Sorting and local wage and skill distributions in France

Pierre-Philippe Combes[*†]
Aix-Marseille School of Economics
Gilles Duranton[*‡]
University of Toronto
Laurent Gobillon[*§]
INED *and* PSE
Sébastien Roux[*‖]
DARES *and* CREST-INSEE
8 November 2011

ABSTRACT: This paper provides descriptive evidence about the distribution of wages and skills in denser and less dense employment areas in France. We confirm that on average, workers in denser areas are more skilled. There is also strong over-representation of workers with particularly high and low skills in denser areas. More generally, inequality is higher within dense areas, even more for wages than for skills. These features are consistent with patterns of migration including negative selection of migrants to less dense areas and positive selection towards denser areas. Nonetheless migration, even in the long-run, is not able to account for all the skill differences between denser and less dense areas, which suggests a role for differences prior to the entry on the labour market or for stronger learning in cities. Finally, we find marked differences across age groups and some suggestions that much of the skill differences across areas can be explained by differences between occupational groups rather than within.

Key words: sorting, wage distribution, skill distribution
JEL classification: J31, J61, R12, R23

[*]This paper is produced as part of the project "European Firms in a Global Economy (EFIGE), a Collaborative Project funded by the European Commission's Seventh Research Framework Programme, Contract number 225551. Any opinions expressed here are those of the author(s) and not those of the European Commission. The access to the data was carried through the CASD dedicated to researchers authorised by the French Comité du secret statistique. We are grateful to Diego Puga for discussions. Thanks also to Jan Eeckhout and Kurt Schmidheiny for motivating us to work on this topic. Finally, this paper is dedicated to Jacques Thisse, his long-standing interest in labour mobility, and his help and support over the years.

[†]GREQAM, 2 Rue de la Charité, 13236 Marseille cedex 02, France (e-mail: `ppcombes@univmed.fr`; website: `http://www.vcharite.univ-mrs.fr/pp/combes/`). Also affiliated with the Centre for Economic Policy Research.

[‡]Department of Economics, University of Toronto, 150 Saint George Street, Toronto, Ontario M5S 3G7, Canada (e-mail: `gilles.duranton@utoronto.ca`; website: `http://individual.utoronto.ca/gilles/default.html`). Also affiliated with the Centre for Economic Policy Research.

[§]Institut National d'Etudes Démographiques, 133 Boulevard Davout, 75980 Paris cedex 20, France (e-mail: `laurent.gobillon@ined.fr`; website: `http://laurent.gobillon.free.fr/`). Also affiliated with the Centre for Economic Policy Research.

[‖]Direction de l'Animation de la Recherche, des Etudes et des Statistiques, Ministère du Travail, de l'Emploi et de la Santé, 39-43, quai André Citroën, 75902 Paris Cedex 15, France (e-mail: `sebastien.roux@travail.gouv.fr`).

1. Introduction

This paper provides descriptive evidence about the distribution of wages and skills in denser and less dense employment areas in France. Our analysis focuses on the sorting of workers by skills. We find strong evidence that workers sort by skills. We confirm that on average, workers in denser areas are more skilled. However, the distribution of skills in denser areas relative to less dense areas is not well described by this difference in their first moment. There is strong over-representation of workers with particularly high and low skills in denser areas. More generally, inequality is higher within dense areas, even more for wages than for skills. These features are consistent with patterns of migration including negative selection of migrants to less dense areas and positive selection towards denser areas. Nonetheless migration, even in the long-run, is not able to account for all the skill differences between denser and less dense areas, which suggests a role for differences prior to the entry on the labour market or for stronger learning in cities. Finally, we find marked differences across age groups and some suggestions that much of the skill differences across areas can be explained by differences between occupational groups rather than within.

These findings are important for a number of reasons. First inequalities between areas receive considerable attention from policy makers, the media, and the general public. Governments devote large sums of money to place-based policies (Glaeser and Gottlieb, 2008, Moretti, 2011). These policies are often criticised on efficiency grounds. Our results show that they also miss their equity target to some extent since the least skilled workers tend to be over-represented in the richest areas that benefit little from these policies.

Second, inequalities between individuals within areas are also an important concern. We show that denser areas are more unequal. This is in part because they host workers drawn from a more unequal distribution of skills. However, this is only part of the story. There is more inequality in denser areas even after conditioning out their more uneven distribution of skills. Higher inequalities are thus an undesirable effect of urbanisation. However, the appropriate response is not to limit urbanisation which brings efficiency gains. Instead, both efficiency and equity require thinking about providing some form of insurance to workers in higher density areas.

Third, the idea of worker sorting across cities by skills has recently caught the interest of urban theorists (Davis, 2009, Behrens, Duranton, and Robert-Nicoud, 2010, Eeckhout, Pinheiro, and Schmidheiny, 2010, Venables, 2011). This paper provides a rich set of facts over which future models of sorting across cities will be able to build.

The approach we develop in this paper looks at the entire distributions of wages and workers skills in denser and less dense areas. There are four reasons why the distributions of wages may differ across areas. First, the initial distribution of workers' skills may differ.

Second, workers may sort by skills. Third, workers may benefit from agglomeration effects that raise their wages for any given level of skills. Fourth, these benefits from agglomeration may differ depending on skills.

To assess the importance and the effects of sorting, we need to overcome four main difficulties. The first is the need for a methodology to compare entire distributions of skills and estimate key parameters that describe how they relate to each other. This is important because the distributions of skills in denser and less dense areas differ by much more than their first (or even second) moment. We extend the approach of Combes, Duranton, Gobillon, Roux, and Puga (2011*b*) so that we estimate how the distribution of skills in denser areas is shifted, dilated, left-truncated, and right-truncated relative to the distribution of skills in less dense areas.

The second main difficulty is to obtain reasonable estimates of workers' skills. We follow the approach of Combes, Duranton, and Gobillon (2008) and use information about workers' wages over time to assess the time- and location- invariant part of their wage which reflects their (permanent) skills. We provide further details about this type of estimation below. We are aware that estimating skills must rely on some identification assumptions. An important benefit of our approach is that it could be applied to alternative measures of skills.

Third, estimating workers skills as we do using a fixed effect approach that allows for time-varying location effects is extremely data intensive since we need to observe workers across the entire distribution of wages. We also need to be able to observe them repeatedly as they move across areas. We use a large scale panel of French workers that satisfies these data requirements.

Fourth, we need to separate sorting from other sources of differences between distributions. While we provide further details about our identification strategy below, we note for now that we rely on the difference between workers who remain in the same employment area (i.e., the 'stayers') and those that move (i.e., the 'movers'). This allows us to separate differences caused by permament skills from those caused by sorting.

Our work is related to several strands of literature. First, the literature about sorting across cities documents the increasing importance of this phenomenon, at least between US cities (Berry and Glaeser, 2005). It also documents that sorting explains a large fraction of the urban wage premium (Combes *et al.*, 2008, Mion and Naticchioni, 2009, Matano and Naticchioni, 2012, Baum-Snow and Pavan, 2011). The complementarity between cities and skills (Glaeser and Resseger, 2010) should lead to the sorting of the most skilled workers in the largest cities. There is however a suspicion that it is also workers with the lowest skills that sort in the largest cities (Glaeser, Kahn, and Rappaport, 2008, Eeckhout *et al.*, 2010). Our work confirms those findings and sheds new lights on them. It also documents how sorting interacts with the life-cycle of workers and their occupational group.

Second, we also contribute to the literature on urban inequalities which has enjoyed a recent revival (Glaeser, Resseger, and Tobio, 2009, Behrens and Robert-Nicoud, 2010, Baum-Snow and Pavan, 2010). Again, we confirm some results from this literature such as greater inequalities in denser areas that cannot be entirely explained by differences in workers' skills. We also shed new lights on urban inequalities by characterising how the distributions of wages differ across areas and look at particular subgroups of interests such as migrants vs. stayers, young vs. old workers, workers in particular occupations, etc.

Third, our work is also related to the literature about migration within countries (see Greenwood, 1997, Etzo, 2008, for reviews) . A key concern in that literature is the selection of workers by skills into migration, a possible counterpart to sorting.[1] We document interesting patterns for the selection into migration such as positive selection for migrants to denser areas and negative selection for workers to less dense areas. In related work, De la Roca (2011) finds results consistent with ours using a very different approach.

The rest of this paper is organised as follows. Section 2 presents the data. Section 3 presents our methodology. Section 4 presents our results for wages and skills. Section 5 presents results for migrants and stayers and explores a variety of other issues related to the age and occupations of workers. Finally, section 6 concludes.

2. Data

Our spatial units are French employment areas ('zones d'emploi'). Continental France is fully covered by 341 employment areas, whose boundaries are defined on the basis of daily commuting patterns. Employment areas are meant to capture local labour markets and most of them correspond to a city and its catchment area or to a metropolitan area. This choice of relatively small areas (on average 1,500 km^2) is consistent with previous findings in the agglomeration literature that differences in area productivity are largely determined by agglomeration effects at small spatial scale (e.g., Rosenthal and Strange, 2008). In addition, extant literature on French data has favoured employment areas (e.g., Combes *et al.*, 2008, Combes, Duranton, Gobillon, and Roux, 2010, Combes *et al.*, 2011*b*) over alternative units such as urban areas which do not fully cover France.

We group areas by employment density, our preferred measure of local scale. Since Ciccone and Hall (1996), density-based measures have often been used to assess overall scale effects. Relative to alternatives measures of scale such as total employment or total

[1] To be clear, sorting is a statement about (differences in) the distribution of skills by location. Selection (in migration) is a statement about choices of location by workers depending on their skills. The two are obviously related but we note that selection is only a sufficient condition for sorting which can occur through other channels such as inherited skills or the local schooling system.

population, density-based measures are more robust to zoning problems. In particular, Greater Paris is divided into a number of employment areas. The true economic scale of these Parisian employment areas is much better captured by their density than any absolute measure of employment.[2]

For wages, we use an extract from the Déclarations Annuelles des Données Sociales (DADS) or Annual Social Data Declarations database from the French statistical institute (INSEE) as in Combes *et al.* (2008) and Combes *et al.* (2010).[3] The DADS are collected for pension, benefits, and tax purposes. Establishments must fill a report for each of their employees every year. An observation thus corresponds to an employee-establishment-year combination. The data are restricted to employees in manufacturing and services working in France. For years before 2002, all workers born in October of even-numbered years are sampled. From 2002 onwards, the sampling is extended to all workers born in Octobre of any year.

For each observation, we know the age, gender, location of birth (at the 'département' level), occupation at the two-digit level, and annual earnings.[4] Education is missing from the data. We restrict ourselves to years after 1993 for which we know the number of hours worked and construct the hourly wage. We keep only full-time employees for whom hours are set by law to obtain a homogenous sample. The data also contains basic establishment level information such as the three-digit sector and a municipality identifier which is used to determine the employment area.

The raw data contains 27,130,477 employee-establishment-year observations between 1993 and 2007. For reasons of computational tractability, we keep only five points in time separated by three-year intervals: 1995, 1998, 2001, 2004 and 2007, and consider only workers born in October of an even year for consistency across time. We are left with 6,299,337 observations. When restricting the sample to full-time jobs outside the public sector, the sample goes down to 3,159,817 observations. Considering jobs for which workers are aged between 15 and 65, we are left with 3,113,571 observations. Excluding jobs for which the wage or the number of hours is zero, or for which the identifier of the worker is mentioned to be misreported, leads to 2,989,444 observations. Some workers occupy several jobs during a given year. We only keep their main job, defined as the job to which are associated the highest earnings, and our sample decreases to 2,667,490 observations. After deleting jobs outside mainland France, and those with missing or misreported sector or employment area code, we are left with 2,630,436 observations. We also delete jobs in small sectors (such as spatial transport), in non-profit sectors, and in

[2]See Briant, Combes, and Lafourcade (2010) for further discussion about this issue.

[3]The preparation of the wage data follows that of these two papers closely but we use a different set of years. See Combes *et al.* (2008) for further details about these data.

[4]The département is a geographic unit such that mainland France is fully covered by 93 départements.

