Spatial Externalities and Wage Distribution: the Role of Sorting*

Alessia Matano† and Paolo Naticchioni‡

July 2009

Abstract

Recent literature has shown that sorting plays an important role when the impact of spatial externalities on disparities in average wages between locations is investigated. The aim of the paper is to show that sorting matters also when addressing the relationship between spatial externalities (employment density and specialization) and wage distribution, i.e. across workers located at different percentiles of the wage distribution. Previous empirical papers did not control for observed and unobserved heterogeneity since they used aggregate data (Wheeler, 2004, 2007, Moller and Haas, 2003). We can control for observed individual and firm heterogeneity since we make use of the Italian employer-employee panel data. By means of standard quantile estimates, we find an increasing impact of spatial externalities on wage distribution. Furthermore, in order to control also for unobserved individual heterogeneity, we perform quantile fixed effect estimates (Koenker, 2004). Results show that the sorting matters since it dampens all the spatial externality effects along the wage distribution, effects that remain positive only in the upper tail of the wage distribution. Moreover, the impact of sorting is even stronger when using the sample of displaced workers, which represents a random sample of the workforce.

JEL Classification: J31, J61, R23, R30.
Keywords: Spatial Externalities, Sorting, Wage distribution, Quantile Fixed Effects.

---

* We thank the research partnership between ISFOL - Area Mercato del Lavoro (Rome) - and Dipartimento di Economia - University of Rome "La Sapienza" - for the access to the databases. We are very grateful to Roger Koenker for his comments and the availability of his procedure, to Christian Dahl for his precious advices and for the access to the routines, and to Giles Duranton for many useful suggestions. We also thank the participants of the AIEL conference (Brescia 2008), CIDE econometrics conference (Ancona, 2009) and GUNLAB II workshop (Rimini 2009) for helpful advices.
† Alessia Matano, University of Rome "La Sapienza", Dipartimento di Economia. Email: alessia.matano@uniroma1.it.
‡ Paolo Naticchioni, University of Rome "La Sapienza", Dipartimento di Economia and Université Catholique de Louvain. Email: paolo.naticchioni@uniroma1.it.

1. Introduction

The relationship between spatial externalities and differences in average wages between locations has been widely investigated in the literature, while the impact of spatial externalities on the wage distribution is still an open field of research. The theoretical models, which have widely analyzed the role of spatial externalities in fostering growth and productivity, have not yet deeply investigated the distributional effects of spatial externalities, as well as the related empirical evidence is lacking. A notably exception is Wheeler (2004, 2007) who tried to assess empirically how spatial externalities affect workers' wage distribution. Using aggregate data for the US, he shows that spatial externalities, i.e. employment density and industrial specialization, decrease wage inequality. Also Moller and Haas (2003) analyze the relationship between density and wage differentials at different percentiles of the wage distribution in Germany, estimating quantile regressions and using aggregated data derived from a set of observed individual characteristics. They find out that the impact of density increases with the deciles of wage distribution. However, since these empirical studies have been carried out using aggregate data, they cannot control for the relevance of the heterogeneity of workers and firms. Actually, agents' and firm heterogeneity have been proved to be relevant and to generally dampen the magnitude of spatial externalities impacts, i.e., it has been proved that much of the impact of spatial externalities on disparities in average wages between locations is actually due to the sorting of workers and firms (Combes et al., 2008, Mion and Naticchioni, 2009). To the best of our knowledge, there are no papers that investigate the impact of spatial externalities on the wage distribution controlling for the heterogeneity of workers and firms.

This paper aims at filling this gap in the empirical literature. Using individual panel data, we investigate the impact of spatial externalities, in terms of employment density and industrial specialization, along different percentiles of the wage distribution. By doing so, we can control for observed individual and firm heterogeneity. Further, we make use of quantile fixed effects estimates, which allow us to estimate the impact of spatial externalities controlling for unobserved individual heterogeneity, as in Combes et al. (2008) and Mion and Naticchioni (2009).

We focus on the Italian case using a matched employer-employee panel database provided by INPS (Italian Social Security Institute) and elaborated by ISFOL (Italian Institute for the Development of Vocational Training), merged with data on industrial (manufacturing and mining) and services employment provided by INPS, for the period 1991-2001. Using individual panel data we first run standard quantile estimates that allow us to get an estimation of the impact of spatial variables along the wage distribution of Italian workers, controlling also for observed individual and firm heterogeneity. Afterwards, we carry out quantile fixed effects estimates, proposed by Koenker (2004), in order to evaluate whether and how the impacts of spatial externalities change when the unobserved individual heterogeneity is taken into account. As in Mion and Naticchioni (2009) and Combes et al. (2008), our measure of workers' unobserved heterogeneity is related to time-invariant individual skills proxied by an individual fixed effect. This methodology allows us to assess whether the impact of sorting is uniform along the wage distribution. We perform separate estimations for the industry and the service sectors.

Standard quantile estimates computed on individual data display a positive impact of spatial externalities on wages, an impact that increases along the percentiles of the wage distribution. However, when taking into account the unobserved individual heterogeneity, i.e. using quantile fixed effects regressions, all spatial externality coefficients are reduced, and the strongest reductions concern the upper tail of the wage distribution. These findings suggest that sorting matters and that its impact is not uniformly distributed along the wage distribution. Nonetheless, after controlling for the effect of sorting, there is still some evidence of a positive, although smaller in magnitude, impact of spatial externalities on the upper tail of the wage distribution, suggesting that the skilled and high paid workers are those who benefit the most from spatial externalities. This effect might be due either to a higher ability of skilled workers to gain from face-to-face interactions (Glaeser and Marè, 2001), or to complementarity between skilled workers and technological and knowledge spillovers that arise in areas of dense economic activity or high level of specialization. As a robustness check, we carry out the quantile fixed effects estimations on the sample of the displaced workers, which can be considered a random sample of the workforce. Results are even more striking, since spatial externality coefficients are mainly not statistically significant.

The last part of the paper investigates the sorting features in a distributional and sectoral perspectives. First, we show that the high paid and skilled workers self-select themselves into dense and specialized provinces, confirming the sorting results derived in the regression analysis. Second, we analyze the sectoral breakdown of the sorting of workers. Our results show that the sorting of workers follows a not homogeneous pattern among sectors. More specifically, along the density dimension the sorting of workers takes place mainly in the skill-intensive sectors, while along the specialization dimension it is mainly concentrated in the unskill-intensive ones.

The structure of the paper is as follows. In Section 2 we review the theoretical as well as the empirical literature concerning the relationship between spatial externalities, productivity and wages. In Section 3 we describe the data and the indexes of spatial externalities we use throughout the empirical analyses. Section 4 introduces the quantiles estimations, both standard and fixed effects, discusses the empirical specification and presents the main results. Section 5 analyzes the features of the sorting of workers. Section 6 concludes.

2. Related Literature

The role of spatial externalities in fostering growth and local productivity has been a major concern for the theoretical, as well as the empirical, literature. Two of the most investigated spatial factors are the sectoral specialization and the employment density of a specific location.

As for specialization, Marshall (1890) was the first in the literature underlining the productivity gains due to the concentration of a specific industry in a given location. He identified three channels through which these productivity gains may arise: labour market pooling, which allows a more efficient process of matching between workers and firms; input sharing, which allows producers to advantage from a higher level of division of labour; technological and knowledge spillovers. This theoretical approach has been widely investigated and modeled. Among others, Duranton and Puga (2004) and Henderson (1974) have provided the microfoundations of such types of externalities.

As for urbanization economies, the idea that the size of the market, proxied by the employment density, can generate productivity gains goes also back to Marshall (1890) and has been formalized by Abdel Rahman and Fuija (1990) among others. Moreover, Duranton and Puga (2004) highlight the micro foundations of urban agglomeration economies, discussing three channels through which urbanization externalities may arise: matching, learning and sharing. In fact, urbanization increases the probability and the expected quality of matches between workers and firms, which gives raise to productivity gains. Moreover, living in a dense area improves the diffusion and accumulation of knowledge, and hence productivity. Finally, living in an area of dense economic activity brings gains due to the sharing of indivisible facilities by people and to the sharing of a wider variety of input suppliers that can be sustained by a larger final good industry.

Empirically, several works have analyzed the role of spatial externalities in boosting labour, productivity and wages (see among others Combes, 2000, Ciccone and Hall, 1996, Glaeser et al., 1992, Ciccone, 2002).[1] However, these works have mainly used aggregate data to study the relationship between wages and spatial externalities. Therefore, they could not take into account the spatial sorting of workers and firms. Actually, skilled workers concentrate in cities for different reasons. First, cities provide valuable

[1] For a complete empirical review of studies on spatial externalities see Rosenthal and Strange, 2004.

consumption amenities such as cultural activities, museums, theatres etc., which attract skilled workers. Second, as suggested in Moretti (2004), bigger cities offer higher returns to education (private plus social), thus fostering investment in human capital. Third, they provide a better environment to accumulate human capital, thanks to face-to-face interactions (Glaeser and Marè, 2001). As for the spatial sorting of firms, the idea is that bigger and more productive firms locate in large cities. In fact, when the market size expands, labour market competition becomes fiercer and therefore only the most productive firms survive. Hence, they can employ more workers, and thus grow larger (Kim, 1989, Helsley and Strange, 1990, Melitz, 2003).

