European
Firms
In a
Global
Economy

Jorge De la Roca and Diego Puga

Learning by working in big cities

European
Firms
In a
Global
Economy **EFIGE**

SEVENTH FRAMEWORK PROGRAMME

Learning by working in big cities

Jorge De la Roca[*][‡]
CEMFI

Diego Puga[*][§]
CEMFI and *CEPR*

30 July 2012

ABSTRACT: Individual earnings are higher in bigger cities. We consider three reasons: spatial sorting of initially more productive workers, static advantages associated with workers' current location, and learning by working in big cities. Using rich administrative data for Spain, we find that workers in bigger cities do not have higher unobserved initial ability, as reflected in individual fixed-effects. Instead, they obtain an immediate static premium while working in bigger cities and also accumulate more valuable experience, which increases their earnings faster. The additional value of experience accumulated in bigger cities persists even after workers move away and is even stronger for those with higher unobserved initial ability. This combination of effects explains both the higher mean and the greater dispersion of earnings in bigger cities.

Key words: learning, city size, earnings premium, agglomeration economies
JEL classification: R10, R23, J31

[‡]Centro de Estudios Monetarios y Financieros (CEMFI), Casado del Alisal 5, 28014 Madrid, Spain (e-mail: jdelaroca@cemfi.es).

[§]Centro de Estudios Monetarios y Financieros (CEMFI), Casado del Alisal 5, 28014 Madrid, Spain (e-mail: diego.puga@cemfi; website: http://diegopuga.org).
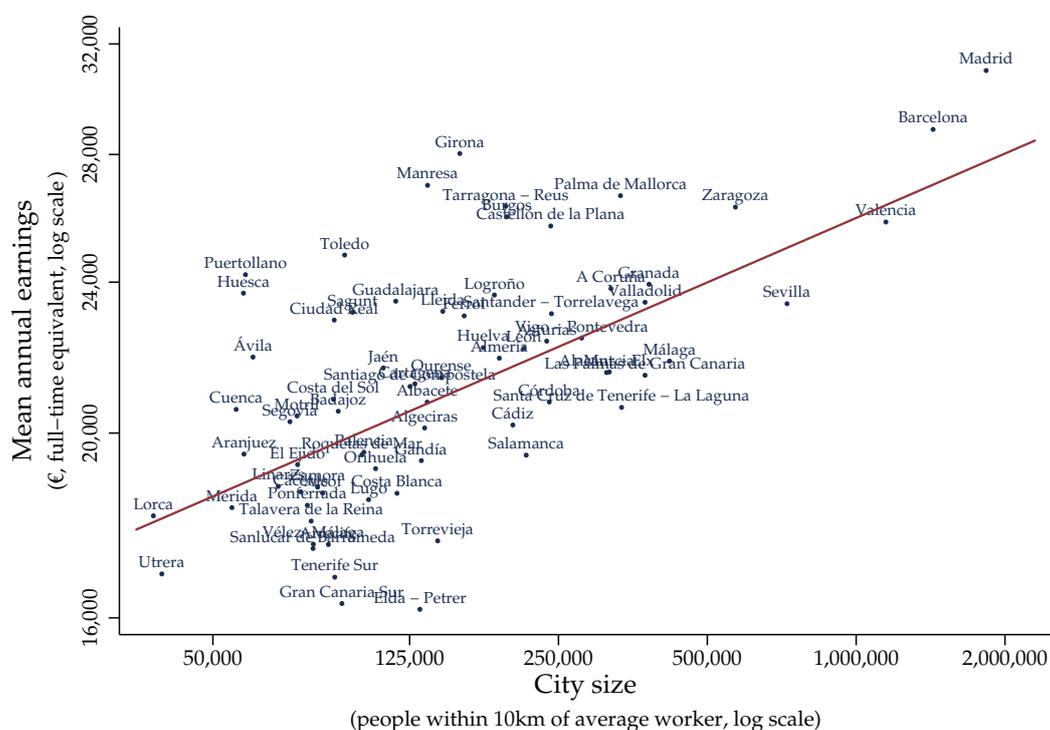
Figure 1: Mean earnings and city size

1. Introduction

Workers in bigger cities earn more than workers in smaller cities and rural areas. Figure 1 plots mean annual earnings for male employees against city size for Spanish urban areas. Workers in Madrid earn €31,000 on average, which is 20% more than workers in Valencia (the country's third biggest city), 46% more than workers in Santiago de Compostela (the median-sized city), and 56% more than workers in rural areas. The relationship between earnings and city size is just as strong in other developed countries.[1] Moreover, differences remain large even when we compare workers with the same education and years of experience and in the same industry. Higher costs of living may explain why workers do not flock to bigger cities, but that does not change the fact that firms are willing to pay higher wages in bigger cities to workers with similar characteristics because of higher productivity. In fact, Combes, Duranton, Gobillon, and Roux (2010) find that establishment-level productivity and wages both increase with city size with similar elasticity.

There are three broad reasons why firms may be willing to pay more to workers in bigger cities. First, there may be some static advantages associated with bigger cities that are enjoyed while working there and lost upon moving away. These static agglomeration economies have received the most attention (see Duranton and Puga, 2004, for a review of possible mechanisms and Rosenthal and Strange, 2004, and Holmes, 2010, for summaries of the evidence). Second, workers

---

[1]In the United States, workers in metropolitan areas with population above one million earn on average 30% more than workers in rural areas (Glaeser, 2011). In France, workers in dense metropolitan areas, i.e. those above the median density, earn on average 19% more than workers in less dense metropolitan areas (Combes, Duranton, Gobillon, and Roux, 2012).

who are inherently more productive may choose to locate in bigger cities. Evidence on such sorting is mixed, but some recent accounts (e.g. Combes, Duranton, and Gobillon, 2008) suggest it may be as important in magnitude as static agglomeration economies. Third, a key advantage of cities is that they facilitate experimentation and learning (Glaeser, 1999, Duranton and Puga, 2001). In particular, bigger cities may provide workers with opportunities to accumulate more valuable experience. Since these dynamic advantages are transformed in higher human capital, they may remain beneficial even when a worker moves away (Glaeser and Maré, 2001, Gould, 2007).

In this paper, we simultaneously consider these three potential sources of the city-size earnings premium: static advantages, sorting based on initial ability and dynamic advantages. We begin in section 2 with a methodological discussion of our approach and how it deals with biases present in earlier estimates in the literature. Then, in section 3, we discuss the rich administrative data for Spain that we use. This follows workers over time and across locations throughout their careers, thus allowing us to compare the earnings of workers in cities of different sizes while controlling for observed and unobserved ability and the experience previously acquired by workers in different cities.

To facilitate a comparison with previous studies, we begin our empirical analysis in section 4 with a simple pooled OLS estimation of the static advantages of bigger cities. This first estimation ignores both the possible sorting of workers with higher unobserved ability into bigger cities as well as any dynamic benefits of bigger cities. As a result, it also produces a biased estimate of the static advantages of bigger cities.

Following Glaeser and Maré (2001) and Combes *et al.* (2008), we introduce worker fixed effects to address the issue of workers sorting on unobservables. This leads to a substantial reduction in the elasticity of earnings with respect to city size, in line with earlier studies. This drop is usually interpreted as evidence of more productive workers sorting into bigger cities (Combes *et al.*, 2008). We show that it is instead the result of ignoring the dynamic benefits that big cities provide.

In section 5 we explicitly consider the dynamic benefits of bigger cities. Taking advantage of being able to follow a large panel of workers throughout their working lives, we let the value of experience vary depending both on where it was acquired and on where it is being used. Our results show that experience accumulated in big cities is more valuable and remains so when workers move elsewhere. We generalize this specification further in section 6, where we show that workers with higher innate ability enjoy greater learning advantages from bigger cities.

Finally, to get a better sense of whether there is sorting of workers with higher innate ability we compare the distribution of ability across cities of different sizes. This exercise is related to recent studies that also compare workers' ability and skills across cities, either by looking at education (e.g., Glaeser and Resseger, 2010), at broader measures of skills (e.g., Bacolod, Blum, and Strange, 2009), or at worker fixed-effects (e.g., Combes *et al.*, 2012). We study worker-fixed effects because we are interested in capturing time-invariant ability beyond that which may be reflected in observable characteristics. However, we show that if we do not estimate worker fixed effects using our full specification, then we end up mixing innate ability with the extra value of big-city experience.

