Spatial Externalities and Wage Distribution: the Role of Sorting*

Alessia Matano† and Paolo Naticchioni‡

January 2009

Preliminary Version

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 this paper is to show that sorting matters also when addressing the relationship between spatial externalities and wage distribution, i.e, across workers located at different points of the wage distribution. Using Italian individual panel data we analyze how the gains in productivity associated with spatial externalities (employment density and specialization) are distributed across different percentiles of the wage distribution. Furthermore, using quantile fixed effect estimates we also show that the sorting of workers matters since it explains most of the impacts of spatial externalities on the wage distribution. After controlling for sorting, spatial externalities entail a slight positive impact on wage inequality, since they positively affect the upper tail of the wage distribution.

JEL Classification: J31, L16, R23, R30.
Keywords: Agglomeration Externalities, Sorting, Wage Inequality, Quantile Fixed Effects.

---

* We thank the research partnership between ISFOL - Area Mercato del Lavoro (Rome) - and Dipartimento di Scienze Economiche - University of Rome "La Sapienza" - for the access to the databases. We are very grateful to Christian Dahl for his precious advices and for the access to the routines and to Roger Koenker for his help and the availability of his procedure.

† Alessia Matano, University of Rome "La Sapienza", DSE Dipartimento di Scienze Economiche. Email: alessia.matano@uniroma1.it.

‡ Paolo Naticchioni, University of Rome "La Sapienza", DSE Dipartimento di Scienze Economiche and Université Catholique de Louvain. Email: paolo.naticchioni@uniroma1.it.

1. Introduction

While the relationship between spatial externalities and differences in average wages between locations has been widely investigated in the literature, 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. Another exception is Moller and Haas (2003), which analyzes 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, these empirical studies suffer from an important drawback. Since they 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 has 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. In this way 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, 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 two different techniques of quantile fixed effects estimates, proposed by Koenker (2004) and Arulampalam et al. (2007), in order to evaluate whether and how the impacts of spatial externalities change when also 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 the estimate of an individual fixed effect. In this way we are able to assess whether sorting entails a not uniform impact on the wage distribution. Moreover, we perform separate estimations for the industry and the service sectors in order to evaluate whether different patterns apply in the two sectors.

Our results show that taking into account unobserved individual heterogeneity significantly changes standard quantile coefficients estimates. In particular, all coefficients of spatial externalities are reduced, and the strongest reductions concern the upper tail of the wage distribution. Moreover, coefficients are no longer always statistically significant. These findings suggest that sorting matters and that its impact is not uniformly distributed along the wage distribution. Further, the impact of sorting is generally higher in the service sector. Nonetheless, even after controlling for the effect of sorting we find out that sectoral specialization has an increasing impact on the wage distribution, even if small in magnitude. Moreover, this is significant only at the right hand side of the wage distribution. As for the impact of density, it is uniformly distributed in the industry sector, while it contributes to increase wage inequality in the service sector since it positively affects the upper tail of the wage distribution.

Our findings therefore confirm in a distributional perspective that sorting matters and captures most of the impact of spatial externalities. Using individual panel data is hence crucial to investigate the impact of spatial externalities. Our findings also show that even after controlling for the role of sorting, there is still some evidence of a not uniformly distributed impact of spatial externalities on wage distribution, which mainly favours workers in the upper tail of the wage distribution. This outcome therefore suggests that the skilled and high paid workers are those who benefit from the spatial externalities. This effect might be due either to a better ability of skilled workers to gain from face-to-face interactions (Glaeser and Marè, 2001) or to the complementarity between skilled workers and technological and knowledge spillovers that arise in areas of dense economic activity or characterized by a high level of sectoral specialization.

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 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 consumption amenities such as cultural activities, museums, theaters 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

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

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. 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.

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 firms 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 the 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 firm is only the 5.6%. Therefore, not taking into account the worker (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.[2]

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.[3] 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.[4]

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.[5]

The dependent variable in our regressions is the (log) real gross weekly wage in euro.[6] 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

[2] Actually, Di Addario and Patacchini, 2007, take into account the issue of sorting finding out that the sorting of workers results to be not a major concern.

[3] 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.

[4] 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.

[5] 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.

[6] 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).

