Spatial wage disparities: Sorting matters![a]

Pierre-Philippe Combes[b]
Greqam - *University of Aix-Marseille*

Gilles Duranton[c]
University of Toronto

Laurent Gobillon[d]
Institut National d'Etudes Démographiques

First version: February 2003
Final version: February 2007

Abstract

Spatial wage disparities can result from spatial differences in the skill composition of the workforce, in non-human endowments, and in local interactions. To distinguish between these explanations, we estimate a model of wage determination across local labour markets using a very large panel of French workers. We control for worker characteristics, worker fixed effects, industry fixed effects, and the characteristics of the local labour market. Our findings suggest that individual skills account for a large fraction of existing spatial wage disparities with strong evidence of spatial sorting by skills. Interaction effects are mostly driven by the local density of employment. Not controlling for worker heterogeneity leads to very biased estimates of interaction effects. Endowments only appear to play a small role.

Key words: local labour markets, spatial wage disparities, panel data analysis, sorting
JEL classification: R23, J31, J61

[a]We are grateful to Lionel Fontagné, Vernon Henderson, Francis Kramarz, Thierry Magnac, Gianmarco Ottaviano, Barbara Petrongolo, Diego Puga, Jean-Marc Robin, Sébastien Roux, Jon Temple, Dan Trefler, two anonymous referees, and especially Henry Overman for fruitful discussions, advice and encouragements. We also acknowledge France Guérin, Francis Kramarz and Sébastien Roux's kind and efficient help with the data. Seminar and conference participants in Bogotá, Bristol, Brown, CREST, Kiel, Lille, LSE, Marseille, Paris, Philadelphia, Rome, Royal Holloway, Stockholm, Stoke-Rochford, Toronto, UCL, and Villars also provided us with very useful feed-back.

[b]Greqam, 2, Rue de la Charité, 13236 Marseille cedex, France, `combes@ehess.univ-mrs.fr`, `http://www.enpc.fr/ceras/combes/`. CNRS researcher also affiliated with Paris-jourdan Sciences-Économiques and the Centre for Economic Policy Research. Hospitality from the economics department at Boston University is gratefully acknowledged.

[c]Corresponding author. Department of Economics, University of Toronto, 150 Saint George Street, Toronto, Ontario, Canada M5S 3G7, `gilles.duranton@utoronto.ca`, `http://individual.utoronto.ca/gilles/default.html`. Also affiliated with the Centre for Economic Policy Research. Funding from the Leverhulme Trust is gratefully acknowledged.

[d]Ined, 133 bd Davout, 75980 Paris Cedex 20, France, `laurent.gobillon@inded.fr`, `http://laurent.gobillon.free.fr/`.

1 Introduction

In many countries, spatial disparities are large and a source of considerable policy concern. In this paper we propose a new approach to account for spatial wage disparities. We implement it on a large panel of French workers.

To explain large spatial wage disparities, three broad sets of explanations can be proposed. First, differences in wages across areas could *directly* reflect spatial differences in the skill composition of the workforce. There are good reasons to suspect that workers may sort across employment areas so that the measured and un-measured productive abilities of the local labour force vary. For instance, industries are not evenly distributed across areas and require different labour mixes so that we expect a higher mean wage in areas specialised in more skill-intensive industries. Such skills-based explanations essentially assume that the wage of worker i is given by $w_i = As_i$, where s_i denotes individual skills and A, the productivity of labour, is independent of location. Consequently, the average wage in area a is the product of average skills, $\overline{s}_a$, by the productivity of labour: $w_a = A\overline{s}_a$.[1]

The second strand of explanations contends that wage differences across areas are caused by differences in local non-human endowments (hereafter endowments). For instance, workers in some areas may have a higher marginal product than in others because of geographical features such as a favourable location (like a port or a bridge on a river), a climate more suited to economic activity, or some natural resources. Arguably, local endowments cannot be restricted to natural features and should also encompass factors of production such as public or private capital, local institutions, and technology. More formally, this type of argument implies that in area a with endowments E_a affecting positively the productivity of labour, the wage is given by $w_a = A(E_a)$.[2]

The third family of explanations argues that some interactions between workers or between firms take place locally and lead to productivity gains. Interactions-based explanations have a wealth of theoretical justifications. Following Marshall (1890), denser input-output linkages between buyers and suppliers, better matching of workers' skills with firms' needs in thicker labour markets, and technological externalities resulting from more intense direct interactions are frequently mentioned (see Duranton and Puga, 2004, for a review).[3] A key issue is whether these benefits stem from the size of the overall market (*urbanisation economies*) or from geographic concentration at the industry level (*localisation economies*). Stated formally, these arguments imply that the mean wage in area a and industry k is given by $w_{a,k} = A(I_a, I_{a,k})$, where I_a and $I_{a,k}$ are two vectors of interaction variables to capture urbanisation and localisation economies.[4]

We are not aware of any work using individual data considering these three strands in a unified framework. This is the main purpose of this paper. In our specification, we allow skills, endowments, and interactions to determine local wages. More formally, our model implies that in equilibrium the wage of worker i in area $a(i)$ and industry $k(i)$ is given by $w_i = A(E_{a(i)}, I_{a(i)}, I_{a(i),k(i)})s_i$.

[1]That sorting could be at the root of systematic wage differences between groups of workers is a long-standing concern of labour economists. They researched this question intensively in the case of wage differences across industries (Krueger and Summers, 1988; Gibbons and Katz, 1992; Abowd et al., 1999) but they have mostly left aside the geographic dimension. On the other hand, scholars interested in regional issues have paid remarkably little attention to this type of explanation. Glaeser and Maré (2001) on the urban wage premium in US cities and Duranton and Monastiriotis (2002) on UK regional convergence stand out as early exceptions.

[2]This (very) broad group of explanations is often at the heart of the work done by growth economists. The literature on this topic is extremely voluminous (see Durlauf and Quah, 1999, and Temple, 1999, for surveys).

[3]The theories relying on input-output linkages and more generally on market access differ starkly with respect to the spatial scale they consider. The traditional focus of urban economics is the city whereas that of the 'New Economic Geography' (Fujita et al., 1999) is more regional and even inter-regional. We pay attention to these issues below.

[4]Interaction-based explanations have received a lot of attention from urban and regional economists. Work on agglomeration economies is usually done at the aggregate level by regressing a measure of local productivity on a set of variables relating to the extent and local composition of economic activity. Results are generally supportive of the existence of both localisation and urbanisation economies. See Rosenthal and Strange (2004) for a review.

A unified framework encompassing skills-, endowments-, and interactions-based explanations should provide us with a sense of magnitudes about the importance of these three types of explanations in determining wage disparities across areas. These magnitudes are crucial to inform policy and to guide future theoretical work. Unfortunately, a unified framework also imposes formidable data requirements. More specifically, to deal properly with skills-based explanations we must control for unobserved worker heterogeneity, which requires a panel of workers. In our empirical analysis, we use a large panel of French workers.

We develop a two-stage approach. The first stage of the regression allows us to assess the importance of skills-based explanations against those highlighting true productivity differences across areas (i.e., between-industry interactions and endowments-based explanations). Formally, we regress individual wages on time-varying worker characteristics, a worker fixed effect, an area-year fixed effect, an industry fixed effect, and a set of variables relating to the local characteristics of the industry (to capture local interactions within industries). The area-year fixed effects can be interpreted as local wage indices after controlling for observed and unobserved worker characteristics and industry effects. Our main result is that differences in the skill composition of the labour force account for 40 to 50% of aggregate spatial wage disparities. This occurs because workers sort across locations according to their measured and unmeasured characteristics: The correlation between the local mean of worker fixed effects and de-trended area fixed effects (which are computed controlling for worker fixed effects) is large at 0.29. This suggests that previous approaches, which typically do not pay much attention to the sorting of workers across areas, are likely to suffer from an important omitted variable problem.

In the second stage of the regression, we use the area fixed effects estimated in the first stage and regress them on a set of time dummies, several variables capturing local interactions between industries, and some controls for local endowments. We use a variety of panel data techniques and instrumental variables approaches to deal with estimation concerns. Our findings point first at substantial local interactions despite the importance of sorting. Urbanisation economies (measured by the density of local employment) play the most important role. Market access plays a less important part, while endowments play a weak role. Second, controlling for sorting halves standard estimates of the intensity of agglomeration economies. Our favourite estimate for the elasticity of wages with respect to employment density is at 3%. Third, after controlling for skills and interactions, residual spatial wage disparities are smaller than disparities in mean wages by a factor of around three. This result is consistent with a major role for skills-based explanations, a moderate role for interactions, and a weak role for endowments.

The rest of the paper is structured as follows. We first document wage disparities between French employment areas in the next section. Then, in Section 3 we propose a general model of spatial wage disparities. In Section 4, this model is estimated on individual data to assess the importance of skills-based explanations. In Sections 5 and 6, we discuss the issues relating to endowments- and interactions-based explanations and assess their importance. In Section 7, we reproduce our regressions using aggregate data. Finally some conclusions are given in Section 8.

2 Wage disparities across French employment areas

The data is extracted from the Déclarations Annuelles des Données Sociales (DADS) or Annual Social Data Declarations database. The DADS are collected by the French Institute for Statistics (INSEE) from all employers and self-employed in France for pension, benefits and tax purposes. A report must be filled by every establishment for each of its employees so that there is a unique record for each employee-establishment-year combination. The extract we use covers all employees in manufacturing and services working in France and born in October of even-numbered years.

The raw data contains 19,675,740 observations running from 1976 to 1998. For each observation,

Table 1: Some simple correlations
Mean local wage in 1998 ($\log w_{a,98}$) as a function of:

	(1) $\log Density_{a,98}$	(2) $\log Emp_{a,98}$	(3) $\log Diversity_{a,98}$	(4) $Skill_{a,98}$
Intercept	5.720^a	5.147^a	5.329^a	5.352^a
	(0.014)	(0.025)	(0.037)	(0.006)
Coefficient	0.049^a	0.049^a	0.047^a	1.763^a
	(0.003)	(0.004)	(0.012)	(0.085)
R^2	0.51	0.34	0.04	0.56

341 observations. Standard error between brackets. $Density_{a,t}$ is the density of employment in employment area a and year t; $Emp_{a,t}$ is total employment; $Diversity_{a,t}$ is the diversity of employment as measured by an inverse-Herfindhal index, $Diversity_{a,t} = Emp^2_{a,t}/\sum_k Emp^2_{a,k,t}$ where subscript k denotes the industries; and $Skill_{a,t}$ is the employment share of professionals. c: significant at 10%, b: significant at 5%, and a: significant at 1%.

we have some basic personal data (age, gender, occupation at the one-digit level but not education), basic establishment level data (including location and firm industry at the three-digit level), number of days worked, and various measures of earnings. For consistency with the model below, we focused only on total labour costs for full-time employees deflated by the French consumer price index. We refer to the real 1980 total labour cost per full working day as the wage.