Once we isolate innate ability from the value of experience accumulated in bigger cities, we find

sorting to be much less important than previously thought. Workers in big and small cities are not particularly different to start with, it is working in cities of different sizes that makes their earnings diverge. They attain a static earnings premium upon arrival in a bigger city and accumulate more valuable experience as they spend more time working there.[2] Because these gains are stronger for workers with higher unobserved initial ability, this combination of effects explains not just the higher mean but also the greater dispersion of earnings in bigger cities that Eeckhout, Pinheiro, and Schmidheiny (2010), Combes *et al.* (2012) and Baum-Snow and Pavan (2012*b*) emphasize.

2. Methodology

Suppose the log wage of worker i in city c at time t, w_{ict}, is given by

$$w_{ict} = \sigma_c + \mu_i + \sum_{j=1}^{C} \delta_{jc} e_{ijt} + x'_{it}\beta + \varepsilon_{ict} \,, \tag{1}$$

where σ_c is a city fixed-effect, μ_i is a worker fixed-effect, e_{ijt} is the experience acquired by worker i in city j up until time t, x_{it} is a vector of time-varying individual and job characteristics, the scalars δ_{jc} and the vector β are parameters, and ε_{ict} is an error term uncorrelated with the explanatory variables.[3]

Equation (1) allows for a static earnings premium associated with currently working in a bigger city, if the city fixed effect σ_c is positively correlated with the city's size. It also allows for the sorting of more productive workers into bigger cities, if the worker fixed-effect μ_i is positively correlated with the city's size. Finally, we conjecture that one of the advantages of bigger cities is that they let workers accumulate more valuable experience, so equation (1) allows experience accumulated in city j to have a different value which may be positively correlated with the city's size. This value of experience δ_{jc} is indexed by both j (the city where experience was acquired) and c (the city where the worker currently works) to allow the value of experience to vary depending not only on where it was acquired but also on where it is being used. In our estimations, we also include terms in e_{ijt}^2, which are relevant but left out of the equations in this section to simplify the exposition.

Static pooled estimation

Imagine that, instead of estimating equation (1), we ignore both unobserved worker heterogeneity and any dynamic benefits of working in bigger cities, and estimate the following relationship:

$$w_{ict} = \sigma_c + x'_{it}\beta + \eta_{ict} \,. \tag{2}$$

---

[2]This is consistent with the counterfactual simulations of the structural model in Baum-Snow and Pavan (2012*a*), which suggest that returns to experience and wage-level effects are the most important mechanisms contributing to the overall city-size wage premium. Baum-Snow and Pavan (2012*a*) address unobserved ability by using a three-type mixture model where the probability of a worker being of certain type depends on the city where he enters the labor market. Since we have a much larger sample (150,000 men observed monthly compared with 1,700 men observed annually) we are able to follow a fixed-effect random-coefficient approach to obtain individual-specific returns to experience and individual specific abilities — and, hence, obtain the distribution of ability for each type of city. In this way we avoid making assumptions on the relationship between observables and unobservables.

[3]The city fixed-effect σ_c could also be time-varying and written σ_{ct} instead. We keep it time-invariant here for simplicity. In our estimations, we have tried both having time-varying and time-invariant city fixed-effects and find that it makes no practical difference for our estimates.

Compared with equation (1), in equation (2) the worker fixed-effect μ_i and the urban experience terms $\sum_{j=1}^{C}\delta_{jc}e_{ijt}$ are missing. Equation (2) can be estimated by ordinary least squares with a cross section of workers or a pooled panel.

Assuming for simplicity that $\text{Cov}(x_{it}, \mu_i + \sum_{j=1}^{C}\delta_{jc}e_{ijt}) = \mathbf{0}$, the resulting pooled OLS estimate of σ_c would be unbiased if and only if

$$\text{Cov}(\iota_{ict}, \eta_{ict}) = 0 \,, \tag{3}$$

where ι_{ict} is a city indicator variable that takes value 1 if worker i is in city c at time t and value 0 otherwise. However, if the richer wage determination of equation (1) holds, the error term of equation (2) includes the omitted variables:

$$\eta_{ict} = \mu_i + \sum_{j=1}^{C}\delta_{jc}e_{ijt} + \varepsilon_{ict} \,. \tag{4}$$

Hence,

$$\text{Cov}(\iota_{ict}, \eta_{ict}) = \text{Cov}(\iota_{ict}, \mu_i) + \text{Cov}(\iota_{ict}, \sum_{j=1}^{C}\delta_{jc}e_{ijt}) \neq 0 \,. \tag{5}$$

Equation (5) shows that a static cross-section or pooled OLS estimation of σ_c suffers from two key potential sources of bias. First, it ignores sorting, and thus the earnings premium for city c, σ_c, is biased upwards if individuals with high unobserved ability, μ_i, are more likely to work there, so that $\text{Cov}(\iota_{ict}, \mu_i) > 0$ (and biased downwards in the opposite case). Second, it ignores dynamic effects, and thus the earnings premium for city c, σ_c, is biased upwards if individuals with more valuable experience, $\sum_{j=1}^{C}(\delta_{jc} - \delta)e_{ijt}$, are more likely to work there, so that $\text{Cov}(\iota_{ict}, \sum_{j=1}^{C}\delta_{jc}e_{ijt}) > 0$ (and biased downwards in the opposite case).[4]

To see how these biases work more clearly, it is useful to consider a simple example. Suppose there are just two cities, one big and one small. Everyone working in the big city enjoys an instantaneous (static) log wage premium of σ. Workers in the big city have higher unobserved ability, which increases their log wage by μ. Otherwise, all workers are initially identical. Over time, experience accumulated in the big city increases log wage by δ per period relative to having worked in the small city instead. For now, assume there is no migration. If there are n time periods, then the pooled OLS estimate of the static big city premium σ has probability limit $\text{plim}\,\hat{\sigma}_{\text{pooled}} = \sigma + \mu + \frac{n+1}{2}\delta$. Thus, a pooled OLS regression overestimates the actual premium by the value of higher unobserved worker ability in the big city (μ) and the higher average value of accumulated experience in the big city ($\frac{n+1}{2}\delta$).

---

[4]Strictly speaking, the actual bias in the pooled OLS estimate of σ_c, $\hat{\sigma}_{c\,\text{pooled}}$, is more complicated because it is not necessarily the case that $\text{Cov}(x_{it}, \mu_i + \sum_{j=1}^{C}\delta_{jc}e_{ijt}) = \mathbf{0}$, as we have assumed. For instance, even if we do not allow the value of experience to vary by city, we may have overall experience, $e_{it} \equiv \sum_{j=1}^{C}e_{ijt}$, as one of the explanatory variables included in x_{it} in equation (2). In this case, δ_{jc} measures the differential value of the experience acquired in city j when working in city c relative to the general value of experience, which we may denote γ. Then $\text{plim}\,\hat{\sigma}_{c\,\text{pooled}} = \sigma_c + \text{Cov}(\iota_{ict}, \mu_i)/\text{Var}(\iota_{ict}) + \sum_{j=1}^{C}\delta_{jc}\text{Cov}(\iota_{ict}, e_{ijt})/\text{Var}(\iota_{ict}) + (\gamma - \hat{\gamma}_{\text{pooled}})\text{Cov}(\iota_{ict}, e_{it})/\text{Var}(\iota_{ict})$. Relative to the simpler example discussed in the main text, the bias incorporates an additional term $(\gamma - \hat{\gamma}_{\text{pooled}})\text{Cov}(\iota_{ict}, e_{it})/\text{Var}(\iota_{ict})$. In practice, this additional term is negligible if $\text{Cov}(\iota_{ict}, e_{it})$ is close to zero, that is, if the total number of days of work experience (leaving aside where it was acquired) is not systematically related to workers' location. In our sample, this is indeed the case: the correlation between mean experience and log city size is not significantly different from 0.

Static fixed-effects estimation

Following Glaeser and Maré (2001) and, more recently, Combes *et al.* (2008), a possible approach to address the issue of workers sorting on unobservables is to introduce worker fixed-effects. Suppose we consider unobserved worker heterogeneity in this way, but still ignore a dynamic city size premium and estimate the following relationship:

$$w_{ict} = \sigma_c + \mu_i + x'_{it}\beta + \zeta_{ict} . \tag{6}$$

Compared with equation (1), the city-specific experience terms $\sum_{j=1}^{C}\delta_{jc}e_{ijt}$ are missing from equation (6). Compared with equation (2), the worker fixed-effect μ_i is included. To estimate σ_c we now need a panel. The worker fixed-effect μ_i can be eliminated by subtracting from equation (6) the time average for each worker:

$$(w_{ict} - \bar{w}_i) = \sum_{j=1}^{C}\sigma_c(\iota_{ict} - \bar{\iota}_{ic}) + (x'_{it} - \bar{x}'_i)\beta + (\zeta_{ict} - \bar{\zeta}_i) . \tag{7}$$

Note that σ_c is now estimated only on the basis of migrants — for workers who are always observed in the same city $\iota_{ict} = \bar{\iota}_{ic} = 1$ every period.

