Sorting and Local Wage and Skill Distributions in France

Pierre-Philippe Combes
Gilles Duranton
Laurent Gobillon
Sébastien Roux

April 2012

DISCUSSION PAPER SERIES

I Z A

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Sorting and Local Wage and Skill Distributions in France

Pierre-Philippe Combes
*GREQAM, Aix-Marseille School of Economics,
CNRS and CEPR*

Gilles Duranton
University of Toronto and CEPR

Laurent Gobillon
INED, PSE, CEPR and IZA

Sébastien Roux
DARES and CREST (INSEE)

ABSTRACT

Sorting and Local Wage and Skill Distributions in France[*]

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. 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, accounts for little of the skill differences between denser and less dense areas. 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.

JEL Classification: J31, J61, R12, R23

Keywords: sorting, wage distribution, skill distribution

Corresponding author:

Laurent Gobillon
Institut National d'Etudes Démographiques (INED)
133 Boulevard Davout
75980 Paris Cedex 20
France
E-mail: laurent.gobillon@ined.fr

1. Introduction

This paper provides descriptive evidence about the distribution of wages and skills in denser and less dense employment areas in France. Taking skills as being fixed over time, 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. These features are consistent with the patterns of migration that we evidence, including negative selection of migrants to less dense areas and positive selection towards denser areas. Nonetheless migration, even in the long-run, accounts for little of the skill differences between denser and less dense areas. This suggests a role for differences in skill formation prior to the entry on the labour market (through either better parental education or better schooling in denser areas) or for stronger worker 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 in terms of wages. 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. While some may view such inequalities as an undesirable effect of urbanisation, our results show that all workers enjoy higher wages in denser areas, albeit to a different extent.

Third, the idea of worker sorting across cities by skills has recently caught the interest of urban theorists (Nocke, 2006, 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. In particular, we note that while models of sorting often predict a strict segmentation of workers by skill or ability across cities, we find little evidence of such perfect sorting (except perhaps at the very top and the very bottom of the distribution of skills). Instead, sorting is more partial and denser areas host both a greater proportion of workers with low skills and a greater proportion of workers with high skills.

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 more than their first (or even second) moment. We use the approach of Combes, Duranton, Gobillon, Roux, and Puga (2011*b*) and estimate how the distribution of skills in denser areas is shifted, dilated, and left-truncated relative to the distribution of skills in less dense areas. Extending this approach, we also estimate a parameter capturing any possible right-truncation in one type of area relative to the other. While we find that a shift and a truncation parameter are often enough to capture the bulk of the differences in the distribution of wages and worker effect between denser and less dense area, using these four parameters allows us to account for close to all the differences in most of our estimations.

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 in our most demanding estimations 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').

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, Baum-Snow and Pavan, 2010). 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, 2012). 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 suggestions 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, Baum-Snow and Pavan, 2010, Behrens and Robert-Nicoud, 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,

and workers in particular occupations.

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. While our results about migrations are consistent with existing patterns of sorting, migration is unlikely to be a major driver of skill differences between areas.

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 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, social 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 em-

---

[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.

[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.

ployees 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 October 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 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 3,529,297 observations. Considering jobs for which workers are aged between 20 and 59, we are left with 3,308,205 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 3,159,867 observations. Some workers occupy several jobs during a given year. We only retain their main job, defined as the job with the highest total earnings for that year. Our sample decreases to 2,773,484 observations. After deleting jobs outside mainland France, and those with missing or misreported sector or employment area code, we are left with 2,734,085 observations. We also delete jobs in small sectors (such as spatial transport), in non-profit sectors, and in 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 2,525,918 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 2,520,855 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 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 and high density are strongly associated and correspond to the largest French cities. Table 1 provides further descriptive statistics for mean hourly wages and employment density for French employment areas. Hourly wages are about 20% higher in employment areas in the top decile of mean wages

---

[4]The département is an administrative unit such that mainland France is fully covered by 93 départements. There is no direct mapping between départements and employment areas.

[5]Another small fraction of extreme observations is removed in the implementation of the procedure to compare distributions described below.

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

Panel (a): Employment density

Panel (b): Mean hourly wage
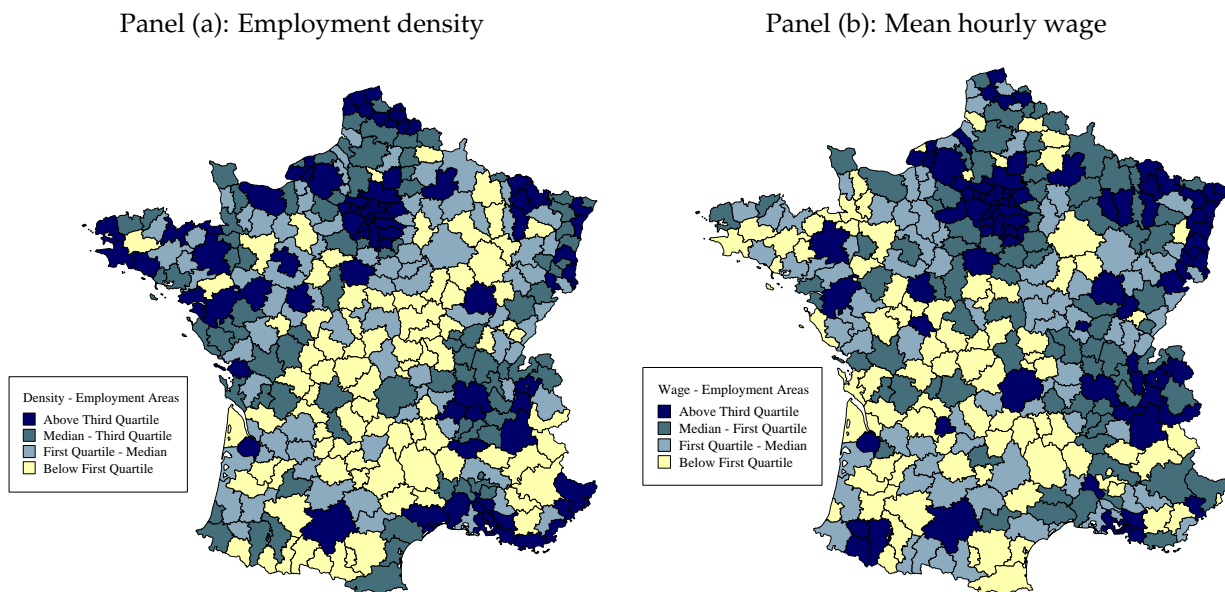
Table 1: Descriptive statistics about spatial wage disparities

	Mean	Std. dev.	1st decile	1st quartile	Median	3rd quartile	9th decile
Hourly wage	13.64	1.31	12.49	12.84	13.34	14.10	14.89
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.

Table 2: Descriptive statistics about wage disparities within groups of areas

	Mean	Std. dev.	1st decile	1st quartile	Median	3rd quartile	9th decile
All areas	15.07	9.09	9.08	10.23	12.32	16.53	24.03
Above median density	15.47	9.48	9.10	10.31	12.58	17.11	25.03
Below median density	13.08	6.48	8.96	9.96	11.38	14.02	18.64

Statistics over workers wages within groups of employment areas in 2007. Hourly wages in 2007 Euros. For example: 9.96 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.

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 table 1 and prior evidence (e.g., Combes *et al.*, 2008, 2010), workers in denser areas receive on average higher wages. It also shows that denser areas are more unequal. Interest-

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	305,615	(97.7%)	7,260	(2.3%)
Below median density in 2004	7,847	(11.9%)	57,938	(88.1%)
Panel (b): Movers only				
Above median density in 2004	55,580	(88.5%)	7,260	(11.5%)
Below median density in 2004	7,847	(69.4%)	3,457	(30.6%)

Row percentages between brackets.

ingly, 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. Figure 3(a) below also provides a plot of the underlying distributions. Finally, panels (a) and (b) of table 3 document patterns of mobility both within and between 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, sector k, and year t. Its profit is

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

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

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

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

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

where $B_{akt} \equiv b(1-b)^{\frac{(1-b)}{b}} \left(p_{akt} A_{akt} r_{at}^{1-b} \right)^{\frac{1}{b}}$. A direct implication of this equation is that wage differences across areas reflect both differences in individual skills and differences in area 'productivity' for sector k. 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 for sector k (B_{akt}) need to be specified. We assume that productivity for sector, area, and year can be expressed as

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

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.[6] 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. 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.

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 (sector-year fixed effect). We interpret worker effects as unobserved skills. Because we only have age and its square as time-varying individual characteristics, we expect most of the variation not related to area-year effects to be captured by worker fixed effects. We also note that cross-section difference in area effects are estimated from movers (who keep the same worker fixed effect in different areas) while the time variation of area effects is estimated from stayers (who again keep the same worker fixed effect).

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

---

[6]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 (2011a).

meet this requirement. The identification of sector-year and area-year effects also requires the normalisation of the effect of one sector and that of one area for one year. In addition, our separate identification of personal skills and area productivity relies on five 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 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 the productivity of all workers equally. 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, 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, our specification assumes perfect substitutability across workers (particularly within firms). There is nonetheless good evidence of imperfect substitutability across skill groups. See for instance Ciccone and Peri (2006) for a related application to US states. The first difficulty in this case is that employment in each skill group now matters to determine wages in an area (that is, B_{akt} becomes a function of several measures of employment). As in the literature on human capital externalities (e.g., Moretti, 2004) and consistent with the previous point, one may also suspect that each group generates its own external effects. Separating agglomeration and composition effects under such circumstances is difficult and we consider it an open problem. Given that our focus is on worker effects and not on local effects, we leave these issues aside here.