Workplace location is identified at the level of 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. Most employment areas correspond to a city and its catchment area or to a metropolitan area. Although the data is of high quality, we carefully avoided a number of pitfalls. After cleaning the data (see Appendix A for details), we ended up with $8,826,422$ observations. For reasons of computational tractability, we keep only six points in time (every four years: 1976, 1980, 1984, 1988, 1992, and 1996). This left us with $2,664,474$ observations when estimating the model on individual data.

Using this data, we can briefly document the extent and persistence of wage disparities between employment areas in France. Typically, in and around Paris wages are on average 15% higher than in large French cities such as Lyon or Marseille, 35% higher than in mid-sized French cities, and 60% higher than in predominantly rural employment areas. To be more systematic, we computed a series of inequality measures between employment areas. The ratio of the highest average to the lowest across all French employment areas remains between 1.62 and 1.88 during the $1976-1996$ period. The ratio of the ninth to the first decile is between 1.19 and 1.23. Finally, the coefficient of variation also remains between 0.08 and 0.09. All this points to rather large and persistent wage disparities between French employment areas.

Table 1, columns $1-4$ reports ordinary least squares (OLS) estimates suggesting that local wages are strongly linked to the structural attributes of their employment area. Column 1 regresses the log of the mean local wage in 1998 on the log of the local density of employment in the same year. The coefficient indicates an elasticity of 4.9% (as typically found in the literature). The explanatory power of this single variable is very strong since the R^2 is 51%. Similar results are obtained in column 2 when using total employment instead of density. In column 3, local wages are regressed on an index of industrial diversity. The effect of this variable is also highly significant but its explanatory power is much weaker. Finally, regressing local wages in column 4 on the share of workers in professional occupations also yields very good results.

3 Theory and estimation

The model

The profit of a competitive representative firm operating in employment area a and industry k in year t is:

$$\pi_{a,k,t} = p_{a,k,t}\, y_{a,k,t} - \sum_{i\in(a,k,t)} w_{i,t}\, \ell_{i,t} - r_{a,k,t}\, z_{a,k,t}, \tag{1}$$

where $p_{a,k,t}$ is the price of its output $y_{a,k,t}$. For any worker i employed in this firm in year t, $w_{i,t}$ and $\ell_{i,t}$ are the daily wage and the number of working days, respectively. Finally, $z_{a,k,t}$ represents the other factors of production and $r_{a,k,t}$ their price. Note that this specification allows for inputs and output markets to be segmented or integrated (when $p_{a,k,t} = p_{k,t}$ and/or $r_{a,k,t} = r_{k,t}$). Output is Cobb-Douglas in effective labour and the other factors of production:

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

where the coefficient b is such that $0 < b \leq 1$, $s_{i,t}$ denotes the skills of worker i in year t, and $A_{a,k,t}$ is the total factor productivity in (a, k, t). At the competitive equilibrium, worker i employed in employment area $a(i, t)$ and industry $k(i, t)$ in year t receives a wage equal to her marginal product:

$$w_{i,t} = b\, p_{a(i,t),k(i,t),t}\, A_{a(i,t),k(i,t),t} \left(\frac{z_{a(i,t),k(i,t),t}}{\sum_{i\in(a,k,t)} s_{i,t}\, \ell_{i,t}} \right)^{1-b} s_{i,t}. \tag{3}$$

Using the first-order condition for profit maximisation with respect to the other factors and inserting it in equation (3) yields:

$$\begin{aligned} w_{i,t} &= b(1-b)^{\frac{(1-b)}{b}} \left(p_{a(i,t),k(i,t),t} \frac{A_{a(i,t),k(i,t),t}}{\left(r_{a(i,t),k(i,t),t}\right)^{1-b}} \right)^{\frac{1}{b}} s_{i,t} \\ &= B_{a(i,t),k(i,t),t}\, s_{i,t}. \end{aligned} \tag{4}$$

Wage differences across areas can reflect differences in individual skills or alternatively they can also reflect true productivity differences caused by endowments and local interactions. Skills (using this word as a shorthand for all the fixed individual attributes which are rewarded on the labour market) are captured by the last term, $s_{i,t}$, in equation (4) whereas the other two explanations enter the term $B_{a,k,t}$ in equation (4). As made clear by this latter term, 'true productivity differences' can work through total factor productivity, $A_{a,k,t}$, or through the price of outputs, $p_{a,k,t}$, or even through the price of non-labour inputs, $r_{a,k,t}$. This implies that we cannot identify price and technology effects separately.[5] Note further that some local characteristics like employment density may have a positive effect on $B_{a,k,t}$ (e.g., agglomeration economies) as well as a negative effect (e.g., congestion). We are not able to identify these effects separately. We can only estimate the overall effect of a variable.

[5] To understand this point better, consider for instance employment area a, which is located in a mountainous region, and industry k. Mountains may have a negative effect on wages in (a, k) because shipping the final output of the industry to the main consumer markets is more expensive, which depresses *f.o.b.* prices. Mountains may have another direct negative effect on wages in (a, k) because operating a plant is more difficult when land is not flat. Finally mountains may have a positive effect on wages because some raw materials such as wood may be more readily available. In this toy example, the first effect works through $p_{a,k,t}$, the second through $A_{a,k,t}$, whereas the third goes through $r_{a,k,t}$. With our approach, we can only estimate the overall effect of local characteristics, the presence of mountains say, in area a and industry k. In other words, we can *identify the determinants* of spatial wage disparities (i.e., endowments, interactions, and skills) but not *the exact channel* through which agglomeration economies percolate. See Duranton and Puga (2004) and Rosenthal and Strange (2004) for further discussion of this classic problem in the agglomeration literature.

A micro-econometric specification

To take equation (4) to the data, we need a specification for both the skill term, $s_{i,t}$, and the 'local industry productivity' term, $B_{a,k,t}$. Assume first that the skills of worker i are given by:

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

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

Turning to $B_{a,k,t}$, which reflects true productivity differences in equation (4), we assume that it is given by:

$$\log B_{a,k,t} = \beta_{a,t} + \mu_{k,t} + I_{a,k,t}\gamma_k, \tag{6}$$

where $\beta_{a,t}$ is an area-year fixed effect, $\mu_{k,t}$ is an industry-year fixed effect, and γ_k is the vector of coefficients associated with $I_{a,k,t}$, the vector of within-industry interactions variables for each area-industry-year.[6]

Combining equations (4), (5), and (6) yields:

$$\log w_{i,t} = \beta_{a(i,t),t} + \mu_{k(i,t),t} + I_{a(i,t),k(i,t),t}\gamma_{k(i,t)} + X_{i,t}\varphi + \delta_i + \epsilon_{i,t}. \tag{7}$$

In equation (7) the interpretations of $I_{a,k,t}\gamma_k$ and $X_{i,t}\varphi$ are problematic. For instance, an industry may employ younger workers. If wages increase with age, this industry will pay lower wages all else equal. We want to think of such systematic industry component as being part of the 'industry effect'. As a consequence, we centre $I_{a(i,t),k(i,t),t}$ and $X_{i,t}$ around their industry mean. The systematic industry components in $I_{a,k,t}\gamma_k$ and $X_{i,t}\varphi$ are added to the industry fixed effect to form a 'total industry effect'. For tractability, we also need to limit the number of coefficients in the model and assume that the time trend is the same for all industries so that this total industry effect can be decomposed into an industry fixed effect and a year effect (which can be normalised to zero for all years since the temporal evolution is also captured by the area-year fixed effect).[7] The final specification for the first stage of the analysis is thus:

$$\log w_{i,t} = \beta_{a(i,t),t} + \mu_{k(i,t)} + \widetilde{I}_{a(i,t),k(i,t),t}\gamma_{k(i,t)} + \widetilde{X}_{i,t}\varphi + \delta_i + \epsilon_{i,t}. \tag{8}$$

where $\widetilde{I}_{a(i,t),k(i,t),t}$ is the centred vector of within-industry interactions variables and $\widetilde{X}_{i,t}$ is the centred vector of individual time-varying characteristics.

Equation (8) corresponds to an inverse labour demand equation.[8] To sum up, we estimate the wages of workers (expressed in constant 1980 francs) as a function of their observed and unobserved

[6]Note that in equation (6), it might seem simpler to use area-industry-year fixed effects rather than area-year fixed effects *plus* industry-year fixed effects. However there would be two problems with doing this. First, it would force us to include more than 200,000 fixed effects in the model (341 employment areas $\times$ 99 industries $\times$ 6 years). These would come in addition to the worker fixed effects introduced in equation (5). Estimating such a large number of worker and area-industry fixed effects is computationally too demanding. Furthermore, many of these fixed effects would be estimated with a very small number of workers (if any at all). This would raise some problems of both identification and statistical significance.

[7]Formally, the effects of within-industry interactions, $I_{a,k,t}\gamma_k$, can be decomposed into an industry specific component independent of location, $I_{\bullet,k,t}\gamma_k$, and a component net of national industry effects, $\widetilde{I}_{a,k,t}\gamma_k \equiv (I_{a,k,t} - I_{\bullet,k,t})\gamma_k$ where $I_{\bullet,k,t}$ is the mean of the $I_{a,k,t}$ weighted by local employment in the industry ($I_{\bullet,k,t} = \frac{1}{N_{k,t}}\sum_{a\in(k,t)} N_{a,k,t}I_{a,kt}$ where $N_{a,k,t}$ is employment in area a, industry k and year t and $N_{k,t}$ is total employment in industry k in year t). Similarly the effect of age can be decomposed into an industry specific component $X_{\bullet,k(i,t),t}\varphi$ and a component net of national industry effect $\widetilde{X}_{i,t}\varphi \equiv (X_{i,t} - X_{\bullet,k(i,t),t})\varphi$. The total industry effect is thus $\mu_{k,t} + I_{\bullet,k,t}\gamma_k + X_{\bullet,k,t}\varphi$. This consists of the industry effect as defined above, plus a national average industry interaction effect and a national average composition effect (in terms of workers' observable characteristics). Then we assume: $\mu_{k,t} + I_{\bullet,k,t}\gamma_k + X_{\bullet,k,t}\varphi = \mu_k + \rho_t$. Finally, since it is not possible to identify ρ_t and $\beta_{a,t}$ separately, we normalise ρ_t to zero for all years.