Assuming again for simplicity that $\text{Cov}(x_{it}, \sum_{j=1}^{C}\delta_{jc}e_{ijt}) = 0$, the resulting fixed-effects estimate of σ_c is unbiased if

$$\text{Cov}\left((\iota_{ict} - \bar{\iota}_{ic}), (\zeta_{ict} - \bar{\zeta}_i)\right) = 0 . \tag{8}$$

However, if the richer wage determination of equation (1) holds,

$$(\zeta_{ict} - \bar{\zeta}_i) = \sum_{j=1}^{C}\delta_{jc}(e_{ijt} - \bar{e}_{ij}) + (\varepsilon_{ict} - \bar{\varepsilon}_i) , \tag{9}$$

and thus

$$\text{Cov}\left((\iota_{ict} - \bar{\iota}_{ic}), (\zeta_{ict} - \bar{\zeta}_i)\right) = \text{Cov}\left((\iota_{ict} - \bar{\iota}_{ic}), \sum_{j=1}^{C}\delta_{jc}(e_{ijt} - \bar{e}_{ij})\right) \neq 0 . \tag{10}$$

Worker fixed effects take care of unobserved worker heterogeneity. However, the estimate of σ_c is still biased because dynamic effects are ignored. The earnings premium for city c is biased upwards if the value of workers' experience tends to be above their individual averages in the periods when they are located in city c. It is biased downwards when the reverse is true.

Again, to see how this bias works more clearly, it is instructive to use the same simple two-city example as for the pooled OLS estimate. Like before, everyone working in the big city enjoys an instantaneous (static) log wage premium of σ. Workers in the big city have higher unobserved ability, which increases their log wage by μ. Otherwise, all workers are initially identical. Over time, experience accumulated in the big city increases log wage by δ per period relative to having worked in the small city instead. Since with worker fixed-effects σ_c is estimated only on the basis of migrants, we add migration to the example. Consider two cases.

First, suppose all migration is from the small to the big city and takes place after migrants have worked in the small city for the first m periods of the total of n periods. The fixed-effects estimate of the static big city premium σ is now estimated by comparing the earnings of migrants before and after moving and has probability limit plim $\hat{\sigma}_{\text{FE}} = \sigma + \frac{n-m+1}{2}\delta$. With all migrants moving from the small to the big city, the fixed-effects regression overestimates the actual static premium (σ)

by the average extra value of the experience migrants accumulate by working in the big city after moving ($\frac{n-m+1}{2}\delta$). The estimation of equation (6) forces the earnings premium to be a pure jump at the time of moving, while in the example it actually has both static and dynamic components. Not trying to separately measure the dynamic part not only hides this, but makes the static part seem larger than it is.

Consider next the case where all migration is from the big to the small city and takes place after migrants have worked in the big city for the first m periods of the total of n periods. Now, we also need to know whether the extra value of experience accumulated in the big city is fully portable or only partially so. Assume only a fraction θ is portable. The fixed-effects estimate of the static big city premium σ then has probability limit plim $\hat{\sigma}_{FE} = \sigma - \frac{m-1}{2}\theta\delta$. With all migrants moving from the big to the small city, the fixed-effects regression underestimates the actual static premium (σ) by the average extra (but depreciated) value of the experience migrants acquired in the big city prior to moving ($\frac{m-1}{2}\theta\delta$). By forcing both the static and dynamic premium to be captured by a discrete jump, the jump now appears to be smaller than it is. The dynamic part is still not separately measured.

This example shows that the estimation with worker fixed-effects deals with the possible sorting of workers across cities on time-invariant unobservable characteristics. However, the estimates of city fixed-effects are still biased due to not considering dynamic benefits. This, in turn, biases any estimate of the earnings premium associated with bigger cities. Migrants from low to big cities tend to bias the static city size premium upwards (their average wage difference across cities is 'too high' because when in big cities they benefit from the more valuable experience they are accumulating there). Migrants from big to small cities tend to bias the static city size premium downwards (their average wage difference across cities is 'too low' because when in small cities they still benefit from the more valuable experience accumulated in big cities). In practice, the bias is likely to be small if the sample is more or less balanced in terms of migration flows across different types of cities, and the learning benefits of bigger cities are highly portable (in the example, if θ is close to 1). The first condition, that migration is balanced, is likely to be true given that gross migration flows are generally large relative to net flows.[5] The second condition, that the learning benefits of bigger cities are highly portable, is one that we cannot assess without actually estimating the fully-fledged specification of equation (1).

The static earnings premium associated with working in bigger cities has been found to be about twice as large when estimated using either pooled or aggregate data, such as that of equation (2), than when estimated using a specification with worker fixed-effects, such as that of equation (6). This is shown by Combes *et al.* (2008), who interpret the difference as evidence of the importance of sorting by more productive workers into bigger cities.

In this section, we have shown that, if learning effects such as those included in equation (1) are important, then the estimation of equation (6) affects not just one, but two sources of bias present in the estimation of equation (2). By including worker fixed-effects, equation (6)

---

[5]In the sample of 150,375 workers that we use in this paper, between 2004 and 2009 there are 8,356 migrations from the five biggest cities to smaller cities in Spain, 8,362 migrations from smaller cities to the five biggest cities, and another 20,725 moves between cities of similar sizes.

addresses the bias arising in equation (2) from workers' possibly sorting on the basis of unobserved idiosyncratic ability. However, including worker fixed-effects also affects the magnitude of the bias in the estimated static city size premium arising from ignoring the dynamic component of the premium. It will not formally eliminate it but, under certain conditions, it can greatly reduce it. The lower static earnings premium found when using worker fixed-effects could thus reflect either the importance of sorting by workers across cities in a way that is systematically related to unobserved ability, or the importance of learning by working in big cities, or a combination of both. We cannot know unless we simultaneously consider the static and the dynamic components of the earnings premium while allowing for unobserved worker heterogeneity. This requires estimating a specification such as (1), where the worker fixed-effect μ_i can be again eliminated by subtracting the time average for each worker:

$$(w_{ict} - \bar{w}_i) = \sum_{j=1}^{C} \sigma_c (\iota_{ict} - \bar{\iota}_{ic}) + \sum_{j=1}^{C} \delta_{jc} (e_{ijt} - \bar{e}_{ij}) + (x'_{it} - \bar{x}'_i)\boldsymbol{\beta} + (\varepsilon_{ict} - \bar{\varepsilon}_i) . \qquad (11)$$

However, the main reason to estimate a specification that allows workers to accumulate more valuable experience in bigger cities and also to put this to use elsewhere is not to see how accurate are current estimates of the static advantages of bigger cities. The main reason is that those static advantages may only be part of what bigger cities provide. Thus, after estimating the restricted specifications of equations (2) and (6) for comparison with earlier studies, we estimate an expression like equation (1). This allows us to separately estimate the static advantages associated with workers' current location, and the dynamic advantages arising from the more valuable experience individuals acquire by working in big cities. We are also able to investigate the extent to which the learning benefits of bigger cities are portable and move with the workers if they relocate. Finally, we can also re-examine the importance of sorting based on initial unobserved ability.

Further econometric concerns

While our approach addresses several econometric biases present in previous studies, there are still two key sources of concern in our estimation. First, the static earnings premium of bigger cities is estimated on the basis of migrants alone and they may not be a representative sample of the population. This is less of an issue for the learning advantages, since the extra value of experience acquired in bigger cities is estimated on the basis of both migrants and stayers.

Second, workers do not migrate randomly. As long as they choose their location based on their fixed-effects (i.e. time-invariant ability) or the explanatory variables in equation (11), the estimation of σ will remain unbiased. However, any unobserved time-varying factor that is correlated with earnings will bias our results. For instance, it could be that workers can anticipate whether they will get a large benefit or a small one from working in bigger cities and only those who get a large gain migrate to bigger cities. A solution to this concern would be to instrument for workers' location. Unfortunately, we are not aware of an instrument that can simultaneously explain migration decisions, exhibit spatial and temporal variation, while being uncorrelated with wage determination. Something we can do is to check whether coefficients are different for workers moving from small to big cities and for workers moving from big to small cities. If workers

migrate based on the magnitude of their gain, workers moving from small to big cities should exhibit large gains while workers moving from big to small cities should exhibit small gains. The fact that we find similar gains for migrants moving either way somewhat alleviates our concerns.