Fourth, we assume that worker effects are fixed over time. In particular, we do not allow area productivity to change 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). Empirical research is suggestive that such learning effects may be at work (Glaeser and Maré, 2001, De La Roca and Puga, 2012). When estimating equation (6) we only estimate an 'average' (over time and locations) worker skill. We also keep this issue in mind when interpreting our results. In particular, what we call 'sorting' need not arise solely from the movement of workers with greater innate abilities. Greater skills might be acquired as a result of working in denser areas or prior to entering the labour market. This is also why, as argued above, comparisons between stayers and movers within each area is of particular interest. These workers are arguably more comparable in terms of (urban) experience.

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. For instance, having workers move to areas that offer high expected wages is not a source of bias since this is captured by the area effect for the year, β_{at}, and is uncorrelated with the error term. This said, we expect workers to take jobs in response to specific wage offers (as in, for instance, Baum-Snow and Pavan, 2012) and there might be some correlation between the error term in equation (6) and either the skills of workers or their choice of area. Nonetheless, there are reasons why such biases are likely to be attenuated. First, we expect location decisions to be also 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 expected wages not short term wage shocks. This argument is consistent with many of our findings below.

Finally, let us add that 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 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 type of area is truncated, shifted, and dilated relative to the distribution of the same variable in another type of area. It extends the approach of Combes *et al.* (2011*b*) to allow for both a left and a right truncation instead of only left truncation. For simplicity, we now present the case when only a shift and a dilation are allowed. The extension that also includes a left and a right truncation is relegated to Appendix A. We present our approach referring to wages, but its implementation with worker effects is identical.

We focus on these four transformations (shift, dilation, left-, and right-truncation) because for the distributions of wages and worker effects they allow us to account for nearly all the difference between denser and less dense areas. We postpone the economic interpretation of these four parameters we estimate until next section.

Consider a situation where any wage w in distribution 1 is transformed linearly into a wage $Dw + A$ in distribution 2, where A is a shift parameter and D is a dilation parameter. Parameter A can be positive or negative, depending on whether there is a shift to the right or to the left, and parameter D can be above or below one, depending on whether there is a dilation or a compression. Note that to be able to interpret $D > 1$ as a dilation and $D < 1$ as a compression, we need to normalise the mean of distribution 1 to zero (by subtracting the mean of distribution 1 to all wages).

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

$$\lambda_2(u) = D\lambda_1(u) + A, \quad \text{for } u \in [0,1] . \tag{7}$$

The estimation procedure is based on this relationship between quantiles. Denote $\theta = (A,D)$ the set of parameters to be estimated. Equation (7) can be rewritten as a continuum of equalities of

the form $m_\theta (u) = 0$ for $u \in [0,1]$ where:

$$m_\theta (u) = \lambda_2 (u) - D\lambda_1 (u) - A .\qquad(8)$$

It is possible to estimate θ minimising the sum of squared empirical counterparts of $m_\theta (u)$, $u \in [0,1]$ obtained when replacing the quantile functions by some estimators. However, nothing guarantees that our minimisation criterion, when applied to the transformation of distribution 2 into distribution 1 (the opposite transformation), leads to the same estimates of A and D. Thus, we also use a continuum of equalities obtained when switching distributions 1 and 2:

$$\lambda_1 (u) = \frac{1}{D}\lambda_2 (u) - \frac{A}{D} \ , \quad \text{for } u \in [0,1] \ .\qquad(9)$$

This equation yields the new set of equalities $\widetilde{m}_\theta (u) = 0$, where:

$$\widetilde{m}_\theta (u) = \lambda_1 (u) - \frac{1}{D}\lambda_2 (u) + \frac{A}{D} \ .\qquad(10)$$

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 minimising the sum of squared values of both $\widehat{m}_\theta (u)$ and $\widehat{\widetilde{m}}_\theta (u)$:

$$\widehat{\theta} = \arg\min_\theta M(\theta) \ , \quad \text{where} \quad M(\theta) = \int_0^1 [\widehat{m}_\theta(u)]^2 \mathrm{d}u + \int_0^1 \left[\widehat{\widetilde{m}}_\theta(u)\right]^2 \mathrm{d}u \ .\qquad(11)$$
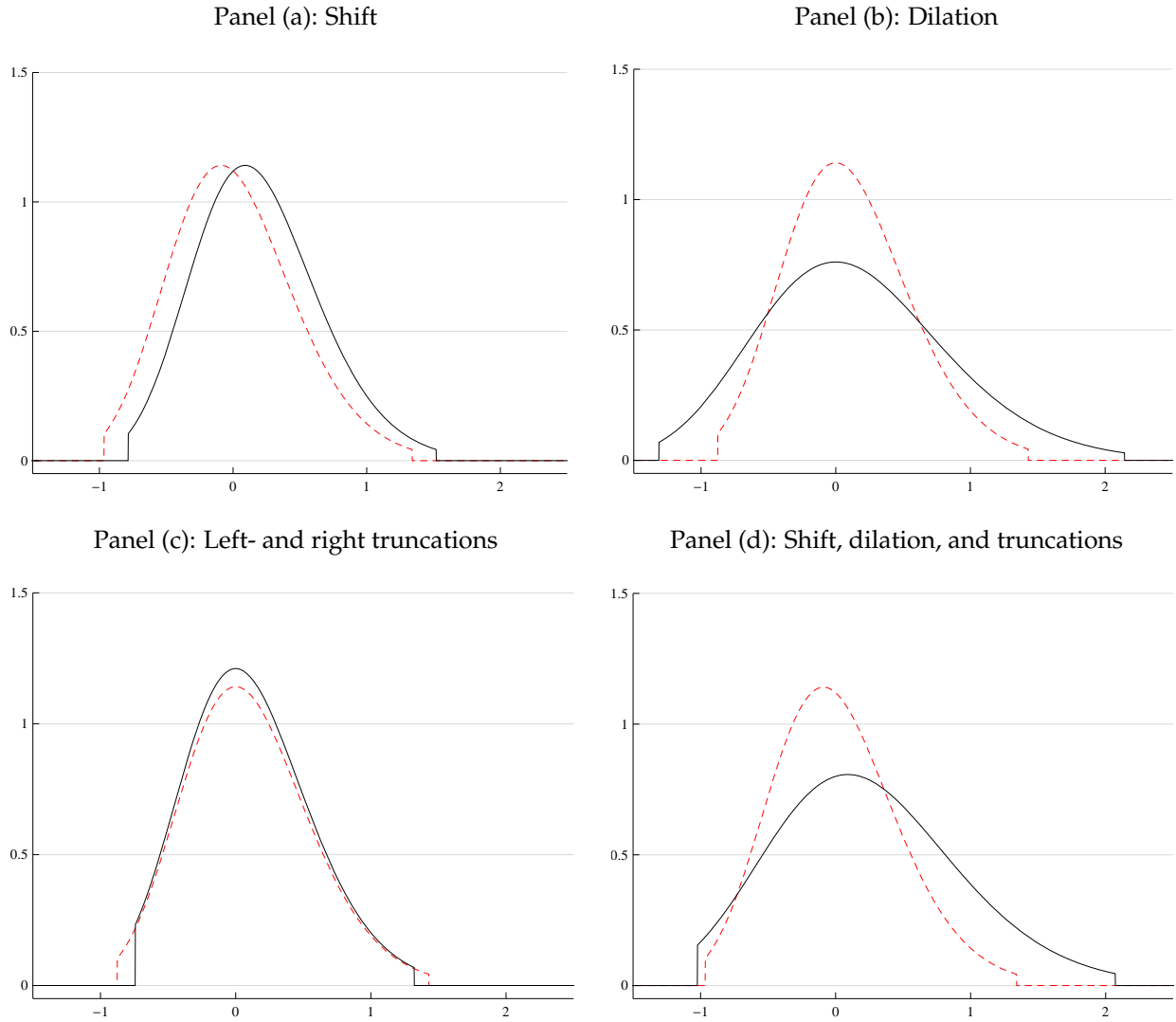
The confidence intervals are obtained by bootstrap. The implementation details are given in Combes *et al.* (2011*b*). Finally, we construct a measure of fit that uses the ratio of mean-squared quantile differences between the transformed distribution 1 and distribution 2 with mean-squared quantile differences between distributions 1 and 2. This measure of fit is a Pseudo-R^2 defined as $R^2 = 1 - M(\widehat{\theta})/M(0,1)$.

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 the hypothetical distribution of log wages in denser areas obtained from a transformation of the distribution in less dense 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 with a shift and a dilation to approximate the distribution of log wages in denser areas. Following the extension we develop in Appendix A, we also estimate the left and right truncations of distribution 2 relative to distribution 1 that can occur beyond a shift and a dilation. We denote $\underline{S}$ the rank at which left truncation occurs and $1 - \overline{S}$ the rank at which right truncation occurs.

Panel (a) considers the case of a simple positive shift only ($A > 0$, $D = 1$, $\underline{S} = \overline{S} = 0$). 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 only ($D > 1$, $A = \underline{S} = \overline{S} = 0$). We can see that the distribution of wages in denser areas spreads further to the left and to the right relative to the

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

Panel (a): Shift

Panel (b): Dilation

Panel (c): Left- and right truncations

Panel (d): Shift, dilation, and truncations

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

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.[7]

Panel (c) considers the case of both a left- and a right-truncation but no shift nor dilation ($\underline{S} > 0$, $\overline{S} > 0$, $A = 0$, $D = 1$). 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 right-shifted, dilated, and left- and right-truncated relative to the same distribution in less dense areas.

---

[7]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.

An important thing to note is that dilation and truncation work 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

Before describing our results, here is how we proceed. In this section, we provide a detailed comparison of the distribution of wages and worker effects in denser vs. less dense employment areas in 2007. In section 5, we provide more detailed results about particular sub-samples of workers. The distribution of wages reflects both differences in the distribution of the 'quantity' of skills and differences in the distribution of the returns to skills. As such, it is informative to look at our results about wages in the conjunction with the corresponding results about worker effects which measure only the quantity of skills.

