirms our intuition that if the correlation between individual effects and spatial variable is not taken into account, as in the random effect, the spatial coefficients ends up to be upper biased. This difference is particularly important for the density variable, where the within coefficient is much lower (from 0.0034 to 0.023) than in the OLS, showing clearly that high skill workers tend to be located in agglomerated regions.

In table 1 we have also added the key variable of our paper, the potential market access. As predicted by theoretical models, the impact of the potential internal market access on wages is significant and positive. Also for this variable the within coefficient (0.027) is much smaller than both the OLS (0.09), confirming also for this variable a clear correlation with the individual effects. It is also worth noting that among the spatial variables the potential market access entail the stronger impact on wage dynamics.

In table 2 we point out that the spatial variable impact is stronger if we consider in the estimates only the male prime age workers, aged between 24 and 39. For instance, the potential market access coefficient is almost twice higher (0.047) than the one related to all workers (0.027), in the within estimate. The choice to consider only the prime age workers is standard in the wage equation setting. First of all because women wage dynamics is often affected by non-economic factors, meaning that standard economic and spatial covariates are less important in explaining women carrier. Second, workers aged 24-39 are the ones most interested in economic and spatial incentives, both for carrier and personal (familiar) reasons.

In table 3 we carry out the same estimates for the manifacturing sector. Results confirm previous analysis. Agglomeration and specification coefficients are positive and significant, as well as the potential market access potential coefficient. When considering prime age workers (Table 4) it is possible to derive results consistent to table 2.

6 Conclusions and Directions for Further Research

References

[1] L. Anselin, Spatial Econometrics: Methods and Models, Kluwer Academic Publishers, Boston 1988.

[2] L. Anselin, H. H. Kelejian, Testing for spatial error autocorrelation in the presence of endogenous regressors,.International Regional Science Review 20 (1997) 153-182.

[3] Anselin, L., A. Varga, and Z. Acs, (1993), "Local geographic spillovers between university research and high technology innovations", *Journal of Urban Economics*, 42, pp. 422-448.

[4] M. Arellano, S. R. Bond, Some tests of Specification for panel data: Monte Carlo evidence and an application to employment equations, Review of Economic Studies 58 (1991) 277-297.

[5] S. Bentolila, J. J. Dolado, Labour Flexibility and Wages: Lessons from Spain Economic Policy 18 (1994) 55-99.

[6] S. Bentolila, J. F. Jimeno, Regional Unemployment Persistence (Spain, 1976-1994), Labour Economics 5 (1998) 25-51.

[7] D. Black, V. Henderson, A Theory of Urban Growth, Journal of Political Economy 107 (1999) 252-284.

[8] A. Ciccone, R. Hall, Productivity and the Density of Economic Activity, American Economic Review 86 (1996) 54-70.

[9] P. P. Combes, M. Lafourcade, Transport Costs Decline and Regional Inequalities: Evidence from France, CEPR Discussion Papers 2894 (2001).

[10] D. R. Davis, D. E. Weinstein, Market access, economic geography and comparative advantage: an empirical test, Journal of International Economics 59 (2003) 1-23.

[11] A. C. Disdiez, K. Head, Exaggerated Reports on the Death of Distance: Lessons from a Meta-Analysis, Mimeo, University of Paris I.

[12] B. Eichengreen, Labor markets and European monetary unification, in P. R. Masson, M. Taylor (Eds.), Policy issues in the operation of currency unions, Cambridge University Press, Cambridge, 1993, pp. 130–162.

[13] R. C. Feenstra, New Product Varieties and the Measurement of International Prices, American Economic Review 84 (1994) 157-177.

[14] M. Fujita, P. Krugman, When in the Economy Monocentric? von Thunen and Chamberlin Unified, Regional Science and Urban Economics 25 (1995) 505-528.

[15] M. Fujita, J. F. Thisse, Economics of agglomeration, Journal of the japanese and international economies 10 (1996), 339-378.

[16] M. Fujita, P. Krugman, A. J. Venables, The spatial economy. Cities, regions and international trade, MIT Press, Cambridge 1999.

[17] M. Fujita, J. F. Thisse, Economics of Agglomeration, Cambridge University Press, Cambdridge 2002.

[18] A. Ginzburg, G. Solinas, M. Scaltriti, R. Zoboli, Un nuovo autunno caldo nel mezzogiorno? Note in margine al dibattito sui differenziali territoriali, Politica Economica, vol.14, no.3, 1998.

[19] G. H. Hanson, Market potential, increasing returns, and geographic concentration, NBER Working Paper 6429 (1998).

[20] G. H. Hanson, Scale Economies and the Geographic Concentration of Industry, Journal of Economic Geography 1 (2001) 255-276.

[21] C. D. Harris, The Market as a Factor in the Localization of Industry in the United States, Annals of the Association of American Geographers 44 (1954) 315-348.