[8]A competitive wage-setting mechanism is assumed. Any imperfect competition framework where the wage is a mark-up on marginal productivity would lead to similar results since in a log specification this mark-up would enter the constant or the industry fixed effects if such mark-ups vary between industries but not between areas. In France, there is some empirical support for the competitive/fixed-mark-up assumption (see Abowd et al., 1999).

characteristics (age and its square plus a worker fixed effect), the area in which they are employed (area-year fixed effects), their industry (industry fixed effects), and the local characteristics of their industry: log share of employment, log number of establishments, and share of workers in professional occupations. The local share of employment and the number of establishments are standard variables appearing in most models of localisation economies (Rosenthal and Strange, 2004). The share of professionals in the industry is a proxy for the average education locally in the industry. This should capture the external effects of human capital in the local industry in the spirit of the literature on human capital externalities (Moretti, 2004).

This estimation allows us to identify separately the effects of 'people' (skills-based explanations) versus those of 'places' (endowments- and interactions-based explanations).[9] It also allows us to assess the respective explanatory power of the effects of skills ($\widetilde{X}_{i,t}\varphi + \delta_i$), of within-industry interactions ($\widetilde{I}_{a,k,t}\gamma_k$), and the joint explanatory power of endowments and between-industry interactions ($\beta_{a,t}$). The second stage of the estimation then uses $\beta_{a,t}$ as dependent variable. It is presented in detail in Section 5.

Identification, estimation method and estimation issues

To identify the sector fixed effects in equation (8), we need enough mobility across sectors so that all industries are 'connected' with each other (at least indirectly) through worker flows. The identification of area-year effects is slightly more subtle. Workers that move across areas provide the identification of the differences between areas over time. Workers that stay identify changes over time for their area. Hence to identify area-year effects we need (i) some workers remain in each of the employment areas between any two consecutive dates and (ii) there is no area or group of areas with no worker flow to the rest of the country. Given the amount of data we have, all these conditions are easily met.[10] Since area-year fixed effects are identified only relative to each other (just like industry fixed effects), some identification constraints are necessary. We set the coefficients for Central Paris in 1980 and that for the meat industry to zero.

Although helpful for identification, our very large number of observations (with a very large number of worker fixed effects) restricts us to a simple estimation procedure for this first stage. We estimate equation (8) using the *within* estimator.[11]

In our econometric specification, the choice of area and industry is assumed to be strictly exogenous. Nonetheless, since our specification contains both area-year fixed effects and industry fixed effects, this assumption should not be too restrictive. It is discussed in Appendix B. In essence, our results will be biased if we have spatial or industry sorting based on the errors but they will not be biased if sorting is based on the explanatory variables, *including individual, area-year, and industry fixed effects*. More concretely, there is a bias when the location decision is driven by the exact wage that the worker can get at locations in a given year but there is no bias when workers base their location decision on the average wage of other workers in an area and their own fixed effects, i.e., when they make their location decision on the basis of their expected wages.[12]

[9] We do not consider the case where individuals may benefit differently from local labour markets depending on their abilities. An analysis of specific benefits from worker-area matches would require to define some individual fixed effects that are area-specific. This is beyond the scope of this paper. Our aim is only to capture the average benefits from locating in a given place through area fixed effects. Provided mobility is exogenous, our results will be unbiased. The broader issue of how endogenous worker mobility may affect our results is discussed below.

[10] See the working paper version of this article and Abowd et al. (1999) for further details about identification.

[11] Since there is a very large number of fixed effects to estimate, we proceed as follows. We first estimate (8) 'within' individual, that is all variables being centred with respect to their mean for each individual. This gives us the coefficients on all variables except the worker fixed effects. Next, we can recover an estimator of each worker fixed effect by computing his or her mean prediction error. By the Frish-Waugh theorem, this is the OLS estimator for the individual fixed effect. Note that only workers appearing at least twice in the panel contribute to the estimation. This leaves us with 653,169 workers representing 2,221,156 observations.

[12] As in standard Roy models, a bias will also arise if the returns to the time-varying unobserved characteristics

If this selection bias is relevant, we can think of several reasons why it is likely to be much attenuated. First, in a country like France with numerous barriers to internal mobility, we expect migration to be driven mostly by long-term considerations. Provided the local shocks are uncorrelated over time, there is then no bias since workers migrate on the basis of future expected wages rather than the wage they can get today (Topel, 1986). Second, we also expect location decisions to be driven by factors unrelated to wages such as idiosyncratic preferences. Using the European Household Panel Survey, Gobillon and Le Blanc (2003) report that only 22% of long-distance moves in France are related to a new job. Third, with time-varying local effects and industry fixed effects, we expect much of the variation caused by the environment to be captured. This should limit the scope of selection. Finally, Dahl (2002) proposes a new approach to deal with selection problems with many possible choices, but this can be applied to cross-section data only and we do not know of any method to correct for such selection biases in panel. He shows that this type of selection bias has only minimal effects on the estimates of the returns to education across US states.

Some concerns also arise with the characteristics of the local industries in $I_{a,k,t}$. As discussed by Ciccone and Hall (1996) and Ciccone (2002), some local characteristics like a high level of specialisation in an industry could be endogenous to high wages in this industry. We leave these concerns aside here on the ground that these variables only have a small explanatory power (see below). Similar concerns with respect to between-industry interactions will be tackled in the second-stage estimation.

Finally, according to Abowd et al. (1999) a wage equation with industry fixed effects should also contain establishment fixed effects. This is because these fixed effects may be correlated with industry fixed effects. This also applies to area fixed effects. Such a correlation would bias the estimates when establishment fixed effects are omitted. However the method developed by Abowd et al. (1999) to deal with large scale matched employer-employee data (using both worker and plant fixed effects) would not allow us to compute the standard deviations for the estimated area fixed effects that are necessary to perform the second stage of the estimation correctly. This approach would also lack theoretical foundations since area fixed effects would then have to be computed by calculating a weighted average of establishment fixed effects by location. A final problem with this alternative approach is that establishment fixed effects are constrained by the estimation to be constant over time. The resulting area fixed effects constructed by aggregating time-invariant establishment fixed effects can then evolve only through the entry and exit of establishments and internal changes in employment and not by changes in interactions and endowments.

4 Skills and sorting across employment areas using individual data

This section presents the results for the estimation of equation (8). Recall that the explanatory variables are the area-year fixed effects, the industry fixed effects, the worker fixed effects, the worker's age and its square, the log share of local industry employment, the log number of establishments, and the share of professionals. Note that in absence of education data, worker fixed effects will capture all the permanent characteristics of workers including their education. Since we are interested in the effects of skills rather than their determinants, this is not an issue provided the coefficients are properly interpreted. We first present a variance analysis and our results about sorting before commenting on the coefficients.

Table 2: Summary statistics for the variance decomposition — estimation of equation (8)

Effect of	Std dev	Simple correlation with: $\log w$	δ	$\beta - \theta$
log real wage ($\log w$)	0.367	1.00	0.78	0.26
residuals (ϵ)	0.166	0.45	0.00	0.00
worker effects ($\delta + X\varphi$)	0.294	0.80	0.98	0.09
worker fixed effects (δ)	0.284	0.78	1.00	0.10
age ($X\varphi$)	0.058	0.23	0.08	0.00
industry fixed effects (μ)	0.043	0.25	0.16	0.05
within-industry interactions ($\widetilde{I}_k \gamma_k$)	0.024	-0.01	0.00	-0.45
within-industry share of professionals	0.011	0.16	0.12	0.29
within-industry establishments	0.019	-0.13	-0.08	-0.62
specialisation	0.017	0.03	0.02	-0.13
area fixed effects (β)	0.140	0.34	-0.05	0.55
de-trended area fixed effects ($\beta - \theta$)	0.065	0.26	0.10	1.00
time (θ)	0.118	0.26	-0.11	0.10

$2,221,156$ observations. All correlations between the effects that are not orthogonal by definition are significant at 1%.

The effect of within-industry share of professionals is that of the share of professional times its coefficient (in vector γ_k). The effect of within-industry establishments is that of the log of the number of establishments times its coefficient. The effect of specialisation is that of the log of the industry share in employment times its coefficient. Area fixed effects are de-trended using the time fixed effects (θ) estimated in the second stage.

The importance of workers' skills

Our first set of results suggests, unsurprisingly, that workers' skills are of fundamental importance and play a much greater role than the local environment and the industry in the determination of individual wages. To show this, we perform a complete variance analysis as in Abowd et al. (1999). Table 2 shows the explanatory power of the different variables for the baseline regression. For each variable or group of variables, the Table reports the standard deviation *of their effect* and their correlation with wages, worker fixed effects and de-trended area fixed effects.

To construct this Table, we computed *the effect of each variable* by multiplying its coefficient by its value for each observation. For instance, consider worker i in (a, k, t). The effect of specialisation is equal to the estimated coefficient on this variable for industry k times the specialisation of area a in this industry. For a group of variables, the sum of the effects is computed. Then, the variability of the effect of each variable across workers can be calculated. When the effect of a variable has a large standard deviation and it is highly correlated with wages, this variable has a large explanatory power. When on the contrary the effect of variable has a small standard deviation and a small correlation with wages, this variable explains only a small fraction of the variations of wages.

Worker fixed effects have by far the largest explanatory power. Their standard deviation (0.284) is close to that of log wages (0.364) and the correlation between worker fixed effects and wages is very high at 0.78. For no other variable, or group of variables, are the standard deviation and the correlation with wages as high. When looking at the effects of observable worker characteristics, it is worth noting that age and its square also have a moderate explanatory power with a standard deviation of 0.058 and a correlation of 0.23 with log wages. Altogether, with a standard deviation of

differ across areas and workers choose their location accordingly. In this respect note that a primary objective of our paper is to decompose spatial disparities. Considering that spatial differences in individual productivity could have multiple dimensions would make such decomposition much more cumbersome and far less transparent. We believe that it is better to consider only one dimension for a first pass on the issue.

Table 3: Spatial wage disparities, 1976 − 1996 average

	Mean wage	Net wage
(Max-Min)/Min	0.74	0.38
(P90−P10)/ P10	0.21	0.14
(P75−P25)/ P25	0.11	0.06
Coefficient of variation	0.08	0.05

Mean wage refers to the de-trended mean wage by employment area. Net wages are calculated as in equation (9). Max, Min, P10, P90, P25, and P75 are the max, the min, the first decile, the last decile, the first quartile, and the last quartile, respectively.

0.294 and a correlation of 0.80 with wages, the combined effect of individual observed and unobserved characteristics is of overwhelming importance.

Turning to within-industry interactions, their explanatory power is very small. The standard deviation of the effect of all within-industry interaction variables together is less than a tenth of that of worker fixed effects. Furthermore, the correlation between log wages and the effect of within-industry interactions is close to zero. Within this group of variables, neither the share of professionals, the number of establishments nor specialisation particularly stands out.