Figure 1: Geographic distribution of log employment density and mean wages by quartile in France, 2007

Panel (a): Mean hourly wage Panel (b): Employment density

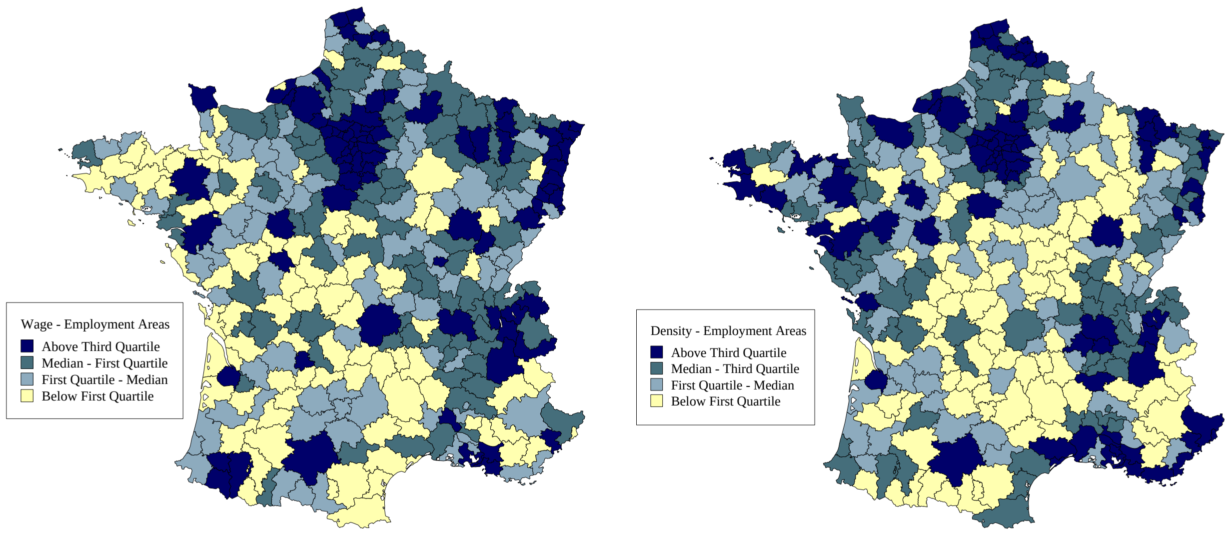


Table 1: Descriptive statistics about spatial wage disparities

	Mean	Std. dev.	1st decile	1st quartile	Median	3rd quartile	9th decile
Hourly wage	13.65	1.35	12.45	12.81	13.39	14.15	14.95
Employment density	1.36	9.50	0.10	0.17	0.28	0.53	1.21

Statistics over the means for 341 employment areas in 2007. Hourly wages in 2007 Euros. Employment density in worker per hectare.

sectors where firms are allowed to aggregate their employment report at the regional level (such as finance - see Combes *et al.*, 2008, for more details). This leads to a sample of 2,311,112 observations. Finally, we trim 0.1% observations on each side of the wage distribution by year to avoid extreme values coming from coding mistakes.[5] Our final sample includes 2,306,486 observations.

Our dataset is complemented with data on aggregate employment by employment area from the 2007 census. Using information about land area from the 1988 municipality census, we can construct the employment density for each employment area. In most of our analysis we distinguish between denser areas (with above-median employment density) and less dense areas (with below-median employment density).

Figure 1 provides maps of employment density in panel (a) and mean wages in panel (b) for French employment areas in 2007. These two maps clearly show that high wages

[5]More extreme observations are removed in the implementation of the procedure to compare distributions described below.

Table 2: Descriptive statistics about wage disparities within areas

	Mean	Std. dev.	1st decile	1st quartile	Median	3rd quartile	9th decile
All areas	15.20	9.56	9.09	10.22	12.28	16.64	24.55
Above median density	15.60	9.97	9.11	10.29	12.54	17.24	25.54
Below median density	13.10	6.70	9.00	9.99	11.34	13.99	18.71

Statistics over workers wages within groups of employment areas in 2007. Hourly wages in 2007 Euros. For example: 9.99 in row 3 and column 4 corresponds to the wage of workers at the first quartile of the wage distribution in areas with below-median density.

Table 3: Descriptive statistics about mobility across areas

	Above median density in 2007	Below median density in 2007
Panel (a): All workers		
Above median density in 2004	273,003 (97.9%)	5,954 (2.1%)
Below median density in 2004	6,504 (11.7%)	48,947 (88.3%)
Panel (b): Movers only		
Above median density in 2004	50,054 (89.4%)	5,954 (10.6%)
Below median density in 2004	6,504 (70.5%)	2,720 (29.5%)

Row percentages between brackets.

and high density are strongly associated and correspond to the largest French cities. Table 1 provides further descriptive statistics for hourly wages and mean employment density. Hourly wages are about 20% higher in employment areas in the top decile of mean wages relative to the bottom decile. Employment density in the top decile of employment areas is 12 times as high as in the bottom decile.

Table 2 provides descriptive statistics about the distribution of wages within areas. It indicates that, consistent with prior evidence (e.g., Combes *et al.*, 2008, 2010), workers in denser areas receive higher wages. It also shows that denser areas are more unequal. Interestingly, wages at the bottom decile of their distribution are only marginally higher in denser areas than in less dense areas whereas for the top decile, wages in denser areas are much higher. Finally, panels (a) and (b) of table 3 document patterns of mobility both within and between areas. Workers in denser areas move nine times as much within these areas relative to movements to less dense areas. Conversely, workers in less dense areas move less than half as much within these areas relative to movements to denser areas.

3. Methodology

Estimating workers' skills

We first explain our methodology to estimate the skills of workers in each area. Our approach is standard and further developed in Combes *et al.* (2008) and Combes *et al.* (2010). Consider a competitive representative firm in area a and year t. Its profit is

$$\pi_{at} = p_{at}\, y_{at} - \sum_{i\in(at)} w_{it}\, \ell_{it} - r_{at}\, z_{at}\,. \tag{1}$$

In this equation, p_{at} is the unit price of the output y_{at}. For any worker i employed by this firm, w_{it} is her wage rate and ℓ_{it} her labour supply. Finally, z_{at} represents the other factors of production and r_{at} their price.[6] Output produced by the representative firm is given by

$$y_{at} = A_{at}\left(\sum_{i\in(at)} s_{it}\, \ell_{it}\right)^{b} (z_{at})^{1-b}\ , \tag{2}$$

where $0 < b \leq 1$, s_{it} is the skills of worker i in year t, and A_{at} is the total factor productivity in the area. At the competitive equilibrium, profit maximisation implies

$$w_{it} = B_{a(it)t}\, s_{it}. \tag{3}$$

where $B_{at} \equiv b(1-b)^{\frac{(1-b)}{b}} \left(p_{at} A_{at} r_{at}^{1-b}\right)^{\frac{1}{b}}$. A direct implication of this equation is that wage differences across areas reflect both differences in individual skills and differences in area 'productivity'. In turn, differences in area productivity will occur because of differences in prices, other factors and input costs, and technology. Differences in prices occur, for instance, when producers in areas with better access to markets can charge higher prices relative to producers in areas with worse market access. Differences in the price of other (fixed) factors such as land can occur because of differences in demand for these other factors for reasons unrelated to production. For instance, areas with better consumption amenities will face a stronger demand for land and, as a result, higher land prices. Finally, differences in technology can reflect localised natural advantages or local interactions across workers. These latter interactions are usually referred to as agglomeration economies and can occur through a variety of channels and mechanisms (Duranton and Puga, 2004).

To take equation (3) to the data, both skills (s_{it}) and area productivity (B_{at}) need to be specified. The model above ignores sectoral effects to keep notations simple. Our

[6]Our model is not indexed by sector but sectoral effects are taken into account in our empirical specification below.

empirical analysis accounts for them in a simple way. We assume that productivity for sector, area, and year can be expressed as

$$\log B_{at} = \beta_{at} + \mu_{kt}, \tag{4}$$

where β_{at} is an area and year fixed effect and μ_{kt} is a sector and year fixed effect. Much of previous literature is concerned with disentangling the various determinants of area productivity. For instance, Combes *et al.* (2010) regress β_{at} after averaging it over time on a set of local characteristics including employment density, market access, and various measures of localised natural advantage. As already argued, our preferred measure of scale is the local density of employment. Assessing whether this measure of agglomeration has a causal effect on area productivity is a fundamental question and an important theme in the literature. We do not press on this issue here. We separate between denser areas and less dense areas and compare the distribution of skills in both types of areas.[7] The exact nature of productivity differences between areas does not matter to us here.

Turning to skills, we assume that for worker i

$$\log s_{it} = X_{it}\varphi + \delta_i + \epsilon_{it}\,, \tag{5}$$

where X_{it} is a vector of time-varying worker characteristics, δ_i is a worker fixed effect, and ϵ_{it} is a measurement error. The errors are assumed to be i.i.d. across periods and workers.

Combining equations (3), (4), and (5) yields the following inverse demand for labour equation

$$\log w_{it} = \beta_{a(it)t} + \mu_{k(it)t} + X_{it}\varphi + \delta_i + \epsilon_{it}\,. \tag{6}$$

That is, we estimate the wages of workers as a function of their observed and unobserved characteristics (age and its square plus a worker fixed effect), the type of area in which they are employed (area-year fixed effects), and their sector and time. We interpret worker effects as unobserved skills.

Being able to separately estimate individual, sector, and area effects requires multiple observations for each worker and enough mobility between sectors and areas. The data we use easily meet this requirement. The identification of sector and year and area and year effects also requires the normalisation of the effect of one area and that of one sector for one year. In addition, the separate identification of personal skills and area productivity relies on four assumptions. First we assume that workers differ in a single dimension for skills. In a series of papers, Bacolod, Blum, and Strange (2009*a*, 2009*b*, and 2010) provide evidence that various dimensions of skills interact differently with urban

[7]Using similar data, Combes *et al.* (2010) provide evidence that employment density has a causal effect on area productivity. Broader evidence about the causal effect of agglomeration is discussed in Combes, Duranton, and Gobillon (2011*a*).

location. In absence of further information about workers, we estimate only a single cumulative effect of skills. Although sorting in cities may be strongly determined by specific elements of skills, we are unable to explore this question further and leave it for future work.

Second, our estimation strategy for skills relies on area productivity affecting worker productivity independently of skills. However, there is evidence of stronger complementarity between worker skills and area productivity than postulated above. Wheeler (2001) and Glaeser and Resseger (2010) find stronger agglomeration effects for more educated workers and Bacolod *et al.* (2009*a*) report a similar finding for individuals with better cognitive and people skills. We need to keep this important issue in mind when comparing the distribution of estimated skills across areas. These empirical findings also suggest that it is more appropriate to compare workers within areas. For this reason, we focus much of our analysis on the comparison between 'movers' and 'stayers' within each area. More precisely and to preview our identification strategy, we can safely compare movers and stayers across quantiles of the estimated distribution of skills when the effect of agglomeration on skills is monotone.

Third, we assume that worker effects are fixed over time. That is, we do not allow area productivity to affect worker skills. The idea that workers might learn more in more productive areas was suggested a long time ago by Alfred Marshall (1890). It has received theoretical attention more recently (Glaeser, 1999, Duranton and Puga, 2004). Recent empirical research is suggestive that such learning effects may be at work (Glaeser and Maré, 2001, De La Roca and Puga, 2011). When estimating equation (6) we only estimate an 'average' worker skill. We also keep this issue in mind when interpreting our results.

Our last important assumption is to treat migration between areas as exogenous. This assumption is less restrictive than it seems since the specification in equation (6) contains area and year effects. Having workers move to denser areas knowing that wages are particularly high there on that year is not a problem since it is captured by β_{at} and is uncorrelated with the error term. Still some problems might arise if, for instance, workers with particularly high (or low) skills move more in response to good job offers. This would create a correlation in equation (6) between (unobserved) skills and the error term. Although this selection bias could be at play, there are reasons why it is likely to be much attenuated. First, we expect location decisions to be driven by factors unrelated to wages such as idiosyncratic preferences. Second, France has high barriers to internal mobility so that we also expect migration to be driven by long-term wages not short term wage shocks.[8] This argument is consistent with many of our findings below.

[8] There is no bias since workers migrate on the basis of future expected wages rather than the exact wage they can get today.

This said, the rest of our approach does not depend on the specifics of the estimation of worker effects. It could be applied to alternative and perhaps more sophisticated estimates of worker effects. While not devoid of potential problems, as just discussed, our approach to the estimation of worker effects offers two benefits. First, it is simple and transparent. Second, it is similar to the approach we have taken in previous work which eases comparisons.

Comparing distribution of skills and wages across areas

We now turn to the second part of our methodology which compares the distributions of worker effects and wages between two groups of areas denoted $i \in \{1,2\}$. More precisely, our approach assesses to what extent the distribution of worker effects or that of wages in one group is a truncated, shifted, and dilated relative to the same distribution in the other group. It extends the approach of Combes *et al.* (2011*b*) to allow for both left and right truncations instead of only left truncations. We present the approach referring to wages, but its implementation with worker effects is identical.