All this literature has focused on the relationship between spatial externalities and the disparities in average wages between locations. As for the relationship between spatial externalities and wage distribution, there are not yet theoretical frameworks that analytically deal with this issue. However, some authors fostered the idea that spatial externalities could entail a not uniform impact along the wage distribution. For instance, it has been argued that more educated workers could benefit more from spatial externalities since they are better able to learn from face-to-face interactions (Glaeser and Marè, 2001). An opposite explanation might suggest that less educated workers could gain more since they have a lower stock of human capital and therefore enjoy increasing returns from the face-to-face interactions with skilled workers (Glaeser and Marè, 2001, Wheeler, 2007).

As for the empirical evidence, there is also a lack of studies that investigate the distributional effects of spatial externalities. A notably exception is Wheeler (2004, 2007) who has empirically investigated the impact of both industrial specialization and density on wage inequality within locations in the US at an aggregate level (metropolitan and state level), using different measures of wage inequality (the $90^{th}/10^{th}$ wage percentile ratio, the residual $90^{th}/10^{th}$ index and the wage difference by educational groups). His findings show that the impact of spatial externalities is not uniformly distributed through different categories of workers. In particular, both density and local industrial specialization reduce wage inequality. Another related work is Moller and Haas (2003) that performs a quasi quantile regression approach (Chamberlain, 1994) to analyze the relationship between density and wage differential at different percentiles of the wage distribution. Actually their focus is on the spatial wage differentials between locations and the estimation is carried out aggregating individuals in cells according to observable characteristics. Their findings show that the spatial wage differential depends on the skill level, is higher in the manufacturing sector, and increases with the deciles of wage distribution. Hence, agglomeration increases wage inequality, since the related wage premium raises along the wage distribution.[2]

However, using aggregate data the detected relationship between spatial externalities and wage inequality is likely to suffer from an omitted variable bias since it does not control for workers and firm heterogeneity. Actually, it has been proved that the sorting of workers and firms is able to explain most of the supposed impact of spatial externalities on the disparities in average wages between locations. For instance, Combes et al. (2008) show that not taking into account the role of the sorting of workers brings to an overestimation of the spatial externalities coefficients from the 70% to 200%. Mion and Naticchioni (2009) show that roughly 75% of the differences in wages between high density and low density provinces is explained by unobserved skills, i.e. differences in individual fixed effects, while the share explained by the sorting of firms is only the 5.6%. Therefore, not taking into account the workers' (observed and unobserved) and firm heterogeneity can cause severe biased estimates of the spatial externalities impact on wages.

In this paper we use individual panel data and quantile fixed effects regression techniques in order to investigate the impact of spatial externalities on wage distribution controlling for workers' and firms heterogeneity. To the best of our knowledge in the literature there is no paper that addresses the role played by sorting in explaining the relationship between spatial externalities and wage distribution. We fill this gap in this paper focusing on the Italian case.

Previous empirical studies concerning the Italian case have investigated the impact of spatial externalities on the average wage levels of Italian workers (see Signorini 1994, 2000, Tattara, 2001, for industry concentration and Di Addario and Patacchini, 2007, for urban density). On the whole the results of these papers have shown a positive and significant impact of spatial externalities on both productivity and wages. However, they did not take into account the spatial sorting of workers and firms. An exception is Mion and Naticchioni (2009) who investigate the impact of spatial externalities on the wages of Italian workers taking into account the relevance of sorting that dampens the magnitude of the spatial variables effects.[3]

[2] Another related paper that focuses on the causes of urban inequality, rather than on the impact of the agglomeration externalities on inequality, is Glaeser et al. (2008). Analysing the US case and using aggregate measures of inequality, they underline that individual skills (in terms of education) explains one third of the variation in income inequality, in line with our findings on the relevance of sorting.

[3] Actually, Di Addario and Patacchini (2007) take into account the issue of the sorting of workers. However, according to their findings, the endogenous sorting of workers into cities turns out to be not a major concern in their analysis.

3. Description of the Data and Definition of Spatial Variables

We use a panel version of the administrative database provided by INPS (Italian Social Security Institute) and elaborated by ISFOL.[4] It is a matched employee-employer dataset, constructed merging the INPS employee information for the period 1985-2002 with both the INPS employer information database from 1985 to 1998 and the ASIA database from 1999 to 2002.[5]

The sample units are industrial (manufacturing and mining) and service dependent workers, both part-time (converted in full-time equivalent) and full-time. We exclude workers in apprenticeship status in order to concentrate the analysis on standard labour market contracts: blue collar and white collar workers. Moreover, we focus on prime age male workers, male workers aged between 25 and 49 (when they first enter in the database), as standard in this literature (see for instance Topel, 1991, Mion and Naticchioni, 2009). Further, we consider only workers with at least three observations in the period in order to be able to get reliable within estimations in our analysis. In this way we end up with an unbalanced panel of 36,121 workers for 283,760 observations for the industry and with an unbalanced panel of 20,902 individuals for 140,428 observations for the service sector.[6] The dependent variable in our regressions is the (log) real gross weekly wage in euro.[7] The base year is 2001. As far as workers' characteristics are concerned, the database contains individual information such as age, gender, occupation, workplace, date of beginning and end of the current contract (if any), the social security contributions, the worker status (part-time or full-time), the real gross yearly wage and the number of worked weeks and days. As for firms, we have the plant location (province), the number of employees and the sector (Ateco81 and Ateco91).

We merge the INPS dataset with data on industrial and service employment provided by INPS for the period 1991-2001. Hence, the focus of our analysis will be on period 1991-2001 for which all individuals and spatial variables are available. Using this database we can define the spatial variables used in the empirical analysis, where the territorial

[4] ISFOL stands for "Institute for the Development of Vocational Training". In particular, the panel version has been constructed considering only one observation per year for each worker. For those workers who display more than one observation per year we selected the longest available contract in terms of weeks worked. We further eliminated those extreme observations below (above) the 0.5th (99.5th) percentile of the wage distribution. The sample scheme has been set up to follow individuals born on the 10th of March, June, September and December and therefore the proportion of this sample on the Italian employees' population is approximately of 1/90.

[5] ASIA stands for "Italian Statistical Archive of Operating Firms". It is provided by ISTAT. This database has been used since 1999, because the INPS employer database was not available after 1998. However, the two databases provide the same set of information that we use in our analysis.

[6] We carry out separate estimations for the industry and the service sector in order to look at possibly different outcomes of the impact of spatial externalities on wage distribution.

[7] Wages have been deflated using as deflator the Consumer Price Index specific for blue collars and white collars (FOI index, *Indice dei Prezzi al Consumo per le Famiglie di Operai e Impiegati*, ISTAT).

breakdown is the province, classified in 95 territorial units. The index of local-sectoral specialization has been computed from the INPS employment data and it is defined as:[8]

$$Spec_{p,s,t} = \ln\left[\frac{empl_{p,s,t} / empl_{p,t}}{empl_{s,t} / empl_{t}}\right]$$

where subscript p refers to province, s to sector and t to time.[9] It is the ratio between the share of sectoral employment on total industrial (services) employment in any province p and the corresponding share at national level. As for urbanization externalities, we define the density as:[10]

$$Dens_{p,t} = ln\left[\frac{empl_{p,t}}{area_{p}}\right]$$

where subscript p refers to province and t to time (province area is measured in square km)[11] Since all the spatial variables are defined in logarithm, we estimate elasticities. Table 1 shows the descriptive statistics of the variables of the analysis.

[8] As in Combes (2000), Mion and Naticchioni (2008) and De Blasio and Di Addario (2002).

[9] We define the index of specialization by sectors following the Ateco81 classification, two digits level, since it provides higher data accuracy.

[10] As in Combes (2000), Mion and Naticchioni (2008) and Ciccone and Hall (1996).

[11] Both the indexes have been computed separately for the industry and the service sectors. However, we also carry out the same estimations using the indexes defined over all workers, finding out similar outcomes. Moreover, the correlation between the two indexes computed separately for the two sectors and for all the economy is 0.97 for specialization and 0.98 for density.