Finally, the explanatory power of area-year fixed effects is substantial, albeit much less so than that of worker fixed effects. Because wages increased everywhere in real terms between 1976 and 1996, a good fraction of the area fixed effects is explained by the time trend over the period. After taking away this trend however, area fixed effects still have an explanatory power more important than that of industry, age, or within-industry interactions. Although this result was to be expected, this is rather interesting in light of the small amount of attention location factors have received so far in the labour literature relative to industry and age.

Spatial wage disparities and sorting

To evaluate the importance of workers' skills on spatial wage disparities, we can also study the variations of a wage index net of worker and industry effects. This 'net wage' is computed from the results of equation (8). It corresponds to the local wage obtained by an 'average' worker in an 'average' industry. We can define such an index $w_{\text{net},a,t}$, which we refer to as the net wage, in the following way:

$$\log w_{\text{net},a,t} \equiv W_t + \widehat{\beta}_{a,t}, \tag{9}$$

where W_t is a normalising time-dependant term such that $w_{\text{net},a,t}$ can be interpreted as a wage.[13]

These net wages can then be compared with the real mean wages per area computed in Section 2. Table 3 compares systematically disparities in mean and net wages. Depending on the inequality measure taken, disparities in net wages may be as low as half of those in mean wages. Put differently, workers' skills explain 40 to 50% of spatial wage disparities.

This result is caused by a strong sorting pattern whereby workers with high fixed effects tend to live in the same areas. To go further on this issue, it is interesting to correlate the average worker fixed effects within each areas with de-trended area fixed effects. The correlation between the two is large at 0.29. Hence, areas where workers with high individual fixed effects work are also areas where the productivity of labour (after controlling for skills) is high. An immediate implication is that large spatial wage disparities reflect true productivity differences across areas that are magnified by the sorting of workers by skills.

[13]Formally, we have $W_t \equiv \frac{1}{N}\sum_{j=1}^{K} N_j \widehat{\mu}_j + \frac{1}{N_{t_0}}\sum_{i\in t_0} \widehat{\delta}_i + \frac{1}{N_{t_0}}\sum_{m=1}^{Z} N_{m,t_0}\widehat{\beta}_{m,t_0} - \frac{1}{Z}\sum_{m=1}^{Z} \widehat{\beta}_{m,t}$ where $t_0 = 1980$ and Z is the number of areas.

Table 4: Summary statistics for the coefficients estimated in equation (8)

Variable	Number of coefficients	Percentage > 0 at 5%	Percentage < 0 at 5%	P90−P10
area fixed effects (de-trended)	2046	10%	78%	0.16
industry fixed effects	99	58%	33%	0.11
age	1	100%	0%	–
squared age	1	0%	100%	–
specialisation	99	95%	0%	0.02
share of professionals	99	81%	3%	0.20
industry establishments	99	1%	85%	0.02

For area fixed effects, significance is calculated relative to the weighted national mean for the period. For industry fixed effects, significance is calculated relative to the weighted national mean. P90− P10 is the difference between the ninth and the first decile.

Analysis of the coefficients

Table 4 reports some summary statistics regarding the coefficients of equation (8).[14] Note first that 88% of the area fixed effects differ significantly from the national mean (weighted for the period). Moreover, this distribution is skewed since only 10% of these area fixed effects are significantly higher than the mean whereas 78% are significantly lower. This is because a few populous employment areas (Paris, its suburbs, and other large French cities) offer significantly higher wages than the national mean.

In line with previous findings in the literature, we find that most specialisation elasticities are positive and significant. The average for all industries is at 2.1%, which is at the lower bound of the estimates found in the literature (Henderson, 1986; Rosenthal and Strange, 2004). The largest specialisation coefficients are found for business services (3.6%) and for two high-tech industries, namely medical instruments (3.9%) and artificial fibres (4.3%). At the other end of the spectrum, the five industries with a coefficient not significantly different from zero are oil refinery, air transport, tobacco, production of weapons and bullets, and production of steel. Given the reliance of most of these industries on localised natural advantage (or some localised infrastructure), these results are not very surprising. The average coefficient on the share of professionals across industries is large at 11.8%. This is in line with the findings in the literature on human capital externalities (see Rauch, 1993, and his followers). Finally, the elasticity with respect to the number of industry establishments is on average at -1.4%. This coefficient is highest in industries such as machine tools and various instrument industries that produce very differentiated goods. The smallest coefficients are obtained in industries where instead efficient plant size is expected to be very large like various extractive industries, naval construction, and energy or water utilities.

[14] Our identification constraints ($\mu_1 = 0$ and $\beta_{\text{Paris},1980} = 0$) imply that standard Student's tests about the significance of the industry and area effects with respect to 0 are not very informative because they depend on the choice of references. We instead test the significance of the coefficients with respect to their weighted industry mean or their weighted area mean for a given year. That is, we test the equalities: $\mu_k = \frac{1}{N}\sum_{j=1}^{K} N_j \mu_j$ and $\beta_{a,t} = \frac{1}{N_t}\sum_{j=1}^{Z} N_{j,t}\beta_{j,t}$, where $N_{j,t}$ is the number of workers in employment area j in period t, N_t denotes the total number of workers in year t, N_j is the total number of workers in industry j across all years, K is the number of industries, and Z is the number of employment areas. These tests can easily be implemented from the estimated coefficients and their covariance matrix. Directly constraining the mean of all area or industry fixed effects to zero in the estimation would have been computationally too demanding.

5 The determinants of area fixed effects: estimation

So far we have assessed the relative importance of 'people' versus 'places' to explain spatial wage disparities. The objective of the second stage of the estimation is to assess the relative importance of endowments and between-industry interactions in explaining the area-year fixed effects.

Specification

The area fixed effects estimated in equation (8) are assumed to be a function of a year fixed effect, of local interactions between industries, and endowments. The econometric specification is:

$$\beta_{a,t} = w_0 + \theta_t + I_{a,t}\gamma + E_{a,t}\alpha + \upsilon_{a,t}. \tag{10}$$

where the θ_t are time fixed-effects and α is a vector of coefficients associated with the endowments variables, $E_{a,t}$. γ is the vector of coefficients associated with local *between-industry* interactions, $I_{a,t}$. The error terms $\upsilon_{a,t}$ that reflect local technology shocks are assumed to be i.i.d. across areas and periods. Finally, we take 1980 as reference so that the coefficient for this year is set to zero.

To capture *between-industry* interactions, we follow the literature (e.g., Ciccone and Hall, 1996) and use the log of the density of local employment ($\log Density$) as main explanatory variable. To distinguish density effects from pure scale effects, we also use the log of land area ($\log Area$).[15] The diversity of the local composition of economic activity may also matter (Glaeser etal., 1992). To capture this, we use the log of the inverse of a Herfindhal index ($\log Diversity$, which is calculated as in Table 1). Finally, it could well be that wage differences across areas are driven by the proximity to markets for intermediate and final goods. These markets may have a spatial scale larger than employment areas as argued by much of the recent literature (Fujita et al., 1999). Hence, we also constructed and experimented with a series of market access variables. The one we retained ($\log Potential$) is the log of the market potential computed from the density of neighbouring areas: $Potential_{a,t} = \sum_{a' \neq a} \frac{Den_{a',t}}{d(a,a')}$ where $d(a,a')$ is the great-circle distance between areas a and a'.

Turning to productive endowments, note that they can raise wages through one of the three channels highlighted above (lower exporting costs, cheaper supplies, or higher productivity). There are many possible endowments that may work through these channels. One can think about airports, high-speed train lines, a favourable climate, closeness to a navigable river or a deep-sea harbour, etc. However, using a complete set of endowments would raise serious endogeneity concerns (more on that below). To avoid this, we only considered four (exogenous) endowment variables, the percentage of municipalities in each employment area with the following location attributes: a sea shore, mountains, lakes and water, and 'outstanding cultural or architectural heritage' (coming from an inventory of monuments made by the central government).

This last explanatory variable is of course unlikely to have a direct effect on local productivity. However, recall that equation (4) shows that the price of non-labour inputs matters in the determination of local wages. As highlighted first by Roback (1982), better consumption amenities (i.e., amenities unrelated to production like an architectural heritage) increase the willingness of consumers to pay for land and thus imply higher local land rents. As a result, firms use relatively less land. In turn, this lowers the marginal product of labour when land and labour are imperfect substitutes in the production function. Put differently, wages may capitalise the effect of non-production

[15] Note that to be consistent we use the log values of the share of employment by industry (in the first stage) and of density and land area (in the second stage). This allows us to estimate the effect of a change in composition of activity keeping all else constant, a change in population keeping land area and composition constant, and a change in land area keeping density and composition constant (i.e., an increase in population keeping density constant). The effects of other changes can be easily computed by summing the coefficients. Alternative specifications using for instance industry employment, density, and total employment are certainly possible. However one must be careful with respect to the interpretation of the coefficients (Combes, 2000).

variables. Some of these variables are missing in our specification as they are not observed. This is an issue only when such consumption amenities affect an explanatory variable like employment density — an issue that we discuss in detail below. Otherwise, this only implies more noisy estimates for the wage effects (as observationally identical employment areas end up paying different wages).

Estimation method

Note that equations (8) and (10) constitute the full econometric specification. We speak of a two-stage estimation because in equation (10), the second stage, we use as dependent variable the area fixed effects estimated in equation (8), the first stage. The alternative is to perform a single-stage estimation and use all the explanatory variables at once.

Such a single-stage estimation is problematic because it does not allow us to compute the variance of local shocks, $\upsilon_{a,t}$.[16] In turn, we cannot distinguish local shocks from purely idiosyncratic shocks at the worker level, which is important with missing endowment variables. Moreover, in a single-stage estimation, the variance of local shocks has to be ignored when computing the covariance matrix of estimators. As shown by Moulton (1990), this creates large biases in the standard errors for the estimated coefficients of aggregate explanatory variables.[17] Our estimation method avoids these pitfalls. As robustness check, we nonetheless ran a single-stage estimation and found qualitatively similar results for estimated coefficients (see Section 6).

Estimation issues

In the estimation of equation (10), note first that the true value of the dependent variable, $\beta_{a,t}$, is unknown. We use instead the unbiased and consistent estimators $\widehat{\beta}_{a,t}$ provided by the first-stage results. However, the fixed effects for areas with few workers are less precisely estimated than those for areas with many workers. Thus, the use of $\widehat{\beta}_{a,t}$ as dependent variable introduces some heteroscedasticity through sampling errors. This can be dealt with by computing a feasible generalised least-square (FGLS) estimator. The procedure is detailed in Appendix C.