Consider first a situation where distribution 2 is left- and right-truncated relative to distribution 1. We denote $\underline{S}$ the rank at which left truncation occurs and $1-\overline{S}$ the rank at which right truncation occurs. We can compare the two distributions by comparing quantiles of distribution 1 at ranks in $\left[\underline{S}, 1-\overline{S}\right]$ with quantiles of distribution 2 at ranks in [0,1]. A restriction on the range of ranks for distribution 1 occurs because ranks outside $\left[\underline{S}, 1-\overline{S}\right]$ do not have any counterpart in distribution 2 which is truncated. Put differently the wage at rank u in distribution 2 is compared to the wage at rank $\underline{S} + \left(1-\overline{S}-\underline{S}\right) u$ in distribution 1. We can interpret $\underline{S}$ as a measure of left truncation and $\overline{S}$ as a measure of right truncation.

Any wage w in group 1 not discarded when truncating the distribution is transformed linearly into a wage $Dw + A$ in group 2, where A is a shift parameter and D is a dilation parameter. Parameter D can be above or below one, depending on whether there is a dilation or a compression, and parameter A can be positive or negative.

More formally, let $\lambda_i(u)$ be the quantile of distribution i at rank u. The transformations of ranks and values described above yield the following relationship between the two distributions:

$$\lambda_2(u) = D\lambda_1\left(\underline{S} + (1-\overline{S}-\underline{S})u\right) + A, \quad \text{for } u \in [0,1] \tag{7}$$

Now consider a situation where distribution 1 is a left- and right-truncated relative to distribution 2. We denote $\underline{T}$ the rank at which left truncation occurs and $1-\overline{T}$ the rank at which right truncation occurs. By the same argument as previously, the relationship between the two distributions is:

$$\lambda_2\left(\underline{T} + (1-\overline{T}-\underline{T})u\right) = D\lambda_1(u) + A, \quad \text{for } u \in [0,1] \tag{8}$$

This equation can be rewritten in a form similar to (7). We make the change of variable $u \to \frac{u-\underline{T}}{1-\overline{T}-\underline{T}}$ in equation (8) which becomes:

$$\lambda_2(u) = D\lambda_1\left(\frac{u-\underline{T}}{1-\overline{T}-\underline{T}}\right) + A, \quad \text{for } u \in \left[\underline{T}, 1-\overline{T}\right] \tag{9}$$

Introducing $\underline{S} = \frac{-\underline{T}}{1-\overline{T}-\underline{T}}$ and $\overline{S} = \frac{\overline{T}}{1-\overline{T}-\underline{T}}$, equation (9) can be rewritten as:

$$\lambda_2(u) = D\lambda_1\left(\underline{S} + (1-\overline{S}-\underline{S})u\right) + A, \quad \text{for } u \in \left[\frac{-\underline{S}}{1-\overline{S}-\underline{S}}, \frac{1-\underline{S}}{1-\overline{S}-\underline{S}}\right] \tag{10}$$

It is possible to nest the two cases of left and right-truncation (7) and (10) into one equation:

$$\lambda_2(u) = D\lambda_1\left(\underline{S} + (1-\overline{S}-\underline{S})u\right) + A, \quad \text{for } u \in \left[\max\left(0, \frac{-\underline{S}}{1-\overline{S}-\underline{S}}\right), \min\left(1, \frac{1-\underline{S}}{1-\overline{S}-\underline{S}}\right)\right] \tag{11}$$

This equation also nests two other cases since it is also satisfied when distribution 1 is left-truncated and distribution 2 is right-truncated, as well as when distribution 2 is left-truncated and distribution 1 is right-truncated.

When $\underline{S} > 0$, the distribution of group 2 is left-truncated relative to the distribution of group 1. Conversely, when $\underline{S} < 0$, the distribution of group 1 is left-truncated relative to the distribution of group 2. At the other end of the distribution when $\overline{S} > 0$, the distribution of group 2 is right-truncated relative to the distribution of group 1 whereas when $\overline{S} < 0$, the distribution of group 1 is right-truncated relative to the distribution of group 2.

Equation (11) cannot be used directly in the estimation as the range of ranks $\left[\max\left(0, \frac{-\underline{S}}{1-\overline{S}-\underline{S}}\right), \min\left(1, \frac{1-\underline{S}}{1-\overline{S}-\underline{S}}\right)\right]$ depends on the two parameters $\underline{S}$ and $\overline{S}$ which are unknown. We thus make the additional change of variable $u \to r_S(u)$, where $r_S(u) = \max\left(0, \frac{-\underline{S}}{1-\overline{S}-\underline{S}}\right) + \left[\min\left(1, \frac{1-\underline{S}}{1-\overline{S}-\underline{S}}\right) - \max\left(0, \frac{-\underline{S}}{1-\overline{S}-\underline{S}}\right)\right] u$ to obtain the specification that we bring to the data:

$$\lambda_2(r_S(u)) = D\lambda_1\left(\underline{S} + (1-\overline{S}-\underline{S})r_S(u)\right) + A\ , \quad \text{for } u \in [0,1] \tag{12}$$

The estimation procedure is based on this equation and matches that of Combes *et al.* (2011*b*) with an additional parameter $\overline{S}$ that we also estimate here. Denote $\theta = \left(A, D, \underline{S}, \overline{S}\right)$ the set of parameters to be estimated. Equation (12) can be rewritten as a continuum of equalities of the form $m_\theta(u) = 0$ for $u \in [0,1]$ where:

$$m_\theta(u) = \lambda_2(r_S(u)) - D\lambda_1\left(\underline{S} + (1-\overline{S}-\underline{S})r_S(u)\right) - A \tag{13}$$

It is possible to estimate θ minimising the squared empirical counterparts of $m_\theta(u)$, $u \in [0,1]$ obtained when replacing the quantile functions by some estimators. However distributions 1 and 2 play an asymmetric role in the continuum of equalities. Thus, we also use a continuum of equalities obtained when switching groups 1 and 2. Making the change of variable $u \to \frac{u-\underline{S}}{1-\overline{S}-\underline{S}}$ in (11), we get:

$$\lambda_1(u) = \frac{1}{D}\lambda_2\left(\frac{u-\underline{S}}{1-\overline{S}-\underline{S}}\right) - \frac{A}{D}, \quad \text{for } u \in \left[\max\left(0,1-\underline{S}\right), \min\left(1-\overline{S},1\right)\right] \tag{14}$$

Introducing $\widetilde{r}(u) = \max(0,1-\underline{S}) + \left[\min\left(1-\overline{S},1\right) - \max\left(0,\underline{S}\right)\right] u$ and making the final change of variable $u \to \widetilde{r}(u)$ yield the new set of equalities $\widetilde{m}_\theta(u) = 0$, where:

$$\widetilde{m}_\theta(u) = \lambda_1(\widetilde{r}(u)) - \frac{1}{D}\lambda_2\left(\frac{\widetilde{r}(u)-\underline{S}}{1-\overline{S}-\underline{S}}\right) + \frac{A}{D} \tag{15}$$

Denote $\widehat{m}_\theta(u)$ the empirical counterpart of $m_\theta(u)$ and $\widehat{\widetilde{m}}_\theta(u)$ the empirical counterpart of $\widetilde{m}_\theta(u)$ obtained when replacing the quantile functions $\lambda_i(u), i \in \{1,2\}$ by their sample estimator. The estimator of θ is obtained minimizing the sum of squared values of both $\widehat{m}_\theta(u)$ and $\widehat{\widetilde{m}}_\theta(u)$:

$$\widehat{\theta} = \arg\min_\theta M(\theta), \quad \text{where} \quad M(\theta) = \int_0^1 \left[\widehat{m}_\theta(u)\right]^2 \mathrm{d}u + \int_0^1 \left[\widehat{\widetilde{m}}_\theta(u)\right]^2 \mathrm{d}u \tag{16}$$

The confidence intervals are obtained by bootstrap. The implementation details are given in Combes *et al.* (2011*b*). Finally, we construct a measure of fit that uses the ratio of mean-squared quantile differences between distributions 1 and 2 with mean-squared quantile differences between the transformed distribution 1 and distribution 2.

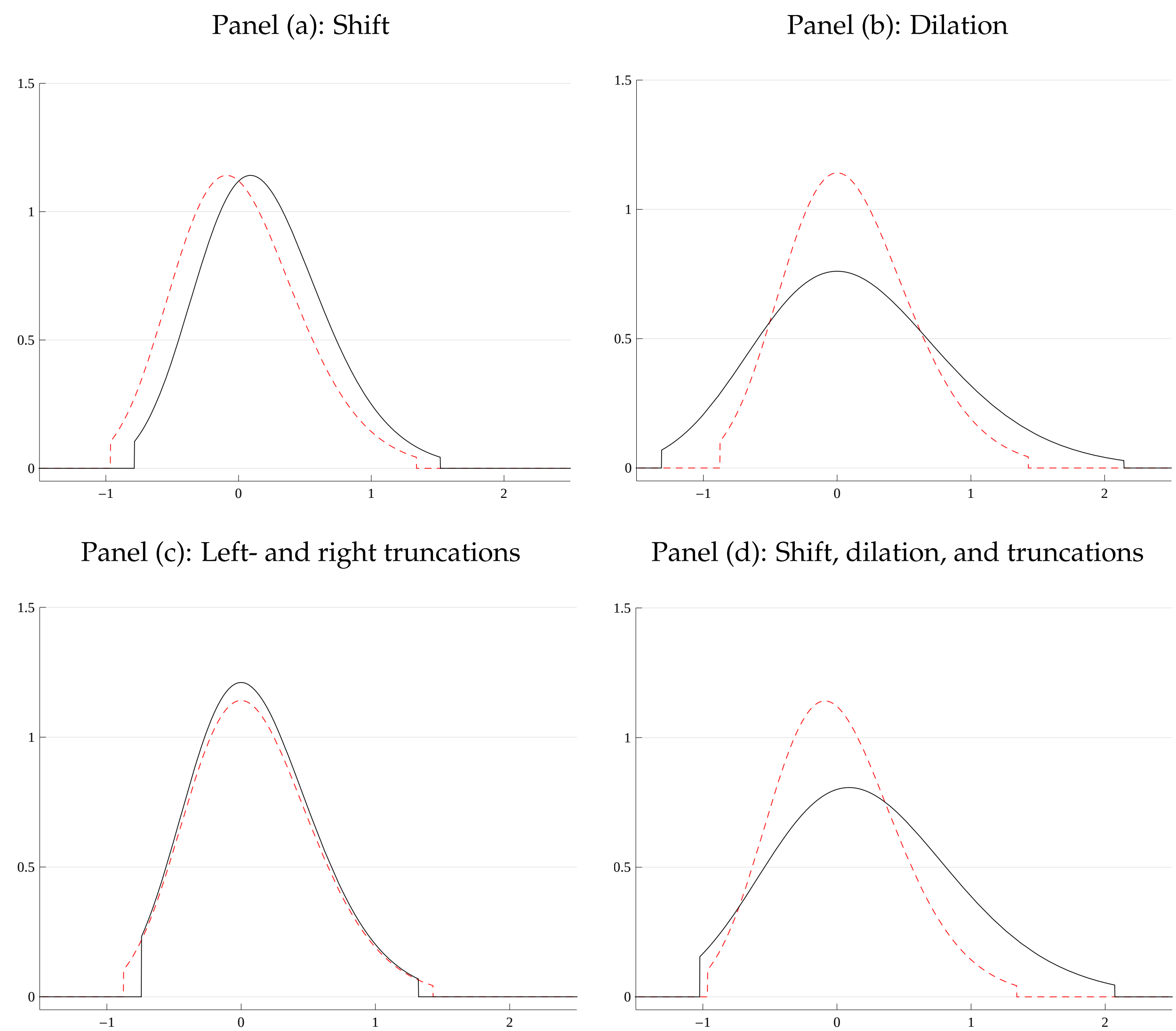
To develop intuitions about our approach, panels (a) to (d) of figure 2 plot the four transformations discussed above. In all panels, the dashed curve represents an hypothetical distribution of log wages in less dense areas and the plain curve represents another hypothetical distribution of log wages in denser areas. For the sake of representation, we assume a Fisher distribution for wages in less dense areas. Our approach allows us to estimate how the distribution of log wages in less dense areas (which we use as reference) can be transformed to approximate the distribution of log wages in denser areas.

Panel (a) considers the case of a simple positive shift ($A > 0$, $D = \underline{S} = \overline{S} = 0$). The two distributions of wages have thus the same shape. The distribution of wages in denser areas is to the right of that in less dense areas. As visual cue, we note that the two peaks of density have the same height.

Panel (b) considers the case of a dilation ($D > 0$, $A = \underline{S} = \overline{S} = 0$). We can see that the distribution of wages in denser areas spreads further to the left and to the right relative to the distribution of wages in less dense areas. A second visual cue is given by the lower peak of the distribution of wages in denser areas.[9]

[9]We also note that the two peaks are not exactly aligned. This is because the skewness of the Fisher implies that the mode of the distribution is below the mean and gets dilated just like the rest of the distribution.