3. Data

Employment histories and earnings

Our main data set is Spain's Continuous Sample of Employment Histories (*Muestra Continua de Vidas Laborales* or MCVL). This is an administrative data set with longitudinal information obtained by matching social security, income tax, and census records for a 4% non-stratified random sample of the population who on a given year have any relationship with Spain's Social Security (individuals who are working, receiving unemployment benefits, or receiving a pension). The criterion for inclusion in the MCVL (based on the individual's Social Security number) is maintained across MCVL waves.[6] We combine five editions of the MCVL, beginning with the first produced, for 2004, so as to have data on an approximately 4% random sample of all individuals who have worked, received benefits or a pension in Spain at any point in 2004–2009.

A crucial feature of the MCVL for our purposes is that workers can be tracked across space based on their workplace location. Social Security legislation requires employers to keep separate contribution account codes for each province in which they conduct business. Furthermore, within a province, a municipality identification code is provided if the workplace establishment is located in a municipality with population greater than 40,000 inhabitants in 2001.

The unit of observation in the source social security data is any change in the individual's labour market status or any variation in job characteristics (including changes in occupation or contractual conditions within the same firm). The data record all changes since the date of first employment, or since 1981 for earlier entrants. Using this information, we construct a panel with monthly observations tracking the working life of individuals in the sample. On each date, we know the individuals's labour market status and, if working, the occupation and type of contract, working hours expressed as a percentage of a full-time equivalent job, the establishment's sector of activity at the NACE 3-digit level, and the establishment's location. Furthermore, exploiting the panel dimension, we can construct precise measures of tenure and experience, calculated as the actual number of days the individual has been employed, respectively, in the same establishment and overall. We can also track cumulative experience in different locations or sets of locations.

Earnings are derived from income tax data for the year of each MCVL edition, where each source of labour income recorded in income tax records is matched to social security records based on both employee and employer (anonymized) identifiers. Gross labour earnings and tax withholdings are

---

[6]More recent editions add individuals who enter the labour force for the first time while they lose those who cease affiliation with the Social Security. Since individuals who stop working remain in the sample while they receive unemployment benefits or a retirement pension, most exits occur when individuals are deceased or leave the country permanently.

recorded separately for each job. This allows us to compute monthly labour earnings, expressed as euros per day of full-time equivalent work, during the period 2004–2009.[7]

The MCVL also provides individual characteristics contained in social security records, such as age and gender, and also characteristics contained in Spain's Continuous Census of Population (Padrón Continuo), such as country of birth, nationality, and educational attainment.[8]

Urban areas

We use official urban area definitions, constructed by Spain's Department of Housing in 2008 and maintained unchanged since then. The 85 urban areas account for 68% of Spain's population and 10% of its surface. Four urban areas (Madrid, Barcelona, Valencia and Sevilla) have populations above one million, Madrid being the largest with 5,966,067 inhabitants in 2009. At the other end, Teruel is the smallest with 35,396 inhabitants in 2009. Urban areas contain 747 municipalities out of the over 8,000 that exhaustively cover Spain. There is large variation in the number of municipalities per urban area. The urban area of Barcelona is made up of 165 municipalities while 21 urban areas contain a single municipality.

Six urban areas (Denia - Jávea, Valle de la Orotava, Blanes - Lloret de Mar, Sant Feliú de Guixols, Soria, and Teruel) have no municipality with a population of at least 40,000 in 2001, and are not included in the analysis since they cannot be identified in the MCVL. We must also exclude the four urban areas in the Basque Country and Navarre (Bilbao, San Sebastián, Vitoria and Pamplona) because we lack earnings from tax returns data since the Basque Country and Navarre collect taxes independently. Last, we exclude Ceuta and Melilla given their special enclave status in continental Africa. This leaves 73 urban areas for which we carry out our analysis.

To measure the scale of each urban area, we calculate the number of people within 10 kilometres of the average person. We do so starting with population counts at the level of individual municipalities from Spain's Continuous Census of Population (Padrón Continuo). We then allocate population within the municipality more finely on the basis of LandScan (Oak Ridge National Laboratory, 2009), a global population data set developed for the United States Department of Defense with a resolution of approximately 1 square km. ($30{\times}30$ arc-seconds) showing spatial distribution patterns of ambient population (average over 24 hours). Finally, we take each $30{\times}30$ arc-seconds cell in the urban a area, trace a circle of radius 10 km. around the cell (encompassing both areas inside and outside the urban area), count population in that circle, and average this count over all cells in the urban area weighting by the population in each cell. This yields the number of people within 10 km. of the average person in the urban area.

---

[7]The MCVL also contains earnings data from social security records going back to 1981 but, unlike the uncensored income tax data that we use to compute monthly earnings, these are either top or bottom coded for about 8% of observations.

[8]A complete national update of the educational attainment of individuals recorded in the Continuous Census of Population was performed in 1996, with a subsequent update by most municipalities in 2001. Beyond that year, any updates happen when individuals complete their registration questionnaire at a new municipality upon moving (a pre-requisite for access to local health and education services) or voluntarily communicate to their municipality a change in their highest level of education.

Our measure of city size is highly correlated with a simple population count (0.94), but deals more naturally with unusual urban areas, in particular those that are polycentric.[9] Most urban areas in Spain comprise a single densely populated urban centre and contiguous areas that are closely bound to the centre by commuting and employment patterns. However, a handful of urban areas are made up of multiple urban centres. A simple population count for these polycentric urban areas tends to exaggerate their scale, because to maintain contiguity they incorporate large intermediate areas that are often only weakly connected to the various centers. For instance, the urban area of Asturias incorporates the cities of Gijón, Oviedo, Avilés, Mieres, and Langreo as well as large areas in between. A simple population count would rank the urban area of Asturias seventh in Spain in terms of its 2009 population (835,231), just ahead of Zaragoza (741,132). Our measure of scale ranks Asturias twenty-third in terms of people within 10 km. of the average person (242,099) and Zaragoza sixth (585,921), which is arguably a more accurate characterization of their relative scale.

Sample restrictions

Our starting sample is a monthly data set for men born in Spain between 1963 and 1991 (i.e., aged 18–46 during the period 1981–2009) and employed at any point between January 2004 and December 2009.[10] This initial sample has 249,227 workers and 11,803,962 monthly observations.

We track workers over time throughout their working life, but study them only when employed in an urban area in 2004–2009. Job spells in the Basque Country and Navarre are excluded because these autonomous regions collect income taxes independently from Spain's national government and we do not have earnings data from income tax records for them. We also exclude job spells in six small urban areas because workplace location is not available for municipalities with population below 40,000 in 2001. Nevertheless, the days worked in urban areas within the Basque Country or Navarre, in the six small excluded urban areas, or in rural areas anywhere in the country are still counted when computing cumulative experience. All of this restricts the sample to 183,447 workers and 7,154,764 monthly observations.

Job spells in agriculture, fishing, mining and other extractive industries are excluded because these activities are typically rural and are covered by special social security regimes where workers tend to self-report earnings and the number of working days recorded is not reliable. Job spells in the public sector, international organizations, and in education and health services are also excluded because earnings in these sectors are heavily regulated by the national and regional governments. Apprenticeship contracts and certain rare contract types are also excluded. Finally, we drop workers who have not worked at least 30 days in any year. This yields our final sample of 150,375 workers and 5,821,846 monthly observations.

[9]Nevertheless, we have tried reestimating our results using a simple population count as an alternative measure of city scale and coefficients remain similar, although they are estimated less precisely.