Table 1: Descriptive Statistics of the Variables of the Analysis

Variable	Observations	Mean	Std. Dev.	Min	Max
Industry					
Log Real Weekly Wage	283,760	6.00	0.39	3.78	7.92
Age	283,760	39.73	7.65	25	59
Age^2	283,760	1,636.83	616.52	625	3,481
Bc	283,760	0.74	0.44	0	1
Wc	283,760	0.26	0.43	0	1
Log Firm Size	283,760	4.56	2.60	0	11.63
Log Specialization Index	283,760	0.15	0.92	-6.94	5.68
Log Density	283,760	3.34	1.19	0.08	5.70
dNorth East	283,760	0.24	0.43	0.00	1.00
dNorth West	283,760	0.37	0.48	0	1
dCentre	283,760	0.17	0.38	0	1
dSouth	283,760	0.16	0.37	0	1
dIsland	283,760	0.06	0.24	0	1
Sectors	283,760	37.35	10.07	11	50
Services					
Log Real Weekly Wage	140,428	6.08	0.48	3.79	7.92
Age	140,428	39.41	7.66	25	59
Age^2	140,428	1,611.75	617.81	625	3,481
Bc	140,428	0.50	0.50	0	1
Wc	140,428	0.50	0.50	0	1
Log Firm Size	140,428	4.99	2.94	0	12.11
Log Specialization Index	140,428	-0.09	0.94	-7.81	3.90
Log Density	140,428	3.32	1.35	-0.24	5.80
dNorth East	140,428	0.20	0.40	0.00	1.00
dNorth West	140,428	0.32	0.47	0	1
dCentre	140,428	0.23	0.42	0	1
dSouth	140,428	0.16	0.37	0	1
dIsland	140,428	0.09	0.28	0	1
Sectors	140,428	73.92	10.42	61	98

Source: Panel ISFOL on INPS data and INPS data.

4. Empirical Analysis: the Impact of Spatial Externalities on Wage Inequality

In this section we estimate the impact of spatial externalities along the wage distribution of Italian workers. In order to accomplish this task, we make use of different econometrics techniques. First, we run standard quantile estimations to see how the impact of spatial externalities varies with the different percentiles of the wage distribution of Italian workers. Second, we estimate quantile fixed effects regressions (Koenker, 2004) in order to estimate the impact of spatial variables on the wage distribution taking into account the unobserved individual heterogeneity. More specifically, we proxy the unobserved individual heterogeneity introducing individual fixed effects in the quantile regressions that capture time invariant worker characteristics such as ability and education (as in Mion and Naticchioni, 2009 and Combes et al., 2008).[12] The aim of the quantile fixed effects estimations, whose technique are explained in detail below, is to understand whether the impact of spatial externalities detected in standard quantile regressions can at least partially be related to the effect of sorting. As in the case of disparities in average wages between locations (Mion and Naticchioni, 2009 and Combes et al., 2008), if spatial externalities coefficients resulted to be reduced in quantile fixed effects estimations, it would mean that sorting is relevant also for the analysis of the relationship between spatial externalities and the wage distribution. Furthermore, if sorting were the main source of the spatial externality impacts, our analysis would suggest that it is crucial to use individual longitudinal data in order to separate the relevance of sorting from that of spatial externalities. Using aggregate analysis would produce biased estimates.

4.1. Quantile regressions and the impact of spatial externalities on the wage distribution

We begin our analysis using standard quantile regression as follows:

$$Q_{\theta,i,t} = \alpha + B'_1 * I\_Char_{i,t} + \beta_1 * Firmsize_{i,t} + \gamma_1 * Spec_{p,s,t} + \gamma_2 * Dens_{p,t} + \varphi_s + \lambda_a + \delta_t + \varepsilon_{it}$$

where subscript θ refers to the percentile, i to individual, s to sector, p to province, a to area and t to time. The percentiles θ estimated are the 10th, 25th, 50th, 75th and 90th percentiles. In this first part of the analysis, we do not take into account the longitudinal dimension of our data since we estimate the cross sectional impact of spatial variables

[12] Our main focus here is on unobserved individual heterogeneity. However, we want to point out that both in standard quantile regressions and in quantile fixed effects regressions we take into account also firm heterogeneity, which we proxy using the firm size, since firm productivity is usually positively related with firm size (Postel-Vinay and Robin, 2006).

along the wage distribution of Italian workers. We carry out separate estimations for the industry and service sectors.

The dependent variable is the logarithm of the real gross weekly wage. The term $I\_Char_{i,t}$ is a set of individual characteristics (age, age squared, blue collar dummy). $Spec_{p,s,t}$ is the index of specialization and $Dens_{p,t}$ is the density of province p defined as in section 3. Moreover, $Firmsize_{i,t}$ is used to control for firm heterogeneity.[13] Finally φ_s, λ_a, δ_t are sectoral, area and time dummies respectively. Since all the variables of interest are in logarithms, we estimate elasticities. Table 2 and 3 show the quantile estimations for the industry and service sectors respectively.

Table 2: Quantile Regressions of Wages on the Spatial Variables. Industry

	q10	q25	q50	q75	q90
Specialization	0.0011	0.0024	0.0054	0.0088	0.0133
	[0.0009]	[0.0007]***	[0.0006]***	[0.0007]***	[0.0009]***
Density	0.0128	0.0128	0.0143	0.0181	0.0208
	[0.0007]***	[0.0007]***	[0.0004]***	[0.0005]***	[0.0007]***
Age	0.0315	0.0292	0.0294	0.0287	0.0227
	[0.0008]***	[0.0006]***	[0.0006]***	[0.0009]***	[0.0013]***
Age Squared	-0.0003	-0.0003	-0.0003	-0.0003	-0.0002
	[0.0000]***	[0.0000]***	[0.0000]***	[0.0000]***	[0.0000]***
Blue Collar Dummy	-0.2397	-0.2761	-0.3555	-0.4829	-0.6428
	[0.0020]***	[0.0017]***	[0.0019]***	[0.0030]***	[0.0039]***
Firm Size	0.0426	0.0402	0.0384	0.0365	0.0341
	[0.0003]***	[0.0002]***	[0.0002]***	[0.0003]***	[0.0005]***
Constant	4.9728	5.1844	5.3799	5.6322	6.0096
	[0.0149]***	[0.0119]***	[0.0115]***	[0.0200]***	[0.0276]***
Area, Time and Sector dummies	yes	yes	yes	yes	yes
N. Observations	283,760	283,760	283,760	283,760	283,760
N. Individuals	36,121	36,121	36,121	36,121	36,121
R squared	0.26	0.29	0.33	0.36	0.40

Notes: Standard Errors in Parenthesis with ***,** and * denoting significance at 1%, 5% and 10% respectevely. Standard errors are given in parenthesis.

Results point out that the impact of both density and specialization increases along the wage distribution of Italian workers. Moreover, these impacts are higher for the service sector. In particular the coefficients of the local sectoral specialization passes from an

[13] Widely has been the literature concerning the positive relationship between firm size and wages. See for instance Krueger and Summers (1988), Brown and Medoff (1989) and Abowd, Kramarz and Margolis (1999).

elasticity of 0.1% at the 10th percentile to 1.3% at the 90th percentile for the industry, and from a negligible elasticity at the 10th percentile to 4.4% at the 90th percentile for the service sector, with a statistically significant the difference between the coefficients in these two deciles in both estimations.

Table 3: Quantile Regressions of Wages on the Spatial Variables. Services

	q10	q25	q50	q75	q90
Specialization	-0.0081	0.0155	0.0281	0.0369	0.0444
	[0.0017]***	[0.0007]***	[0.0008]***	[0.0010]***	[0.0017]***
Density	0.0089	0.0046	0.0134	0.0229	0.0241
	[0.0013]***	[0.0007]***	[0.0008]***	[0.0010]***	[0.0014]***
Age	0.0358	0.0346	0.0339	0.0352	0.0342
	[0.0014]***	[0.0011]***	[0.0012]***	[0.0015]***	[0.0014]***
Age Squared	-0.0003	-0.0003	-0.0003	-0.0003	-0.0002
	[0.0000]***	[0.0000]***	[0.0000]***	[0.0000]***	[0.0000]***
Blue Collar Dummy	-0.2758	-0.2241	-0.2390	-0.3399	-0.5107
	[0.0043]***	[0.0031]***	[0.0019]***	[0.0032]***	[0.0065]***
Firm Size	0.0447	0.0441	0.0408	0.0389	0.0376
	[0.0006]***	[0.0005]***	[0.0004]***	[0.0005]***	[0.0007]***
Constant	4.8170	4.9570	5.0974	5.2627	5.5598
	[0.0289]***	[0.0206]***	[0.0224]***	[0.0298]***	[0.0296]***
Area, Time and Sector dummies	yes	yes	yes	yes	yes
N. Observations	140,428	140,428	140,428	140,428	140,428
N. Individuals	20,902	20,902	20,902	20,902	20,902
R squared	0.24	0.27	0.33	0.34	0.35

Notes: Standard Errors in Parenthesis with ***,** and * denoting significance at 1%, 5% and 10% respectevely. Standard errors are given in parenthesis.