Figure 2: Four possible transformations of the distribution of log wages



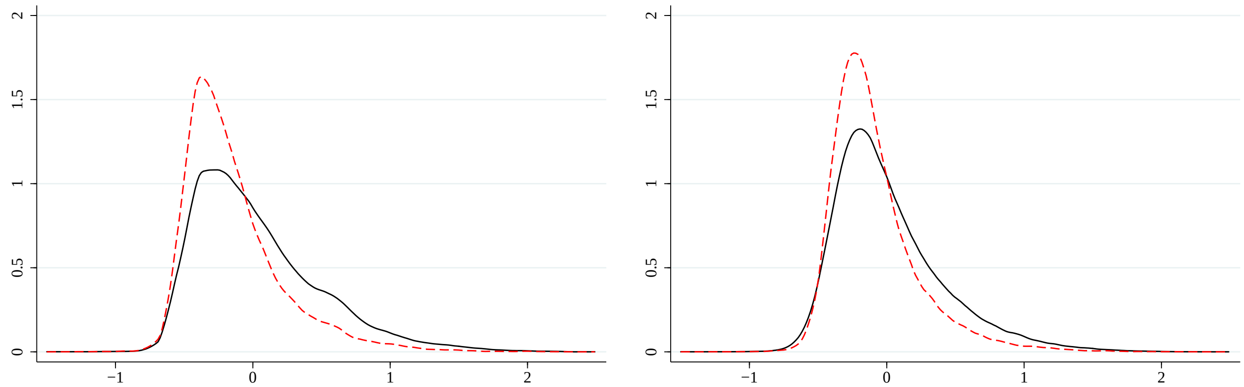
Continuous black line: hypothetical denser areas.
Dashed grey (red) line: hypothetical less dense areas.

Panel (c) considers the case of both a left- and a right-truncation ($\underline{S} > 0$, $\overline{S} > 0$, $A = D = 0$). Visually, the distribution of wages in denser areas has shorter tails. If left-truncation is greater than right truncation ($\underline{S} > \overline{S}$) as in the example we plot, this creates some asymmetry between the two distributions. Truncating the tails also leads to a higher peak. The case of 'negative' truncation corresponds to a situation where the distribution in less dense areas would be more truncated than the distribution in denser areas.

Finally panel (d) considers a situation where the distribution of wages in denser areas is righted-shifted, dilated, and left- and right-truncated relative to the same distribution in less dense areas. An important thing to note is that dilation and truncation work

Figure 3: Wages and worker effects in denser vs. less dense areas

Panel (a): Log wages | Panel (b): Worker effects



Continuous black line: denser areas.
Dashed grey (red) line: less dense areas.

partly in opposite directions. While truncation raises the peak of the distribution and reduces its spread by cutting the tails, dilation lowers the peak of the distribution and increases its spread. However, these transformations do not fully offset each other. To speak loosely, if a tail of the distribution expands less than its body, this will be captured by both dilation and truncation. If it expands more, it will be captured by both dilation and negative truncation. Hence, when considering these transformations together we must keep in mind that the estimated coefficients need to be interpreted in conjunction with each other. We also note that, visually, it is difficult to detect dilation and truncation separately. Estimation results are needed.

4. First results

Panel (a) of figure 3 plots the distributions of log wages in denser and less dense areas in 2007.[10] Table 4 reports estimation results regarding how these two distributions relate to each other. The first row of this table only allows for a shift in the distribution of log wages in less dense areas to approximate the same distribution in denser areas. The estimated shift coefficient is 0.126 indicating that wages are 13.4% higher ($= \exp(0.126) - 1$) in denser areas relative to less dense areas. An average employee in a French employment area with above-median density benefits from a density that is 2.8 log points higher than an average employee in an area with below-median density. This difference implies that this estimate of 0.126 for the shift in log wage is equivalent to an (arc) elasticity of wages

[10]Unless specified otherwise, all figures are for 2007. In addition all the figures are centred around the sample mean of all areas for the relevant groups of workers (all the workers in 2007 here but the distributions for more specific groups are analysed below).

Table 4: Log wages in denser vs. less dense areas

	$\hat{A}$	$\hat{D}$	$\hat{\underline{S}}$	$\hat{\overline{S}}$	R^2
w_L to w_H (shift only)	0.126 (0.001) *	-	-	-	0.609
w_L to w_H (shift and dilation)	0.127 (0.001) *	1.375 (0.007) *	-	-	0.982
w_L to w_H (shift and left- and right-truncations)	0.113 (0.005) *	-	−0.124 (0.010) *	−0.084 (0.002) *	0.935
w_L to w_H (shift, dilation, and left-truncation)	0.125 (0.001) *	1.379 (0.007) *	0.003 (0.001) *	-	0.983
w_L to w_H (shift, dilation, left- and right-truncations)	0.136 (0.002) *	1.631 (0.012) *	0.047 (0.002) *	0.029 (0.001) *	0.999

*: for $\hat{A}$, $\hat{\underline{S}}$, and $\hat{\overline{S}}$ significantly different from 0 at 5%; for $\hat{D}$ significantly different from 1 at 5%.
w_L: distribution of log wages in less dense areas.
w_H: distribution of log wages in denser areas.
85,812 observations in less dense areas. 443,584 observations in denser areas.

with respect to employment density of $0.126/2.8 = 0.045$. Using different years of the same data and a standard regression framework, Combes *et al.* (2008) report a comparable number of 0.049 for the unconditional elasticity wages with respect to employment density. We also note that the shift parameter alone only explains about 60% of the mean-squared quantile difference between these two wage distributions. Put differently, there is more than a difference in the first moment to explain the differences between the distributions of wages in denser and less dense areas.

The second row of table 4 adds a dilation parameter to the estimation. We obtain an estimate of 1.375 for this coefficient and the shift coefficient is essentially unchanged. That the distribution of log wages in denser areas is dilated relative to less dense areas can be readily seen from the plot of panel (a) of figure 3 where the distribution of log wages in denser areas has a lower peak and spreads more relative to that in less dense areas. From the last column of the same row, it is clear that this additional coefficient leads to a much better fit since 98% of the mean-squared quantile difference between the two distributions can now be accounted for. This estimate of 1.375 for dilation is large. Together with the shift parameter of 0.127, it suggests that a worker at the first decile of the wage distribution in denser areas is predicted to make 1.4% more than a worker at the first decile of the wage distribution in less dense areas. For workers at the first quartile of both distributions, the predicted difference is larger at 5.5%. For workers at the median, the predicted difference is again larger at 10.6%. For workers at the third quartile, the predicted difference is even larger at 23.2%. Finally, for workers at the top decile, the predicted difference is 36.5%. In the data, these difference are 1.2%, 2.9%, 10.5%, 19.7%, and 33.5%, respectively. This suggests that the predicted wage distribution in denser areas matches its empirical counterpart rather well but tends to slightly over-predict wage differences.

To improve on these results and reach our preferred estimation, we proceed in stages. In row 3 of table 3 we start with estimating a specification with a shift parameter together with left- and right- truncation parameters. We find strong negative truncation for the distribution of log wages in less dense areas to approximate that in denser areas. More intuitively, there are far fewer workers with very high or very low wages at both tails of the distribution in less dense areas relative to denser areas. This can be captured by (positive) truncation at both tails of the distribution of log wages in denser areas to approximate the same distribution in less dense areas. Conversely, if we use distribution of log wages in area with below-median density as reference, that transformation becomes a negative truncation at both tails. It is also important to note that the same difference is approximated by a dilation in row 2.

In row 4 of table 3, we estimate a left-truncation coefficient jointly with a shift and a dilation coefficient. Interestingly, this estimation corresponds to the main estimation of Combes *et al.* (2011*b*) but compares distributions of log wages instead of distributions of log total factor productivity. For all sectors combined, Combes *et al.* (2011*b*) find a shift parameter of 0.091, a dilation parameter of 1.227, and an insignificant left-truncation parameter of 0.001 for total factor productivity. While we find stronger shift and dilation for wages here, we note that these two sets of results are of similar magnitude.

Finally in row 5, we jointly estimate both left- and right- truncation parameters jointly with a shift and a dilation parameter. Allowing for these four parameters enables us to approximate the distribution of log wages in denser areas from the distribution of log wages in less dense areas nearly perfectly. Interestingly, our estimation results indicate that the distribution of log wages in denser areas is both left- and right-truncated relative to less dense areas. This may seem contradictory with the results of row 3. However the truncation coefficients must be understood in relation to a large dilation coefficient. That the distribution of log wages in less dense areas must be both truncated and dilated to approximate that of wages in denser areas is telling us that dilation is less important in the tail of the distribution. Overall, despite positive truncation, the dilation coefficient 'dominates' and there is over-representation of workers at both tails of the distribution of log wages in denser areas once the shift between the two distributions is taken into account. This can be observed directly from the plot of panel (a) of figure 3. In addition, this interpretation is consistent with the fact that dilation is even more important when we allow for truncation.

Panel (b) of figure 3 plots the distributions of worker effects in denser and less dense areas. Table 5 mirrors table 4 for these two distributions of worker effects. The first row, which only allows for a shift in the distribution of worker effects in less dense areas to

Table 5: Worker effects in denser vs. less dense employment areas

	$\hat{A}$	$\hat{D}$	$\hat{\underline{S}}$	$\hat{\overline{S}}$	R^2
δ_L to δ_H (shift only)	0.051 (0.001) *	-	-	-	0.351
δ_L to δ_H (shift and dilation)	0.052 (0.001) *	1.304 (0.005) *	-	-	0.977
δ_L to δ_H (shift and left- and right-truncations)	0.035 (0.005) *	-	-0.050 (0.029)	-0.056 (0.026) *	0.942
δ_L to δ_H (shift, dilation, and left-truncation)	0.040 (0.002) *	1.338 (0.008) *	0.019 (0.003) *	-	0.982
δ_L to δ_H (shift, dilation, left- and right-truncations)	0.040 (0.002) *	1.340 (0.020) *	0.020 (0.004) *	-0.001 (0.002)	0.982

*: for $\hat{A}$, $\hat{\underline{S}}$, and $\hat{\overline{S}}$ significantly different from 0 at 5%; for $\hat{D}$ significantly different from 1 at 5%.
δ_L: distribution of worker effects in less dense areas.
δ_H: distribution of worker effects in denser areas.
85,812 observations in less dense areas. 443,584 observations in denser areas.

approximate the same distribution in denser areas reports a coefficient of 0.051. This corresponds to an (arc) elasticity of worker effects with respect to density of 0.018. Subject to the caveats given above about the estimation of worker effects, this indicates that the unobserved skills of workers are on average higher in denser areas. The difference between the elasticity of wages with respect to density of 0.045 estimated above and this number is 0.027 and may be interpreted as the elasticity of wages with respect to density after conditioning out individual effects. Combes *et al.* (2010) provide a direct estimate of the same elasticity that is closely comparable to this one and equal to 0.033. Put slightly differently, this result indicates that slightly less than half of the differences in wages across areas can be accounted for by differences in worker effects. We note nonetheless that this shift parameter only accounts for 35% of the mean-squared quantile difference between distributions of worker effects in denser vs. less dense areas.

The addition of a dilation parameter in row 2 improves the fit of the specification even more than for wages. As with wages, there is considerable dilation of the distribution of worker effects in less dense areas to approximate the corresponding distribution in denser areas. Like with wages again, adding truncation parameters instead of dilation in row 3 indicates significant negative truncation. Negative truncation and dilation both capture the over-representation of workers with particularly high or low effects in denser areas. However, the dilation parameter introduced in row 2 offers a better fit with the data than the two truncation parameters introduced in row 3. Considering truncation parameters and dilation jointly with a shift in rows 4 and 5 leaves the results of row 2 mostly unchanged. In row 6, we end up with a shift parameter of 0.040 that is about a third of the corresponding coefficient for wages, a dilation parameter of 1.34 that is lower than for wages, modest left-truncation and no right-truncation.

We draw a number of conclusions from these results. First, looking only at the first moment of the distribution of log wages or worker effects – the main focus of prior literature – hides considerable heterogeneity. We nonetheless confirm two key findings from this literature. Both wages and worker effects are higher in denser areas.[11] Even a focus on the second moment of the distribution like that of the recent literature on urban inequalities is not enough. This literature (e.g. Glaeser *et al.*, 2009, Behrens and Robert-Nicoud, 2010, Baum-Snow and Pavan, 2010) shows that wage dispersion is stronger in larger cities. We confirm this for French employment areas grouped by density but also show that this greater dispersion occurs everywhere in the distribution of log wages and arises because of the over-representation of workers with very high wages and, to a lesser extent, the over-representation of workers with very low wages in denser areas. Interestingly, the dilation of the wage distribution in denser areas is also greater in the middle of the distribution than in the tails. Consistent with the conclusions of Glaeser *et al.* (2009) we also find that the greater dispersion of wages in denser areas can be partly accounted for by the greater dispersion of worker effects in those areas. However the dilation coefficient is lower for worker effects than for wages. This is consistent with the notion that higher density also generates greater wage disparities even after conditioning out worker effects.