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 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:[7]

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

where subscript p refers to province, s to sector and t to time.[8] 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:[9]

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

where subscript p refers to province and t to time (province area is measured in square km)[10]

Since all the spatial variables are defined in logarithm, we estimate elasticities. Table 1 shows the descriptive statistics of the variables of the analysis.

[7] As in Combes (2000), Mion and Naticchioni (2008) and De Blasio and Di Addario (2002).
[8] We define the index of specialization by sectors following the Ateco81 classification, two digits level, since it provides higher data accuracy.
[9] As in Combes (2000), Mion and Naticchioni (2008) and Ciccone and Hall (1996).
[10] 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	1636.83	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	1611.75	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 regression 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, Arulampalam et al., 2007) in order to estimate the impact of spatial variables on the wage distribution taking into account 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).[11] The aim of the quantile fixed effects estimations, whose technique are explained in details below, is to understand whether the detected impact of spatial externalities found 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 within locations. 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 estimates of the impact of spatial externalities

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 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.[12] 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.

[11] 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).

[12] 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).

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

As for density, the elasticity estimates go from 1.3% at the 10^{th} percentile to 2.1% at the 90^{th} in the industry and from 0.9% at the 10^{th} percentile to 2.4% at the 90^{th} 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 the 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.[13]

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

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.

[13] 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 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.

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 two different techniques for the quantile fixed effects estimations. The first one is 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.[14] 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 α_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 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).[15]

---

[14] 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).

[15] 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

The second estimation technique we use is suggested 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, it is carried out a simultaneous quantile estimation adding as explanatory variables the fixed effects estimated in the first stage. Though the asymptotic properties of this estimator are not known, it seems to have a good performance and it is quite simple to implement (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]

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 reductions. Therefore, sorting captures most of the impact of specialization on wage inequality and its impact is increasing 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 section 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

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).

[16] We again use bootstrapping to get the coefficients and standard errors estimates.

[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.

[18] The differences among coefficients at different percentiles are not statistically different from zero while they were significant 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

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. Hence, when taking sorting into account, density entails just a slightly positive impact on wage inequality in the service sector, while it does no longer affect wage inequality in the industry sector.

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.

Table 6 and 7 show the results of the two stages procedure. Results are substantially the same as those obtained in the Koenker procedure, thus confirming the previous results.[20]

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] We also carry out the fixed effects quantile estimations adopting a finer decomposition for the firm size in order to further analyze the impact of the sorting of firms on the wage inequality. In particular, we create an interaction effect between regional dummies and the firm size that allows to capture different performance of firms located in different regions (since Italy is characterized by strong regional differences). The outcomes of these estimations

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.

Hence, our findings show that the wage premium associated with spatial variables is actually a mixture between ability bias and the spillover advantages stemming from agglomeration, in line with Glaeser and Marè (2001), Mion and Naticchioni (2009) and Combes et al. (2008). However, by means of a quantile fixed effects approach we can go a step further, emphasizing two main new findings. First, we can show that the importance of the sorting of workers increases along the wage distribution. Second, once taken into account the role of sorting, there is evidence of a positive impact of spatial externalities in the upper tail of the wage distribution. This latter finding can be consistent with various possible explanations. For instance, Glaeser and Marè (2001) suggest that skilled workers are attracted by cities and, at the same time, cities make skilled workers more productive. Hence, skilled workers are better able to gain from the face-to-face interactions or from the technological and knowledge

show that density does not anymore entail a significant impact along the wage distribution of Italian workers in the service sector, while results do not change significantly in the industry sector. As for the sectoral specialization, results are similar to the basic fixed effects quantile regressions. Hence, there is a some evidence of the sorting of firms in explaining the impact of agglomeration externalities on wages. However, this is less relevant compared to that of the sorting of workers. This result is in line with Mion and Naticchioni (2009). These estimates are available upon request.

spillovers that arise in areas of dense economic activity or characterized by a relatively high degree of sectoral specialization.