As for density, the elasticity estimates go from 1.3% at the 10th percentile to 2.1% at the 90th in the industry and from 0.9% at the 10th percentile to 2.4% at the 90th in the service sector, with again statistically significant differences. These findings suggest that, as also pointed out by other authors (Wheeler, 2007), the impact of spatial externalities is not uniform along the wage distribution. In particular, in the case of Italy, specialization and density seem to increase wage inequality. This finding is at odd with Wheeler (2007) that finds a reduction of wage inequality related to spatial externalities, while it is in line with Moller and Hass (2003) that finds out an increasing impact of spatial externalities along the wage distribution.[14]

[14] As for the control variables in the regressions, they have the expected signs: wages shows a concave shape in age, that is generally higher at the lowest percentiles. The blue collar dummies negatively

To sum up, cross sectional quantile estimations point out that the impact of industrial localization and employment density is significant and increases along the wage distribution of Italian workers.

4.2. The role of sorting using quantile fixed effects estimates

The estimations computed in the previous section could be biased since they do not take into account the unobserved individual heterogeneity. In fact, as shown by the abovementioned works (Mion and Naticchioni, 2009 and Combes et al., 2008), it might be argued that skilled individuals are likely to sort in those provinces that provide consumption amenities, have higher returns to skills and speed up face-to-face interactions. As a consequence, it is likely that the estimated coefficients for the spatial variables are biased, since they can actually incorporate the effect due to the sorting of workers. Therefore, we perform quantile fixed effects estimations that are able to control for such an element.

We apply the technique elaborated by Koenker (2004) and implemented by Bache et al. (2008) and Bargain and Melly (2008) among others. Koenker estimates quantile regressions adding individuals' dummies in the estimations. By this means he is able to estimate the impact of an explanatory variable on the dependent variable taking into account the unobserved individual heterogeneity. However, since by using this technique the number of estimating parameters significantly increases, Koenker adds to the minimization algorithm a penalty term that takes into account the variability problem that arises estimating a so large number of parameters.[15] Hence, his technique minimizes the following expression:

$$\min_{\alpha,\beta} \sum_{k=1}^{q} \sum_{i=1}^{n} \sum_{j=1}^{t_i} w_k \rho_{\theta_k}(y_{ij} - \alpha_i - x_{ij}^{'} \beta(\theta_k)) + \lambda \sum_{i=1}^{n} \left| \alpha_i \right|$$

where, $\rho_{\theta k}(u)=u(\theta-I(u<0))$ is the piecewise linear quantile loss function of Koenker and Bassett (1978). The weights w_k control for the relative influence of the q quantiles on the estimation of the a_i parameters, which we set equal in each quantile since we are mainly interested in controlling for the fixed effects (as in Bache et al., 2008). The last term in the

impact wages, while the coefficients for firm size are positive and significant. Further, we also perform the same quantile estimations using lagged values of the spatial variables for two main reasons. First, to capture an eventual time lag in the adjustment of the wages with respect to spatial variables. Second, to control for possible endogeneity problems arising from simultaneous determination of spatial variables and wages. Results are largely the same. These estimations are available upon request.

[15] Indeed, the introduction of a so large number of individual fixed effects can significantly inflate the variability of estimates of other covariates effects (Koenker, 2004).

above expression represents the penalty term and λ describes the importance of the penalty term in the minimization formula. We set it equal to 1, as in Koenker (2004) and Bache et al. (2008).[16] Table 4 and 5 show the coefficients computed with the Koenker procedure. Due to computational problems we had to extract random samples of our data in order to run these estimations. Hence, the samples are constituted by 69,355 observations for 8,961 individuals for the industry and by 69,308 observations for 10,236 individuals for the service sector.[17]

Table 4: Quantile Regressions of Wages on the Spatial Variables using the Koenker Procedure. Industry

	q10	q25	q50	q75	q90
Specialization	0.0005	0.0021	0.0031	0.0046	0.0061
	[0.0023]	[0.0020]	[0.0019]	[0.0019]**	[0.0022]***
Density	0.0117	0.0109	0.0109	0.0113	0.0126
	[0.0039]***	[0.0038]***	[0.0038]**	[0.0038]***	[0.0039]***
Age	0.0381	0.0318	0.0273	0.0265	0.0256
	[0.0034]***	[0.0167]*	[0.0031]***	[0.0048]***	[0.0037]***
Age Squared	-0.0003	-0.0003	-0.0002	-0.0002	-0.0002
	[0.0000]***	[0.0002]	[0.0000]***	[0.0001]***	[0.0000]***
Blue Collar Dummy	-0.0937	-0.0972	-0.1002	-0.1104	-0.1326
	[0.0417]**	[0.0414]**	[0.0410]**	[0.0399]***	[0.0371]***
Firm Size	0.0249	0.0232	0.0229	0.0224	0.0210
	[0.0022]***	[0.0023]***	[0.0023]***	[0.0024]***	[0.0026]***
Constant	8.2506	8.4576	8.6083	8.6919	8.7987
	[4.8376]*	4.8486]*	[4.8369]*	[4.8361]*	[4.8383]*
Area, Time and Sector dummies	yes	yes	yes	yes	yes
N. Observations	69,355	69,355	69,355	69,355	69,355
N. Individuals	8,961	8,961	8,961	8,961	8,961

Notes: Standard Errors in Parenthesis with ***,** and * denoting significance at 1%, 5% and 10% respectevely. Bootstrapped standard errors are given in parenthesis. The bootstrapping was done using samples of 10,000 observations and 2,000 iterations.

[16] It is worth noting that in this model if lambda equal to zero we obtain the generic fixed effects estimator (the penalty term disappears), while if lambda tends to infinity we obtain an estimate of the model purged of the fixed effects. Moreover, as it can be seen from the formula, this model requires the simultaneous estimation of all quantiles, since individuals fixed effects are supposed to be constant among different quantiles in such a way to reduce the number of parameters estimates. Koenker (2004) shows the consistency of this estimation technique, while for the standard errors it requires bootstrap estimations (see Koenker, 2004, for further details on this technique). However, since we are dealing with longitudinal data the standard bootstrap estimation cannot be applied here. Instead a subsampling bootstrap approach is used where random samples of individuals are drawn repeatedly with replacements (Abrevaya and Dahl, 2006, Bache et al., 2008).

[17] The samples are basically the same in terms of the observable characteristics with respect to the original dataset. We also use other random samples from our data in order to check the robustness of our results, which are largely the same.

Table 5: Quantile Regressions of Wages on the Spatial Variables using the Koenker Procedure. Services

	q10	q25	q50	q75	q90
Specialization	0.0006	0.0025	0.0043	0.0062	0.0078
	[0.0049]	[0.0045]	[0.0041]	[0.0037]*	[0.0034]**
Density	0.0067	0.0069	0.0070	0.0080	0.0097
	[0.0045]	[0.0044]	[0.0043]	[0.0042]*	[0.0041]**
Age	0.0457	0.0397	0.0356	0.0335	0.0305
	[0.0061]***	[0.0103]***	[0.0042]***	[0.0061]***	[0.0064]***
Age Squared	-0.0004	-0.0003	-0.0003	-0.0003	-0.0002
	[0.0001]***	[0.0001]***	[0.0001]***	[0.0001]***	[0.0001]***
Blue Collar Dummy	-0.1250	-0.1225	-0.1227	-0.1269	-0.1405
	[0.0555]***	[0.0561]***	[0.0560]***	[0.0554]***	[0.0527]***
Firm Size	0.0179	0.0163	0.0166	0.0167	0.0145
	[0.0048]***	[0.0051]***	[0.0050]***	[0.0050]***	[0.0056]***
Constant	7.7204	7.9180	8.0536	8.1464	8.2857
	[1.1935]***	[1.2324]***	[1.2326]***	[1.2413]***	[1.2572]***
Area, Time and Sector dummies	yes	yes	yes	yes	yes
N. Observations	69,308	69,308	69,308	69,308	69,308
N. Individuals	10,236	10,236	10,236	10,236	10,236

Notes: Standard Errors in Parenthesis with ***,** and * denoting significance at 1%, 5% and 10% respectevely. Bootstrapped standard errors are given in parenthesis. The bootstrapping was done using samples of 10,000 observations and 2,000 iterations.