5. Further results

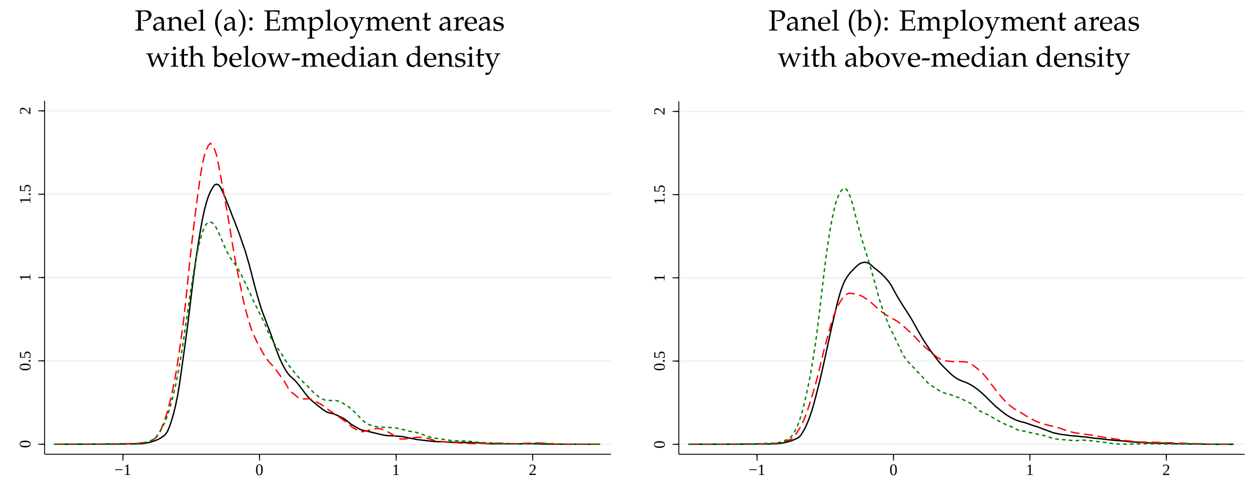
Migrants and stayers, 2004–2007

After removing differences across areas that apply to all workers as we do above, there are three reasons why the distributions of worker effects may differ across areas. The first is sorting in the traditional sense with workers actively going to different locations depending on their skills in a situation where worker effects are properly estimated. Such a movement might be due to workers with different skills having different preferences for locations or facing different costs of living. The second reason for having different distributions of worker effects is that the distribution of 'innate' skills may differ across areas.[12] We can call this 'sorting at birth'. The third is that agglomeration may have heterogeneous effects on workers' wages depending on their skills. This is still a case of sorting in the traditional sense but the fixed effects estimated above will no longer accurately reflect the

[11] Even though the empirical literature broadly agrees on the fact that higher density causes higher wages (Combes *et al.*, 2011*a*), again, we refrain from making any causal statement here.

[12] For these differences to persist, we must (plausibly) appeal to imperfect mobility. A possible story may emphasise that historically some areas had a disproportionate share of high skill workers caused for instance by earlier development and investments in education. If skills can be transmitted across generations, the skill distribution of 'native' workers may still differ across areas in 2007.

Figure 4: Log wages of stayers vs. migrants in denser and less dense areas



Continuous black line: stayers within the same group of areas.
Long-dashed grey (red) line: migrants within the same groups of areas.
Short-dashed grey (green) line: migrants to the other groups of areas.

distribution of skills but instead will reflect a mixture of skills and agglomeration effects. For instance, if more skilled workers benefit more from agglomeration, the fixed effects of these workers overestimate their true (unobserved) skills.

To go deeper into the analysis, we now focus on movers and stayers within denser and less dense areas. Within an area, movers and stayers face similar agglomeration effects within their quantile. In addition, if we also assume that heterogenous agglomeration effects are monotone in skills, we can at least identify which are the quantiles of the distribution which are affected by sorting and to which extent. This holds despite workers effects being possibly inappropriately estimated. That the transformation of worker productivity caused by agglomeration should be monotone in skills is a reasonable assumption to make given prior results that there may be a complementarity between cities and skills where more skilled workers benefit more from being in denser areas (Wheeler, 2001, Bacolod *et al.*, 2009*a*, Glaeser and Resseger, 2010).

To identify stayers and movers we compare the employment area of work in 2004 with that in 2007. Panel (a) of figure 4 plots the distribution of log wages of stayers in less dense areas, that of movers between less dense areas, and that of movers from a less dense area to a denser area. Panel (b) of figure 4 plots the corresponding distributions for workers initially employed in denser areas. Table 6 reports some estimation results comparing these distributions.

The first row of table 6 assesses how the distribution of the log wages of migrants between less dense areas is best approximated by shifting, dilating, and truncating the

Table 6: Log wages of stayers vs. migrants in denser and less dense areas

	$\hat{A}$	$\hat{D}$	$\underline{\hat{S}}$	$\hat{\hat{S}}$	R^2	Obs. 1	Obs. 2
w_{LL} to $w_{L\to L}$	-0.145	0.619	-0.098	-0.036	0.999	49,829	9,314
	(0.003) *	(0.023) *	(0.014) *	(0.007) *			
w_{LL} to $w_{L\to H}$	-0.073	0.966	-0.028	-0.019	0.984	49,829	27,333
	(0.004) *	(0.017) *	(0.002) *	(0.004) *			
$w_{L\to L}$ to $w_{L\to H}$	0.078	1.470	0.008	0.010	0.980	9,314	27,333
	(0.005) *	(0.044) *	(0.003) *	(0.004) *			
w_{HH} to $w_{H\to H}$	-0.174	0.770	-0.024	-0.048	0.991	250,966	165,285
	(0.003) *	(0.009) *	(0.002) *	(0.004) *			
w_{HH} to $w_{H\to L}$	-0.276	0.428	-0.129	-0.070	0.999	250,966	26,669
	(0.002) *	(0.009) *	(0.012) *	(0.005) *			
$w_{H\to H}$ to $w_{H\to L}$	-0.123	0.515	-0.090	-0.040	0.989	165,285	26,669
	(0.004) *	(0.017) *	(0.018) *	(0.005) *			
$w_{H\to L}$ to $w_{L\to H}$	0.072	1.395	0.002	0.012	0.982	26,669	27,333
	(0.004) *	(0.032) *	(0.001)	(0.004) *			

*: for $\hat{A}$, $\underline{\hat{S}}$, and $\hat{\hat{S}}$ significantly different from 0 at 5%; for $\hat{D}$ significantly different from 1 at 5%.
w_{ii}: distribution of log wages of stayers in less dense areas ($i = L$) or denser areas ($i = H$).
$w_{i\to j}$: distribution of log wages of migrants from an area of type i ($\in \{L,H\}$) to an area of type j ($\in \{L,H\}$).
Obs. 1 and Obs. 2: number of observations in the first (reference) group and the second.

distribution of the log wages of stayers in less dense areas. The results indicate that migrants receive on average lower wages than stayers. The distribution of the wages of these migrants is also more compressed although negative truncation indicates that this compression is less pronounced at both tails. The second row of table 6 compares the distribution of log wages of the same stayers to that of migrants to denser areas. The results go in the same direction as those of the first row but the coefficients are smaller. In row 3, the comparison of the distribution of the log wages of migrants to denser areas with the same distribution for migrants within less dense areas reveals a higher mean wage for the former and a more dilated distribution of their log wages. This is consistent with the results of the first two rows.

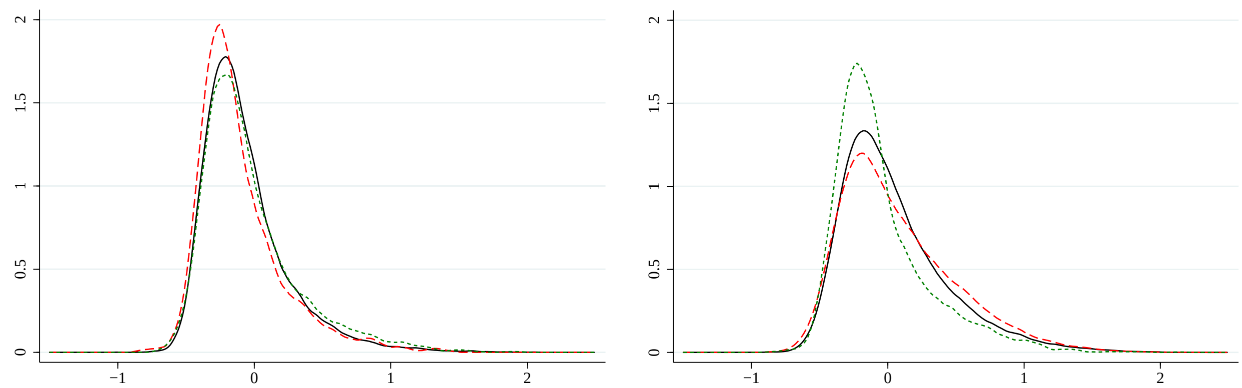
Rows 4 to 6 perform a similar exercise for the distributions of the log wages of workers initially located in denser areas. The results reveal that migrants have their wages drawn from a more compressed distribution with a lower mean than that of stayers. The wage difference with stayers is lower for migrants remaining in a denser area than for those going to a less dense area. Again the compression is less important at both tails.

Comparing rows 6 with row 3 reveals that movers within denser areas have a higher wage than movers from denser to less dense areas, while movers within less dense areas have a lower wage than movers from less dense to denser areas. Consistent with this result, row 7 compares the distribution of log wages of workers that migrated towards a denser area with that of workers that migrated towards a less dense area and shows that

Figure 5: Worker effects of stayers vs. migrants in denser and less dense areas

Panel (a): Employment areas with below-median density

Panel (b): Employment areas with above-median density



Continuous black line: stayers within the same group of areas.
Long-dashed grey (red) line: migrants within the same groups of areas.
Short-dashed grey (green) line: migrants to the other groups of areas.

Table 7: Worker effects of stayers vs. migrants in denser and less dense areas

	$\hat{A}$	$\hat{D}$	$\underline{\hat{S}}$	$\hat{\overline{S}}$	R^2	Obs. 1	Obs. 2
δ_{LL} to $\delta_{L\to L}$	0.035 (0.011) *	0.701 (0.087) *	−0.275 (0.086) *	−0.001 (0.015)	0.984	49,829	9,314
δ_{LL} to $\delta_{L\to H}$	0.021 (0.004) *	0.908 (0.026) *	−0.135 (0.017) *	−0.007 (0.003) *	0.990	49,829	27,333
$\delta_{L\to L}$ to $\delta_{L\to H}$	−0.003 (0.003)	1.120 (0.022) *	0.004 (0.002)	−0.020 (0.005) *	0.978	9,314	27,333
δ_{HH} to $\delta_{H\to H}$	−0.059 (0.002) *	0.854 (0.016) *	−0.078 (0.011) *	−0.005 (0.003)	0.993	250,966	165,285
δ_{HH} to $\delta_{H\to L}$	−0.037 (0.016) *	0.536 (0.078) *	−0.340 (0.121) *	−0.016 (0.018)	0.991	250,966	26,669
$\delta_{H\to H}$ to $\delta_{H\to L}$	−0.010 (0.003) *	0.821 (0.017) *	−0.006 (0.003) *	0.020 (0.005) *	0.976	165,285	26,669
$\delta_{H\to L}$ to $\delta_{L\to H}$	0.013 (0.002) *	1.108 (0.016) *	−0.004 (0.003)	−0.007 (0.003) *	0.980	26,669	27,333

*: for $\hat{A}$, $\underline{\hat{S}}$, and $\hat{\overline{S}}$ significantly different from 0 at 5%; for $\hat{D}$ significantly different from 1 at 5%.
δ_{ii}: distribution of worker effects for stayers in less dense areas ($i = L$) or denser areas ($i = H$).
$\delta_{i\to j}$: distribution of worker effects for migrants from an area of type i ($\in \{L,H\}$) to an area of type j ($\in \{L,H\}$).
Obs. 1 and Obs. 2: number of observations in the first (reference) group and the second.

the wages of the former are on average higher than the wages of the latter.