[22] K. Head, J. Ries, Increasing Returns Versus National Product Differentiation as an Explanation for the Pattern of US-Canada Trade, American Economic Review.91 (2001) 858-876.

[23] K. Head, T. Mayer, Market Potential and the Location of Japanese Investment in the European Union, CEPR Discussion paper 3455 (2002).

[24] E. Helpman, The Size of Regions, in D. Pines, E. Sadka, I. Zilcha (Eds.), Topics in Public Economics, Cambridge University Press, Cambridge 1998, pp. 33-54.

[25] J. V. Henderson, Marshall's Scale Economies, Journal of Urban Economics 53 (2003) 1-28.

[26] J. V. Henderson, A. Kuncoro, M. Turner, Industrial Development and Cities, Journal of Political Economy 103 (1995) 1067-1081.

[27] A. B. Jaffe, M. Trajtenberg, and R. Henderson (1993), "Geographic localization of knowledge spillovers as evidenced by patent citations", Quarterly Journal of Economics, 108, pp. 577-598.

[28] P. Krugman, Increasing returns and economic geography, Journal of political economy 99 (1991) 483-499.

[29] P. Krugman, Development, geography, and economic theory, MIT Press, Cambridge 1995.

[30] P. Krugman, A. J. Venables, Globalization and the inequalities of nations, Quaterly Journal of Economics 110 (1995) 857-880.

[31] R. E. Lucas, On the mechanics of economic development, Journal of monetary economics 22 (1988) 3-22.

[32] G. I. P. Ottaviano, D. Puga, Agglomeration in the global economy: A survey of the New Economic Geography, The World Economy 21 (1998) 707-731.

[33] G. I. P. Ottaviano, T. Tabuchi, J. F. Thisse, Agglomeraion and Trade Revisited, International Economic Review 43 (2002) 409-436.

[34] H. G. Overman, D. Puga, Unemployment clusters across European regions and countries, Economic Policy 34 (2002) 115-147.

[35] J. Pinkse, M. E. Slade, C. Brett, Spatial Price Competition: A Semiparametric Approach, Econometrica 70 (2002) 1111-1155.

[36] D. Puga, The rise and fall of regional inequalities, European Economic Review 43 (1999) 303-334.

[37] J. Roback, Wages, Rents, and the Quality of Life, Journal of Political Economy 90 (1982) 1257-1278.

[38] S. Rosen, Wage-Based Indexes of the Urban Quality of Life, in P. Mieszkowski, M. Straszheim (Eds.), Current Issues in Urban Economics, Johns Hopkins University Press, Baltimore 1979, 74-104.

[39] S. Rosenthal, W. Strange, The Determinants of Agglomeration, Journal of Urban Economics 50 (2001) 191-229.

[40] S. Rosenthal, W. Strange, Geography, Industrial Organization, and Agglomeration, Review of Economics and Statistics 85 (2003) 377-393.

[41] S. Rosenthal, W. Strange, Evidence on the nature and sources of agglomeration economies, in V. Henderson, J. F. Thisse (Eds.), Handbook of Regional and Urban Economics, Vol. 4, North Holland, Amsterdam, 2004, forthcoming.

[42] A. C. F. Teixeira, Transport Policies in the Light of the New Economic Geography: The Portuguese Experience, Mimeo, CORE.

Table 1: Regression Results for all industries and all workers

Dependent Variables	(1) OLS	(2) GLS	(3) Within
$Gender(female)$	-0.2038*	-0.1894*	
	(0.0008)	(0.0018)	
Age	0.0305*	0.0293*	0.0194*
	(0.0002)	(0.0003)	(0.0030)
Age^2	-0.0003*	-0.0003*	-0.0003*
	(0.0000)	(0.0000)	(0.0000)
F_Size	0.0451*	0.0335*	0.0174*
	(0.0001)	(0.0002)	(0.0003)
Blue collar	-0.8350*	-0.3682*	-0.1752*
	(0.0031)	(0.0026)	(0.0029)
White collar	-0.5412*	-0.1595*	-0.1071*
	(0.0031)	(0.0023)	(0.0023)
$Spec$	0.0061*	0.0005	0.0025*
	(0.0004)	(0.0005)	(0.0006)
Den	0.0234*	0.0205*	0.0034*
	(0.0003)	(0.0005)	(0.0007)
$Migr$	.0004***	-0.0001	0.0023*
	(.0010)	(0.0008)	(0.0009)
$\widetilde{W}$	0.0900*	0.0729*	0.0273*
	(0.0013)	(0.0024)	(0.0042)
$Yearly : Dummies$	Yes	Yes	Yes
$Cov(u_i X\hat{\beta})$	-	-	-0.252
R^2	0.543	0.520	0.295
$Numb.Obs.$	560040	560040	560040
$Numb.Indiv.$	92579	92579	92579

Note: Standard errors in brackets. *, **, *** denote significance at respectively the 1%, 5% and 10% levels.