As shown below, the second-stage results using the FGLS correction are very close to those obtained with simpler estimation techniques without any correction. This shows that the effects of the sampling errors on the coefficients estimated at the second stage are negligible.[18] Consequently, when dealing with endogeneity problems, we will ignore them to keep the econometrics reasonably simple.

The second main estimation issue is that some local characteristics are likely to be endogenous to local wages. For instance, employment areas receiving a positive technology shock may attract migrants. This leads to a positive correlation between the second-stage residuals and the density of employment. In this particular case, reverse-causality is going to bias the estimates upwards. Alternatively, as argued above, missing consumption amenities may imply a negative correlation between employment density and the residuals and thus bias the estimates downwards. Hence, endogeneity is potentially a serious concern for the second stage of the estimation (and all the more so since the direction of the bias is unclear).

To deal with this issue, we consider two solutions. Following Ciccone and Hall (1996), the first one is to argue that endogeneity may be caused by 'contemporaneous' local shocks. Considering that these shocks did not have any effect on the distribution of the population in the past, we can instrument employment density between 1976 and 1998 by long-lagged population variables. This strategy rests on the hypothesis that population agglomeration in the past is not related to

[16] This is because (i) the model is projected in the *within* dimension and (ii) workers can move between areas.

[17] Alternative approaches like standard robust clustering methods do not work here because the covariance matrix of error terms is too complex for the reasons already mentioned in the previous footnote.

[18] This is because we have a very large number of observations with many stayers and large flows of movers between areas. This allows us to estimate the area-year fixed effects very precisely.

Table 5: Summary statistics for the variance decomposition — estimation of equation (10)

Effect of	Std dev	Simple correlation with: $\log w$	δ	$\beta - \theta$
between-industry interactions ($I\gamma$)	0.077	0.22	0.12	0.90
density	0.067	0.20	0.12	0.84
land area	0.024	−0.15	−0.08	−0.62
diversity	0.002	−0.04	−0.06	−0.31
market potential	0.036	0.19	0.08	0.78
amenities ($E\alpha$)	0.011	−0.10	−0.06	−0.48
residuals (η)	0.029	0.04	−0.08	0.03

2,221,156 observations. Variables in the first column are all centred around their year mean.

modern differences in productivity, an hypothesis that is more likely to hold for very long lags. Our instruments are the log density of urban population in 1831, 1861, 1891, and 1921. We also use the log market potential calculated using 1831 population data and a peripherality index (the log mean-distance to all other employment areas). Resting on several instruments (instead of only 1831 urban population) offers two additional benefits. Since the population is taken in log, using a multiplicity of census dates is equivalent to instrumenting by past levels *and* long-run historical growth rates. Furthermore, having multiple instruments allows us to instrument not only for employment density but also for the market potential, diversity, and even land area.[19] We can also conduct exogeneity and over-identification tests.

The second strategy is to assume that areas have permanent characteristics affecting their productivity and introduce area fixed effects in (10). First-differencing will then remove these fixed effects together with observed permanent characteristics such as land area and amenities. With this strategy, contemporaneous shocks may nonetheless bias the results since a rise in productivity may lead to an increase in employment density. We can then instrument the changes in employment density (rather than their level). The instruments we use are the same as above since past levels may drive current growth (be it only through a mean-reversal effect) just like long-run population growth rates. We also use a bunch of variables from the 1968 population census. These variables refer mostly to the demographics, average education, composition of employment and state of the housing stock of each employment area in 1968 (see below for details). If we obtain similar coefficients with these two strategies, we can be reasonably confident about our results.

6 The determinants of area fixed effects: results

The importance of employment density

We first perform a variance decomposition. The results are reported in Table 5 for the complete OLS regression (i.e., column 3 in Table 6 below). Employment density clearly stands out. Its effect and that of local fixed effects are very correlated at 0.84. Their standard errors are nearly equal. Market potential comes second in importance with land area. The explanatory power of the diversity of local industrial composition and amenity variables is close to nil. This suggests a small explanatory power for local endowments. It could be that our amenity variables do not capture all endowments well but the relatively small variance of the second-stage residuals also points at a small explanatory power for endowments.[20]

[19]The reason why land area needs to be instrumented is because areas were defined depending on employment density so that any bias affecting density is likely to affect land area as well.

[20]Note that we perform our variance analysis on the complete OLS specification rather than our preferred specification where interactions variables are instrumented. However the results for the variance analysis on our preferred estimation

Analysis of the coefficients

The coefficients obtained in the estimation of equation (10) are given in Table 6. The first column reports results for the baseline specification where density, land area and diversity are used as explanatory variables.[21] At 3.7%, the coefficient on density is at the lower bound of previous estimates in the literature (Rosenthal and Strange, 2004). This suggests that worker heterogeneity was captured in part by density in previous work (see Section 7 for more on this). The coefficient on land area is smaller than that on density by a factor of three. An increase in population through a higher density has a much larger wage effect than the same population increase obtained by a larger land area keeping density constant.[22]

Column 2 in Table 6 performs the same regression as the baseline but uses the FGLS correction discussed above, which corrects for heteroscedasticity. The differences with the baseline are minimal. This reflects the fact that the area fixed effects are precisely estimated in the first stage.

In column 3, we added some controls for productive endowments and amenities (seaside, lake, mountains and architectural heritage) and market potential to the baseline regression. Comparing with column 1, the addition of these extra controls slightly lowers the coefficient on density and increases that on land area. The coefficient on the diversity of the composition of activity becomes negative and significant. Among the added variables, the coefficient on market potential is positive and highly significant. Its magnitude is comparable to that on density. If the market potential of an area doubles (e.g., employment density doubles in all other areas) wages increase by 3.5%. Turning to the four amenity variables, recall that they can have both a direct effect as productive endowments and an indirect effect of opposite sign as consumption amenities (through land prices affecting the quantity of land used by firms and thus the marginal product of labour). We expect the presence of an outstanding heritage to have a minimal direct productive effect and a much larger amenity effect. This is what we observe. The same holds for the presence of a lake for which the productivity benefits are also likely to be very small. The coefficients on sea and mountains are positive. In the case of the sea variable, the positive productivity effect slightly dominates the amenity effect. The case of mountains is more ambiguous since the expected sign of both the direct and indirect effects is unclear. In any case, note that the net effects for all four variables are significant but small.

Column 4 is our preferred specification. Density, land area, diversity and market potential are instrumented by long-lagged population variables dating back to 1831 and the peripherality of the area. Comparing the results to the previous column, endogeneity appears to be a serious concern. It can be noted first that the coefficient on density decreases again. Our coefficient on density, at 3.0%, is below most estimates in the literature, which are in the $4 - 8\%$ range. To repeat, the major reason for this difference is the failure of previous literature to control properly for unobserved individual heterogeneity. After instrumenting, the coefficient on land area becomes insignificant. It turns out that the endogeneity bias is much larger for this variable. Similarly, after instrumenting, the coefficient on market potential also declines from 3.5 to 2.4%. Overall we find that endogeneity is a more serious concern than previously concluded. In part, this is because we consider more

are very similar. The standard deviations for the effects of employment density and market potential decrease slightly but the standard deviation for all interaction effects (when jointly considered) is unchanged.

[21] It is likely that employment density does not affect all industries with the same intensity (Henderson, 2003). The two-step estimation prevents us from exploring this issue further. We leave it for future work.

[22] When using the same variables directly in equation (8) to perform a single-stage estimation (whose results are available upon request), we find very similar values for the effects of industry characteristics. The average coefficient of industry specialisation is 2.2% (against 2.1% in the two-stage estimation). The coefficient on employment density is also very close: 3.2% (against 3.7% in the two-stage estimation). That on land area shows a larger discrepancy at 2.1% (against 1.1%). The insignificant coefficient on industrial diversity changes sign. These differences between the two-stage and single-stage estimations find their sources in the correlations between the individual explanatory variables and the aggregate error terms (recall that the error structure in the two-step estimation differs from that of a single step estimation). In any case, the explanatory power of both land area and diversity remains small so that these changes in the coefficients do not alter our conclusions.

Table 6: Estimation results for equation (10)

Regression	(1) Levels OLS 1	(2) Levels FGLS	(3) Levels OLS 2	(4) Levels 2SLS	(5) First-Dif OLS	(6) First-Dif 2SLS
$\log Density$	0.0371^a	0.0357^a	0.0322^a	0.0302^a	0.0349^a	0.0289
	(0.0008)	(0.0010)	(0.0007)	(0.0063)	(0.0043)	(0.0175)
$\log Area$	0.0113^a	0.0106^a	0.0218^a	0.0041	-	-
	(0.0014)	(0.0016)	(0.0013)	(0.0154)		
$\log Diversity$	0.0020	0.0006	-0.0046^b	-0.0407^c	-0.0047	-0.0296
	(0.0023)	(0.0025)	(0.0020)	(0.0208)	(0.0032)	(0.0200)
$\log Potential$			0.0351^a	0.0244^a	0.1385^a	0.1427^c
			(0.0014)	(0.0042)	(0.0474)	(0.0715)
Sea			0.0111^a	0.0004	-	-
			(0.0033)	(0.0046)		
Mountain			0.0333^a	0.0209^a	-	-
			(0.0032)	(0.0041)		
Lake			-0.0254^a	-0.0263^a	-	-
			(0.0054)	(0.0088)		
Heritage			-0.0091^b	-0.0202^a	-	-
			(0.0043)	(0.0068)		
Time dummies	Yes	Yes	Yes	Yes	Yes	Yes
R^2 (within time)	60%	-	72%	-	-	-

2,046 observations. Standard error between brackets. c: significant at 10%, b: significant at 5%, and a: significant at 1%. In column 4, density, land area, and diversity are instrumented by urban population density in 1831, 1861, 1891, and 1921 together with market potential computed using 1831 urban population data and mean distances to other areas. The R^2 for the instrumental regressions are 0.64 for density, 0.35 for area, 0.17 for diversity, and 0.92 for market potential. A test of overidentifying restrictions shows that our instruments are valid even at a 10% level. Diversity and market potential are clearly endogenous while density and land area are only marginally exogenous. In column 6, we instrument the changes in log density, log diversity and log area with the same variables as in column 4 plus a set of variables from the 1968 population census: mean age, mean age when leaving education, shares of the different occupational groups, share of population born in France, share of workers employed in the public sector, share of population living in an accommodation with hot water, with flushing toilet, with toilet inside, share of people living in a 'normal accommodation' (apartment or house as opposed to second residence, flat-share, etc), and mean deterioration of accommodation. The R^2 for the instrumental regressions are 0.35 for changes in density, 0.05 for changes in diversity, and 0.89 for changes in market potential.

variables (density, land area, diversity, and market potential) and more instruments than previous work. This may also be caused by the fact that French employment areas are rather small so that the effects of local shocks are easier to pick up.