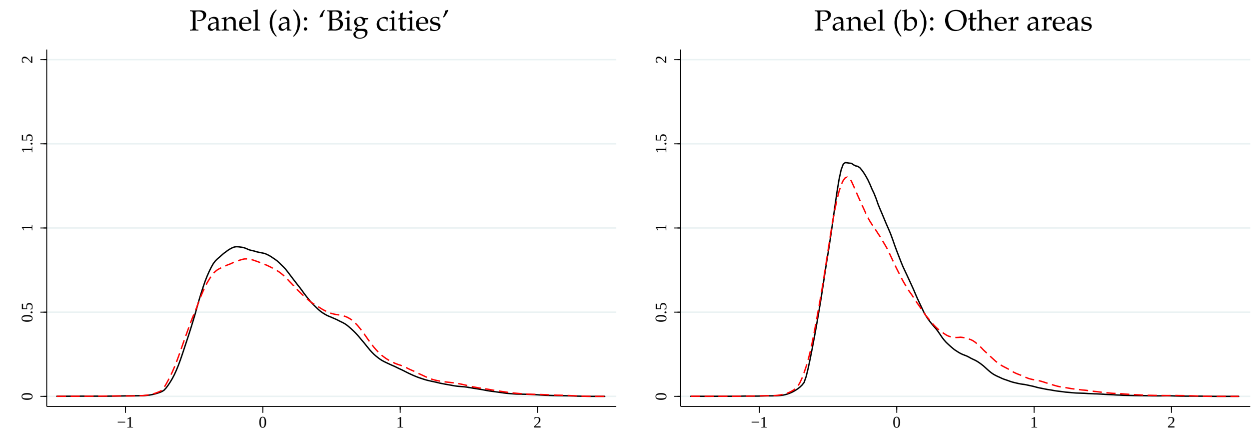
Panels (a) and (b) of figure 5 mirror panels (a) and (b) of figure 4 for workers effects instead of log wages. Panel (a) plots the distribution of worker effects of stayers within less dense areas, that of movers between less dense areas, and that of movers from a less dense to a denser area. Panel (b) plots the corresponding distributions from workers initially employed in denser areas. Table 7 reports some estimation results comparing these distributions.

The results from the first two rows of table 7 indicate that the distribution of worker effects for migrants from less dense areas is both greatly compressed and negatively truncated relative to the distribution of worker effects of stayers in less dense areas. However, these large coefficients only capture the fact that the distribution for migrants is more compressed in the middle and about the same in both tails. Overall these distributions remain close as can be visually confirmed by the plot of panel (a) of figure 5. The most noteworthy results here is that migrants from less dense areas have on average higher worker effects. In addition, the difference in mean worker effect between migrants going to a denser area and those going to another less dense area is not significant as shown in row 3.

The results from rows 4 and 5 for stayers and migrants from denser areas are in sharp contrast with those from less dense areas. First, although there is again evidence of compression and negative truncation, for migrants to less dense areas the compression effect clearly dominates, even in the tails as can be see from panel (b) of figure 5. Relative to stayers in denser areas, migrants also have on average lower worker effects. In addition, migrants that move to a less dense area have on average a modestly higher worker effect relative to those going to another denser area as shown by the results of row 6. Finally row 7 shows that migrants to denser areas have on average higher worker effects relative to those that migrate in the opposite direction.

We draw a number of conclusions from these results. Relative to stayers, migrants have on average lower wages regardless of where they go to and where they come from. This confirms well-established results from the migration literature (Greenwood, 1997, Etzo, 2008). The distribution of the wage of migrants is also compressed in the middle but not always in the tails relative to that of stayers. At the same time, migrants from less dense areas have on average higher worker effects than stayers from the same areas whereas the opposite holds true for denser areas. In addition, migrants towards denser areas have on average higher worker effects relative to migrants towards less dense areas. That is, there is on average positive selection of migrants from less dense areas and negative selection of migrants from denser areas. Overall, these patterns in the selection into migration offers strong evidence regarding the sorting of workers with higher worker effects into denser areas. We also observe that workers with the lowest and highest worker effects in denser areas have a greater propensity to stay which is consistent with the over-representation documented above of these workers in those areas.

Figure 6: Log wages of natives vs. non-natives in 'big cities' and other areas



Continuous black line: natives within the same group of areas.
Dashed grey (red) line: non-natives within the same groups of areas.

Table 8: Log wages of natives vs. non-natives in 'big cities' and other areas

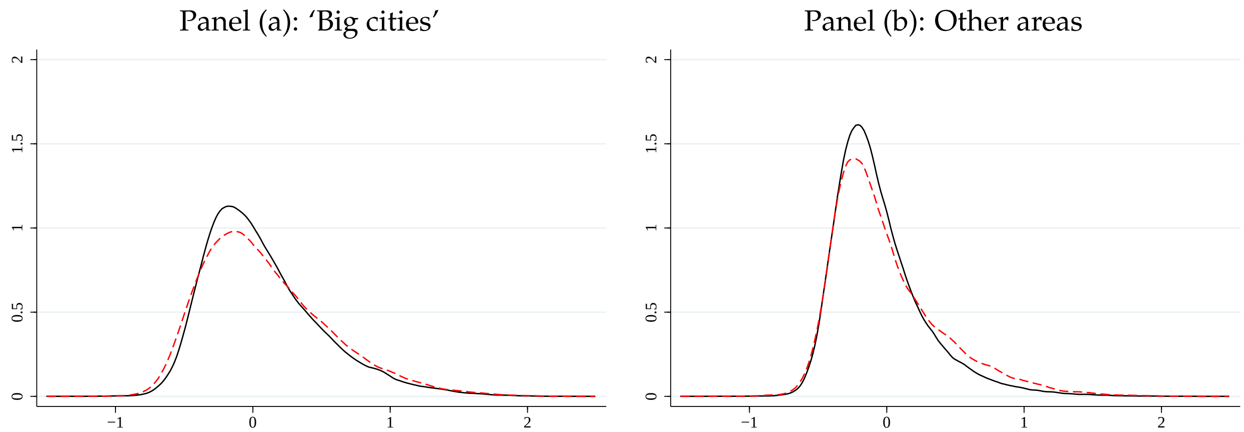
	$\hat{A}$	$\hat{D}$	$\underline{\hat{S}}$	$\hat{\hat{S}}$	R^2	Obs. 1	Obs. 2
w_S to w_B	0.220 (0.003) *	1.756 (0.017) *	0.036 (0.003) *	0.051 (0.002) *	0.998	336,010	193,386
$w_{B \to B}$ to $w_{S \to B}$	0.052 (0.004) *	1.121 (0.009) *	0.015 (0.002) *	0.010 (0.003) *	0.981	103,531	89,855
$w_{S \to S}$ to $w_{B \to S}$	0.028 (0.003) *	1.255 (0.015) *	0.050 (0.005) *	0.005 (0.002) *	0.982	298,923	37,087

*: for $\hat{A}$, $\underline{\hat{S}}$, and $\hat{\hat{S}}$ significantly different from 0 at 5%; for $\hat{D}$ significantly different from 1 at 5%.
w_i: distribution of log wages of all workers working in big cities ($i = B$) or other areas ($i = S$).
$w_{i \to j}$: distribution of log wages of workers born in an area of type i ($\in \{B,S\}$) and working in an area of type j ($\in \{B,S\}$).
Obs. 1 and Obs. 2: number of observations in the first (reference) group and the second.

'Sorting at birth'

Rather than compare migrants and stayers over a three-year period from 2004 to 2007, we now take a longer perspective and define migration status by the difference between the place of work in 2007 and the place of birth. Because we do not know the municipality of residence at birth but only the 'département', we need to change zoning. We refer to the départements hosting the four biggest French cities (75, 78, 91, 92, 93, 94, and 95, Paris; 69, Lyon; 13, Marseille; and 59, Lille) as the 'big cities' and refer to everything else as 'other areas'. We call 'natives' workers who were born in a big city and worked in one in 2007. All the other workers working in big cities are referred to as 'non-natives'. We use the same rule to define natives and non-natives in other areas.

Figure 7: Worker effects of natives vs. non-natives in 'big cities' and other areas



Continuous black line: natives within the same group of areas.
Dashed grey (red) line: non-natives within the same groups of areas.

Panel (a) of figure 6 plots the distribution of log wages of natives and non-natives in the four largest French cities while panel (b) does the same for workers in other areas. Table 8 reports some estimation results comparing these distributions.

The first row of table 8 starts with results regarding the distribution of log wages of all workers (natives and non-natives) in big cities relative to the same distribution in other areas. We find that the distribution of log wages in big cities is right-shifted, dilated, and left- and right-truncated relative to that in other areas. These results generally confirm those obtained above for density. The higher coefficients for the shift and dilation also suggest that what is described above is particularly acute in the four largest French cities.[13] The second and third rows of table 8 show that in both big cities and other areas, the distribution of the wages of non-natives is right-shifted, dilated, and modestly right- and left-truncated relative to the distribution of the wages of natives.

Turning to the distribution of worker effects, panel (a) of figure 7 plots it for natives and non-natives in big cities while panel (b) of the same figure does the same for other areas. Table 9 mirrors table 8 for worker effects instead of wages and reports estimation results that correspond to these plots. Comparing the distribution of worker effects in big cities and other areas, row 1 of table 9 shows that the former is right-shifted, dilated, left- and right-truncated relative to the former. This results essentially confirm the same

[13]In the data, workers at the first decile of the wage distribution in big cities make 1.6% more than a worker at the first decile of the wage distribution in other areas. For workers at the first quartile of both distributions, the difference is at 5.0%. For workers at the median, the difference is again larger at 17.1%. For workers at the third quartile, the difference is even larger at 32.6%. Finally, for workers at the top decile, the difference is 42.8%. These differences are larger than those between denser and less dense areas reported above.

Table 9: Worker effects of natives vs. non-natives in 'big cities' and other areas

	$\hat{A}$	$\hat{D}$	$\underline{\hat{S}}$	$\hat{\bar{S}}$	R^2	Obs. 1	Obs. 2
δ_S to δ_B	0.044 (0.004) *	1.524 (0.025) *	0.033 (0.008) *	0.011 (0.002) *	0.991	336,010	193,386
$\delta_{B\to B}$ to $\delta_{S\to B}$	0.012 (0.003) *	1.346 (0.016) *	0.033 (0.002) *	0.026 (0.003) *	0.987	103,531	89,855
$\delta_{S\to S}$ to $\delta_{B\to S}$	0.014 (0.004) *	1.188 (0.038) *	0.033 (0.012) *	-0.008 (0.005)	0.976	298,923	37,087

*: for $\hat{A}$, $\underline{\hat{S}}$, and $\hat{\bar{S}}$ significantly different from 0 at 5%; for $\hat{D}$ significantly different from 1 at 5%.
δ_i: distribution of worker effects for all workers working in big cities ($i = B$) or other areas ($i = S$).
$\delta_{i\to j}$: distribution of worker effects for workers born in an area of type i ($\in \{B,S\}$) and working in an area of type j ($\in \{B,S\}$).
Obs. 1 and Obs. 2: number of observations in the first (reference) group and the second.

comparison made for the distribution of worker effects in denser and less dense areas. As with wages, we find stronger results when focusing on the four biggest French cities relative to other areas than with our comparison of denser and less dense areas. In the second and third row, we also find that in both types of areas the distribution of worker effects for non-natives is right-shifted and dilated relative to the same distribution for natives.

We draw a number of conclusions from these results. First, even though we documented above that migration is associated with lower wages in the short-run, non-natives have on average higher wages than natives in the long-run. A caveat to this result is that we can evidence it for only the four départements where the biggest French cities are located.[14] There is also more dispersion in the wages of non-natives. For worker effects, we also find that non-natives have on average higher worker effects but there is also more dispersion. However, differences between natives and non-natives are small and the higher and more dispersed worker effects in the four largest cities cannot be explained entirely by the non-natives. To see this consider the following heuristic calculation. According to the first row of table 9, worker effects are on average 4.4 points higher in big cities. According to the second row, non-natives, which form 46% of the workforce of big cities have worker effects which are only 1.2 points higher on average. These numbers establish the existence of sorting in the traditional sense over a longer time horizon but they also suggest an important role for 'sorting at birth' or of some stronger learning in the largest cities.

[14] Looking at smaller cities is not really an option because we can only use French départements which are large relative to their main cities except with the four largest cities. In results not reported here, we obtain the result that 2004-2007 migrants to these four départements have on average lower wages.