Consider an hypothetical situation where the distribution of worker effects is the same in denser and less dense areas. A positive shift ($A > 0$) in the distribution of wages in denser areas relative to less dense areas as in figure 2(a) will be interpreted as the outcome of agglomeration economies or localised natural advantage that benefit all workers equally. Dilation ($D > 1$) as in figure 2(a) will reflect increasing wage differences for workers in denser areas as lower and higher skills are considered. A compression is the opposite of dilation ($D < 1$) and implies instead lower wage differences at both tails of the distribution in denser areas.

Consider now a situation where the distribution of worker effects differs across denser and less dense area. When comparing distributions of worker effects instead of distribution of wages, the shift and dilation parameters take a different meaning and reflect different forms of imperfect sorting. A positive shift in the distribution of worker effects in denser areas relative to less dense areas reveal the over-representation of skilled workers and the under-representation of less skilled workers. Dilation will instead imply the over-representation of workers at both ends of the skill distribution in denser areas.

Finally, we are also interested in truncation, in both the upper and lower tails. Truncation can arise for two reasons. The first is what we may call 'true truncation'. For wages, this may reflect institutional differences such as a minimum wage which may be more binding in less dense areas relative to denser areas. For worker effects, left- and right-truncations are natural outcomes of many models of sorting. For instance, Nocke (2006) and Venables (2011) both predict that the distribution of worker effects in denser areas will be left-truncated and negatively right-truncated relative to less dense areas. Sorting in Eeckhout *et al.* (2010) implies a positive dilation in denser

Figure 3: Wages and worker effects in denser vs. less dense areas in 2007

Panel (a): Log wages   Panel (b): Worker effects
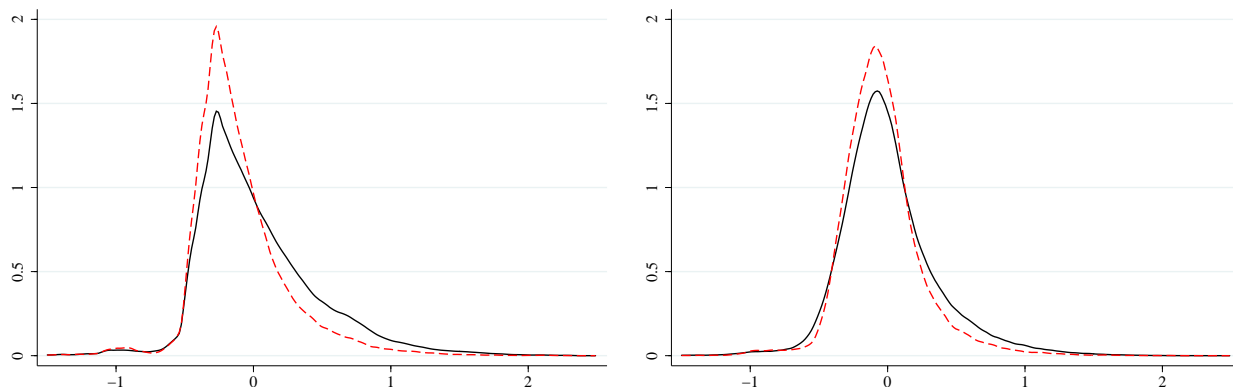
Continuous black line: denser areas.
Dashed grey (red) line: less dense areas.

Table 4: Log wages in denser vs. less dense areas, 2007

	$\hat{A}$	$\hat{D}$	$\underline{\hat{S}}$	$\bar{\hat{S}}$	R^2
w_L to w_H (shift only)	0.123 (0.001) *	-	-	-	0.631
w_L to w_H (shift and dilation)	0.127 (0.001) *	1.354 (0.006) *	-	-	0.987
w_L to w_H (shift and left- and right-truncations)	0.236 (0.270)	-	-0.456 (1.804)	-0.065 (0.078)	0.946
w_L to w_H (shift, dilation, and left-truncation)	0.125 (0.001) *	1.358 (0.006) *	0.003 (0.002) *	-	0.987
w_L to w_H (shift, dilation, left- and right-truncations)	0.129 (0.002) *	1.572 (0.013) *	0.051 (0.003) *	0.025 (0.002) *	0.999

*: for $\hat{A}$, $\underline{\hat{S}}$, and $\bar{\hat{S}}$ 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.
94,945 observations in less dense areas. 471,172 observations in denser areas.

areas whereas Behrens *et al.* (2010) predict a positive shift and (Davis, 2009) predicts a mixture of shift, dilation, and truncation.

Related to the description of figure 2(d), truncation can occur for a second reason. Imagine for instance that dilation in the distribution of skills in dense areas is stronger in the middle of the distribution and less strong at both tails. In our estimation this will be reflected in dilation and in truncations at both tails. Put differently, truncation can mitigate the effect of dilation and capture more subtle transformations of distributions. We call this form of truncation, 'mitigating truncation'. To distinguish between true and mitigating truncation, we note that mitigating truncation occurs together with dilation and should maintain a 'larger' support for the transformed distribution. True truncation implies instead a 'smaller' support for the transformed distribution. Symmetrically, negative truncations can also occur to mitigate the effects of compression when the

latter is less important in the tails of the reference distribution.

Panel (a) of figure 3 plots the distributions of log wages in denser and less dense areas in 2007.[8] Table 4 reports estimation results regarding how these two distributions relate to each other. Row 1 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 0.123 indicating that wages are 13.1% higher ($= \exp(0.123) - 1$) 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 0.123 for the shift in log wage is equivalent to an (arc) elasticity of wages with respect to employment density of $0.123/2.8 = 0.044$. 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 slightly more than 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.

Row 2 of table 4 adds a dilation parameter to the estimation. We obtain an estimate of 1.354 for this coefficient and the shift coefficient marginally increases to 0.127. 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 close to 99% of the mean-squared quantile difference between the two distributions can now be accounted for. This estimate of 1.354 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 0.8% more than a worker at the first decile of the wage distribution in less dense areas. For workers at the first quartile of the two distributions, the predicted difference is larger at 4.7%. For workers at the median, the predicted difference is again larger at 9.7%. For workers at the third quartile, the predicted difference is even larger at 18.2%. Finally, for workers at the top decile, the predicted difference is 30.7%. In the data, these difference are 1.6%, 3.6%, 10.5%, 22.0%, and 34.3%, 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 around the first quartile and under-predict for the other other quartiles.

To improve on these results and reach our most complete estimation, we proceed in steps. In row 3 of table 4 we start with estimating a specification with a shift parameter together with left- and right- truncation parameters. We find strong negative truncation, especially left-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 distri-

---

[8]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).

bution 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 less dense areas as reference, that transformation becomes a negative truncation at both tails. We note that the same difference between our two wage distributions is approximated by a dilation in row 2. Importantly, the R^2 for this specification is lower than with a dilation parameter in row 2. In addition, the truncation parameters we estimate are not significant. All this suggests that the dilation parameter captures the transformation of the distribution of log wages between less dense and denser areas much better than the two truncation parameters.

In row 4 of table 4, we estimate a left-truncation coefficient jointly with a shift and a dilation coefficient. 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 together 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 tails of the distribution. Overall, despite positive truncation, the dilation coefficient 'dominates' and we are in a situation of mitigating truncation with an 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. Row 1, which only allows for a shift in the distribution of worker effects in less dense areas to approximate the same distribution in denser areas reports a coefficient of 0.050. 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 0.044 estimated above and this number is 0.026 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

Table 5: Worker effects in denser vs. less dense employment areas, 2007

	$\hat{A}$	$\hat{D}$	$\hat{\underline{S}}$	$\hat{\bar{S}}$	R^2
δ_L to δ_H (shift only)	0.050 (0.001) *	-	-	-	0.403
δ_L to δ_H (shift and dilation)	0.052 (0.001) *	1.275 (0.005) *	-	-	0.983
δ_L to δ_H (shift and left- and right-truncations)	0.040 (0.003) *	-	−0.053 (0.007) *	−0.050 (0.006) *	0.949
δ_L to δ_H (shift, dilation, and left-truncation)	0.044 (0.002) *	1.299 (0.008) *	0.014 (0.003) *	-	0.987
δ_L to δ_H (shift, dilation, left- and right-truncations)	0.044 (0.002) *	1.273 (0.018) *	0.009 (0.005) *	0.003 (0.002) *	0.988

*: for $\hat{A}$, $\hat{\underline{S}}$, and $\hat{\bar{S}}$ 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.
94,945 observations in less dense areas. 471,172 observations in denser areas.

result indicates that a little less than half of the difference in wages across areas can be accounted for by differences in worker effects. We note nonetheless that this shift parameter only accounts for 40% 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 again 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 0.044 that is about a third of the corresponding coefficient for wages, a dilation parameter of 1.373 that is lower than for wages, and tiny left- and right-truncations to mitigate the effect of dilation for both tails.

In results not reported here, we also considered four types of areas divided by density quartiles. We find similar results with a right-shift, a dilation, and modest mitigating truncation between any two consecutive density quartiles for both wages and worker effects. For the transformations from the bottom to the second density quartile and from the second to the third density quartile, the shift and dilation parameters are lower than reported in tables 4 and 5. The differences between the third and the top density quartile are larger. This reflects the fact that density is on average 12 times as high in the top density quartile relative to the third quartile whereas it is only about twice as high in the third density quartile relative to the second and in the second relative to the first.