Table 2: Regression Results for male prime age workers (24-39) for all industries

Dependent Variables	(1) OLS	(2) GLS	(3) Within
Age	0.0481*	0.0485*	0.0466*
	(0.0015)	(0.0010)	(0.0080)
Age^2	-0.0005*	-0.0005*	-0.0005*
	(0.0000)	(0.0000)	(0.0000)
F_Size	0.0424*	0.0319*	0.0194*
	(0.0002)	(0.0004)	(0.0004)
Blue collar	-0.7617*	-0.3567*	-0.2148*
	(0.0049)	(0.0037)	(0.0041)
white collar	-0.4889*	-0.1771*	-0.1466*
	(0.0049)	(0.0031)	(0.0031)
Spec	0.0068*	-0.0005***	0.0002***
	(0.0006)	(0.0007)	(0.0008)
Den	0.0221*	0.0187*	0.0056*
	(0.0005)	(0.0008)	(0.0011)
Migr	0.0025**	0.0006***	0.0012***
	(0.0015)	(0.0012)	(0.0012)
$\widetilde{W}$	0.1203*	0.0911*	0.0474*
	(0.0021)	(0.0039)	(0.0059)
$Yearly : Dummies$	Yes	Yes	Yes
$Cov(u_i X\hat{\beta})$	-	-	0.29
R^2	0.54	0.51	0.41
Numb.Obs.	175700	175700	175700
Numb.Indiv.	24353	24353	24353

Note: Standard errors in brackets. *, **, *** denote significance at respectively the 1%, 5% and 10% levels.

Table 3: Regression Results only for Manufacturing

Dependent Variables	(1) OLS	(2) GLS	(3) Within
$Gender(female)$	-0.2166*	-0.2067*	-
	(0.0010)	(0.0022)	-
Age	0.0282*	0.0245*	0.0144*
	(0.0003)	(0.0004)	(0.0030)
Age^2	-0.0003*	-0.0002*	-0.0002*
	(0.0000)	(0.0000)	(0.0000)
F_Size	0.0398*	0.0335*	0.0215*
	(0.0002)	(0.0003)	(0.0004)
Blue collar	-0.8939*	-0.4257*	-0.1927*
	(0.0041)	(0.0037)	(0.0044)
white collar	-0.5896*	-0.1874*	-0.1191*
	(0.0041)	(0.0033)	(0.0033)
$Spec$	-0.0034*	0.0016**	0.0054*
	(0.0005)	(0.0007)	(0.0009)
Den	0.0259*	0.0235*	0.0026**
	(0.0004)	(0.0007)	(0.0011)
$Migr$	0.0047*	0.0045*	0.0072*
	(0.0014)	(0.0013)	(0.0014)
$\widetilde{W}$	0.1059*	0.0916*	0.0223*
	(0.0018)	(0.0033)	(0.0070)
$Yearly : Dummies$	Yes	Yes	Yes
$Cov(u_i X\hat{\beta})$	-	-	0.237
R^2	0.57	0.550	0.308
$Numb.Obs.$	279754	279754	279754
$Numb.Indiv.$	49323	49323	49323

Note: Standard errors in brackets. *, **, *** denote significance at respectively the 1%, 5% and 10% levels.

Table 4: Regression Results only for Manufacturing and for prime age workers

Dependent Variables	(1) OLS	(2) GLS	(3) Within
Age	0.0550*	0.0461*	0.0485*
	(0.0019)	(0.0013)	(0.0074)
Age^2	-0.0006*	-0.0005*	-0.0005*
	(0.0000)	(0.0000)	(0.0000)
F_Size	0.0378*	0.0312*	0.0220*
	(0.0004)	(0.0005)	(0.0007)
Blue collar	-0.8212*	-0.3854*	-0.2294*
	(0.0067)	(0.0052)	(0.0058)
white collar	-0.5261*	-0.1877*	-0.1525*
	(0.0068)	(0.0044)	(0.0044)
$Spec$	-0.0046*	0.0011***	0.0021***
	(0.0009)	(0.0012)	(0.0013)
Den	0.0229*	0.0206*	0.0058*
	(0.0007)	(0.0012)	(0.0016)
$Migr$	0.0031***	0.0035**	0.0050*
	(0.0021)	(0.0017)	(0.0018)
$\widetilde{W}$	0.1241*	0.0903*	0.0272*
	(0.0030)	(0.0057)	(0.0094)
$Yearly : Dummies$	Yes	Yes	Yes
$Cov(u_i X\hat{\beta})$	-	-	0.19
R^2	0.5304	0.5023	0.39
$Numb.Obs.$	89756	89756	89756
$Numb.Indiv.$	13519	13519	13519

Note: Standard errors in brackets. *, **, *** denote significance at respectively the 1%, 5% and 10% levels.