In column 5, we report the results for a simple first-difference estimation. Interestingly, the results are not very different from those of Column 3. The main exception is the coefficient for market potential, for which the standard error is much larger. This suggests that controlling for permanent unobserved characteristics of employment areas does not affect much the results. In column 6, when instrumenting the changes in density, diversity and market potential, we find again results close to those of our IV regression in levels (column 4). The coefficient on density is just below 3% while that on diversity is also negative. The coefficient on market potential remains positive with again a large standard error. Furthermore, our instruments for the first differences are weak (and we consequently do not give much weight to the results in this column).

Residual spatial wage disparities

To examine spatial wage disparities, we can now compute a 'residual wage', that is a local wage controlling for skills and all interactions, from the results of the baseline regression for the second stage. We can define such index $w_{\text{resid},a,t}$ (or residual wage) as:

$$\log w_{\text{resid},a,t} \equiv W + \widehat{\eta}_{a,t}, \tag{11}$$

where W is defined in a similar way as after equation (9). This residual wage corresponds to the local wage obtained by an 'average' worker employed in an 'average' industry and in an area with 'average' interactions.

The difference between highest and the lowest residual wage divided by the lowest residual wage across all employment areas is 0.23 instead of 0.38 for the de-trended net wage (i.e., the wage after controlling for skills and industry) and 0.74 for the de-trended mean wage. The same ratio for the first and the last decile is 0.07 instead of 0.14 and 0.21 for net and mean wages, respectively. For the first and last quartile, we find 0.04, 0.06 and 0.11 for residual, net, and mean wages respectively. Finally, the coefficient of variation for residual wages is 0.03 against 0.05 for net wages and 0.08 for mean wages. The salient result is thus that once skills and interactions are controlled for, about two thirds of the wage disparities between employment areas disappear.

7 Aggregate wage differences across employment areas

Research is often restricted in the data it can use. Existing studies on regional disparities typically use mean wages (or output per worker) by industry and location. It is of course impossible to directly implement our micro-founded specifications (8) and (10) with aggregate data. In this section, we first show how the simple model introduced above (where wages are determined at the worker level) can be aggregated and estimated at the level of each employment area and industry. We then compare the aggregate data results with those obtained above using individual data.

Aggregation issues

Once we abstract from the longitudinal dimension of the panel, and in absence of information about education, we can use the information about occupations (self-employed, professional, skilled, unskilled white-collar, unskilled blue-collar) to proxy for skills. Since occupations may change over time, we assume that worker fixed effects are such that $\delta_i = \sum_{k,c} d_{i,k,c,t}\delta_{c,k} + \iota_{i,t}$ where $d_{i,k,c,t}$ is an occupation dummy taking value one when worker i is in occupation c and industry k at date t, $\delta_{c,k}$ is the corresponding coefficient, and $\iota_{i,t}$ is a residual term. Averaging (7) over all $N_{a,k,t}$ workers in the same local industry (a,k) in year t yields:

$$\log w_{a,k,t} = \frac{1}{N_{a,k,t}} \sum_{i\in(a,k,t)} \log w_{i,t} = \beta_{a,t} + \mu_{k,t} + I_{a,k,t}\gamma_k + \frac{1}{N_{a,k,t}} \sum_{i\in(a,k,t)} \left(X_{i,t}\varphi + d_{i,k,c,t}\delta_{c(i,t),k}\right) + \varsigma_{a,k,t}, \tag{12}$$

where $\varsigma_{a,k,t} = \frac{1}{N_{a,k,t}} \sum_{i\in(a,k,t)}(\epsilon_{i,t} + \iota_{i,t})$.

If there is some sorting across space or industries that leads the mean of the residual term $\iota_{i,t}$ to be correlated with some of the explanatory variables at the (a,k,t) level, the estimated coefficients are biased. This is a first major limitation when using aggregate data. Another aggregation problem in equation (12) regards data availability. Typically, one may have access to the mean wage in an industry and area but not to the mean of log-wages. Hence the mean of log-wages must be proxied by the log of mean wages. A similar problem arises among the explanatory variables when using (as we do) the squared age of workers. Again the mean of squared individual ages requires individual level data. With aggregate data, it can only be proxied by the square of the mean age. This implies some measurement problems for wages and squared age.[23]

We can again centre within-industry interactions and worker time-varying characteristics so that all systematic industry components can be brought together with the industry fixed effect.[24] We obtain:

$$\begin{cases} \log w_{a,k,t} = \mu_k + \beta_{a,t} + \widetilde{I}_{a,k,t}\gamma_k + \widetilde{X}_{a,k,t}\varphi + \sum_c \widetilde{q}_{c,a,k,t}\delta_{c,k} + \varsigma_{a,k,t}, \\ \beta_{a,t} = w_0 + \theta_t + E_{a,t}\alpha + I_{a,t}\gamma + \upsilon_{a,t}. \end{cases} \tag{13}$$

These two equations mirror equations (8) and (10). As argued above, the share of workers in professional occupations in industry and employment areas should be used as one of the regressors in the vector $\widetilde{I}_{a,k,t}$ to capture human capital interactions within industries. However this variable also now appears independently in equation (13) following the aggregation of individual skills. Hence the coefficient on the share of professionals captures both skill composition effects *and* local interactions in the industry. The two cannot be separately identified. This constitutes another limitation of aggregate data. Finally, the first stage equation must be estimated by weighting each observation by the square-root of its number of workers to avoid heteroscedasticity (Coelho and Ghali, 1973). Turning to the second stage (and as previously), we do not know the true values of the area fixed effects, $\beta_{a,t}$. Hence, we use $\widehat{\beta}_{a,t}$ rather than $\beta_{a,t}$ keeping a similar estimation method as before (again see Appendix C). We also impose the same identification conditions: $\mu_1 = 0$ and $\theta_{1980} = 0$.

Results

At the aggregate level, we perform the two-stage estimation using all the twenty years of data available as we are not limited by sample size. The first stage of the regression with all the variables ($7,514$ in total) has a R^2 of 81% compared with 31% for the same regression with individual data without the worker fixed effects. This difference is obviously explained by the considerable variation in individual wages that is averaged out by aggregation.

As with individual data, we then perform a detailed variance analysis of the first stage of the estimation. The main finding is that the effect of all the explanatory variables we consider is much larger than previously.[25] With respect to the share of the various occupations, a higher explanatory

[23]However, these measurement problems are very minor. The correlations between mean-log-wage and log-mean-wage by industry and location and between mean-squared-age and squared-mean-age by location are both 0.99.

[24]Define the centred share of occupation c in (a,k,t): $\widetilde{q}_{c,a,k,t} \equiv q_{c,a,k,t} - q_{c,.,k,t}$ where $q_{c,a,k,t} \equiv \frac{1}{N_{a,k,t}} \sum_{i\in(a,k,t)} d_{i,k,c,t}$ is the share of occupation c in (a,k,t) and $q_{c,.,k,t}$ its weighted mean across all employment areas. To mirror the approach developed in Section 4, we assume $\mu_{k,t} + I_{.,k,t}\gamma_k + X_{.,k,t}\varphi + \sum_c q_{c,.,k,t}\delta_{c,k} = \mu_k + \rho_t$, that is the sum of all the industry effects can be decomposed into a time-invariant industry effect and a time effect (which is again normalised to zero).

[25]The standard deviation for the wages is at 0.258 (against 0.367 with individual data). The standard deviation for the de-trended area fixed effect is at 0.074 (against 0.065). That for the effect of age and its square is unchanged at

Table 7: Estimation results for the second stage of equation (13)

Regression	(1) Levels OLS 1	(2) Levels FGLS	(3) Levels OLS 2	(4) Levels 2SLS	(5) First-Dif OLS	(6) First-Dif 2SLS
log *Density*	0.0625^a	0.0618^a	0.0584^a	0.0562^a	0.0336^a	-0.0281
	(0.0005)	(0.0005)	(0.0005)	(0.0041)	(0.0031)	(0.0274)
log *Area*	0.0344^a	0.0359^a	0.0419^a	0.0245^b	-	-
	(0.0008)	(0.0008)	(0.0008)	(0.0100)		
log *Diversity*	0.0007	-0.0008	-0.0033^a	-0.0507^a	-0.027	-0.0588
	(0.0014)	(0.0014)	(0.0012)	(0.0136)	(0.0021)	(0.0301)
log *Potential*			0.0279^a	0.0192^a	-0.0627	0.2527^b
			(0.0008)	(0.0027)	(0.0474)	(0.1259)
Sea			0.0151^a	0.0059^b	-	-
			(0.0020)	(0.0029)		
Mountain			0.0435^a	0.0307^a	-	-
			(0.0019)	(0.0026)		
Lake			-0.0143^a	-0.0154^a	-	-
			(0.0033)	(0.0055)		
Heritage			-0.0266^a	-0.0389^a	-	-
			(0.0027)	(0.0042)		
Time dummies	Yes	Yes	Yes	Yes	Yes	Yes
R^2 (within time)	77%	-	82%	-	-	-

6,820 observations. Standard error between brackets. c: significant at 10%, b: significant at 5%, and a: significant at 1%. In columns 4, density, land area, and diversity are instrumented by urban population density in 1831, 1861, 1891, and 1921 together with market potential computed using 1831 data and mean distances to other areas. The R^2 for the instrumental regressions are 0.64 for density, 0.35 for area, 0.17 for diversity, and 0.92 for market potential. A test of overidentifying restrictions shows that instruments are valid at 5%. All our instrumented variables are endogenous at 5%. In column 6, we instrument the changes in log density, log diversity and log area with the same variables as in column 4 plus a set of variables from the 1968 population census: mean age, mean age when leaving education, shares of the different occupational groups, share of population born in France, share of workers employed in the public sector, share of population living in an accommodation with hot water, with flushing toilet, with toilet inside, share of people living in a 'normal accommodation' (apartment or house as opposed to second residence, flat-share, etc), and mean deterioration of accommodation. The R^2 for the instrumental regressions are 0.35 for changes in density, 0.05 for changes in diversity, and 0.89 for changes in market potential.