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. 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.[9] Interestingly, given the presence of mitigating truncation at both tails 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. We can call this 'sorting at birth'.[10] 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 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, even when agglomeration effects are heterogeneous and provided they are monotone in skills, we can at least identify 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

---

[9]We focus our discussion on the sorting on 'unobservables' which in our case accounts for most of the variation in wages that we can explain. To a large extent this is because we only have age and its square as time-varying observable characteristics. Sorting on age goes in the opposite direction relative to sorting on unobservables since less dense areas have a higher proportion of older workers and wages tend to increase with age. However such sorting is modest in magnitude and age accounts only for a small fraction of wages so that the overall effect of such sorting is tiny.

[10]Even if such sorting takes place when entering the labour market. Typically, it can occurs from higher parental skills or better education in denser areas.

Figure 4: Log wages of stayers vs. migrants in denser and less dense areas, 2007
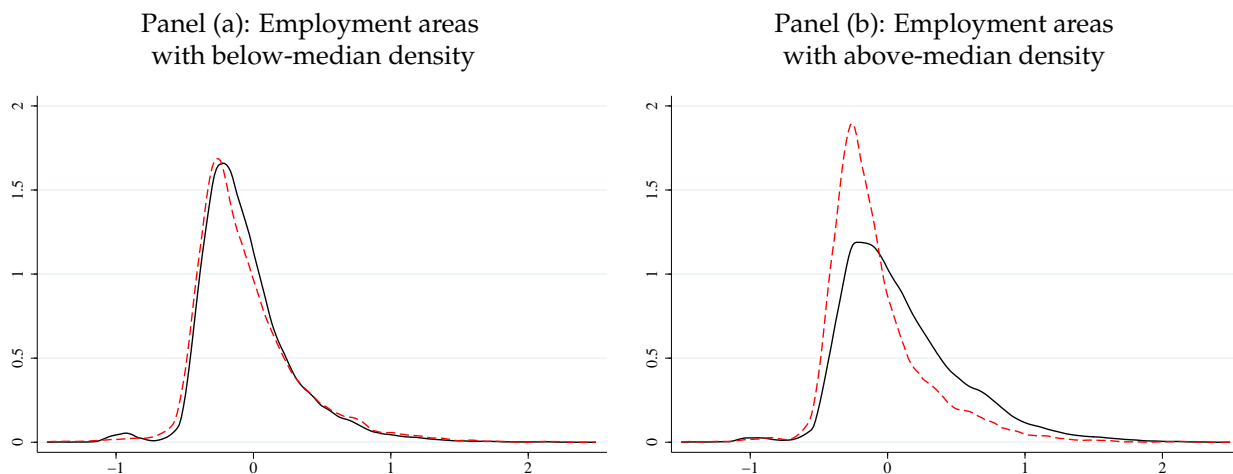
| Panel (a): Employment areas with below-median density | Panel (b): Employment areas with above-median density |

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

Table 6: Log wages of stayers vs. migrants in denser and less dense areas, 2007

	$\hat{A}$	$\hat{D}$	R^2	Obs. 1	Obs. 2
w_{LL} to $w_{L \to H}$	-0.005 (0.005)	1.100 (0.015) *	0.890	57,938	7,847
w_{HH} to $w_{L \to H}$	-0.160 (0.004) *	0.815 (0.011) *	0.982	305,615	7,847
w_{HH} to $w_{H \to L}$	-0.183 (0.005) *	0.784 (0.012) *	0.970	305,615	7,260
w_{LL} to $w_{H \to L}$	-0.028 (0.005) *	1.056 (0.017) *	0.755	57,938	7,260

*: for $\hat{A}$ 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.

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 density of the employment area of work in 2004 with that in 2007. Panel (a) of figure 4 plots the distribution of log wages of stayers within 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 with a shift and a dilation parameter. Table 14 in Appendix B provides a more complete set of results with two additional truncation parameters.

Row 1 of table 6 assesses how the distribution of the log wages of migrants from less dense to denser areas is best approximated by shifting and dilating the distribution of the log wages of stayers in less dense areas. The results indicate that migrants from less dense areas receive

Figure 5: Worker effects of stayers vs. migrants in denser and less dense areas, 2007
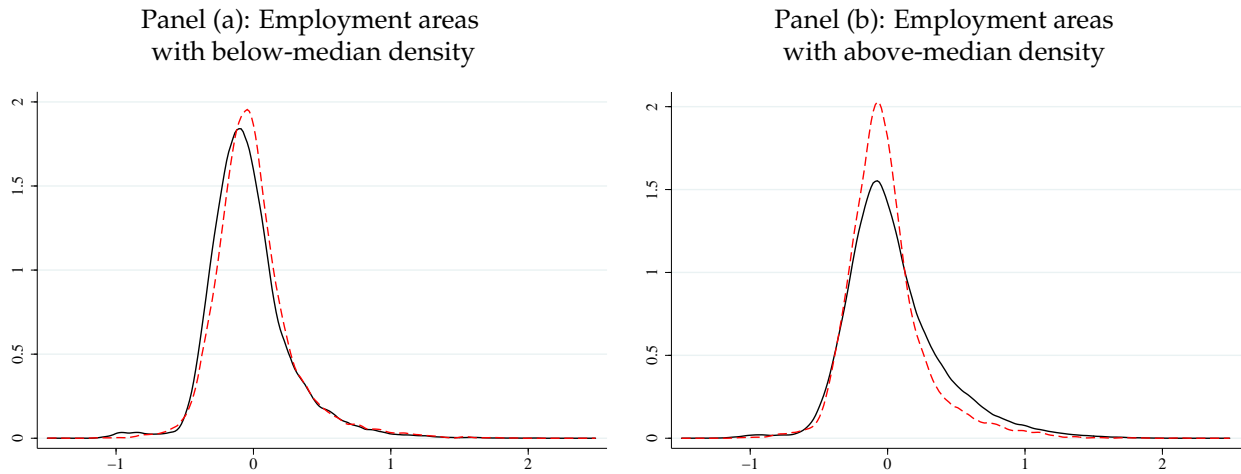
Panel (a): Employment areas
with below-median density

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

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

on average marginally lower wages than stayers in less dense areas but the difference is not significant. The distribution of the wages of these migrants is also mildly dilated relative to that of stayers.

Row 2 compares stayers in denser areas to migrants arriving from less dense areas. The wage difference between them is large, even larger than the average wage difference between workers in denser and less dense areas. The wage distribution of migrants to denser areas is also more compressed than that of stayers in denser areas.

Row 3 reports results regarding the distribution of the log wages of migrants from denser to less dense areas relative to the distribution of the log wages of stayers in denser areas. The results indicate that migrants to less dense areas receive much lower wages than stayers in denser areas. The distribution of the wages of these migrants is also compressed relative to that of stayers.

Finally, row 4 provides a comparison between stayers in less dense areas and migrants from denser areas to less dense areas. Migrants also receive lower wages, albeit only modestly so. Their wage distribution is also sightly dilated relative to the stayers in less dense areas.

Panels (a) and (b) of figure 5 mirror panels (a) and (b) of figure 4 for worker effects instead of log wages. Panel (a) plots the distribution of worker effects of stayers within 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 with a shift and a dilation parameter. Table 15 in Appendix B provides a more complete set of results with two additional truncation parameters.

Row 1 of table 7 assesses how the distribution of worker effects for migrants from less dense to denser areas is best approximated by shifting and dilating the distribution of worker effects for stayers in less dense areas. The results indicate that migrants from less dense areas have on average higher worker effects than stayers. The distribution of the worker effects of these migrants to denser areas is also modestly compressed relative to that of stayers in less dense areas.

19

Table 7: Worker effects of stayers vs. migrants in denser and less dense areas, 2007

	$\hat{A}$	$\hat{D}$	R^2	Obs. 1	Obs. 2
δ_{LL} to $\delta_{L \to H}$	0.028 (0.003) *	0.973 (0.014) *	0.942	57,938	7,847
δ_{HH} to $\delta_{L \to H}$	−0.054 (0.003) *	0.756 (0.011) *	0.965	305,615	7,847
δ_{HH} to $\delta_{H \to L}$	−0.063 (0.004) *	0.770 (0.013) *	0.966	305,615	7,260
δ_{LL} to $\delta_{H \to L}$	0.019 (0.004) *	0.992 (0.017) *	0.796	57,938	7,260

*: for $\hat{A}$ 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.

Row 2 compares stayers in denser areas with migrants from less dense to denser areas and shows that the latter have on average lower worker effects. They also come from a more compressed distribution. In row 3, the comparison of stayers in denser areas with, this time, migrants from denser to less dense areas leads to broadly similar results.
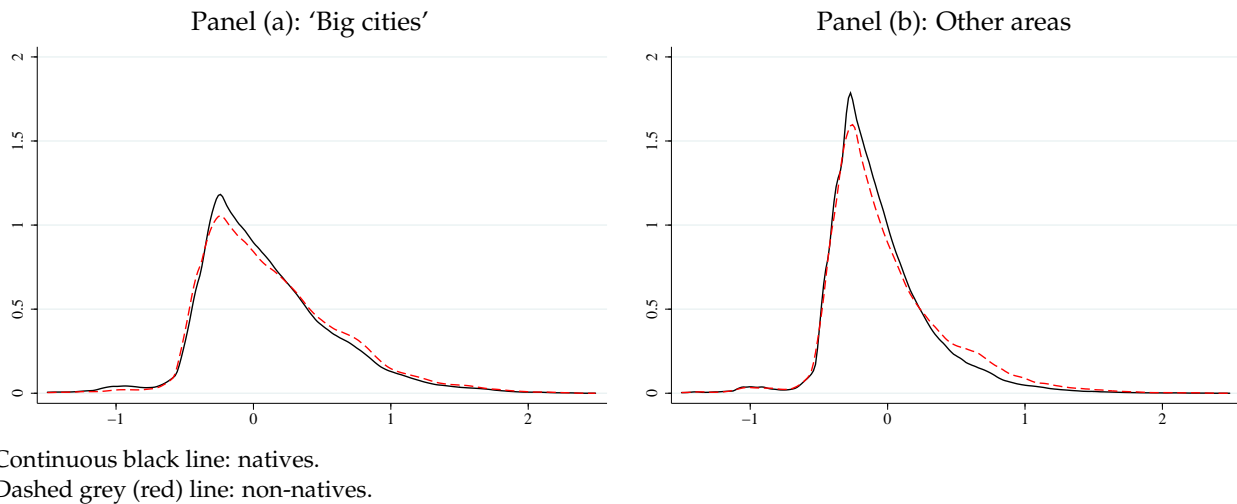
Finally, row 4 reports results regarding how the distribution of the worker effects of migrants from denser to less dense areas compares to the distribution of the log wages of stayers in less dense areas. The results indicate that migrants to less dense areas have lower worker effects than stayers in less dense areas. The distribution of the worker effects of these migrants is also strongly compressed relative to that of stayers.