Table 10: Log wages of young and old stayers vs. young and old migrants in denser and less dense areas

	$\hat{A}$	$\hat{D}$	$\underline{\hat{S}}$	$\hat{\bar{S}}$	R^2	Obs. 1	Obs. 2
Panel (a): Young workers (below 35)							
w_L to w_H	0.142 (0.002) *	1.703 (0.013) *	0.011 (0.001) *	0.031 (0.002) *	0.994	166,406	97,359
w_{LL} to $w_{L\to L}$	-0.071 (0.004) *	0.680 (0.027) *	-0.109 (0.019) *	-0.017 (0.006) *	0.998	17,898	6,217
w_{LL} to $w_{L\to H}$	-0.034 (0.004) *	0.977 (0.025)	-0.047 (0.003) *	-0.021 (0.007) *	0.991	17,898	17,877
$w_{L\to L}$ to $w_{L\to H}$	0.019 (0.006) *	1.101 (0.125)	0.003 (0.006)	$-$ 0.051 (0.027)	0.914	6,217	17,877
w_{HH} to $w_{H\to H}$	-0.075 (0.003) *	0.867 (0.012) *	-0.034 (0.002) *	-0.025 (0.004) *	0.992	98,879	105,813
w_{HH} to $w_{H\to L}$	-0.144 (0.008) *	0.496 (0.029) *	-0.140 (0.066) *	-0.032 (0.012) *	0.998	98,879	17,081
$w_{H\to H}$ to $w_{H\to L}$	-0.086 (0.003) *	0.609 (0.017) *	-0.016 (0.02) *	-0.008 (0.005)	0.978	105,813	17,081
Panel (b): Old workers (above 45)							
w_L to w_H	0.322 (0.005) *	1.606 (0.015) *	0.001 (0.011)	0.058 (0.002) *	0.998	84,426	47,682
w_{LL} to $w_{L\to L}$	-0.085 (0.017) *	0.922 (0.046)	0.011 (0.031)	0.002 (0.006)	0.992	16,865	1,440
w_{LL} to $w_{L\to H}$	0.063 (0.011) *	1.315 (0.042) *	0.027 (0.014) *	0.015 (0.007) *	0.993	16,865	4,283
$w_{L\to L}$ to $w_{L\to H}$	0.157 (0.022) *	1.482 (0.108) *	0.024 (0.026)	0.022 (0.027)	0.995	1,440	4,283
w_{HH} to $w_{H\to H}$	-0.133 (0.013) *	1.096 (0.018) *	0.089 (0.017) *	0.003 (0.004)	0.996	79,751	25,567
w_{HH} to $w_{H\to L}$	-0.308 (0.007) *	0.602 (0.018) *	-0.003 (0.016)	-0.052 (0.012) *	0.999	79,751	4,202
$w_{H\to H}$ to $w_{H\to L}$	-0.218 (0.007) *	0.564 (0.015) *	-0.050 (0.02) *	-0.056 (0.005) *	0.999	25,567	4,202

*: for $\hat{A}$, $\underline{\hat{S}}$, and $\hat{\bar{S}}$ significantly different from 0 at 5%; for $\hat{D}$ significantly different from 1 at 5%.
w_i: distribution of log wages of all workers in less dense areas ($i = L$) or denser areas ($i = H$).
w_{ii}: distribution of log wages of stayers in less dense areas ($i = L$) or denser areas ($i = H$).
$w_{i\to j}$: distribution of log wages of migrants from an area of type i ($\in \{L,H\}$) to an area of type j ($\in \{L,H\}$).
Obs. 1 and Obs. 2: number of observations in the first (reference) group and the second.

Sorting by age: young and old workers

An issue with the results reported so far is that they may hide some important heterogeneity across different types of workers. In particular, life-cycle issues seem particularly important in migration decisions (see Etzo, 2008, for further references and discussion). Here we replicate some of the estimations performed above but separately perform them on 'young' workers (aged below 35) and 'old' workers (aged above 45).

Table 10 reports estimation results that correspond to the last row of table 4 and table 6 for the log wages of young workers in panel (a) and the log wages of old workers in panel (b). Comparing the two panels of table 10, a number of interesting features emerge. First, from the first line of each panel, we find that for both young and old workers the distribution of log wages in denser areas is right-shifted, dilated, and modestly truncated at both tails relative to the same distribution in less dense areas. These two groups essentially behave like the overall sample in table 4. The main difference is that the higher mean for wages in denser areas is much higher for old workers relative to young workers.

Turning to migrants from less dense areas in row 2-4 of both panels of table 10, we find more important differences, in particular for migrants going to denser areas. The results for young worker are similar to those for the overall sample: a lower mean wage and a small amount of negative truncation but no dilation. For older workers, those that migrate to denser areas have instead higher mean wages drawn from a more dilated distribution than the stayers. When looking at migrants from denser areas in rows 5 to 7, the opposite pattern emerges. Old migrants to less dense areas have a much lower mean wage relative to old stayers than is the case for young migrants relative to young stayers.

To shed further light on this question, table 11 reports estimation results for worker effects of old and young workers that complement the wage results of table 10. The first row of both panels of table 11 contains estimation results for the comparison of all young and old workers in denser areas relative to the same group in less dense areas. The differences are stark. Young workers in denser areas have a lower mean wage relative to young workers in less dense areas. The difference in mean worker effect is -0.035. We find the opposite for old workers and the positive difference is large at 0.218. This suggests that the higher average worker effects in denser areas for the entire sample reported in table 5 is mostly accounted for by old workers. At the same time, for both types of workers the distribution of worker effects is more dilated in denser areas. In turn, this suggests that both types of workers contribute to the over-representation of workers with very high and very low worker effects in denser areas. Interestingly, combining the results of the first row of both panels of table 10 with those of the first row of both panels of table 11 indicates that log wages are higher by 0.177 for young workers in denser areas against only 0.104 for old workers after conditioning out worker effects.

Turning to migrants from less dense areas in rows 2 to 4 of both panels of table 11, we find that the young migrants to a denser area have a slightly lower mean worker effect whereas old migrants to a denser area have on average higher mean worker effects. The opposite holds for migrants within less dense areas. For migrants from denser areas in rows 3 to 5 of both panels, we find that young migrants to a less dense area or another denser area do not differ much from the stayers. For old workers from a denser area,

Table 11: Worker effects of young and old stayers vs. young and old migrants in denser and less dense areas

	$\hat{A}$	$\hat{D}$	$\underline{\hat{S}}$	$\hat{\hat{S}}$	R^2	Obs. 1	Obs. 2
Panel (a): Young workers (below 35)							
δ_L to δ_H	-0.035 (0.002) *	1.296 (0.022) *	0.012 (0.002) *	-0.015 (0.004) *	0.984	166,406	97,359
δ_{LL} to $\delta_{L\to L}$	0.021 (0.004) *	1.075 (0.035) *	-0.031 (0.005) *	0.035 (0.010) *	0.993	17,898	6,217
δ_{LL} to $\delta_{L\to H}$	-0.006 (0.003) *	1.194 (0.040) *	-0.025 (0.007) *	0.010 (0.006)	0.989	17,898	17,877
$\delta_{L\to L}$ to $\delta_{L\to H}$	-0.033 (0.004) *	1.150 (0.028) *	0.011 (0.003) *	-0.014 (0.005) *	0.982	6,217	17,877
δ_{HH} to $\delta_{H\to H}$	-0.023 (0.001) *	1.113 (0.013) *	-0.015 (0.002) *	0.015 (0.002) *	0.992	98,879	105,813
δ_{HH} to $\delta_{H\to L}$	0.005 (0.002) *	0.971 (0.014) *	-0.022 (0.002) *	0.045 (0.005) *	0.982	98,879	17,081
$\delta_{H\to H}$ to $\delta_{H\to L}$	0.030 (0.002) *	0.865 (0.013) *	-0.008 (0.002) *	0.025 (0.004) *	0.982	105,813	17,081
Panel (b): Old workers (above 45)							
δ_L to δ_H	0.218 (0.006) *	1.555 (0.021) *	-0.008 (0.005)	0.044 (0.003) *	0.999	84,426	47,682
δ_{LL} to $\delta_{L\to L}$	-0.082 (0.013) *	0.920 (0.041)	-0.003 (0.019)	0.001 (0.008)	0.994	16,865	1,440
δ_{LL} to $\delta_{L\to H}$	0.026 (0.008) *	1.245 (0.036) *	0.001 (0.011)	0.009 (0.005) *	0.995	16,865	4,283
$\delta_{L\to L}$ to $\delta_{L\to H}$	0.108 (0.024) *	1.333 (0.072) *	-0.006 (0.037)	0.006 (0.009)	0.996	1,440	4,283
δ_{HH} to $\delta_{H\to H}$	-0.103 (0.005) *	1.065 (0.015) *	0.030 (0.008) *	0.002 (0.002)	0.998	79,751	25,567
δ_{HH} to $\delta_{H\to L}$	-0.241 (0.005) *	0.660 (0.019) *	0.001 (0.009)	-0.034 (0.006) *	0.999	79,751	4,202
$\delta_{H\to H}$ to $\delta_{H\to L}$	-0.146 (0.007) *	0.626 (0.018) *	-0.021 (0.014)	-0.035 (0.006) *	0.999	25,567	4,202

*: for $\hat{A}$, $\underline{\hat{S}}$, and $\hat{\hat{S}}$ significantly different from 0 at 5%; for $\hat{D}$ significantly different from 1 at 5%.
δ_i: distribution of worker effects for all workers in less dense areas ($i = L$) or denser areas ($i = H$).
δ_{ii}: distribution of worker effects for stayers in less dense areas ($i = L$) or denser areas ($i = H$).
$\delta_{i\to j}$: distribution of worker effects for migrants from an area of type i ($\in \{L,H\}$) to an area of type j ($\in \{L,H\}$).
Obs. 1 and Obs. 2: number of observations in the first (reference) group and the second.

the differences are large. Old migrants have markedly lower worker effects, particularly those that migrate to a less dense area. Interestingly, simple computations on the results of tables 10 and 11 show that the differences in wages between migrants and stayers for old and young workers reported in table 10 are much more similar once the differences in worker effects are conditioned out.

We draw a number of conclusions from these results. First, looking at all workers

Table 12: Log wages of stayers vs. migrants in professional occupations in denser and less dense areas

	$\hat{A}$	$\hat{D}$	$\hat{\underline{S}}$	$\hat{\overline{S}}$	R^2	Obs. 1	Obs. 2
w_L to w_H	0.084 (0.004) *	1.149 (0.002) *	0.021 (0.004) *	−0.002 (0.004)	0.997	4,283	28,391
w_{LL} to $w_{L\to L}$	−0.124 (0.029) *	0.904 (0.131)	−0.036 (0.034)	−0.006 (0.035)	0.992	3,114	362
w_{LL} to $w_{L\to H}$	−0.141 (0.013) *	1.088 (0.062)	0.008 (0.012)	−0.008 (0.012)	0.996	3,114	2,579
$w_{L\to L}$ to $w_{L\to H}$	−0.019 (0.031)	1.182 (0.148)	0.038 (0.029)	−0.006 (0.027)	0.958	362	2,579
w_H to $w_{H\to H}$	−0.191 (0.004) *	1.022 (0.017)	−0.008 (0.003) *	−0.002 (0.003)	0.997	38,301	23,262
w_H to $w_{H\to L}$	−0.229 (0.011) *	1.013 (0.013)	−0.031 (0.024)	0.019 (0.012)	0.999	38,301	1,316
$w_{H\to H}$ to $w_{H\to L}$	−0.045 (0.010) *	0.986 (0.047)	−0.018 (0.016)	0.017 (0.008) *	0.987	23,262	1,316

*: for $\hat{A}$, $\hat{\underline{S}}$, and $\hat{\overline{S}}$ significantly different from 0 at 5%; for $\hat{D}$ significantly different from 1 at 5%.
w_i: distribution of log wages of all workers in less dense areas ($i = L$) or denser areas ($i = H$).
w_{ii}: distribution of log wages of stayers in less dense areas ($i = L$) or denser areas ($i = H$).
$w_{i\to j}$: distribution of log wages of migrants from an area of type i ($\in \{L,H\}$) to an area of type j ($\in \{L,H\}$).
Obs. 1 and Obs. 2: number of observations in the first (reference) group and the second.

within one type of area hides considerable heterogeneity between different age groups. Second, the mean wage premium for locating in denser areas is lower for young workers. However, after conditioning out worker effects, it turns out that young workers benefit more from density than old workers. Third, the sorting of workers with particularly high or low worker effects in denser areas appears to take place for all age groups. On the other hand, the higher mean worker effects in denser areas is mostly accounted for by old workers. Whether this is an effect of sorting that gradually increases over the life-cycle or a consequence of workers learning more in larger cities is an open question (De La Roca and Puga, 2011).

Sorting by occupations

We interpret worker effects as unobserved skills. When we estimate these effects above we only use age and its square as individual controls for workers. Consequently our 'unobserved' skills will capture nearly all permanent skills of workers. We do not observe education nor other traditional measures of skills. However, we know occupation, a crude measure of skills. To have another peak at heterogeneity across workers, we now replicate some of the estimations performed above but separate workers by their one-digit occupational category.