From Table 4 and 5 it comes out that previous results change when taking into account the relevance of the sorting of workers, proxied by the individual fixed effects. The coefficients of the spatial variables are considerably reduced compared to previous quantile estimations and, in some cases, they are even no longer statistically significant. In particular, specialization still has an increasing impact (even if strongly reduced) along the wage distribution of Italian workers. Moreover, this impact results to be significant only at the right-hand tail of the wage distribution. As for the service sector, the impact of sorting is even stronger, entailing a greater estimates reduction. Therefore, sorting captures most of the impact of specialization on wage inequality and its impact increases along the wage distribution indicating that skilled workers are sorted in high specialized provinces. Actually, not taking into account the sorting of workers generates an overestimation of the specialization coefficient up to 100% at the 90th percentile in the industry sector and up to 500% in the service sector. Nonetheless, there is still evidence of an increasing impact of sectoral specialization along the wage distribution that suggests that industrial specialization contributes to increase wage inequality both in the industry and the service sectors.

As for the impact of density in the industry sector, coefficients computed using quantile fixed effect estimates are basically stable along the wage distribution, suggesting that the impact of sorting is strongest at the highest quantiles since in cross sectional estimations the impact was increasing along the wage distribution.[18] As far as the service sector is concerned, coefficients estimates remain statistically different from zero only at the right hand side of the wage distribution, entailing a positive, even if small, impact on wages.[19] Also in this case sorting matters, and again it mostly affects the highest percentiles of the wage distribution. Moreover, again, the coefficients reduction is striking: the elasticity for the employment density is reduced by almost 100% in the industry sector at the 90th percentile and by more than 100% in the service sector. In this framework, density entails only a slight positive impact on wage inequality in the service sector, while it does no longer affect wage inequality in the industry sector. [20]

Results of quantile fixed effects estimates point out that sorting matters and explains most of the impact of spatial externalities. However, after controlling for sorting there is some evidence of a slight positive impact of spatial externalities on the upper tail of wage distribution. These latter findings are consistent with the idea in Glaeser and Marè (2001)

---

[18] The differences among coefficients at different percentiles are not statistically different from zero while they were significantly different from zero using cross sectional quantile regressions.

[19] We want also to point out the impact of the coefficients in terms of standardized elasticities. Such standardized (or beta) coefficients are obtained by multiplying the coefficients estimates by the standard deviation of the explanatory variable and dividing them by the standard deviation of the dependent variable. In this way the regression coefficients are converted into units of sample standard deviation and give us a measure of how much variability can be explained by the explanatory variable (see Wooldridge, 2003, section 6.1). For instance, considering the coefficient of specialization in column 5 (90th wage percentile) of table 4, we get a standardized elasticity of ((0.0061*0.918)/0.393)= 0.014. This means that a one standard deviation increase in the logarithm of the specialization index implies an increase of 0.014 standard deviation in the logarithm of the 90th wage percentile of Italian workers. Likewise, we get a standardized elasticity of ((0.0126*1.186)/0.393)=0.038 for density.

[20] As a robustness check we also carry out the quantile fixed effects estimates using another technique proposed by Arulampalam et al. (2007) and also implemented by Bache et al. (2008). It is a two stage regression where in the first stage a standard within panel regression is performed to produce an estimation of the fixed effects. In the second stage, a simultaneous quantile estimation is carried out, adding as explanatory variables the fixed effects estimated in the first stage. Though the asymptotic properties of this estimator are still unknown, it seems to perform well and it is quite simple to implement (Bache et al., 2008). We rely on bootstrap for the coefficients and standard errors estimates. Results of this two stage technique widely confirm the outcomes of the Koenker's one. We do not show these estimates for sake of synthesis. They are available upon request. Furthermore, we carry out other two estimations as robustness checks for our analysis focusing on the firm size. First of all, we perform the fixed effects quantile estimations (Koenker, 2004) adopting a finer specification for the firm size. In particular, we run the estimations with an interaction effect between regional dummies and the firm size in order to better capture the heterogeneity in the firm size effect, heterogeneity that depends on regional differences (since Italy is characterized by strong regional differences). Second, we perform the quantile fixed effects estimations neglecting the influence of the firm heterogeneity (we drop the firm size) in order to check whether the possible collinearity between the firm size and the density dampens the impact of density on wages. Results of both these robustness checks are largely the same, thus confirming previous outcomes (moreover the correlation between the firm size and the density in Italy is only 0.10). All these estimates are available upon request.

that skilled workers are attracted by cities and cities make skilled workers more productive since they are better able to gain from face-to-face interactions or from the technological and knowledge spillovers that arise in areas of dense economic activity or high level of specialization.[21]

Finally, it could be argued that also the quantile fixed effects estimates might be biased. Actually, fixed effects estimates are mainly identified by movers, i.e. those workers who change location and/or industry. However, movers are likely not to be a random sample of the workforce, since their mobility choices can be due to different reasons, such as improving their jobs, changing jobs because they have been fired by an on-going firms or because their firms closed down. These patterns of movers might generate heterogeneity when investigate mobility issues and, hence, fixed effects estimates could be biased. Therefore, we carry out the same estimates on the sample of the displaced workers (as in Dustmann and Meghir, 2005, and Mion and Naticchioni, 2009) since, by assuming that firm closure is exogenous conditional on observables, it represents a random sample of the workforce: the job changes of displaced workers are not related to their past choices. Table 6 and table 7 show the results of these estimates respectively for the industry and the service sectors.

[21] As a robustness check we perform a two stage estimation in order to look at the impact of the spatial variables on an aggregate measure of inequality, the Gini index. Therefore, we first run a fixed effects estimation of wages on the observable workers characteristics (age, age squared, blue collar dummy) and on the firm size. Then we compute the Gini index at provincial-sectoral level by using the residuals ($\varepsilon_{i,t}$) of this first stage regression, and we regress them on the spatial variables by using fixed effects estimations. We also run the same procedure using the joint residuals (the residuals plus the fixed effects estimates, $\varepsilon_{i,t}+u_i$) to look at the differences of the spatial variable impact on inequality when taking or not into account the individual unobserved heterogeneity. Results widely confirm previous outcomes. In fact, as for the industry sector, when not taking into account the unobserved heterogeneity (using the joint residuals, $\varepsilon_{i,t}+u_i$), spatial variables entail a strong, positive and significant impact on wage inequality, while, when considering the role of sorting (using the $\varepsilon_{i,t}$ residuals), density does not longer entail any significant impact on wage inequality, while specialization still does. As for the service sector, spatial variables have a positive and significant impact on the Gini index in both cases. However, it results to be generally reduced when taking into account the unobserved individual heterogeneity. We do not show these estimates for a sake of synthesis. They are available upon request.

Table 6: Quantile Fixed Effects Regressions of Wages on the Spatial Variables using the Koenker Procedure. Sample of Displaced Workers. Industry

	q10	q25	q50	q75	q90
Specialization	-0.0019	0.0005	0.0021	0.0026	0.0024
	[0.0036]	[0.0027]	[0.0025]	[0.0026]	[0.0031]
Density	0.0069	0.0071	0.0068	0.0069	0.0086
	[0.0051]	[0.0049]	[0.0048]	[0.0048]	[0.0049]*
Age	0.0357	0.0300	0.0266	0.0229	0.0168
	[0.0048]***	[0.0039]***	[0.0033]***	[0.0034]***	[0.0042]***
Age Squared	-0.0003	-0.0003	-0.0002	-0.0002	-0.0001
	[0.0001]***	[0.0000]***	[0.0000]***	[0.0000]***	[0.0000]**
Blue Collar Dummy	-0.0981	-0.0929	-0.0902	-0.0911	-0.1004
	[0.0258]***	[0.0259]***	[0.0261]***	[0.0261]***	[0.0260]***
Firm Size	0.0139	0.0131	0.0132	0.0133	0.0118
	[0.0031]***	[0.0029]***	[0.0027]***	[0.0027]***	[0.0028]***
Constant	10.4932	10.6557	10.7596	10.8745	11.0665
	[4.3339]**	[4.3320]**	[4.3284]***	[4.3280]***	[4.3279]***
Area, Time and Sector dummies	yes	yes	yes	yes	yes
N. Observations	26,306	26,306	26,306	26,306	26,306
N. Individuals	5,922	5,922	5,922	5,922	5,922

Notes: Standard Errors in Parenthesis with ***,** and * denoting significance at 1%, 5% and 10% respectevely. Bootstrapped standard errors are given in parenthesis. The bootstrapping was done using samples of 10,000 observations and 2,000 iterations.