Tables 14 and 15 in Appendix B which also consider two truncation parameters generally confirm these results. For wages in table 14, positive and negative truncations occur to mitigate the effects of dilation or compression. Asymmetric truncations also sometimes occur to correct for asymmetries of the transformed distribution with respect to the dilation or compression of the reference distribution. For worker effects in table 15, negative truncations occur to mitigate the effects of the estimated compressions.

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).

At the same time, migrants from less dense areas have on average higher worker effects than stayers from the same areas. Migrants from denser areas have instead on average lower fixed effects than stayers from the same areas. That is, there is on average positive selection of migrants from less dense areas and negative selection of migrants from denser areas. At their destination in denser areas, migrants from less dense areas have lower worker effects relative to the stayers in denser areas. Migrants from denser areas to less dense areas have instead higher worker effects relative to the stayers in less dense areas. Migrants from both type of areas are thus 'in-between' denser and less dense areas in terms of fixed effects. Put differently, since in-migrants and out-migrants do not differ much, short-run migration is unlikely to account for much of the sorting evidenced above. We also observe that workers with the lowest and highest worker effects in

Figure 6: Log wages of natives vs. non-natives in 'big cities' and other areas, 2007

Panel (a): 'Big cities'   Panel (b): Other areas

Continuous black line: natives.
Dashed grey (red) line: non-natives.

denser areas have a greater propensity to stay which is consistent with the over-representation documented above of these workers at both tails in denser areas.

'Sorting at birth'

Rather than compare migrants and stayers over a three-year period from 2004 to 2007, we now take a long-term perspective and define migration status by the difference between the place of work in 2007 and the place of birth. Because at birth we only know 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.

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 of wages with a shift and a dilation parameter. Table 16 provides a more complete set of results with two additional truncation parameters.

Row 1 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 and dilated relative to that in other areas. These results generally confirm those obtained above for denser and less dense employment areas. The higher coefficients for the shift and dilation also suggest that what is described above is

Table 8: Log wages of natives vs. non-natives in 'big cities' and other areas, 2007

	$\hat{A}$	$\hat{D}$	R^2	Obs. 1	Obs. 2
w_S to w_B	0.171 (0.001) *	1.380 (0.003) *	0.982	364,666	201,451
w_{SS} to $w_{S \to B}$	0.196 (0.002) *	1.458 (0.0075 *	0.984	324,420	94,773
w_{BB} to $w_{S \to B}$	0.036 (0.002) *	1.065 (0.004) *	0.932	106,678	94,773
w_{BB} to $w_{B \to S}$	−0.107 (0.002) *	0.873 (0.004) *	0.934	106,678	40,246
w_{SS} to $w_{B \to S}$	0.052 (0.002) *	1.196 (0.007) *	0.966	324,420	40,246

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

particularly acute in the four largest French cities.[11]

Row 2 compares the natives of other areas who work there to those that now work in big cities. We find that the wage differences between them are even larger than when we compare all workers in big cities and other areas as in the previous row. The same occurs for the dilation parameter.

Row 3 reports results for the comparison of wages between natives and non-natives in big cities. It shows that the distribution of wages for non-natives is slightly right-shifted and dilated relative to the same distribution for natives. Row 4 compares instead natives from big cities who still work there with those that now work in other areas. The latter receive on average lower wages which comes from a more compressed distribution.

Finally, row 5 compares natives and non-natives in other areas. It shows that the wages of non-natives are on average higher and more dispersed relative to those of natives. The results of table 16 in Appendix B generally confirm all these findings. It also provides further evidence of mitigating truncation.

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. Table 17 provides a more complete set of results with two additional truncation parameters.

Comparing the distribution of worker effects in big cities and other areas, row 1 of table 9 shows that the former is right-shifted and dilated relative to the latter. This results essentially confirm the same comparison made for the distribution of worker effects in denser and less dense areas. For

---

[11]In the data, workers at the first decile of the wage distribution in big cities make 1.9% 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 5.7%. For workers at the median, the difference is again larger at 16.6%. For workers at the third quartile, the difference is even larger at 31.1%. Finally, for workers at the top decile, the difference is 41.1%. These differences are larger than those between denser and less dense areas reported above.

Figure 7: Worker effects of natives vs. non-natives in 'big cities' and other areas, 2007

Panel (a): 'Big cities'   Panel (b): Other areas
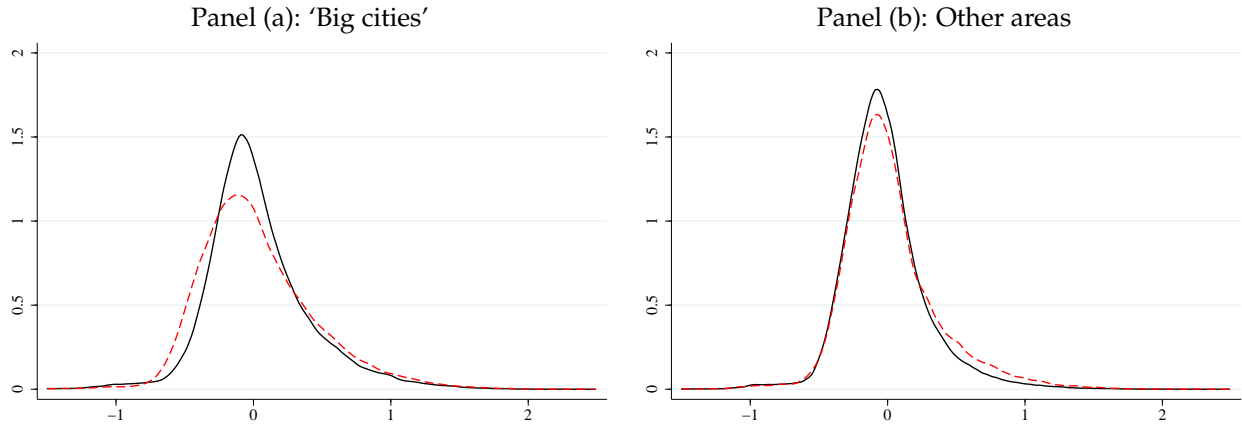
Continuous black line: natives.
Dashed grey (red) line: non-natives.

Table 9: Worker effects of natives vs. non-natives in 'big cities' and other areas, 2007

	$\hat{A}$	$\hat{D}$	R^2	Obs. 1	Obs. 2
δ_S to δ_B	0.050 (0.001) *	1.361 (0.004) *	0.985	364,666	201,451
δ_{SS} to $\delta_{S \to B}$	0.050 (0.001) *	1.498 (0.006)	0.978	324,420	94,773
δ_{BB} to $\delta_{S \to B}$	-0.009 (0.002) *	1.166 (0.004) *	0.901	106,678	94,773
δ_{BB} to $\delta_{B \to S}$	-0.017 (0.002) *	0.922 (0.005) *	0.876	106,678	40,246
δ_{SS} to $\delta_{B \to S}$	0.041 (0.002) *	1.185 (0.007)	0.925	324,420	40,246

*: for $\hat{A}$ 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 ($i = S$).
δ_{ii}: distribution of worker effects for workers born and working in an area of type i ($\in \{B,S\}$)
$\delta_{i \to j}$: 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.

worker effects, unlike with wages, we do not 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.

Row 2 of table 9 compares the worker effects of the natives of other areas who work there to those that now work in big cities. We find that the latter have on average higher fixed effects. The shift parameter is the same as in the comparison between all workers in big cities and other areas as in the previous row. The distribution of worker effects of natives of other areas that work in large cities is also highly dilated relative to those that still work in other areas.

Row 3 reports results for the comparison of worker effects between natives and non-natives in big cities. It shows that the distribution of worker effects for non-natives has a marginally lower mean and is dilated relative to the same distribution for non-natives. Row 4 compares instead

natives from big cities who still work there with those that now work in other areas. The latter have on average lower worker effects which come from a more compressed distribution.

Finally, row 5 compares natives and non-natives in other areas. It shows that the worker effects of non-natives are on average higher and more dispersed relative to those of natives. The results of table 16 in Appendix B confirm these findings while providing further evidence of mitigating truncation.

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.[12] There is also more dispersion in the wages of non-natives. For worker effects, we observe positive selection among those that left big cities and negative selection among those that left other areas. These 'long-run' results for natives and non-natives mirror the 'short-run' results obtained above for migrants and non-migrants in denser and less dense areas. This said, the differences between in-migrants and out-migrants are larger in the long run than in the short run both in terms of their average worker effect and dispersion of these worker effects. Hence migration, both in the short and the long run is a determinant of the sorting of more skilled workers in denser areas / big cities and the larger dispersion of their skills. However, the small differences between in- and out-migrants suggest a modest role for migration in accounting for sorting. In turn, this suggests an important role for 'sorting at birth' through either better parental skills transmitted to their offsprings or better education in big cities and denser areas. Alternatively, this can also suggest the existence of stronger workplace learning in big cities and denser areas.