Table 6: Quantile Regressions of Wages on the Spatial Variables using the Plugin Fixed Effects Procedure. Industry

	q10	**q25**	**q50**	**q75**	**q90**
Specialization	0.0008	0.0021	0.0031	0.0047	0.0061
	[0.0106]	[0.0101]	[0.0102]	[0.0104]**	[0.0110]**
Density	0.0126	0.0123	0.0127	0.0132	0.0150
	[0.0027]***	[0.0024]***	[0.0023]***	[0.0024]***	[0.0026]***
Age	0.0403	0.0343	0.0309	0.0294	0.0282
	[0.0026]***	[0.0020]***	[0.0019]***	[0.0020]***	[0.0025]***
Age Squared	-0.0004	-0.0003	-0.0003	-0.0002	-0.0002
	[0.0000]***	[0.0000]***	[0.0000]***	[0.0000]***	[0.0000]***
Blue Collar Dummy	-0.0579	-0.0723	-0.0787	-0.0896	-0.1190
	[0.0106]***	[0.0101]***	[0.0102]***	[0.0104]***	[0.0110]***
Firm Size	0.0246	0.0228	0.0224	0.0214	0.0197
	[0.0017]***	[0.0016]***	[0.0016]***	[0.0016]***	[0.0016]***
Constant	-0.3365	-0.0628	0.0697	0.1867	0.4232
	[0.0406]***	[0.0198]***	[0.0134]***	[0.0219]***	[0.0420]***
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.

Table 7: Quantile Regressions of Wages on the Spatial Variables using the Plugin Fixed Effects Procedure. Services

	q10	q25	q50	q75	q90
Specialization	-0.0029	-0.0014	0.0001	0.0024	0.0047
	[0.0028]	[0.0024]	[0.0023]	[0.0023]	[0.0026]*
Density	0.0067	0.0075	0.0078	0.0086	0.0107
	[0.0048]	[0.0046]	[0.0046]*	[0.0046]*	[0.0046]**
Age	0.0526	0.0455	0.0418	0.0391	0.0356
	[0.0027]***	[0.0021]***	[0.0020]***	[0.0021]***	[0.0028]***
Age Squared	-0.0005	-0.0004	-0.0004	-0.0003	-0.0003
	[0.0000]***	[0.0000]***	[0.0000]***	[0.0000]***	[0.0000]***
Blue Collar Dummy	-0.0902	-0.0979	-0.1007	-0.1037	-0.1210
	[0.0108]***	[0.0106]***	[0.0107]***	[0.0107]***	[0.0114]***
Firm Size	0.0139	0.0108	0.0102	0.0098	0.0067
	[0.0022]***	[0.0020]***	[0.0020]***	[0.0020]***	[0.0021]***
Constant	-0.4399	-0.1129	0.0215	0.1422	0.4150
	[0.0407]***	[0.0184]***	[0.0112]***	[0.0185]***	[0.0414]***
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.

5. Conclusions

In this paper we aim at deepening the relevance of sorting when investigating the relationship between spatial externalities (in terms of industrial specialization, market potential and density) and wage distribution. In order to accomplish this task, we use individual panel data and quantile fixed effects estimations, which allow us to get an estimation of the impact of the spatial variables at different percentiles of the wage distribution not affected by unobserved individual (and firm) heterogeneity. Our results for the Italian case show that sorting matters and its impact increases along the wage distribution of workers. Moreover, sorting actually explains most of the outcome of spatial externalities detected 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.

Nonetheless, even after controlling for sorting, there is still some evidence of a not uniformly distributed impact of spatial externalities on the wage distribution. In particular, workers located in the upper tail of the wage distribution benefit from a positive, even if small, wage premium arising from specialization, both in the industry and in the service sector, in turn entailing a positive effect on wage inequality. As for density, only in the service sector there is evidence of a wage premium for those workers located in the upper tail of the wage distribution, while in the industry the benefits are substantially equally distributed along the wage distribution. To sum up, most of the impact of spatial externalities is actually captured by the sorting of workers, and the residual effect of these externalities favours workers located in the upper tail of the wage distribution, entailing a positive effect on wage inequality.

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

18. 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.

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

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

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

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

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

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

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

26. Mion G., Naticchioni P., (2009), "The Spatial Sorting and Matching of Skills and Firms", forthcoming in *Canadian Journal of Economics*.

27. 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.

28. 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.

29. 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.

30. 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.

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

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

33. 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.

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

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

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

37. Wooldridge J.M., (2003), “Introductory Econometrics”, 2nd Ed., South-Western.