Table 7: Quantile Fixed Effects Regressions of Wages on the Spatial Variables using the Koenker Procedure. Sample of Displaced Workers. Services

	q10	q25	q50	q75	q90
Specialization	-0.0050	-0.0035	-0.0017	0.0009	0.0020
	[0.0058]	[0.0049]	[0.0046]	[0.0045]	[0.0050]
Density	0.0072	0.0068	0.0068	0.0063	0.0056
	[0.0071]	[0.0068]	[0.0068]	[0.0069]	[0.0072]
Age	0.0529	0.0507	0.0470	0.0422	0.0373
	[0.0077]***	[0.0068]***	[0.0062]***	[0.0096]***	[0.0075]***
Age Squared	-0.0005	-0.0005	-0.0004	-0.0004	-0.0003
	[0.0001]***	[0.0001]***	[0.0001]***	[0.0001]***	[0.0001]**
Blue Collar Dummy	-0.0981	-0.0929	-0.0902	-0.0911	-0.1004
	[0.0460]	[0.0471]	[0.0469]	[0.0465]	[0.0459]*
Firm Size	0.0126	0.0108	0.0103	0.0102	0.0084
	[0.0044]***	[0.0044]**	[0.0044]**	[0.0044]**	[0.0047]*
Constant	7.2192	7.3142	7.4359	7.5897	7.7757
	[1.1725]**	[1.1549]**	[1.1768]**	[1.2070]***	[1.2223]***
Area, Time and Sector dummies	yes	yes	yes	yes	yes
N. Observations	9,450	9,450	9,450	9,450	9,450
N. Individuals	2,146	2,146	2,146	2,146	2,146

Notes: Standard Errors in Parenthesis with ***,** and * denoting significance at 1%, 5% and 10% respectevely. Bootstrapped standard errors are given in parenthesis. The bootstrapping was done using samples of 2,000 observations and 2,000 iterations.

The results computed on the sample of displaced workers are even more striking. Actually, coefficients estimates are even more reduced in magnitude than in previous quantile fixed effects estimates and they are mainly not statistically significant. Hence, the impact of spatial externalities basically disappears, suggesting that sorting captures all the effects related to spatial externalities. However, it is also worth noting that the sample of displaced workers over-represent some workers' characteristics of the original sample. In fact, the percentage of blue collar workers is higher in the sample of displaced workers, and this is likely to explain part of the different results of these estimations compared to the previous ones. Therefore, even if these estimates widely confirm the importance of the sorting of workers, we consider the quantile fixed effects on the whole sample as our preferred estimates. [22]

[22] Further, one might argue that our estimates might be affected by the endogeneity between wages and spatial variables. Unfortunately we cannot make use of a fixed effects IV quantile procedure, since, to the best of our knowledge, it is not still available. Nonetheless, we run the same quantile fixed effects estimations on the 3-lagged values of our spatial variables to take into account at least simultaneity issues. Results do not significantly change, and in some cases spatial externalities coefficients are even higher.

5. Sorting features

5.1 Sorting distribution

The results in the previous section highlighted that sorting plays a crucial role in explaining the differences in wages among locations characterized by different levels of density and/or specialization, especially for what concern the upper tale of the wage distribution. The aim of this section is to characterize the features of sorting, by looking at the distribution of the individual fixed effects among high and low density provinces, as well as among high and low specialized provinces. Since the individual fixed effects are a proxy for time-invariant individual characteristics that represent our measure of workers' unobserved heterogeneity, the analysis of the spatial distribution of the individual fixed effects can help us to better understand the characteristics of the sorting of workers along the spatial dimensions. We first split provinces in both low density (LD) and high density (HD) and low specialized (LS) and high specialized (HS) on the basis of the median of the (time average of) density and specialization in our database. We then use the individual fixed effects obtained from our preferred specification, the quantile fixed effect estimations of Table 4 (for industry) and 5 (for services). Table 8 reports summary statistics of the distribution of skills in both HD, LD and HS, LS provinces, providing a picture of the sorting of workers across provinces. As for density, workers in HD provinces are much more skilled than those in LD provinces, both in the industry (0.051 vs -0.053) and the service sector (0.047 vs -0.047).[23] Moreover, the difference in skill averages between HD and LD provinces is even higher when these provinces are also HS (0.083 vs -0.077 for the industry and 0.106 vs -0.056 for the service sector), suggesting that -as expected- there is an interrelation between the effects of the two spatial variables. It is also possible to compute a rough measure of how much the difference in the average wages between HD and LD areas (at the denominator) can be explained by the difference in the average fixed effects (at the numerator). We find out that differences in fixed effects account for 67% of row spatial wage variation between HD and LD provinces in the industry sector and 57% in the service sectors, measures similar to those described in Mion and Naticchioni (2009) (75% for all the economy).

As for specialization, table 8 shows that workers in HS provinces are more skilled than those in LS provinces only in the service sector (0.026 HS vs -0.026 LS). As for the industry sector, the difference in skills between HS and LS provinces is negligible (it becomes more relevant when this difference is considered in provinces characterized also by a HD: 0.083 vs 0.023). Further, individual fixed effects accounts account for 39% (74%)

[23] These findings are in line with Mion and Naticchioni (2009) who find out similar results for the two sectors together, by using the fixed effects computed in IV within estimates.

of the differences in row wage variation between HS and LS provinces in the industry (service) sector.[24]

Table 8: Distribution of the Individual Fixed Effects by High and Low Density and High and Low Specialized Provinces						
	Industry			**Services**		
	mean	**std. err.**	**n.obs**	**mean**	**std.err.**	**n.obs.**
HD	0.051	0.0016***	35549	0.047	0.0020***	34470
LD	-0.053	0.0014***	33806	-0.047	0.0019***	34838
HS	-0.003	0.0016*	35073	0.026	0.0020***	34779
LS	0.003	0.0015*	34282	-0.026	0.0019***	34529
HD and HS	0.083	0.0024***	16263	0.106	0.0029***	17718
HD and LS	0.023	0.0021***	19286	-0.015	0.0027***	16752
LD and HS	-0.077	0.0019***	18810	-0.056	0.0025***	17061
LD and LS	-0.023	0.0021***	14996	-0.037	0.0027***	17777

Notes: HD stands for High Density, LD stands for Low Density, HS stands for High Specialization, LS stands for Low Specialization. The difference between High and Low Density (Specialization) is defined by the median value of Density (Specialization).

Let us now go back to the main focus of the econometric analysis, the distributional consequences of the agglomeration externalities and the role of sorting. In order to further characterize the stronger impact of sorting in the upper tail of the wage distribution, we compute summary statistics on the distribution of the individual fixed effects between HD and LD, and between HS and LS provinces, by terciles of the wage distribution (0-33, 33-66, 66-100).

Table 9 reports the related descriptive statistics. First of all, it is worth noting that, as expected, the individual fixed effects are negative for those workers belonging to the lowest wage tercile, negligible for those in the central tercile, and positive for those belonging to the highest tercile. Second, in the third tercile of the wage distribution (Panel C) the number of observations is much higher in the HD provinces than in the LD provinces. Conversely in the first tercile (Panel A) the number of observations in the LD provinces is higher than in the HD provinces. This suggests the presence of a composition effect, i.e. high-paid (low paid) workers are concentrated in HD (LD) provinces. Third, the difference in average skill levels of workers located in HD (LS) and LD (LS) provinces is relevant only when taking into account the highest tercile. In fact, while the panel A and B of table 9 show that the means of the fixed effects are not

[24] Note that the low percentage of explanation for the specialization dimension detected for the industry sector is due to the construction sector, which displays negative fixed effects averages, which turn out to be also more negative in the HS than in the LS provinces. Actually, when excluding this sector the averages of the fixed effects in the HS and LS areas result to be respectively 0.042 and 0.026, whose difference accounts for roughly 90% of the row spatial wage variation.

statistically different in high and low density (or specialized) provinces, panel C of table 9 shows statistically relevant differences for the third tercile (66^{th}-100^{th}) of the wage distribution. In particular, as for density in the industry sector, the difference in skills levels between workers in HD and LD provinces is significant and of about 0.05, while in the service sector this is of 0.01. As for specialization, the difference in skill levels between workers employed in HS and LS provinces is around 0.03 in the industry sector and 0.04 in the service sector and again these differences are statistically significant. This pattern therefore implies that the difference in skill averages observed in table 8 between HD and LD provinces, as well as between HS and LS provinces, is actually driven by the difference in skill levels of the highest skilled workers, i.e. those located in the upper tail of the wage distribution (sorting). This also confirms that sorting is at work and explains the coefficients drop detected in quantile fixed effects estimations for the right hand tale of the wage distribution. Fourth, the averages of the fixed effects are higher in the service sector, i.e. the most skilled workers are employed in the service sector. This outcome allows us to explain why in the cross sectional quantile estimations the impact of spatial externalities was stronger in the service sector and why the service sector experienced the greater coefficients drop in the quantile fixed effects estimates.

These outcomes are consistent with Johnson (1953) who argues that urban workers have higher ability than non-urban workers, and with Glaeser and Marè (2001) and Moretti (2004) that argue that skilled individuals sort in cities because of the advantages these locations can offer.