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). In addition, looking at workers of different age groups may provide some suggestive evidence regarding the importance of workplace learning and help separate it from sorting at birth.

Table 10 reports estimation results for the distribution of wages that replicate some of our above estimations but perform them separately for 'young' workers (aged below 35) in panel (a) and the log wages of 'old' workers (aged above 45) in panel (b). Comparing row 1 of 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, and dilated relative to the same distribution in less dense areas. The main difference is that the mean wage difference between denser and less dense areas is much larger for old workers than for

---

[12]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, 2007

	$\hat{A}$	$\hat{D}$	R^2	Obs. 1	Obs. 2
Panel (a): Young workers (below 35)					
w_L to w_H	0.084 (0.002) *	1.397 (0.009) *	0.985	37,399	198,825
w_{LL} to $w_{L \to H}$	0.027 (0.005) *	1.178 (0.023) *	0.970	16,580	4,518
w_{HH} to $w_{L \to H}$	−0.092 (0.005) *	0.841 (0.014) *	0.990	99,891	4,518
w_{HH} to $w_{H \to L}$	−0.120 (0.004) *	0.760 (0.014) *	0.984	99,891	3,790
w_{LL} to $w_{H \to L}$	−0.002 (0.005)	1.065 (0.022) *	0.716	16,580	3,790
Panel (b): Old workers (above 45)					
w_L to w_H	0.171 (0.002) *	1.312 (0.037) *	0.988	31,270	142,460
w_{LL} to $w_{L \to H}$	0.062 (0.013) *	1.298 (0.027) *	0.966	23,546	1,365
w_{HH} to $w_{L \to H}$	−0.122 (0.013) *	0.986 (0.011)	0.904	111,231	1,365
w_{HH} to $w_{H \to L}$	−0.169 (0.012) *	0.896 (0.026) *	0.953	111,231	1,393
w_{LL} to $w_{H \to L}$	0.014 (0.012)	1.179 (0.035) *	0.852	23,546	1,393

*: for $\hat{A}$ significantly different from 0 at 5%; for $\hat{D}$ significantly different from 1 at 5%.
w_i: distribution of log wages for all workers in less dense areas ($i = L$) or denser areas ($i = H$).
w_{ii}: distribution of log wages for stayers in less dense areas ($i = L$) or denser areas ($i = H$).
$w_{i \to j}$: distribution of log wages 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.

young workers. Nonetheless, taken together these two groups essentially behave like the overall sample in table 4.

In rows 2 to 5 of the two panels of table 10, we provide a number of comparisons between migrants and stayers in denser and less dense areas for young and old workers. Looking first at migrants from less dense areas to denser areas in row 2 of both panels, we find results that confirm those of row 1. Young workers moving to denser areas receive a wage that is higher than the stayers in less dense areas. For old workers moving to denser areas this difference is larger. This suggests that the negative effect found for all workers therefore resulted from the over-representation of young workers among migrants. Moreover for both groups of workers the difference in wage is less than the difference in wage between denser and less dense areas. Turning next to the comparison between migrants from less dense areas and stayers in denser areas in row 3, we find again more pronounced differences for old workers. Similar features are observed in rows 4 and 5 with again larger differences between less dense and denser areas for old workers than for young workers.

To shed further light on this question, table 11 reports estimation results for worker effects of old

Table 11: Worker effects of young and old stayers vs. young and old migrants in denser and less dense areas, 2007

	$\hat{A}$	$\hat{D}$	R^2	Obs. 1	Obs. 2
Panel (a): Young workers (below 35)					
δ_L to δ_H	−0.008 (0.001) *	1.236 (0.008) *	0.969	37,399	198,825
δ_{LL} to $\delta_{L\to H}$	0.010 (0.003) *	1.068 (0.020) *	0.968	16,580	4,518
δ_{HH} to $\delta_{L\to H}$	−0.005 (0.003)	0.855 (0.014) *	0.895	99,891	4,518
δ_{HH} to $\delta_{H\to L}$	−0.019 (0.003) *	0.831 (0.015) *	0.875	99,891	3,790
δ_{LL} to $\delta_{H\to L}$	−0.004 (0.003)	1.038 (0.020)	0.586	16,580	3,790
Panel (b): Old workers (above 45)					
δ_L to δ_H	0.112 (0.002) *	1.301 (0.009) *	0.986	31,270	142,460
δ_{LL} to $\delta_{L\to H}$	0.045 (0.011) *	1.227 (0.037) *	0.915	23,546	1,365
δ_{HH} to $\delta_{L\to H}$	−0.081 (0.011) *	0.939 (0.027) *	0.812	111,231	1,365
δ_{HH} to $\delta_{H\to L}$	−0.085 (0.011) *	0.898 (0.026) *	0.899	111,231	1,393
δ_{LL} to $\delta_{H\to L}$	0.042 (0.011) *	1.174 (0.035) *	0.925	23,546	1,393

*: for $\hat{A}$ 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.

and young workers that complement the wage results of table 10. Row 1 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 marginally lower mean wage relative to young workers in less dense areas. The difference in mean worker effect is −0.008. We find the opposite for old workers and the positive difference is sizeable at 0.112. 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 0.092 for young workers in denser areas against only 0.069 for old workers after conditioning out worker effects.

Turning to migrants from less dense areas in row 2 of both panels of table 11, we find that the young migrants to a denser area have a slightly higher mean worker effect relative to young

stayers in less dense areas. For old migrants to a denser area, the difference in fixed effect is larger. When we now compare these migrants from less dense areas to the stayers in denser areas in row 3, the difference is again small for young workers and large for old workers. For both age groups, the distribution of the fixed effect of migrants is compressed relative to that of stayers in denser areas.

For migrants from denser areas in row 4 of both panels, we find that young migrants to a less dense area have on average a mildly lower fixed effect relative to stayers in denser areas. For old workers from a denser area, the differences are large: Old migrants from denser areas have markedly lower worker effects. Finally in row 5, the comparison of the same migrants to less dense areas but this time with the stayers in less dense areas shows little to no difference for young workers and larger effects again for old workers.

Tables 18 and 19 in Appendix B extend the results of tables 10 and 11 and estimate two truncation parameters in addition to a shift and a truncation parameter. They generally confirm the results we describe here and truncations usually mitigate the effects of dilation or compression. For the wage of old migrants to less dense areas, we find large negative shift parameters and large levels of truncation. We attribute these findings to a specification problem. These are rare cases where the distribution for the group of workers at hand relative to the reference group is not well captured by a shift and a dilation. Adding two truncation parameters can lead to large estimated truncation parameters to mitigate the effect of large estimated dilation and shifts parameters. We also note that the number of old migrants is typically small.

We draw a number of conclusions from these results. First, looking at all workers 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. The fact that the difference in worker effects does not increase much with age when we compare migrants from less dense areas to stayers in less dense areas whereas the difference between these migrants and the stayers in denser areas shows a much more pronounced increase with age is suggestive that stronger learning in denser areas might be at play. However, we need to remain cautious because we cannot exclude that the sorting of migrants becomes more pronounced with age.

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 peek at heterogeneity across workers, we now replicate some of the estimations performed above but separate workers by their one-digit occupational category.

Table 12: Log wages of stayers vs. migrants in professional occupations in denser and less dense areas, 2007

	$\hat{A}$	$\hat{D}$	R^2	Obs. 1	Obs. 2
w_L to w_H	0.077 (0.007) *	1.051 (0.016) *	0.957	5,295	66,629
w_{LL} to $w_{L \to H}$	−0.051 (0.016) *	1.058 (0.036)	0.840	3,525	864
w_{HH} to $w_{L \to H}$	−0.134 (0.015) *	0.996 (0.031)	0.991	50,811	864
w_{HH} to $w_{H \to L}$	−0.145 (0.014) *	0.981 (0.035)	0.977	50,811	698
w_{LL} to $w_{H \to L}$	−0.062 (0.014) *	1.042 (0.042)	0.978	3,525	698

*: for $\hat{A}$ significantly different from 0 at 5%; for $\hat{D}$ significantly different from 1 at 5%.
w_i: distribution of log wages for all workers in less dense areas ($i = L$) or denser areas ($i = H$).
w_{ii}: distribution of log wages for stayers in less dense areas ($i = L$) or denser areas ($i = H$).
$w_{i \to j}$: distribution of log wages 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 second row of table 4 which performs the same comparison for all workers, there is again evidence of a higher mean wage and a dilation of the distribution 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 0.077 for workers in professional occupations instead of 0.127 for all workers. Dilation is also markedly less: 1.051 instead of 1.354.

Turning to migrants in rows 2 to 5, there is evidence of 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 differences between migrants and stayers for workers in professional occupations are not very different. The salient feature of the wage results for migrants in professional occupations relative to all migrants is that there is 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. Row 1 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 close to zero and the dilation coefficient is modest at 1.040. This is in sharp contrast with the corresponding result for all workers in table 5, which showed both a significant positive shift coefficient of 0.050 and a larger dilation coefficient of 1.275. 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

Table 13: Worker effects of stayers vs. migrants in professional occupations in denser and less dense areas, 2007

	$\hat{A}$	$\hat{D}$	R^2	Obs. 1	Obs. 2
δ_L to δ_H	0.001 (0.006)	1.040 (0.014) *	0.663	5,295	66,629
δ_{LL} to $\delta_{L \to H}$	−0.091 (0.013) *	1.054 (0.036)	0.979	3,525	864
δ_{HH} to $\delta_{L \to H}$	−0.094 (0.013) *	0.991 (0.031)	0.987	50,811	864
δ_{HH} to $\delta_{H \to L}$	−0.073 (0.012) *	0.996 (0.035)	0.976	50,811	698
δ_{LL} to $\delta_{H \to L}$	−0.070 (0.012) *	1.060 (0.042)	0.966	3,525	698

*: for $\hat{A}$ 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.

between the same two groups. For instance, for all workers the unconditional wage premium is 0.127 and $0.127 - 0.050 = 0.077$ after conditioning out worker effects. For workers in professional occupations, the unconditional wage premium is 0.077 and $0.077 - 0.001 = 0.076$ 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. Relative to the findings for workers across all occupations in table 7, the contrast is stark. Workers in professional occupations migrating to a denser area have lower worker effects than the stayers in less dense areas, unlike what we find above for all workers. They also have lower worker effects than the stayers in denser areas. In addition, workers in professional occupations who migrate to less dense areas also have much lower worker effects relative to stayers both at their origin and destination. 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 migrants relative to stayers across all occupations in table 7.