[10]We focus on men due to the huge changes experienced by Spain's female labour force during the period over which we track labour market experiences. Most notably, the participation rate for prime-age women (25–54) increased from 30% in 1981 to 77% in 2009. We leave out foreign-born workers and those born before 1963 because we cannot track their full labour histories. We exclude spells workers spend as self-employed because labour earnings and workplace locations are not available during such periods, but still include job spells as employees for the same individuals.
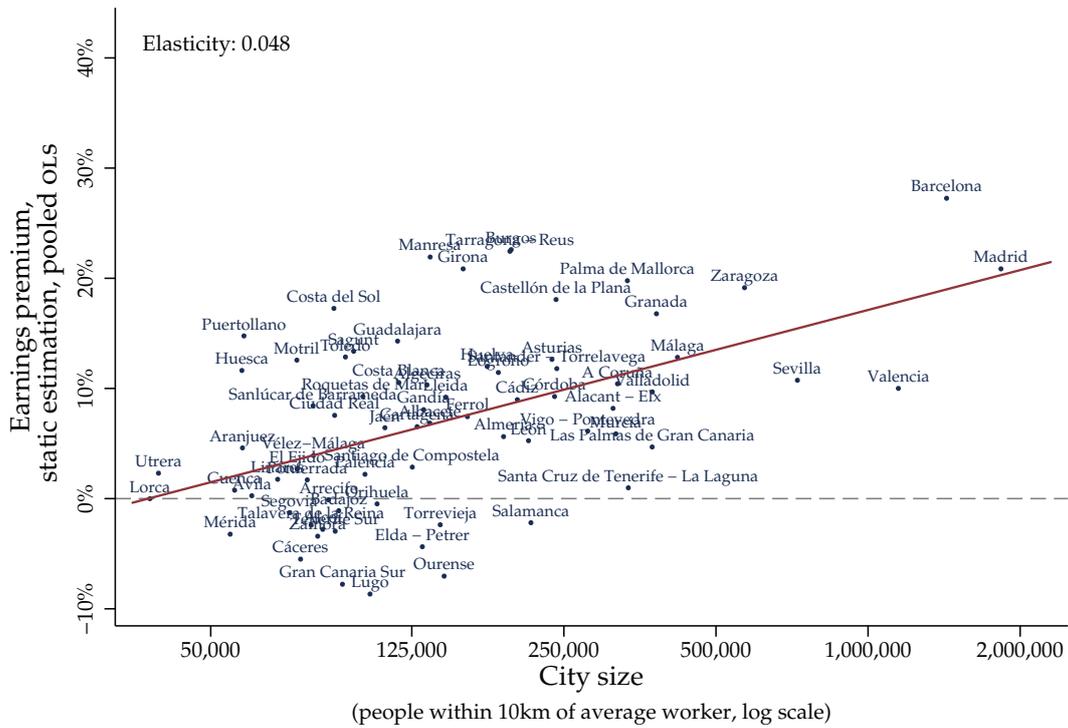
Figure 2: Static OLS estimation of the city-size premium

4. Static benefits of bigger cities

We begin by pooling the data and estimating the static city-size earnings premium without taking into account neither learning effects nor unobserved worker heterogeneity. We do this in a two-stage process. In the first stage we estimate equation (2), regressing log daily earnings on a complete set of city indicators, while controlling for individual and job characteristics. Then, in a second stage, we regress the coefficients of the city indicators on our measure of log size to estimate the elasticity of earnings with respect to city size.

Columns (1) and (2) in table 1 show the results for this two-stage estimation. As we would expect, column (1) shows that log earnings are concave in overall experience and tenure in the firm and increase monotonically with occupational skills.[11] Having tertiary education and working under a full-time and permanent contract are also associated with higher earnings.

In column (2) the estimate of the elasticity of earnings with respect to city size is 0.048. More detail on the numbers behind this estimate can be seen in figure 2, which plots the city indicators estimated in column (1) against log city size. We find sizable geographic differences in earnings even for observationally-equivalent workers. For instance, a worker in Madrid earns 21% more than a worker with the same observable characteristics in Lorca — the smallest city in our sample. The largest earning differential of 36% is found between workers in Barcelona and Lugo. City size

---

[11]Employers assign workers into one of ten social security occupation categories which we have regrouped into seven skill categories. For instance, top managers are assigned to social security category 1 equivalent to our 'very-high-skilled occupation' category.

Table 1: Estimation of the static city-size earnings premium

Dependent variable:	(1) Log earnings	(2) City indicator coefficients column (1)	(3) Log earnings	(4) City indicator coefficients column (3)
Log city size		0.048 (0.008)***		0.025 (0.006)***
City indicators	Yes		Yes	
Worker fixed-effects	No		Yes	
Experience	0.033 (0.001)***		0.107 (0.002)***	
Experience^2	-0.001 (0.000)***		-0.001 (0.000)***	
Firm tenure	0.014 (0.001)***		0.003 (0.000)***	
Firm tenure^2	-0.001 (0.000)***		-0.000 (0.000)***	
Secondary education	0.100 (0.002)***			
University education	0.185 (0.004)***			
Very-high-skilled occupation	0.790 (0.006)***		0.256 (0.006)***	
High-skilled occupation	0.520 (0.005)***		0.195 (0.004)***	
Medium-high-skilled occupation	0.375 (0.006)***		0.127 (0.005)***	
Medium-skilled occupation	0.227 (0.004)***		0.093 (0.003)***	
Medium-low-skilled occupation	0.120 (0.005)***		0.059 (0.005)***	
Low-skilled occupation	0.064 (0.002)***		0.021 (0.002)***	
Observations	5,821,846	73	5,821,846	73
R^2	0.489	0.256	0.118	0.164

Notes: All specifications include a constant term. Columns (1) and (3) include month-year indicators, two-digit sector indicators, and contract-type indicators. Coefficients are reported with robust standard errors in parenthesis, which are clustered by worker in columns (1) and (3). ***, **, and * indicate significance at the 1, 5, and 10 percent levels. The R^2 reported in column (3) is within workers. Worker values of experience and tenure are calculated on the basis of actual days worked and expressed in years.

is a powerful predictor of differences in earnings as it can explain a quarter of the variation that is left after controlling for observable worker characteristics (R^2 of 0.256 in column 2). This pooled OLS estimate of the elasticity of earnings with respect to city size reflects that doubling city size is associated with an approximate increase of 5% in earnings.

We have carried out alternative estimations for this pooled OLS two-stage estimation. First, we have included interactions of city and year indicators in the first-stage to address the possibility of such city effects being time-variant. Then, in the second stage we regress all estimated city-year indicators on time-varying log city size and year indicators. The estimated elasticity remains unaltered at 0.048. Second, we have also estimated the elasticity in a one-stage process by including log city size directly in the Mincerian specification of log earnings. In this case, the estimated elasticity rises slightly to 0.053.[12]

Following our discussion in section 2, the pooled OLS estimate of the elasticity of interest conceals two potential biases: unobserved worker heterogeneity and the omission of more valuable experience accumulated in big cities. In column (3) of table 1 we estimate equation (6) by introducing worker fixed-effects in the first stage of the estimation. This strategy takes care of the first concern i.e., more productive workers (or those with higher unobserved time-invariant ability) sorting into bigger cities. The difference with the Mincerian specification of log earnings in column (1) is that now we estimate city indicators on the basis of migrants. All other coefficients are estimated by exploiting time variation and job changes within workers' lives. In column (4) the estimated elasticity of earnings with respect to city size drops substantially to 0.026.[13]

The pooled OLS estimate of the elasticity of interest, 0.048, is in line with previous estimates that use worker-level data with similar sample restrictions. Combes *et al.* (2010) find an elasticity of 0.051 for France while Glaeser and Resseger (2010) obtain an elasticity of 0.041 for the US.[14] When worker fixed-effects are introduced Combes *et al.* (2010) obtain a drop in the elasticity of 35% to 0.033, while Mion and Naticchioni (2009) report a larger drop of 66% for Italy. Our estimated drop of 46% lies in between both.

5. Dynamic benefits of bigger cities

We now turn to a joint estimation of the static and dynamic advantages of bigger cities while allowing for unobserved worker heterogeneity and sorting. This involves our full specification of equation (1). For this, we need to keep track of the experience a worker has accumulated in one city or group of cities of similar size. In column (1) of table 2 we add to the first-stage specification the experience (calculated in days and then expressed in years) accumulated in the two biggest cities — Madrid and Barcelona — and the square of this to allow for concavity in the effect. We also add experience accumulated in the next three biggest cities — Valencia, Sevilla

---

[12]When we include the share of sector employment in the city in the first stage to allow for localization economies, the estimated elasticity rises marginally to 0.050.

[13]Alternative estimations result in similar magnitudes of this elasticity ranging between 0.025 and 0.027.

[14]Combes *et al.* (2010) aggregate individual data into a city-sector level data to estimate an elasticity analogous to our pooled OLS result. Mion and Naticchioni (2009) find the lowest estimate of this elasticity for Italy (0.022).