Table 8: Correlation between the effects of the variables after aggregation by area and year

	area f.-e.	density	area	diversity	market potential	residuals (agg)
mean worker f.-e.	0.29	0.44	0.22	-0.01	0.17	-0.10
area f.-e.	1	0.77	0.34	-0.23	0.62	0.56
density		1	0.58	-0.21	0.52	0.02
land area			1	0.25	0.49	-0.39
diversity				1	-0.10	-0.42
market potential					1	0.04

2,046 observations computed from the estimations at the individual level (using column 4 of Table 6). Area fixed effects are estimated from (8) and we subtracted time fixed effects estimated from (10). Worker fixed effects are estimated from (8) and then averaged by employment area. The effects of density, land area and diversity are computed using their coefficients as estimated in (10) times the value of the variable.

power was to be expected given that these variables now capture both the skill composition of the local industry and some interactions therein. For the other variables (specialisation in particular), this indicates that some correlation with individual unobserved heterogeneity is present.

As can be seen from Table 7, the same conclusion arises with the second stage of the regression. The R^2 (*within* time) of the second stage of the baseline regression is well above what we obtained with individual data at 77% (against 60%). Hence when workers' unobserved heterogeneity is not controlled for, some of it is captured by aggregate variables.

Consistent with the previous finding, we also find that the first-stage coefficients are much higher than with individual data. Because they capture within-industry interactions together with compositional effects, the coefficients on the share of professionals are much higher than with individual data. More interestingly the specialisation coefficients are also much higher: on average 4.3% against 2.1%. Similar discrepancies occur with regard to the second stage coefficients (see Table 7). In the most basic specification (column 1), the coefficient on density is at 6.3% instead of 3.7% with individual data. That on land area is at 3.4% against 1.1% with individual data. In the aggregate data equivalent of our preferred specification (column 4), we find that the coefficient on employment density is still at 5.6% against 3.0% with individual data.

As can be seen from Table 8, the discrepancies between estimations with aggregate and individual data are easily explained by the sorting of workers by skills. We have already underlined in Section 4 that the correlation between the average worker fixed effect by area and the de-trended area-year fixed effect at 0.29 is high in individual regressions. It is even higher (0.53) when the area-year fixed effects are computed on aggregate data. In conclusion, when sorting is not taken into account the coefficient on density is over-estimated by nearly 100%, that on land area is over-estimated by up to several orders of magnitude whereas those on specialisation are also over-estimated by 100%. These are clearly large biases.

8 Concluding comments

This paper proposes a general framework to investigate the sources of wage disparities across local labour markets: skills, endowments and within- and between-industry interactions. This framework unites different strands of literature that were so far mostly disjoint. It shows that the research about the 'estimation of agglomeration economies' is closely intertwined with those dealing with 'regional disparities', 'local labour markets' and 'migration'. Empirically, the main novelty of the paper is to use a very large panel of workers and a consistent approach to exploit it. This allows us to assess precisely the effects of unobserved worker heterogeneity. We find that the effect of individual skills is quantitatively very important in the data. Up to half of the spatial wage disparities can be traced back to differences in the skill composition of the workforce. Workers with better labour market characteristics tend to agglomerate in the larger, denser and more skilled local labour market. We believe more work is now needed to understand the nature of this sorting.[26]

We also pay considerable attention to the issues of simultaneity. When correcting for possible biases, our estimates for economies of density, at around 3.0%, are lower than in previous literature. Nonetheless, economies of density still play an important role in explaining differences in local wages.

0.058, that for industry fixed effects is at 0.097 (against 0.043), that for specialisation is at 0.047 (against 0.017), and that for the number of establishments is at 0.035 (against 0.019). Finally with aggregate data the standard deviation for the share of professionals is four times as large at 0.046 (against 0.011). The effect of all the occupations has a standard deviation equal to 0.110.

[26] One explanation could be based on a self-selection effect in internal migrations. As suggested long ago by Alfred Marshall, it may be that "the most enterprising, the most highly gifted, those with the highest physique and strongest character go [to the large towns] to find scope for their abilities" (Marshall, 1890). Nocke (2006) proposes a formalisation of this argument. Alternatively, the largest cities may offer some particular amenities that appeal more to the workers commanding the highest wages. A third hypothesis (Glaeser and Maré, 2001; Wheeler, 2006) is that workers may learn more in larger cities.

We find that the market potential also matters. The evidence on other types of local interactions such as those taking place within particular industries is more mixed. They are significant but do not matter much quantitatively in explaining local wages disparities. Our approach also suggests at best a modest direct role for local non-human endowments in the determination of local wages.

Appendix A Data description and background

A detailed description of the wage data can be found in the working paper version of this article and in Abowd et al. (1999). A detailed description of French employment areas appears in Combes (2000) and in our working paper (Combes et al, 2004). Finally, Cohen et al (1997) provide some background about wage setting in France as well as international comparisons. In this appendix, we briefly describe our treatment of the data.

- *Missing years.* Three years (1981, 1983 and 1990) are missing due to lack of sampling by INSEE during census periods.

- *Wages, earnings and labour costs.* For each observation, and using total net nominal earnings, number of days worked and work status (full-time or part-time), we computed an annualised nominal wage. We then added mandatory payroll taxes for both employees and employers (which differ over time, across wage levels, work status, and for textile workers) to obtain total annualised labour costs.

- *Imputed wages.* The original data contains imputed wages for some workers and missing years. Starting with 19,675,740 observations, we deleted all imputed values and ended up with 18,581,470 observations.

- *Missing values and coding errors.* We deleted all the observations for which one or more variables of interest was missing, the duration of employment was equal to zero, wages are negative, or workers were not born in October of even years. After these deletions, we were left with 17,495,335 observations. We also deleted all the observations for which we could not determine the industry of employment or the employment area. This left us with 16,458,989 observations.

- *Mainland private sector employees of working age.* We excluded all apprentices and workers not employed in the private sector. We also restricted the sample to workers aged 15 to 65 employed in mainland France. Workers employed in Corsica and overseas territories were deleted to end up with 14,067,326 observations.

- *Part-timers.* Because the number of hours is unknown before 1993, we excluded all part-time workers. In case of multiple observations for a worker over a given year (corresponding to more than one job), we kept only one observation (the one with the most working days). This left us with 10,551,810 observations.

- *Excluded industries.* We use a sectoral classification with 114 industries. Agriculture and fishing industries are not normally covered by the extract. Remaining workers in these sectors were excluded. We also excluded all industries with less than 500 observations over the period (Spatial transport, Extraction of uranium, and Extraction of metals). In a few industries, firms with a large number of establishments can aggregate their reporting at the regional level. We excluded these industries (Financial intermediation, Insurance, Financial auxiliaries, Telecommunications, and Postal services). Finally, we also excluded a few non-competitive industries (Public administration, Extra-territorial activities, and Associations). We ended up with 9,389,838 observations across 99 industries.

- *Outliers.* The initial data had a number of outliers with wages either unrealistically high or well below the minimum wage. These seem to be caused by reporting mistakes in the net nominal earnings or in the number of working days. We decided to get rid of the 3% lowest and highest wages for every year.

The final sample contains $8,826,422$ observations. When working with the 6 years we selected (1976, 1980, 1984, 1988, 1992, and 1996), the sample contains $2,664,474$ observations. When we aggregate the data by area, industry, and year we have $378,022$ observations for the 1976-1998 period.

Appendix B Endogeneity of location and industry choices

We examine here the necessary assumptions about migrations and workers flows between industries for the strict exogeneity of the industry and location of employment to be warranted.

Consider worker i having to choose an employment area and an industry in a static framework. We assume that this worker's utility depends only on her level of consumption of a composite good whose price is the same everywhere. Indirect utility can then be written as a function of the wage: $v = v(w)$. Worker i chooses her employment area and industry so as to maximise her wages net of the (monetary) costs of migration. This choice can be decomposed in three steps.

1. At the beginning of period t, any industry k in an employment area a can be characterised by a wage $w_{i,a,k,t}$. This wage depends not only on individual attributes and local characteristics of the industry, but also on a shock noted $\psi_{i,a,k,t}$. Using (4) and (5), the wage satisfies:

$$\log w_{i,a,k,t} = \log B_{a,k,t} + X_{i,t}\varphi + \delta_i + \psi_{i,a,k,t}. \tag{B 1}$$

   We assume that all the explanatory variables in $B_{a,k,t}$ and $X_{i,t}$ are strictly exogenous.

2. The worker then chooses an employment area $a(i,t)$ and an industry $k(i,t)$ so as to maximise her utility. Assume first that the worker knows the distribution of the shocks $\psi_{i,a,k,t}$ without knowing their exact values. The maximisation programme of the worker is then:

$$\max_{(a,k)\in t} E_{\psi_{i,a,k,t}} \left[v \left(w_{i,a,k,t} - c_{a,k} \right) \right], \tag{B 2}$$

   where $E_{\psi_{i,a,k,t}}$ is the expectation operator on the distribution of $\psi_{i,a,k,t}$, and $c_{a,k}$ is a mobility cost equal to zero when $a = a(i,t-1)$ and $k = k(i,t-1)$. In this case, the choice of $a(i,t)$ and $k(i,t)$ is independent from the realisation of $\epsilon_{i,t} = \psi_{i,a(i,t),k(i,t),t}$. The location and industry of employment are thus determined solely on the basis of exogenous variables entering the wage equation and the mobility costs. Hence, when the worker knows only the distribution of the shocks, the assumption of strict exogeneity is satisfied.
   Turning now to the case where the worker can observe all the $\psi_{i,a,k,t}$, the maximisation programme is:

$$\max_{(a,k)\in t} \left[v \left(w_{i,a,k,t} - c_{a,k} \right) \right]. \tag{B 3}$$

   In this case, the choice of $a(i,t)$ and $k(i,t)$ is correlated with the realisation of all shocks $\psi_{i,a,k,t}$, and in particular $\epsilon_{i,t} = \psi_{i,a(i,t),k(i,t),t}$. Hence, the assumption of strict heterogeneity of location and industry choice does not hold.

   There are finally intermediate cases for which only some $\psi_{i,a,k,t}$ are observed by the worker. If these observed shocks are not correlated with $\epsilon_{i,t}$, the exogeneity assumption is satisfied. If they are, the model is misspecified again.

3. After choosing an employment area and industry, the individual shock, $\epsilon_{i,t}$, is known and the worker is paid according to (7). The worker then faces the same decision at period $t+1$.