The results of table 20 and 21 in Appendix B which also estimate two truncation parameters generally confirm these findings. The truncation parameters we estimate in these two tables are generally insignificant and the other two parameters do not change much.

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 larger 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 (2012) 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 the comparison between 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, albeit 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. However, a long-run examination of natives and non-natives in the largest French cities suggests that selection into migration can account for only part of the observed patterns of sorting, if at all. 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. Finally, we find little evidence about the extreme form of segmentation predicted by many models of sorting. In practice, sorting implies that some skills are over- or under-represented in some areas. We failed to find any evidence of 'true' truncations in the distributions of wages and worker effects.

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 denser or less dense areas). 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. Hopefully future research will be able to disentangle further the various sources of sorting.

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. 2012. 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.

Ciccone, Antonio and Giovanni Peri. 2006. Identifying human capital externalities: Theory with an applications. *Review of Economic Studies* 73(2):381–412.

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 (MA): 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. 2012. The dynamic earnings premium of dense cities. Processed, 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. 2004. Human capital externalities in cities. In Vernon Henderson and Jacques-François Thisse (eds.) *Handbook of Regional and Urban Economics*, volume 4. Amsterdam: North-Holland, 2243–2291.

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

Nocke, Volker. 2006. A gap for me: Entrepreneurs and entry. *Journal of the European Economic Association* 4(5):929–956.

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.

Appendix A. Empirical strategy with truncations

In this appendix, we present the second part of our methodology when the distribution 2 is allowed to be translated, dilated and truncated relative to distribution 1.

Consider first a situation where distribution 2 is both 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:

$$\lambda_2(u) = D\lambda_1\left(\underline{S} + (1 - \overline{S} - \underline{S})u\right) + A, \quad \text{for } u \in [0,1] . \tag{A1}$$

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:

$$\lambda_2\left(\underline{T} + (1 - \overline{T} - \underline{T})u\right) = D\lambda_1(u) + A, \quad \text{for } u \in [0,1] . \tag{A2}$$

This equation can be rewritten in a form similar to (A1). We make the change of variable $u \to \frac{u - \underline{T}}{1 - \overline{T} - \underline{T}}$ in equation (A2) which becomes:

$$\lambda_2(u) = D\lambda_1\left(\frac{u - \underline{T}}{1 - \overline{T} - \underline{T}}\right) + A, \quad \text{for } u \in \left[\underline{T}, 1 - \overline{T}\right] . \tag{A3}$$

Introducing $\underline{S} = \frac{-\underline{T}}{1 - \overline{T} - \underline{T}}$ and $\overline{S} = \frac{\overline{T}}{1 - \overline{T} - \underline{T}}$, equation (A3) can be rewritten as:

$$\lambda_2(u) = D\lambda_1\left(\underline{S} + (1 - \overline{S} - \underline{S})u\right) + A, \quad \text{for } u \in \left[\frac{-\underline{S}}{1 - \overline{S} - \underline{S}}, \frac{1 - \underline{S}}{1 - \overline{S} - \underline{S}}\right] . \tag{A4}$$

It is possible to nest the two cases of left and right-truncation (A1) and (A4) into one equation:

$$\lambda_2(u) = D\lambda_1\left(\underline{S} + (1 - \overline{S} - \underline{S})u\right) + A, \quad \text{for } u \in \left[\max\left(0, \frac{-\underline{S}}{1 - \overline{S} - \underline{S}}\right), \min\left(1, \frac{1 - \underline{S}}{1 - \overline{S} - \underline{S}}\right)\right] . \tag{A5}$$

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 (A5) 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:

$$\lambda_2\left(r_S(u) \right) = D\lambda_1\left(\underline{S} + (1 - \overline{S} - \underline{S})r_S(u) \right) + A , \quad \text{for } u \in [0,1] . \tag{A6}$$

The estimation procedure is based on this equation and matches that of Combes *et al.* (2011b) 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 (A6) can be rewritten as a continuum of equalities of the form $m_\theta(u) = 0$ for $u \in [0,1]$ where:

$$m_\theta(u) = \lambda_2\left(r_S(u) \right) - D\lambda_1\left(\underline{S} + (1 - \overline{S} - \underline{S})r_S(u) \right) - A . \tag{A7}$$

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 (A5), we get:

$$\lambda_1(u) = \frac{1}{D}\lambda_2\left(\frac{u - \underline{S}}{1 - \overline{S} - \underline{S}} \right) - \frac{A}{D}, \quad \text{for } u \in \left[\max\left(0, 1 - \underline{S} \right), \min\left(1 - \overline{S}, 1 \right) \right] . \tag{A8}$$

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

$$\widetilde{m}_\theta(u) = \lambda_1\left(\widetilde{r}(u) \right) - \frac{1}{D}\lambda_2\left(\frac{\widetilde{r}(u) - \underline{S}}{1 - \overline{S} - \underline{S}} \right) + \frac{A}{D} . \tag{A9}$$

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 $\widehat{\widetilde{m}}_\theta(u)$:

$$\widehat{\theta} = \arg\min_\theta M(\theta), \quad \text{where} \quad M(\theta) = \int_0^1 \left[\widehat{m}_\theta(u) \right]^2 \mathrm{d}u + \int_0^1 \left[\widehat{\widetilde{m}}_\theta(u) \right]^2 \mathrm{d}u . \tag{A10}$$

The confidence intervals are obtained by bootstrap. The implementation details are given in Combes *et al.* (2011b). Finally, we construct a measure of fit that uses the ratio of mean-squared quantile differences between the transformed distribution 1 and distribution 2 with mean-squared quantile differences between distributions 1 and 2. This measure of fit is a Pseudo-R^2 defined as $R^2 = 1 - M(\widehat{\theta}) / M(0,1,0,0)$.

Appendix B. Estimations with truncations

Stayers vs. migrants

Table 14: Log wages of stayers vs. migrants in denser and less dense areas, 2007

	$\hat{A}$	$\hat{D}$	$\underline{\hat{S}}$	$\bar{\hat{S}}$	R^2	Obs. 1	Obs. 2
w_{LL} to $w_{L \to H}$	-0.025 (0.005) *	1.116 (0.039) *	0.038 (0.016) *	-0.001 (0.006)	0.961	57,938	7,847
w_{HH} to $w_{L \to H}$	-0.186 (0.005) *	0.690 (0.017) *	-0.018 (0.008) *	-0.039 (0.006) *	0.999	305,615	7,847
w_{HH} to $w_{H \to L}$	-0.226 (0.006) *	0.587 (0.032) *	-0.029 (0.018)	-0.070 (0.016) *	0.998	305,615	7,260
w_{LL} to $w_{H \to L}$	-0.060 (0.006) *	0.954 (0.054) *	0.021 (0.025)	-0.024 (0.011) *	0.942	57,938	7,260

*: for $\hat{A}$, $\underline{\hat{S}}$, and $\bar{\hat{S}}$ significantly different from 0 at 5%; for $\hat{D}$ significantly different from 1 at 5%.
w_{ii}: distribution of log wages for stayers in less dense areas ($i = L$) or denser areas ($i = H$).
$w_{i \to j}$: distribution of log wages 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 15: Worker effects of stayers vs. migrants in denser and less dense areas, 2007

	$\hat{A}$	$\hat{D}$	$\underline{\hat{S}}$	$\bar{\hat{S}}$	R^2	Obs. 1	Obs. 2
δ_{LL} to $\delta_{L \to H}$	0.032 (0.003) *	0.904 (0.023) *	-0.026 (0.004) *	-0.009 (0.023) *	0.993	57,938	7,847
δ_{HH} to $\delta_{L \to H}$	-0.038 (0.004) *	0.652 (0.016) *	-0.076 (0.004) *	-0.017 (0.016) *	0.993	305,615	7,847
δ_{HH} to $\delta_{H \to L}$	-0.051 (0.005) *	0.629 (0.025) *	-0.090 (0.006) *	-0.028 (0.025) *	0.994	305,615	7,260
δ_{LL} to $\delta_{H \to L}$	0.017 (0.004) *	0.898 (0.023) *	-0.024 (0.004) *	-0.015 (0.023) *	0.979	57,938	7,260

*: for $\hat{A}$, $\underline{\hat{S}}$, and $\bar{\hat{S}}$ 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.