Table 9: Distribution of the Individual Fixed Effects by High/Low Density, High/Low Specialized Provinces and Wage Percentiles

	Industry			Services		
Wage percentile	**mean**	**std. err.**	**n.obs**	**mean**	**std.err.**	**n.obs.**
<33			**Panel A**			
HD	-0.24	0.0014***	9195	-0.36	0.0027***	8868
LD	-0.25	0.0012***	14901	-0.33	0.0022***	14157
HS	-0.25	0.0013***	12612	-0.34	0.0025***	10510
LS	-0.24	0.0013***	11484	-0.35	0.0023***	12515
33-66			**Panel B**			
HD	-0.04	0.0012***	14901	-0.02	0.0013***	11106
LD	-0.03	0.0010***	10337	0.00	0.0013	11679
HS	-0.04	0.0010***	11225	-0.01	0.0013***	11816
LS	-0.04	0.0010***	11280	-0.01	0.0013***	10969
>66			**Panel C**			
HD	0.31	0.0023***	14186	0.35	0.0023***	14496
LD	0.26	0.0025***	8568	0.34	0.0026***	9002
HS	0.31	0.0025***	11236	0.37	0.0025***	12453
LS	0.28	0.0024***	11518	0.33	0.0024***	11045

Notes: HD stands for High Density, LD stands for Low Density, HS stands for High Specialization, LS stands for Low Specialization. The difference between High and Low Density (Specialization) is defined by the median value of Density (Specialization).

5.2 Sectoral breakdown of sorting

In this section we aim at characterizing the sorting of workers among the different sectors of the economy in order to address two main issues. First, we want to investigate whether the distributional patterns of the sorting of workers is homogeneous across sectors. Second, we aim at better understanding the impact of spatial externalities in the service sector, which represents a relatively unexplored field within the spatial economic literature.[25] To the best of our knowledge, there are no papers that have focused on the sectoral breakdown of the sorting of workers. Our variable of interest is the individual skill, proxied by the individual fixed effects computed in our preferred specification (table 4 and 5 for industry and services, respectively). Instead of carrying out a set of cumbersome descriptive statistics across provinces and sectors, we regress the individual fixed effects on the sectors dummies, dummies related to both high density and specialized provinces (HD and HS), and interaction

[25] Graham (2009), using firm level data, analyzes the urbanization and localization economies across manufacturing and services in UK. He underlines the novelty of studying in detail the service sector in the spatial economic literature.

terms between sector and spatial variables.[26] Moreover, since our main concern is on distributional issues, we run quantile regressions (on the 10th, 50th and 90th percentiles). By doing so, we can first identify the skill intensity by sector, provided by the coefficient of each sector dummy (when both HD and HS are equal to zero). Table 10a shows that at the 90th percentile average fixed effects are very high in the skill-intensive sectors such as energy-chemicals, papers, machinery-electrical-transport equipment for the industry, and financial activities, real estates and informatics for the service sectors.[27]

Secondly, by looking at the differences in skill levels in each quantile when passing from LD (LS) to HD (HS) provinces, we can determine which are the sectors characterized by the highest incidence of sorting of workers, sectors that represent the driving force behind the impact of sorting detected in the previous section. Table 10b shows that there is a general increase in the difference in average fixed effects between LD to HD provinces from the 10th to the 90th percentile: the sorting of workers increases along the wage distribution for what concerns the density dimension. Moreover, the sectors that enjoy the highest increases in the difference in average fixed effects between LD and HD provinces along the percentiles are mainly the skill-intensive ones: 'papers', 'energy-chemicals', 'machinery-electrical and transport equipment', 'financial activities' and 'wholesale and retail trade'.[28]

As for the sorting of workers along the specialization dimension, the patterns of the average fixed effects are more heterogeneous among the different sectors (table 10c). In fact, some sectors show increasing sorting of workers (an increase in the difference of the average fixed effects from the 10th to the 90th percentile), while others show a negligible, or even decreasing, pattern of the sorting of workers. As for the industry, the sectors characterized by an increasing sorting of workers along the percentiles are mainly the unskill-intensive ones: 'mining', 'food', 'textile' and 'leather', sectors which represent most of the industrial districts of the Italian productive system. In the service sector the increase in average fixed effects is mainly concentrated in the 'wholesale and retail trade' and 'other public services'. It is also worth noting that since the 'wholesale and retail trade' is the sector with the highest

[26] Sectors follow the NACE classification, two letters codes.

[27] Also 'gas, electricity and water supply' as well as 'private education, health and social services' show high fixed effects values.

[28] Actually also 'real estates, rent, leasing and informatics' shows a relevant raise in the fixed effects when passing from the 50th to the 90th percentile. However, at the same time there is a wage premium due to skills also at the 10th percentile, which give raise to a negligible difference between the 90th and the 10th percentile.

number of observations within services (24,699 vs 69,308 for all services), it is possible to claim that this sector represents the driving force behind the coefficient drop at the 90th percentile when passing from the cross sectional quantile regressions to the quantile fixed effects regressions in the service sector.

Table 10a: OLS Regression of the Fixed Effects on Sectors, Density and Specialization

	Main Effects (LD and LS)		
	q10	**q50**	**q90**
Industry			
Mining	-0.430	-0.221***	-0.028***
Food	-0.305***	-0.081***	0.186***
Textile	-0.346***	-0.150***	0.204***
Leather	-0.327***	-0.126***	0.064***
Wood, rubber and plastics	-0.341***	-0.113***	0.247***
Paper	-0.242***	0.005	0.424***
Energy-chemicals	-0.174***	0.106***	0.548***
Mineral	-0.336***	-0.085***	0.198***
Metal	-0.324***	-0.089***	0.221***
Machinery, electrical and transport equipment	-0.257***	-0.043***	0.279***
Gas, electricity and water supply	-0.069***	0.134***	0.346***
Construction	-0.344***	-0.149***	0.144***
Services			
Wholesale and retail trade	-0.402***	-0.187***	0.112***
Hotels and restaurants	-0.457***	-0.207***	0.051***
Transports	-0.616***	0.069***	0.276***
Financial activities	0.008	0.234***	0.660***
Real estates, rent, leasing and informatics	-0.829***	-0.114***	0.288***
Private education, health and social services	-0.474***	-0.039***	0.402***
Other public services	-0.505***	-0.090***	0.263***

Notes: Sectors follow the Nace classification, two letters codes. The difference between High and Low Density (Specialization) is defined by the median value of Density (Specialization).

Table 10b: OLS Regression of the Fixed Effects on Sectors, Density and Specialization			
	Interaction with Density (HD-LD)		
	q10	**q50**	**q90**
Industry			
Mining	0.004	0.143***	0.191***
Food	0.025***	0.052***	0.105***
Textile	0.039***	0.074***	0.117***
Leather	0.005	0.058***	0.085***
Wood, rubber and plastics	0.030***	0.050***	0.077***
Paper	0.048***	0.085***	0.178***
Energy-chemicals	-0.029***	0.070***	0.193***
Mineral	0.050***	0.049***	0.074***
Metal	0.063***	0.063***	0.100***
Machinery, electrical and transport equipment	0.032***	0.091***	0.191***
Gas, electricity and water supply	-0.059***	-0.033***	0.027*
Construction	0.002	0.035***	0.080***
Services			
Wholesale and retail trade	0.045***	0.075***	0.210***
Hotels and restaurants	0.009	0.007	0.017
Transports	0.086***	0.038***	0.093***
Financial activities	-0.090***	-0.044***	0.093***
Real estates, rent, leasing and informatics	0.309***	0.150***	0.305***
Private education, health and social services	0.058***	0.019**	0.134***
Other public services	0.141***	0.140***	0.099***

Notes: Sectors follow the Nace classification, two letters codes. The difference between High and Low Density (Specialization) is defined by the median value of Density (Specialization).

Table 10c: OLS Regression of the Fixed Effects on Sectors, Density and Specialization			
	Interaction with Specialization (HS-LS)		
	q10	**q50**	**q90**
Industry			
Mining	0.153***	0.124***	0.277***
Food	-0.011*	0.031***	0.055***
Textile	0.002	0.040***	0.135***
Leather	-0.022*	-0.048***	0.083***
Wood, rubber and plastics	0.040***	-0.014***	0.038***
Paper	0.016*	0.035***	0.130***
Energy-chemicals	0.096***	0.022***	0.004
Mineral	0.030***	0.016***	0.031**
Metal	0.026***	0.061***	0.020***
Machinery, electrical and transport equipment	0.027***	0.006**	0.052***
Gas, electricity and water supply	-0.171***	0.014	0.128***
Construction	-0.076***	-0.047***	-0.063***
Services			
Wholesale and retail trade	0.037***	0.052***	0.131***
Hotels and restaurants	0.044***	0.025***	0.058***
Transports	-0.001	-0.002	0.027***
Financial activities	0.028***	0.013***	-0.008
Real estates, rent, leasing and informatics	0.225***	0.060***	0.034***
Private education, health and social services	-0.032*	-0.008	0.000
Other public services	0.059***	0.020***	0.266***

Notes: Sectors follow the Nace classification, two letters codes. The difference between High and Low Density (Specialization) is defined by the median value of Density (Specialization).