Table 13: Worker effects of stayers vs. migrants in professional occupations in denser and less dense areas

	$\hat{A}$	$\hat{D}$	$\hat{\underline{S}}$	$\hat{\overline{S}}$	R^2	Obs. 1	Obs. 2
δ_L to δ_H	-0.006 (0.004)	1.112 (0.013) *	0.021 (0.003) *	0.001 (0.002)	0.986	4,283	28,391
δ_L to $\delta_{L\to L}$	-0.168 (0.034) *	0.978 (0.160)	-0.036 (0.074)	-0.003 (0.042)	0.993	3,114	362
δ_L to $\delta_{L\to H}$	-0.224 (0.012) *	1.030 (0.051)	0.008 (0.011)	-0.024 (0.010) *	0.999	3,114	2,579
$\delta_{L\to L}$ to $\delta_{L\to H}$	-0.054 (0.031)	1.068 (0.158)	0.012 (0.047)	-0.017 (0.064)	0.917	362	2,579
δ_H to $\delta_{H\to H}$	-0.215 (0.003) *	0.960 (0.012) *	-0.004 (0.003)	-0.017 (0.003)	0.999	38,301	23,262
δ_H to $\delta_{H\to L}$	-0.205 (0.009) *	0.993 (0.035)	0.011 (0.009)	0.001 (0.008)	0.999	38,301	1,316
$\delta_{H\to H}$ to $\delta_{H\to L}$	0.011 (0.009)	1.031 (0.035)	0.013 (0.008)	0.017 (0.008) *	0.889	23,262	1,316

*: for $\hat{A}$, $\hat{\underline{S}}$, and $\hat{\overline{S}}$ significantly different from 0 at 5%; for $\hat{D}$ significantly different from 1 at 5%.
δ_i: distribution of worker effects for all workers in less dense areas ($i = L$) or denser areas ($i = H$).
δ_{ii}: distribution of worker effects for stayers in less dense areas ($i = L$) or denser areas ($i = H$).
$\delta_{i\to j}$: distribution of worker effects for migrants from an area of type i ($\in \{L,H\}$) to an area of type j ($\in \{L,H\}$).
Obs. 1 and Obs. 2: number of observations in the first (reference) group and the second.

Table 12 reports estimation results for the wages of workers in professional occupations. Row 1 compares the distribution of log wages of all workers in professional occupations in denser areas relative to the same distribution in less dense areas. As in the last row of table 4 which performs the same comparison for all workers, there is again evidence of a higher mean wage, a dilation of the distribution, and a left-truncation to approximate the distribution of log wages in denser areas from the same distribution in less dense areas. However the effects are much attenuated. The mean wage premium in denser areas is only 0.084 for workers in professional occupations instead of 0.136 for all workers. Dilation is also markedly less: 1.149 instead of 1.631.

Turning to migrants in row 2-7, there is evidence of much lower wages for migrants regardless of where they are from and where they go to. Comparing with the corresponding results of table 6, the wage difference between migrants and stayers for workers in professional occupation is larger than for all workers. The other salient feature of the results for migrants in professional occupations relative to all migrants is that there is far less dilation of the distribution of log wages of migrants to denser areas and far less compression of the distribution of the log wages of migrants to less dense areas.

To push the analysis further, table 13 reports estimation results for the worker effects of workers in professional occupations. The first row of this table compares the distribution of worker effects for all workers in professional occupations in denser areas with the same distribution in less dense areas. The key result from this row is that the differences between the two distributions are minimal. The negative shift coefficient is insignificant and the dilation coefficient is modest at 1.112. This is in sharp contrast with the corresponding result for all workers in table 5, which showed both a significant positive shift coefficient and a larger dilation coefficient of 1.340. These differences in the patterns of worker effects for workers in professional occupations relative to all workers appear to account for large fraction of the differences in the distribution of log wages between the same two groups. For instance, for all workers the unconditional wage premium is 0.136 and $0.136 - 0.040 = 0.096$ after conditioning out worker effects. For workers in professional occupations, the unconditional wage premium is 0.084 and $0.084 + 0.006 = 0.090$ after conditioning out worker effects.

The second interesting feature of table 13 concerns the worker effects of migrants in professional occupations. They are on average much lower than those of stayers. Again, it appears that worker effects account for much of the significant wage gap between migrants and stayers evidenced in table 12. Relative to the findings for all workers in table 7, the results are very different. Workers in professional occupations migrating to a denser area have lower worker effects than the stayers in those areas, unlike what we find above for all workers. In addition, workers in professional occupations who migrate to less dense areas have much lower worker effects relative to stayers, instead of only modestly lower effects in the case of all workers. In addition, there is only very little dilation or compression in the distribution of worker effects for migrants in professional occupations relative to stayers. This is unlike what we observe for all migrants relative to all stayers in table 7.

We also repeated tables 12 and 13 for other occupational groups such as intermediate occupations and blue-collar occupations. We do not report these results here for the sake of brevity but they resemble those for workers in professional occupations. The difference is that the mean of worker effects for workers in intermediate occupations is now significantly lower in denser areas. This difference becomes even more negative for workers in blue-collar occupations. For migrants, the negative difference in mean worker effect relative to stayers is attenuated for workers in intermediate occupations relative to workers in professional occupations. The attenuation is even stronger for workers in blue-collar occupations.

All these results point to one conclusion. It appears that much of the sorting results evidenced above are driven by differences between occupational groups of workers and much less by what happens within these groups. This result is consistent with the results of Baum-Snow and Pavan (2011) who find that after accounting for a rich set of observables, US workers in large cities have lower unobservable skills. Of course, the strength of our conclusion depends on the extent to which occupational categories truly reflect skills instead of wages.

6. Conclusion

This paper provides a new methodology to examine the sorting of workers across locations by skills. We apply it to denser and less dense employment areas in France. We find that workers in denser areas are on average more skilled. Workers with particularly high skills are more represented in denser areas but so are, to a lesser extent, workers with particularly low skills. Consistent with this, we find that there is positive selection of migrants to denser areas and negative selection of migrants to less dense areas. Migration is also associated with a sizeable wage difference between migrants and stayers. In addition, a long run examination of natives and non-natives in the largest French cities suggests that selection into migration can account for only a fraction of the observed patterns of sorting. Interestingly, both sorting and selection into migration appear stronger for older workers. Finally, we find marked differences across age groups and some suggestions that much of the skill differences across areas can be explained by differences between occupational groups rather than within.

This rich set of results is consistent with a number of previous findings (e.g., migration is costly), strengthens more recent conjectures (e.g., the over-representation of workers at both tails in denser areas), and provides new facts (e.g., selection into migration can be positive or negative depending on whether we consider workers moving to higher or lower density). This set of results also raise a number of further questions. While we pay a lot of attention to life-cycle issues, we have neglected other important dimensions such as gender which remain to be explored. The importance of what we refer to as 'selection at birth' begs further analysis about what is behind the differences in permanent skills of 'natives' between denser and less dense areas: family background or education. The differences in the sorting of young vs. old workers can be explained by workers learning more in denser areas. In turn, this suggests using a more sophisticated approach to estimate workers skills. What we treat here as permanent skills should be able to evolve over time and depend on where workers are. Finally, we say nothing about the sources of sorting. Hopefully future research will be able to disentangle differential returns to density, differences in cost of living, and various types of amenities and public services.

References

Bacolod, Marigee, Bernardo S. Blum, and William C. Strange. 2009*a*. Skills in the city. *Journal of Urban Economics* 65(2):136–153.

Bacolod, Marigee, Bernardo S. Blum, and William C. Strange. 2009*b*. Urban interactions: soft skills versus specialization. *Journal of Economic Geography* 9(2):227–262.

Bacolod, Marigee, Bernardo S. Blum, and William C. Strange. 2010. Elements of skill: Traits, intelligences, education, and agglomeration. *Journal of Regional Science* 50(1):245–280.

Baum-Snow, Nathaniel and Ronni Pavan. 2010. Inequality and city size. Processed, Brown University.

Baum-Snow, Nathaniel and Ronni Pavan. 2011. Understanding the city size wage gap. *Review of Economic Studies* 79. Forthcoming.

Behrens, Kristian, Gilles Duranton, and Frédéric Robert-Nicoud. 2010. Productive cities: Sorting, selection, and agglomeration. Processed, University of Toronto.

Behrens, Kristian and Frédéric Robert-Nicoud. 2010. Survival of the fittest in cities: Agglomeration, polarization, and income inequality. Processed, University of Geneva.

Berry, Christopher R. and Edward L. Glaeser. 2005. The divergence of human capital levels across cities. *Papers in Regional Science* 84(3):407–444.

Briant, Anthony, Pierre-Philippe Combes, and Miren Lafourcade. 2010. Does the size and shape of geographical units jeopardize economic geography estimations? *Journal of Urban Economics* 67(3):287–302.

Ciccone, Antonio and Robert E. Hall. 1996. Productivity and the density of economic activity. *American Economic Review* 86(1):54–70.

Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon. 2008. Spatial wage disparities: Sorting matters! *Journal of Urban Economics* 63(2):723–742.

Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon. 2011*a*. The identification of agglomeration economies. *Journal of Economic Geography* 11(2):253–266.

Combes, Pierre-Philippe, Gilles Duranton, Laurent Gobillon, and Sébastien Roux. 2010. Estimating agglomeration economies with history, geology, and worker effects. In Edward L. Glaeser (ed.) *The Economics of Agglomeration*. Cambridge (MASS.): National Bureau of Economic Research, 15–65.

Combes, Pierre-Philippe, Gilles Duranton, Laurent Gobillon, Sébastien Roux, and Diego Puga. 2011*b*. The productivity advantages of large cities: Distinguishing agglomeration from firm selection. Mimeographed, University of Toronto.

Davis, Donald R. 2009. A spatial knowledge economy. Processed, Columbia University.

De la Roca, Jorge. 2011. Selection in initial and return migration: Evidence from moves across Spanish cities. Processed, CEMFI and IMDEA Social Sciences.

De La Roca, Jorge and Diego Puga. 2011. The dynamic earnings premium of dense cities. Work in progress, CEMFI and IMDEA Sociale Sciences.

Duranton, Gilles and Diego Puga. 2004. Micro-foundations of urban agglomeration economies. In J. Vernon Henderson and Jacques-François Thisse (eds.) *Handbook of Regional and Urban Economics*, volume 4. Amsterdam: North-Holland, 2063–2117.

Eeckhout, Jan, Roberto Pinheiro, and Kurt Schmidheiny. 2010. Spatial sorting: Why New York, Los Angeles and Detroit attract the greatest minds as well as the unskilled. Processed, University Pompeu Fabra.

Etzo, Ivan. 2008. Internal migration: A review of the literature. Processed, University of Cagliari.

Glaeser, Edward L. 1999. Learning in cities. *Journal of Urban Economics* 46(2):254–277.

Glaeser, Edward L. and Joshua D. Gottlieb. 2008. The economics of place-making policies. *Brookings Papers on Economic Activity* (1):155–253.

Glaeser, Edward L., Matthew E. Kahn, and Jordan Rappaport. 2008. Why do the poor live in cities? The role of public transportation. *Journal of Urban Economics* 63(1):1–24.

Glaeser, Edward L. and David C. Maré. 2001. Cities and skills. *Journal of Labor Economics* 19(2):316–342.

Glaeser, Edward L., Matt Resseger, and Kristina Tobio. 2009. Inequality in cities. *Journal of Regional Science* 49(4):617–646.

Glaeser, Edward L. and Matthew R. Resseger. 2010. The complementarity between cities and skills. *Journal of Regional Science* 50(1):221–244.

Greenwood, Michael J. 1997. Internal migrations in developed countries. In Mark R. Rosenzweig and Oded Stark (eds.) *Handbook of Population and Family Economics*, volume 1B. Amsterdam: North-Holland, 647–720.

Marshall, Alfred. 1890. *Principles of Economics*. London: Macmillan.

Matano, Alessia and Paolo Naticchioni. 2012. Spatial externalities and wage distribution: The role of sorting. *Journal of Economic Geography* 12. Forthcoming.

Mion, Giordano and Paolo Naticchioni. 2009. The spatial sorting and matching of skills and firms. *Canadian Journal of Economics* 42(1):683–701.

Moretti, Enrico. 2011. Local labor markets. In Orley Ashenfelter and David Card (eds.) *Handbook of Labor Economics*, volume 4B. Amsterdam: North-Holland, 1237–1313.

Rosenthal, Stuart S. and William C. Strange. 2008. The attenuation of human capital spillovers. *Journal of Urban Economics* 64(2):373–389.

Venables, Anthony J. 2011. Productivity in cities: Self-selection and sorting. *Journal of Economic Geography* 11(2):241–251.

Wheeler, Christopher H. 2001. Search, sorting, and urban agglomeration. *Journal of Labor Economics* 19(4):879–899.