Natives and non-natives in big and other cities

Table 16: Log wages of natives vs. non-natives in 'big cities' and other areas, 2007

	$\hat{A}$	$\hat{D}$	$\hat{\underline{S}}$	$\hat{\bar{S}}$	R^2	Obs. 1	Obs. 2
w_S to w_B	0.212 (0.002) *	1.684 (0.015) *	0.032 (0.003) *	0.044 (0.002) *	0.999	364,666	201,451
w_{SS} to $w_{S\rightarrow B}$	0.228 (0.003) *	1.827 (0.018) *	0.052 (0.003) *	0.048 (0.002) *	0.999	324,420	94,773
w_{BB} to $w_{S\rightarrow B}$	0.033 (0.003) *	1.106 (0.008) *	0.015 (0.003) *	0.007 (0.002) *	0.984	106,678	94,773
w_{BB} to $w_{B\rightarrow S}$	−0.166 (0.006) *	0.700 (0.015) *	0.002 (0.002)	−0.066 (0.008) *	0.991	106,678	40,246
w_{SS} to $w_{B\rightarrow S}$	0.023 (0.005) *	1.263 (0.019) *	0.064 (0.011) *	0.005 (0.002) *	0.985	324,420	40,246

*: for $\hat{A}$, $\hat{\underline{S}}$, 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 for all workers working in big cities ($i = B$) or other areas ($i = S$).
$w_{i\rightarrow j}$: distribution of log wages 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.

Table 17: Worker effects of natives vs. non-natives in 'big cities' and other areas, 2007

	$\hat{A}$	$\hat{D}$	$\hat{\underline{S}}$	$\hat{\bar{S}}$	R^2	Obs. 1	Obs. 2
δ_S to δ_B	0.043 (0.003) *	1.478 (0.019) *	0.029 (0.006) *	0.011 (0.002) *	0.993	364,666	201,451
δ_{SS} to $\delta_{S\rightarrow B}$	0.022 (0.003) *	1.791 (0.018) *	0.076 (0.005) *	0.023 (0.001) *	0.998	324,420	94,773
δ_{BB} to $\delta_{S\rightarrow B}$	−0.002 (0.002)	1.328 (0.015) *	0.029 (0.002) *	0.024 (0.002) *	0.990	106,678	94,773
δ_{BB} to $\delta_{B\rightarrow S}$	−0.029 (0.002) *	0.891 (0.011) *	0.006 (0.002) *	−0.010 (0.003) *	0.979	106,678	40,246
δ_{SS} to $\delta_{B\rightarrow S}$	0.017 (0.002) *	1.169 (0.035) *	0.064 (0.010) *	−0.010 (0.005) *	0.981	324,420	40,246

*: for $\hat{A}$, $\hat{\underline{S}}$, 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 ($i = S$).
$\delta_{i\rightarrow j}$: 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.

Young and Older workers

Table 18: Log wages of young and old stayers vs. migrants in denser and less dense areas, 2007

	$\hat{A}$	$\hat{D}$	$\underline{\hat{S}}$	$\overline{\hat{S}}$	R^2	Obs. 1	Obs. 2
Panel (a): Young workers (below 35)							
w_L to w_H	0.083 (0.005) *	1.453 (0.047) *	0.007 (0.002) *	0.004 (0.008)	0.990	37,399	198,825
w_{LL} to $w_{L \to H}$	0.026 (0.007) *	1.186 (0.051) *	0.003 (0.020)	0.001 (0.006)	0.971	16,580	4,518
w_{HH} to $w_{L \to H}$	−0.095 (0.005) *	0.785 (0.022) *	−0.016 (0.010)	−0.013 (0.005) *	0.998	99,891	4,518
w_{HH} to $w_{H \to L}$	−0.113 (0.006) *	0.598 (0.029) *	−0.111 (0.034) *	−0.045 (0.011) *	0.998	99,891	3,790
w_{LL} to $w_{H \to L}$	−0.011 (0.005) *	0.991 (0.034)	−0.003 (0.008)	−0.016 (0.009)	0.893	16,580	3,790
Panel (b): Old workers (above 45)							
w_L to w_H	0.204 (0.005) *	1.508 (0.024) *	0.018 (0.004) *	0.031 (0.004) *	0.999	31,270	142,460
w_{LL} to $w_{L \to H}$	0.009 (0.066)	1.251 (0.097) *	0.055 (0.082)	−0.014 (0.013)	0.995	23,546	1,365
w_{HH} to $w_{L \to H}$	−0.196 (0.116)	0.812 (0.104)	0.024 (0.113)	−0.053 (0.023) *	0.994	111,231	1,365
w_{HH} to $w_{H \to L}$	−1.045 (0.405) *	1.176 (0.211)	0.791 (0.305) *	−0.003 (0.027)	0.998	111,231	1,393
w_{LL} to $w_{H \to L}$	−0.554 (0.386)	1.354 (0.125) *	0.679 (0.276) *	0.002 (0.006)	0.971	23,546	1,393

*: for $\hat{A}$, $\underline{\hat{S}}$, and $\overline{\hat{S}}$ significantly different from 0 at 5%; for $\hat{D}$ significantly different from 1 at 5%.
w_i: distribution of log wages for all workers in less dense areas ($i = L$) or denser areas ($i = H$).
w_{ii}: distribution of log wages for stayers in less dense areas ($i = L$) or denser areas ($i = H$).
$w_{i \to j}$: distribution of log wages 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 19: Worker effects of young and old stayers vs. migrants in denser/less dense areas, 2007

	$\hat{A}$	$\hat{D}$	$\hat{\underline{S}}$	$\hat{\bar{S}}$	R^2	Obs. 1	Obs. 2
Panel (a): Young workers (below 35)							
δ_L to δ_H	−0.014 (0.002) *	1.239 (0.019) *	0.004 (0.001) *	−0.005 (0.004)	0.994	37,399	198,825
δ_{LL} to $\delta_{L \to H}$	0.009 (0.003) *	1.068 (0.026) *	0.001 (0.004)	−0.001 (0.005)	0.968	16,580	4,518
δ_{HH} to $\delta_{L \to H}$	0.004 (0.003)	0.842 (0.024) *	−0.016 (0.005) *	0.006 (0.006)	0.990	99,891	4,518
δ_{HH} to $\delta_{H \to L}$	−0.008 (0.004) *	0.760 (0.024) *	−0.035 (0.008) *	−0.004 (0.006)	0.985	99,891	3,790
δ_{LL} to $\delta_{H \to L}$	−0.001 (0.003)	0.985 (0.029)	−0.013 (0.007)	−0.005 (0.005)	0.827	16,580	3,790
Panel (b): Old workers (above 45)							
δ_L to δ_H	0.139 (0.004) *	1.422 (0.017) *	0.001 (0.002)	0.021 (0.003) *	0.999	31,270	142,460
δ_{LL} to $\delta_{L \to H}$	−0.009 (0.019)	1.103 (0.078)	0.046 (0.082)	−0.028 (0.013)	0.991	23,546	1,365
δ_{HH} to $\delta_{L \to H}$	−0.153 (0.024) *	0.749 (0.063) *	0.028 (0.113)	−0.058 (0.023) *	0.993	111,231	1,365
δ_{HH} to $\delta_{H \to L}$	−0.160 (0.113)	0.752 (0.114) *	0.051 (0.305)	−0.056 (0.027)	0.981	111,231	1,393
δ_{LL} to $\delta_{H \to L}$	−0.042 (0.160)	1.263 (0.117) *	0.165 (0.276)	0.004 (0.006)	0.976	23,546	1,393

*: for $\hat{A}$, $\hat{\underline{S}}$, 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 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.

Professionals

Table 20: Log wages of stayers vs. migrants in professional occupations in denser and less dense areas, 2007

	$\hat{A}$	$\hat{D}$	$\underline{\hat{S}}$	$\overline{\hat{S}}$	R^2	Obs. 1	Obs. 2
w_L to w_H	0.053 (0.008) *	1.079 (0.037) *	0.025 (0.009) *	−0.005 (0.006)	0.996	5,295	66,629
w_{LL} to $w_{L \to H}$	−0.086 (0.019) *	1.160 (0.122	0.054 (0.028)	0.004 (0.022)	0.977	3,525	864
w_{HH} to $w_{L \to H}$	−0.142 (0.019) *	1.063 (0.091)	0.026 (0.026)	0.010 (0.017)	0.995	50,811	864
w_{HH} to $w_{H \to L}$	−0.113 (0.018) *	0.982 (0.084)	−0.028 (0.025)	0.014 (0.018)	0.998	50,811	698
w_{LL} to $w_{H \to L}$	−0.052 (0.019) *	1.093 (0.101)	0.004 (0.018)	0.014 (0.019)	0.987	3,525	698

*: for $\hat{A}$, $\underline{\hat{S}}$, and $\overline{\hat{S}}$ significantly different from 0 at 5%; for $\hat{D}$ significantly different from 1 at 5%.
w_i: distribution of log wages for all workers in less dense areas ($i = L$) or denser areas ($i = H$).
w_{ii}: distribution of log wages for stayers in less dense areas ($i = L$) or denser areas ($i = H$).
$w_{i \to j}$: distribution of log wages 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 21: Worker effects of stayers vs. migrants in professional occupations in denser and less dense areas, 2007

	$\hat{A}$	$\hat{D}$	$\underline{\hat{S}}$	$\overline{\hat{S}}$	R^2	Obs. 1	Obs. 2
δ_L to δ_H	−0.010 (0.007)	1.058 (0.030)	0.014 (0.008)	−0.001 (0.005)	0.806	5,295	66,629
δ_{LL} to $\delta_{L \to H}$	−0.115 (0.022) *	1.110 (0.129)	0.040 (0.005)	−0.001 (0.022)	0.993	3,525	864
δ_{HH} to $\delta_{L \to H}$	−0.110 (0.018) *	1.002 (0.075)	0.018 (0.027)	−0.004 (0.017)	0.994	50,811	864
δ_{HH} to $\delta_{H \to L}$	−0.064 (0.018) *	1.060 (0.086)	0.008 (0.020)	0.018 (0.018)	0.987	50,811	698
δ_{LL} to $\delta_{H \to L}$	−0.069 (0.027) *	1.146 (0.138)	0.021 (0.044)	0.017 (0.022)	0.978	3,525	698

*: for $\hat{A}$, $\underline{\hat{S}}$, and $\overline{\hat{S}}$ 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.