To sum up, in this section we found out interesting and new findings concerning the sectoral breakdown of sorting of workers. We showed that it follows an heterogeneous pattern among the different sectors of the economy. As for density, the sorting of workers is stronger in the skill-intensive sectors, thus pointing out that high skill individuals employed in such sectors are those who benefit the most from spatial externalities. Conversely, sorting along the specialization dimension is mainly concentrated in unskill-intensive sectors, sectors that typically represent the industrial districts in the Italian productive system.

6. Conclusions

In this paper we investigate the role that sorting plays in the relationship between spatial externalities (in terms of industrial specialization and density) and wage distribution. We use individual Italian panel data and quantile fixed effects estimations, since they allow us to get an estimate of the impact of spatial variables not affected by observed and unobserved individual and firm heterogeneity. Our results show that sorting matters and its impact increases along the wage distribution; it actually captures most of the impact of spatial externalities derived in standard quantile estimations. These findings therefore suggest the need of taking into account the role of sorting when considering the relationship between wage inequality and spatial externalities. Finally, analyzing the sectoral breakdown of sorting, we show that that it is not homogeneous across sectors. More specifically, along the density dimension the sorting of workers takes place in skill-intensive sectors, while along the specialization dimension it is mainly concentrated in the unskill-intensive ones.

References

1. Abowd J., Kramarz F. and Margolis D., (1999), "High wage workers and high wage firms", *Econometrica*, vol.67, n.2, pp. 251-334.

2. Abdel Rahman H.M. and Fujita M., (1990), "Product Variety, Marshallian Externalities and City Size", *Journal of Regional Science*, Vol.30(2), pp.165-183.

3. Abrevaya, J. and Dahl C., (2006), "The effects of Birth Inputs on Birth-weight: Evidence from Quantile Estimation on Panel Data", forthcoming in *Journal of Business and Economic Statistic*.

4. Arulampalam W., Naylor R.A. and Smith J., (2007), "Am I Missing Something? The Effects of Absence from Class on Student Performance", *University of Warwick Working Paper* n.820.

5. Bache S.H., Dahl C. and Tang Kristensen J., (2008), "Determinants of Birthweight Outcomes: Quantile Regression Based on Panel Data", *CREATES Research Paper* 2008-20.

6. Bargain O. and Melly B., (2008), "Public Sector Pay Gap in France: New Evidence Using Panel Data", *IZA Working Paper* n. 3427.

7. Brown G. and Medoff J., (1989), "The Employer Size Wage Effect", *Journal of Political Economy,* Vol. 97, pp.1027-1057.

8. Chamberlain G., (1994), "Quantile Regression, Censoring, and the Structure of Wages". In: Sims, C. (ed.): *Advances in Econometrics: Sixth World Congress.* 6, 1. Cambridge University Press: Cambridge Mass.: 405–437.

9. Ciccone A. and Hall R., (1996), "Productivity and the Density of Economic Activity", *American Economic Review*, vol.86, pp.54-70.

10. Ciccone A., (2002), "Agglomeration Effect in Europe", European Economic Review, vol.46, pp.213-227.

11. Combes P., (2000), "Economic structure and local growth: France, 1984-1993", *Journal of Urban Economics*, Vol.47, pp.329-355.

12. Combes P., Duranton G. and Gobillon L., (2008), "Spatial Wage Disparities: Sorting Matters!", *Journal of Urban Economics*, 63, pp.723-742.

13. De Blasio G. and Di Addario S., (2002), "Laber market pooling: evidence from Italian industrial districts", *Temi di Discussione*, Banca d'Italia, n.453.

14. Di Addario S. and Patacchini E., (2007), "Wages and the Cities: Evidence from Italy", *Centro Studi Luca d'Agliano Development Studies Working Paper* No. 231, forthcoming in *Labour Economics*.

15. Duranton G. and Puga D., (2003), "Microfoundations of urban agglomeration economies", *Handbook of Regional and Urban Economics*, Vol.4, Henderson J.V. and Thisse J.-F. (eds), Elsevier North Holland, Amsterdam.

16. Dustmann C. and Meghir ., (2005), "Wages, Experience and Seniority", *Review of Economic Studies*, n.72, pp.77-108.

17. Glaeser E.L., Resseger M.G. and Tobio C., (2008), "Urban Inequality", NBER Working Paper n.14419.

18. Glaeser E.L., Kallal H.D., Scheinkman J.A. and Shleifer A., (1992), "Growth in Cities", *The Journal of Political Economy*, vol.100, n.6, pp.1126-1152.

19. Glaeser E.L. and Marè D.C., (2001), "Cities and Skills", *Journal of Labor Economics*, vol.19(2), pp.316-342.

20. Graham, D. J. (2009). Identifying urbanisation and localisation externalities in manufacturing and service industries. *Papers in Regional Science* 88 (1), 63{84.

21. Helsley R. and Strange W., (1990), "Matching and Agglomeration Economies in a System of Cities", *Regional Science and Urban Economics*, vol.20, n.2, pp.189-212.

22. Henderson J.V., (1974), "The sizes and types of cities", *American Economic Review*, vol.64, n.4, pp.640-656.

23. Johnson D. Gale, (1953), "Comparability of Labor Capacities of Farm and Nonfarm Labor," *American Economic Review*, June , 43, 296-313.

24. Kim S., (1989), "Labor Specialization and the Extent of the Market", *Journal of Political Economy*, vol.97, n.3, pp.692-705.

25. Koenker R., (2004), "Quantile Regression for Longitudinal Data", *Journal of Multivariate Analysis*, vol.91(1), pp.74-89.

26. Koenker R. and Bassett G., (1978), "Regression Quantiles", *Econometrica*, vol.46, pp.33-50.

27. Krueger A.B. and Summers L.H., (1988), "Efficiency Wages and the Inter-Industry Wage Structure", *Econometrica*, Vol.56(2), pp.259-293.

28. Marshall, A., (1890), "Principles of Economics", London: Macmillan.

29. Melitz M., (2003), "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity", *Econometrica*, vol.71, n.6, pp.1695-1725.

30. Mion G., Naticchioni P., (2009), "The Spatial Sorting and Matching of Skills and Firms", *Canadian Journal of Economics*, Canadian Economics Association, vol. 42(1), pages 28-55, February.

31. Moller J. and Hass A., (2003), "The Agglomeration Differential Reconsidered: an Investigation with German Micro Data 1984-1997", in: Broecker, J., D. Dohse, R. Soltwedel (Eds.). *Innovation Clusters and Interregional Competition*. Berlin: Springer.

32. Moretti E., (2004), "Human Capital Externalities in Cities", in Henderson J.V. and Thisse J.F. (eds.), *Handbook of Regional and Urban Economics*, Elsevier-North Holland, Amsterdam, Vol.4.

33. Postel-Vinay F. and Robin J.M., (2006), "Microeconometric Search-Matching Models and Matched Employer-Employee Data", in R. Blundell, W. Newey and T. Persson (eds.), *Advances in Economics and Econometrics: Theory and Applications*, Ninth World Congress, Volume 2, Cambridge: Cambridge University Press.

34. Rosenthal S.S. and Strange W.C., (2004), "Evidence on the Nature and Sources of Agglomeration Economies". *Handbook of Regional and Urban Economics*, Vol.4, J.V. Henderson and F.Thisse (eds.), Elsevier-North Holland, Amsterdam.

35. Signorini L.F., (1994), "Una verifica quantitativa dell'effetto distretto", *Sviluppo Locale*, vol.1, n.1, pp.31-70.

36. Signorini L.F., (2000), (by) "Lo Sviluppo locale. Un'indagine della Banca d'Italia sui distretti industriali",Donzelli, Roma.

37. Tattara G., (2001), "L'efficienza dei distretti industriali: una ricerca condotta dal servizio studi della Banca d'Italia", *Economia e Società Regionale*, n.4, pp.114-143.

38. Topel R., (1991), "Specific capital, mobility, and wages: Wages rise with job seniority", *Journal of Political Economy*, Vol. 99, 145-176.

39. Wheeler C.H., (2004), "Wage Inequality and Urban Density", *Journal of Economic Geography*, vol.4, pp.421-437.

40. Wheeler C.H., (2007), "Industry Localization and Earnings Inequality: Evidence from US Manufacturing", *Papers in Regional Science*,Vol.86(1),pp.77-100.

41. Wooldridge J.M., (2003), "Introductory Econometrics", 2nd Ed., South-Western.