Table 2: Estimation of the dynamic and static city-size earnings premia

Dependent variable:	(1) Log earnings	(2) Initial premium (city indicator coefficients column (1))	(3) Medium-term premium (initial + 7 years local experience)
Log city size		0.025 (0.006)***	0.049 (0.011)***
City indicators	Yes		
Worker fixed-effects	Yes		
Experience 1st-2nd biggest cities	0.027 (0.001)***		
(Experience 1st-2nd biggest cities)^2	-0.001 (0.000)***		
Experience 1st-2nd biggest cities $\times$ now in smaller	0.002 (0.001)*		
Experience 3rd-5th biggest cities	0.011 (0.001)***		
(Experience 3rd-5th biggest cities)^2	-0.000 (0.000)***		
Experience 3rd-5th biggest cities $\times$ now in bigger	0.001 (0.003)		
Experience 3rd-5th biggest cities $\times$ now in smaller	-0.002 (0.002)		
Experience	0.094 (0.002)***		
Experience^2	-0.001 (0.000)***		
Firm tenure	0.002 (0.000)***		
Firm tenure^2	-0.000 (0.000)***		
Very high skilled occupation	0.251 (0.006)***		
High skilled occupation	0.193 (0.004)***		
Medium-high skilled occupation	0.128 (0.005)***		
Medium skilled occupation	0.094 (0.003)***		
Medium-low skilled occupation	0.060 (0.005)***		
Low skilled occupation	0.022 (0.002)***		
Observations	5,821,846	73	73
R^2	0.120	0.165	0.366

Notes: All regressions include a constant term. Column (1) includes month-year indicators, two-digit sector indicators, and contract-type indicators. Coefficients are reported with robust standard errors in parenthesis, which are clustered by worker in column (1). ***, **, and * indicate significance at the 1, 5, and 10 percent levels. The R^2 reported in column (1) is within workers. Worker values of experience and tenure are calculated on the basis of actual days worked and expressed in years. City medium-term premium calculated for workers' average experience in one city (7.24 years).

and Zaragoza — and the square of this. We still take care of unobserved time-invariant worker heterogeneity by using worker fixed-effects, just as in column (3) of table 1. Our results indicate that experience accumulated in bigger cities is more valuable than overall experience accumulated elsewhere. For instance, the first year of experience in Madrid or Barcelona raises earnings by 2.7% relative to having worked that same year in a city below the top-five. The first year of experience in a city ranked 3$^{\text{rd}}$ to 5$^{\text{th}}$ raises earnings by 1.1% relative to having worked that same year in a city below the top-five. We have also tried finer groupings of cities by size (not reported), but found no significant differences in the value of experience within the reported groupings (e.g., between Madrid and Barcelona).

In our earnings specification we also allow for the value of experience accumulated in big cities to vary depending on where it is used. For this purpose, we include an interaction of years of experience accumulated in the top-two cities and an indicator for being currently working in a smaller city. Similarly, we include interactions of years of experience accumulated in cities ranked 3$^{\text{rd}}$ to 5$^{\text{th}}$ and indicators for currently working in either bigger or smaller cities. We find all these interactions to be either non-significant or of small quantitative importance which suggests that the experience acquired in big cities is highly portable.[15] Glaeser and Resseger (2010) show that workers who reside in US metropolitan areas get a larger wage increase from the same level of potential overall experience than workers in rural areas. However, they find that the effect does not vary across metropolitan areas of different sizes. Our results help understand why this is the case: what matters across metropolitan areas is where experience is acquired and not where it is used. Experience accumulated in bigger cities is more valuable and remains so even when workers move away to cities of different sizes.

Earnings profiles

An illustrative way to present our results is to plot the evolution of earnings for workers in different cities, calculated on the basis of the coefficients estimated in column (1) of table 2. In panel (a) of figure 3, the higher solid line depicts the earnings profile over ten years of an individual working in Madrid during this entire period relative to the earnings of a worker with identical characteristics (both observable and time-invariant unobservable) who instead works in Santiago de Compostela (the median-sized city in our sample). To be clear, the red line does not represent how fast earnings rise while working in Madrid, they represent how much faster they rise when working in Madrid than when working in Santiago. For the worker in Madrid, the profile of relative earnings has an intercept and a slope component. First, we calculate the intercept as the difference in estimated city indicators between Madrid and Santiago. Next, we compute the slope by evaluating the experience accumulated in Madrid and its square at different years and subtracting from this, the value of

---

[15]It is worth noting that city indicators are still estimated on the basis of migrants. Yet, to estimate the value of experience acquired in cities of different sizes we exploit temporal variation in earnings and days of experience within workers' lives for both stayers and migrants. To estimate the value of experience in bigger cities when used elsewhere we need sufficient workers who accumulate experience in different types of cities. This requirement is easily satisfied in the data given that we track workplace locations since 1981 or entry in social security, although our estimation period is 2004–2009.
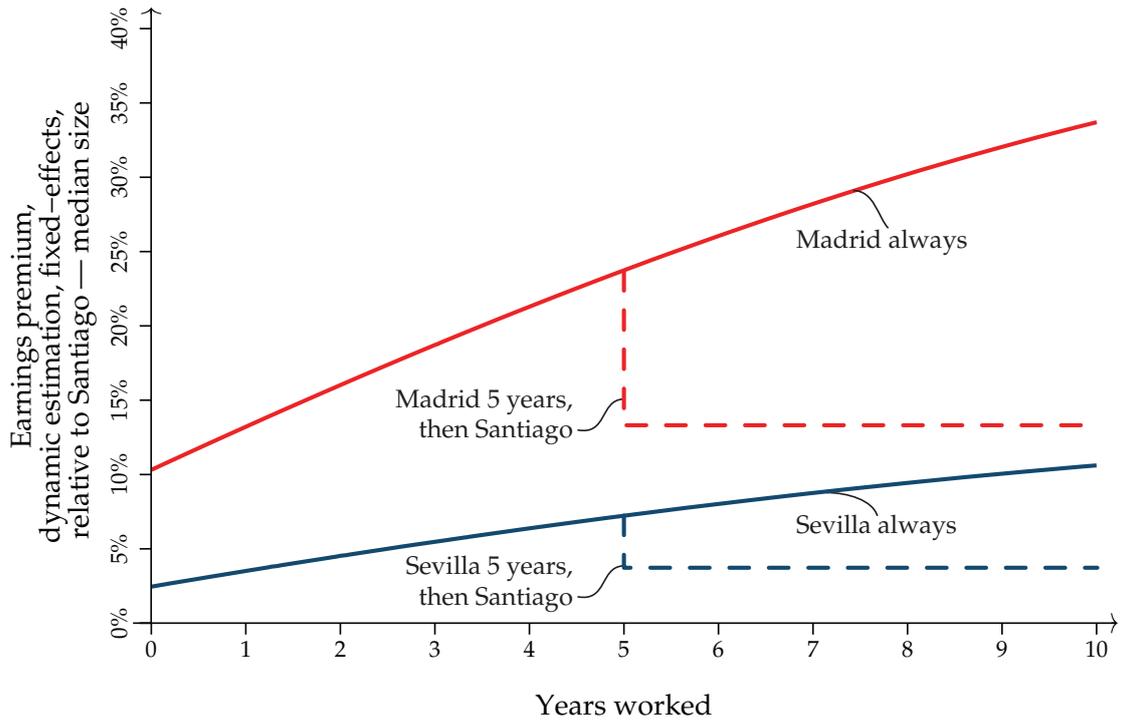
Panel (a) Allowing for learning benefits of bigger cities

Panel (b) Not allowing for learning benefits of bigger cities

Figure 3: Earnings profile relative to median-sized city

experience acquired in Santiago and its square. Initially, a worker in Madrid earns 10% more than a worker in Santiago, but this gap widens considerably, so that after ten years the difference in earnings reaches 34%. The lower solid line depicts the earnings profile over ten years of an individual working in Sevilla relative to the earnings of a worker in Santiago. There is still a substantial gap in the profile of relative earnings, although smaller in magnitude than in the case of Madrid: an initial earnings differential of 2% and of 11% after ten years.

The dashed lines in panel (a) of figure 3 illustrate the portability of the learning advantages of bigger cities. The top dashed line shows the estimated earnings profile for an individual who, after five years of working in Madrid, moves to Santiago. Up until year five, his relative earnings profile is the same as that of a worker who always works in Madrid. At that point, by moving, this worker loses the city fixed-effect associated with currently working in Madrid, which is replaced by the city fixed-effect associated with currently working in Santiago. He also has the value of his experience accumulated in Madrid re-valued according to the estimates of column (1) of table 2. Since the estimated change in the value of experience acquired in Madrid after moving is quantitatively small, the 10% drop in earnings is explained mostly by the difference in city fixed-effects. From then onwards, his relative earning profile appears flat in the plot (meaning earnings thereafter rise at the same pace as for a worker who has always been in Santiago) but above the horizontal axis, reflecting that this migrant earns 13% more than someone who has always been in Santiago, thanks to the more valuable experience accumulated in Madrid. Someone moving to Santiago after five years in Sevilla exhibits a similar qualitatively relative profile, although with smaller magnitudes.