In a dynamic framework

Consider for simplicity that the explanatory variables other than area-year and industry dummies, noted $Y_{i\tau}$, are strictly exogenous. We also ignore savings. At period t, the worker chooses her location and industry taking into account all available information including the observed shocks $\psi_{i,a,k,t}$ and their past evolution. We introduce the following notations: $Y_i^t = \{Y_{i\tau}\}_{\tau \leqslant t}$ and $\psi_i^t = \{\psi_{i,a,k,\tau} \,|a \leqslant Z,\, k \leqslant K,\, \tau \leqslant t,\, \psi_{i,a,k,\tau} \text{ known by } i\}$. The vector of state variables at the beginning of period t is $\left(\psi_i^{t-1},\, a(i,t-1),\, k(i,t-1)\right)$. Past employment area $a(i,t-1)$ and industry $k(i,t-1)$ enter this vector because mobility costs can depend on them. The history of observed shocks ψ_i^{t-1} is included because it can be used to predict the current and future realisations of shocks. The sequences of expected locations and industries are noted $\{a(i,\tau)\}_{t \leqslant \tau \leqslant T}$ and $\{k(i,\tau)\}_{t \leqslant \tau \leqslant T}$, respectively, with T the last period of work for i. Any worker solves:

$$\max_{(a_t,k_t)\in t,\ldots,(a_T,k_T)\in T} E\left[\sum_{\tau=t}^{T} \rho^{\tau} v\left(w_{i,a_\tau,k_\tau,\tau} - c_{a_\tau,k_\tau}\right) \big| Y_i^t,\, \psi_i^{t-1},\, Z(i,t-1),\, K(i,t-1)\right], \tag{B 4}$$

with ρ the discount rate.
We can reach different conclusions depending on the dynamic process determining the shocks $\psi_{i,a,k,t}$. If we first suppose that shocks are idiosyncratic, the same conclusions as in the static case apply. The location $a(i,t)$ and the industry $k(i,t)$ are correlated with $\epsilon_{i,t}$ if and only if the worker can collect information on $\epsilon_{i,t}$ at period t. If we suppose instead that shocks follow an $AR(1)$ process and that the worker can obtain some information on $\epsilon_{i,t}$ through her history of shocks ψ_i^{t-1}, then three issues arise:

1. The location $a(i,t)$ and the industry $k(i,t)$ are correlated with $\epsilon_{i,t}$. This correlation is however weaker than in the static case because workers take into account future wages in their mobility decisions. Indeed, the information related to current shock present in future wage shocks is decreasing with the time horizon and becomes negligible when it grows arbitrarily large.

2. $a(i,t)$ and $k(i,t)$ are correlated with past shocks $\{\epsilon_{i,\tau}\}_{\tau<t}$ as shocks follow an $AR(1)$ process.

3. $a(i,t)$ and $k(i,t)$ are correlated with future shocks $\{\epsilon_{i,\tau}\}_{\tau>t}$. However, the predictive power of the information set at t decreases over time. Thus, the worker can form only inaccurate expectations about future shocks. Thus the correlation between $a(i,t)$ and $k(i,t)$ in the one hand, and $\epsilon_{i\tau}$, for $\tau > t$, in the other hand, decreases when τ increases.

These three remarks suggest that the results may be biased because the explanatory variables can be correlated not only with present shocks, but also with past and future shocks. However, although we may have more sources of bias than in the static case, these correlations are likely to be weak because workers take future wages into account in their mobility decision while having little information about future shocks. Extensions to other dynamic processes for the shocks are straightforward.

Appendix C Two-stage estimation

What follows is a complete description of our two-stage estimation procedure.

Equation (10) can be re-written compactly:

$$\beta = D\Phi + \eta, \tag{C 5}$$

where $\beta = (\beta_{1,1}, ..., \beta_{Z,T})'$, $\Phi = (w_0, \theta_1, ..., \theta_T, \gamma)'$, D is the matrix of all aggregate explanatory variables after vectorisation, and $\eta = (\eta_{1,1}, ..., \eta_{Z,T})'$.

An area-year fixed effect is set arbitrarily to zero to secure identification. Because the exact value of the area fixed effects is unknown, this equation cannot be directly estimated with OLS. It is however possible to compute a consistent and unbiased estimator of β from the first stage results. Note first that (C 5) can be transformed into:

$$\widehat{\beta} = D\Phi + \eta + \Psi, \tag{C 6}$$

where $\widehat{\beta} = (\widehat{\beta}_{1,1},...,\widehat{\beta}_{Z,T})'$ is the estimator of β obtained in the first stage of the regression (with $\widehat{\beta}_{1,1}$ set to zero for convenience) and $\Psi = \widehat{\beta} - \beta$ is a sampling error. Equation (C 6) can then be estimated in the following way:

1. Compute the OLS estimate of Φ from (C 6):

$$\widehat{\Phi}_{OLS} = (D'D)^{-1}D'\widehat{\beta} = \Phi + (D'D)^{-1}D'(\eta + \Psi) \tag{C 7}$$

2. It is then possible to define $\widehat{\sigma}^2$ such that:

$$\widehat{\sigma}^2 = \frac{1}{tr(M_D)} \left\{ \left(\widehat{\eta + \Psi}\right)' \left(\widehat{\eta + \Psi}\right) - tr\left[\widehat{V}\left(\Psi \left|\Omega\right.\right)\right] \right\}, \tag{C 8}$$

   where $M_D = I - D\left(D'D\right)^{-1}D'$, $\widehat{\eta + \Psi} = \widehat{\beta} - D\widehat{\Phi}_{OLS} = M_D(\eta + \Psi)$, Ω is the set of all explanatory variables in the model, and $\widehat{V}\left(\Psi \left|\Omega\right.\right)$ is the estimator of the covariance matrix obtained from the first stage estimation bordered with zeros in the first line and first column. As shown by Gobillon (2004), $\widehat{\sigma}^2$ is an unbiased estimator of σ^2 when η is orthogonal to ϵ. It is also consistent under some reasonable assumptions.

3. We can now compute an unbiased estimator of the covariance matrix $V\left(\eta + \Psi \left|\Omega\right.\right)$:

$$\widehat{V} = \widehat{\sigma}^2 I + \widehat{V}\left(\Psi \left|\Omega\right.\right). \tag{C 9}$$

4. Measurement errors on the dependant variable create some heteroscedasticity. To control for this, the feasible generalised least-square (FGLS) estimator of Φ can be computed. It is given by:

$$\widehat{\Phi}_{FGLS} = \left(D'\widehat{V}^{-1}D\right)^{-1} D'\widehat{V}^{-1}\widehat{\beta}. \tag{C 10}$$

5. Finally, it is possible to compute a consistent estimator of the variance of $\widehat{\Phi}_{FGLS}$:

$$\widehat{V}\left(\widehat{\Phi}_{FGLS} \left|\Omega\right.\right) = \left(D'\widehat{V}^{-1}D\right)^{-1}. \tag{C 11}$$

References

J. M. Abowd, F. Kramarz, D. N. Margolis, High wage workers and high wage firms, Econometrica 67 (2) (1999) 251–333.

A. Ciccone, Agglomeration effects in Europe, European Economic Review 46 (2) (2002) 213–227.

A. Ciccone, R. E. Hall, Productivity and the density of economic activity, American Economic Review 86 (1) (1996) 54–70.

P. R. Coelho, M. A. Ghali, The end of the North-South wage differential: Reply, American Economic Review 63 (4) (1973) 757–762.

D. Cohen, A. Lefranc, G. Saint-Paul, French unemployment: A transatlantic perspective, Economic Policy: A European Forum 0 (25) (1997) 265–285.

P.-P. Combes, Economic structure and local growth: France, 1984–1993, Journal of Urban Economics 47 (3) (2000) 329–355.

P.-P. Combes, G. Duranton, L. Gobillon, Spatial wage disparities: Sorting matters!, Discussion Paper 4240, CEPR (2004).

G. B. Dahl, Mobility and the return to education: Testing a Roy model with multiple markets, Econometrica 70 (6) (2002) 2367–2420.

G. Duranton, V. Monastiriotis, Mind the gaps: The evolution of regional earnings inequalities in the UK 1982–1997, Journal of Regional Science 42 (2) (2002) 219–256.

G. Duranton, D. Puga, Micro-foundations of urban agglomeration economies, in: V. Henderson, J.-F. Thisse (eds.), Handbook of Regional and Urban Economics, volume 4. North-Holland, Amsterdam, 2004, pp. 2063–2117.

S. N. Durlauf, D. T. Quah, The new empirics of economic growth, in: J. B. Taylor, M. Woodford (eds.), Handbook of Macroeconomics volume 1A, North-Holland, Amsterdam, 1999, pp. 231–304.

M. Fujita, P. R. Krugman, A. J. Venables, The Spatial Economy: Cities, Regions, and International Trade, MIT Press, Cambridge, MA, 1999.

R. Gibbons, L. Katz, Does unmeasured ability explain inter-industry wage differentials?, Review of Economic Studies 59 (3) (1992) 515–535.

E. L. Glaeser, H. Kallal, J. A. Scheinkman, A. Schleifer, Growth in cities, Journal of Political Economy 100 (6) (1992) 1126–1152.

E. L. Glaeser, D. C. Maré, Cities and skills, Journal of Labor Economics 19 (2) (2001) 316–342.

L. Gobillon, The estimation of cluster effects in linear panel model, processed, INED (2004).

L. Gobillon, D. Le Blanc, Migrations, income and skills, working Paper 2003-47, CREST-INSEE (2003).

J. V. Henderson, Efficiency of resource usage and city size, Journal of Urban Economics 19 (1) (1986) 47–70.

J. V. Henderson, Marshall's economies, Journal of Urban Economics 53 (1) (2003) 1–28.

A. B. Krueger, L. H. Summers, Efficiency wages and the inter-industry wage structure, Econometrica 56 (2) (1988) 259–293.

A. Marshall, Principles of Economics, Macmillan, London, 1890.

E. Moretti, Human capital externalities in cities, in: V. Henderson, J.-F. Thisse (eds.), Handbook of Regional and Urban Economics, volume 4, North-Holland, Amsterdam, 2004, pp. 2243–2291.

B. R. Moulton, An illustration of a pitfall in estimating the effects of aggregate variables on micro units, Review of Economics and Statistics 72 (2) (1990) 334–338.

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

J. E. Rauch, Productivity gains from geographic concentration of human capital: Evidence from the cities, Journal of Urban Economics 34 (3) (1993) 380–400.

J. Roback, Wages, rents and the quality of life, Journal of Political Economy 90 (6) (1982), 1257–1278.

S. S. Rosenthal, W. C. Strange, Evidence on the nature and sources of agglomeration economies, in: V. Henderson, J.-F. Thisse (eds.), Handbook of Regional and Urban Economics, volume 4, North-Holland, Amsterdam, 2004, pp. 2119–2171.

J. Temple, The new growth evidence, Journal of Economic Literature 37 (1) (1999) 112–156.

R. H. Topel, Local labor markets, Journal of Political Economy 94 (3(2)) (1986) S111–S143.

W. H. Wheeler, Cities and the growth of wages among young workers: Evidence from the NLSY, Journal of Urban Economics 60 (2) (2006) 162–184.