The evolution of earnings portrayed in panel (a) of figure 3 shows that much of the earnings premia that big cities offer are not instantaneous, but instead accumulate over time and are highly portable. This perspective contrasts with the usual static view that earlier estimations of these premia have adopted. This static view is summarized in panel (b) of figure 3. Once again we depict the profile of relative earnings for a worker in Madrid or Sevilla relative to a worker in Santiago, but now on the basis of column (3) of table 1 instead of column (1) of table 2. In this view, implicit in the standard fixed-effects estimation without city-specific experience, relative earnings for a worker in Madrid exhibit only a constant difference with respect to Santiago: a static premium of 10% gained immediately when starting to work in Madrid and lost immediately upon departure. Our findings reveal that the premium of working in big cities has a sizable dynamic component and that workers do not lose this when moving to smaller cities. This latter result strongly suggests that a learning mechanism is indeed behind the accumulation of the premium.

Short-term and medium-term city-size earnings premia

After having addressed the two sources of bias we have emphasized in the first stage of the estimation, we can proceed to estimate the elasticity of the static earnings premium with respect to city size in the second stage. In column (2) of table 2 we regress the city indicators estimated in column (1) on log city size and obtain an elasticity of 0.026. This magnitude is essentially identical to the static fixed-effects estimate in column (4) of table 1. In section 2 we showed that the bias in the static fixed-effects estimate would tend to be small if the direction of migration flows is balanced and the learning benefits of bigger cities are portable. Migration flows between cities

of different size are indeed balanced in our data, as already noted above. And the estimates of our dynamic specification show that experience accumulated in bigger cities remains roughly just as valuable if workers relocate. This is good news, because it implies that existing fixed-effects estimates of the static gains from bigger cities are accurate and robust to the existence of important dynamic effects.

Studying the static earnings premium from currently working in bigger cities alone, however, ignores that there are also important dynamic gains. To study a longer horizon, we can estimate a medium-term premium that incorporates both static and dynamic components. For this purpose, we add to the fixed-effects for each city the estimated value of experience accumulated in that same city evaluated at the average experience in a single location of workers in our sample (7.24 years). The estimated elasticity of this medium-term premium with respect to city size, in column (3) of table 2, is 0.049.

Comparison of the 0.049 elasticity of the medium-term premium with respect to city size in column (3) of table 2 with the 0.026 elasticity of the short-term static premium in column (2) indicates that in the medium term, about half of the gains from working in bigger cities are static and about half are dynamic.

Note also that the 0.049 elasticity of the medium-term premium with respect to city size in column (3) of table 2 is almost identical to the static pooled OLS estimate in column (2) of table 1. This suggests that the drop in the estimated elasticity between a static pooled OLS estimation and a static fixed-effects estimation is not due to sorting being important but to dynamic effects being important. When estimating the medium-term elasticity in column (3) of table 2, we have brought dynamic effects back in, but left sorting on unobserved time-invariant ability out. The fact that this takes us back from the magnitude of the static fixed-effects to the magnitude of the pooled OLS estimate indicates that learning effects can fully account for the difference. This not only underscores the relevance of the dynamic benefits of bigger cities, it also suggests that sorting may not be very important. We return to this issue later in the paper.

While our estimate of the medium-term benefit of working in bigger cities resembles a basic pooled OLS estimate, our methodology allows us to separately quantify the static and the dynamic components and to discuss the portability of the dynamic part. Further, the estimation of the combined medium-term effect is more precise. Figure 4 plots the estimated medium-term premia against log city size. Compared with the plot for the pooled OLS specification in figure 2, log city size explains a larger share of variation in medium-term earnings across cities (R^2 of 0.366 vs. 0.256). In fact, we observe that many small and medium-sized cities now lie closer to the regression line. One reason why some cities are outliers in the pooled OLS estimation is that they have either relatively many or relatively few workers who have accumulated substantial experience in the biggest cities. Workers in cities far above the regression line in figure 2, such as Tarragona-Reus, Girona, Manresa or Puertollano, have accumulated at least 5% of their overall experience in the five biggest cities. Workers in cities far below the regression line in figure 2, such as Santa Cruz de Tenerife-La Laguna, Ourense, Elda-Petrer, Lugo or Gran Canaria Sur have accumulated less than 1.5% of their overall experience in the five biggest cities.

Figure 4: Dynamic fixed-effects estimation of the medium-term city-size premium

Addressing the endogeneity of city sizes

While we have addressed potential sources of bias in the first-stage estimation of column (1) in table 2, an important potential source of bias remains in the second-stage estimation of columns (2) and (3). The association between earnings premia and size is subject to endogeneity concerns. More precisely, an omitted variable bias could arise if some city characteristic simultaneously boosts earnings and attracts workers to the city, thus increasing its size. We may also face a reverse causality problem if higher earnings similarly lead to an increase in city size.

The extant literature has already addressed this endogeneity concern and found it to be of small practical importance (Ciccone and Hall, 1996, Combes *et al.*, 2010). Relative city sizes are very stable over time (Eaton and Eckstein, 1997, Black and Henderson, 2003). If certain cities are large for some historical reason that is unrelated with the current earnings premium (other than through size itself), we need not be too concerned about the endogeneity of city sizes. Thus, following (Ciccone and Hall, 1996) we instrument current city size using historical city size data. In particular, our population instrument counts the number of people within 10 kilometres of the average resident in a city back in 1900.[16] Following Combes *et al.* (2010), we also use land fertility data, with the idea that this was an important driver of relative city sizes back when the country was mostly agricultural, and these relative size differences may have persisted, but land fertility is

---

[16]We obtain historical population data from Goerlich, Mas, Azagra, and Chorén (2006) who construct decennial municipality population series using all available censuses from 1900 to 2001, keeping constant the areas of municipalities in 2001. We replicate our strategy to construct current urban area size, but use instead 1900 municipal population; however, since we lack the equivalent of LandScan information at that time, we distribute population uniformly within the municipality.

not directly important for production today. In particular, we use as an instrument the percentage of land within 25 kilometers of the city centre that has high potential quality. Potential land quality refers to the inherent physical quality of the land resources for agriculture, biomass production and vegetation growth, prior to any modern intervention such as irrigation.[17] In addition to these instruments used in previous studies, we incorporate two additional instruments suggested by the work of Saiz (2010). A city's ability to grow is limited by the availability of land suitable for construction. Saiz studies the geographical determinants of land supply in the United States and shows that land supply is greatly affected by how much land around a city is covered by water or has slopes greater than 15%. Thus, we also use as instruments the percentage of land within 25 kilometers of the city centre that is covered by oceans, rivers or lakes and the percentage that has slopes greater than 15%.[18] The final instrument we include is motivated by the work of Goerlich and Mas (2009). They document how small municipalities with high elevation, of which there are many in Spain, lost population to nearby urban areas over the course of the 20th century. An urban area's current size, for a given size in 1900, could thus be affected by having high-elevation areas nearby. The instrument we use to incorporate this is the log mean elevation within 25 kilometers of the city centre.

Table 3 gives the first and second stages of our instrumental variable estimation. The first-stage results in column (1) show that the instruments are jointly significant and strong. The F-statistic (or robust Kleinberger-Papp rk Wald statistic) for weak identification exceeds all thresholds proposed by Stock and Yogo (2005) for the maximal relative bias and maximal size. The LM test confirms our instruments are relevant as we reject the null that the model is underidentified. We can also rule out potential endogeneity of the instruments: the Hansen-J test cannot reject the null of the instruments being uncorrelated with the error. Lastly, according to the endogeneity test, the data does not reject the use of OLS.[19]

Column (2) of table 3 shows that the elasticity of the initial premium with respect to city size is not substantially affected by instrumenting (it is 0.023, compared with 0.025 in table 2). Similarly, column (3) shows that the elasticity of the medium-term premium with respect to city size is also almost unchanged by instrumenting (it is 0.048, compared with 0.049 in table 2). In fact, a Hausman test fails to reject that instrumental variables are not required to estimate these elasticities. This is

---

[17]The source of the land quality data is the CORINE Project (Coordination of Information on the Environment), initiated by the European Commission in 1985 and later incorporated by the European Environment Agency into its work programme (European Environment Agency, 1990). We calculate the percentage of land within 25 kilometers of the city centre with high potential quality using Geographic Information Systems (GIS). The city centre is defined as the centroid of the main municipality of the urban area (the municipality that gives the urban area its name or the most populated municipality when the urban area does not take its name from a municipality).

[18]Geographic information on the location of water bodies in and around urban areas is computed using GIS and the digital map of Spain's hydrography included with Goerlich *et al.* (2006). Slope is calculated on the basis of elevation data from the Shuttle Radar Topographic Mission(Jarvis, Reuter, Nelson, and Guevara, 2008), which records elevation for points on a grid 3 arc-seconds apart (approximately 90 metres).

[19]The instruments are also individually significant, with the only exception of log mean elevation around the city (which, given the motivation, only makes sense as an instrument after controlling for historical city size) and the percentage of water. Regarding water, note that in addition to the negative effect on land supply, it has a positive effect on land demand through its amenity value. Its overall effect on city size is, thus, ambiguous. The first-stage of the instrumental variable estimation shows a small net positive effect of water bodies around a city.

Table 3: IV estimation of the dynamic city size earnings premium

Dependent variable:	(1) Log size	(2) Initial premium	(3) Medium-term premium
Instrumented log city size		0.023 (0.008)***	0.048 (0.014)***
Log city size 1900	0.702 (0.074)***		
% high-quality land within 25km of city centre	0.016 (0.006)**		
% water within 25km of city centre	0.006 (0.002)**		
% steep terrain within 25km of city centre	-0.014 (0.006)**		
Log mean elevation within 25km of city centre	0.292 (0.086)***		
Observations	73	73	73
R^2	0.687	0.164	0.366
F-test weak ident. (H_0: instruments jointly insignificant)		35.698	35.698
P-value LM test (H_0: model underidentified)		0.008	0.008
P-value J test (H_0: instruments uncorr. with error term)		0.246	0.139
P-value endog. test (H_0: exogeneity of instrumented var.)		0.522	0.680

Notes: All regressions include a constant term. Column (1) is the first-stage regression of log city size on a set of historical population and geographical instruments. Columns (2) and (3) are second-stage regressions of city premia on instrumented log city size. Coefficients are reported with robust standard errors in parenthesis. ***, **, and * indicate significance at the 1, 5, and 10 percent levels. The F-statistic (or robust Kleinberger-Papp rk Wald statistic) reported on the weak instruments identification test exceeds all thresholds proposed by Stock and Yogo (2005) for the maximal relative bias and maximal size.

in line with the consensus among urban economists that the endogeneity of city sizes ends up not being an important issue when estimating the benefits of bigger cities (Combes *et al.*, 2010).

6. The interaction between ability and the learning benefits of bigger cities

Following Baker (1997), a large literature emphasizes that there is substantial heterogeneity in earnings profiles across workers. We have shown that an important part of the advantages associated with bigger cities is that they provide steeper earnings profiles. Given that both higher individual ability and experience acquired in bigger cities can increase earnings faster, we now explore whether there are complementarities between them, i.e. whether more able workers enjoy greater learning advantages from bigger cities.

A simple approach is is to classify workers into different ability groups based on observables, for instance their educational attainment or occupational skills. When we try this, the estimation results (not reported) show no significant differences in the value of experience acquired in different cities across worker types defined by observable indicators of ability. This leads us to use a broader definition of ability that includes both observables and unobservables, as captured by worker fixed-effects.

Table 4: Estimation of the heterogeneous dynamic and static city-size earnings premia

Dependent variable:	(1) Log earnings net of worker fixed-effect	(2) Initial premium (city indicator coefficients column (1))	(3) Medium-term premium (initial + 7 years local experience)
Log city size		0.025 (0.006)***	0.046 (0.010)***
City indicators	Yes		
Experience 1st-2nd biggest cities	0.024 (0.001)***		
(Experience 1st-2nd biggest cities)^2	-0.001 (0.000)***		
Exp. 1st-2nd biggest $\times$ worker fixed-effect	0.020 (0.002)***		
(Exp. 1st-2nd biggest)$^2\times$ worker fixed-effect	-0.000 (0.000)***		
Experience 3rd-5th biggest cities	0.010 (0.001)***		
(Experience 3rd-5th biggest cities)^2	-0.000 (0.000)***		
Exp. 3rd-5th biggest $\times$ worker fixed-effect	0.014 (0.002)***		
(Exp. 3rd-5th biggest)$^2\times$ worker fixed-effect	-0.001 (0.000)***		
Experience	0.100 (0.001)***		
Experience^2	-0.001 (0.000)***		
Experience $\times$ worker fixed-effect	0.059 (0.002)***		
(Experience)$^2\times$ worker fixed-effect	-0.002 (0.000)***		
Firm tenure	0.002 (0.000)***		
Firm tenure^2	-0.000 (0.000)***		
Occupation indicators	Yes		
Observations	5,821,846	73	73
R^2	0.126	0.156	0.334

Notes: All regressions include a constant term. Column (1) also includes month-year indicators, two-digit sector indicators, and contract-type indicators. Coefficients in column (1) are reported with bootstrapped standard errors in parenthesis which are clustered by worker (achieving convergence of coefficients and mean squared error of the estimation in each of the 100 bootstrap iterations). Coefficients in columns (2) and (3) are reported with robust standard errors in parenthesis. ***, **, and * indicate significance at the 1, 5, and 10 percent levels. The R^2 reported in column (1) is within workers. Worker values of experience and tenure are calculated on the basis of actual days worked and expressed in years. City medium-term premium calculated for workers' average experience in one city (7.24 years).

To incorporate our interaction between ability and the learning benefits of bigger cities into our framework, suppose the log wage of worker i in city c at time t, w_{ict}, is given by

$$w_{ict} = \sigma_c + \mu_i + \sum_{j=1}^{C}(\delta_j + \phi_j\mu_i)e_{ijt} + x'_{it}\beta + \varepsilon_{ict} . \tag{12}$$

In this specification we allow the value of experience accumulated in city c to differ for individuals with different levels of unobserved ability μ_i. We can estimate equation (12) recursively. Given a set of fixed-effects (for instance, those coming from estimating equation (11), which corresponds to $\phi_j = 0$), we can estimate equation (12) by ordinary least squares, then obtain a new set of estimates of worker fixed-effects as

$$\hat{\mu}_i = \frac{w_{ict} - \hat{\sigma}_c - \sum_{j=1}^{C}\hat{\delta}_j e_{ijt} - x'_{it}\hat{\beta}}{1 + \sum_{j=1}^{C}\hat{\phi}_j e_{ijt}} , \tag{13}$$

then, given these new worker fixed-effects estimate again equation (12), and so on until convergence is achieved.[20]

Table 4 shows the results of our iterative estimation. Relative to column (1) of table 2 we have added interactions between experience and ability (estimated worker fixed-effects). The interactions are statistically significant and large in magnitude. To get a better sense of the differences implied by the coefficients of table 4, figure 5 replicates the earnings profiles of figure 3 for workers of different ability based on this table. We consider two different workers, a high-ability one (in the 75th percentile of the estimated overall worker fixed-effects distribution) and a low-ability one (in the 25th percentile of the same distribution). The two solid lines depict the earnings profiles over ten years for a high-ability worker in Madrid and in Sevilla, relative to the earnings of an individual with identical observable characteristics and the same level of ability who is working in the median-sized city, Santiago de Compostela. After ten years, the earnings gap between working in Madrid and Santiago is 37% for the high-ability worker and 27% for the low ability worker. The differences between working in Sevilla and Santiago are smaller but still sizable: 13% for the high-ability worker and 8% for the low-ability worker.

These results reveal that there is a large role for heterogeneity in the dynamic benefits of city size. Experience is more valuable when it is acquired in big cities, and that it is workers with higher idiosyncratic ability who experience the greatest gains.

7. Sorting

Our estimations simultaneously consider static advantages associated with workers' current location, learning by working in big cities and spatial sorting. However, we have so far left sorting mostly in the background. We now bring sorting to the fore, by comparing the distribution of worker ability across cities.

Several other papers compare workers' ability and skills across cities. Some focus on education (e.g., Glaeser and Resseger, 2010) while others look at broader measures of skills (e.g., Bacolod *et al.*, 2009). We study worker-fixed effects because we are interested in distinguishing whether

---

[20]The equations in the text are again linear in experience to simplify the exposition and for consistency with our earlier methodological discussion. In our empirical estimations we include terms in both experience and its square.
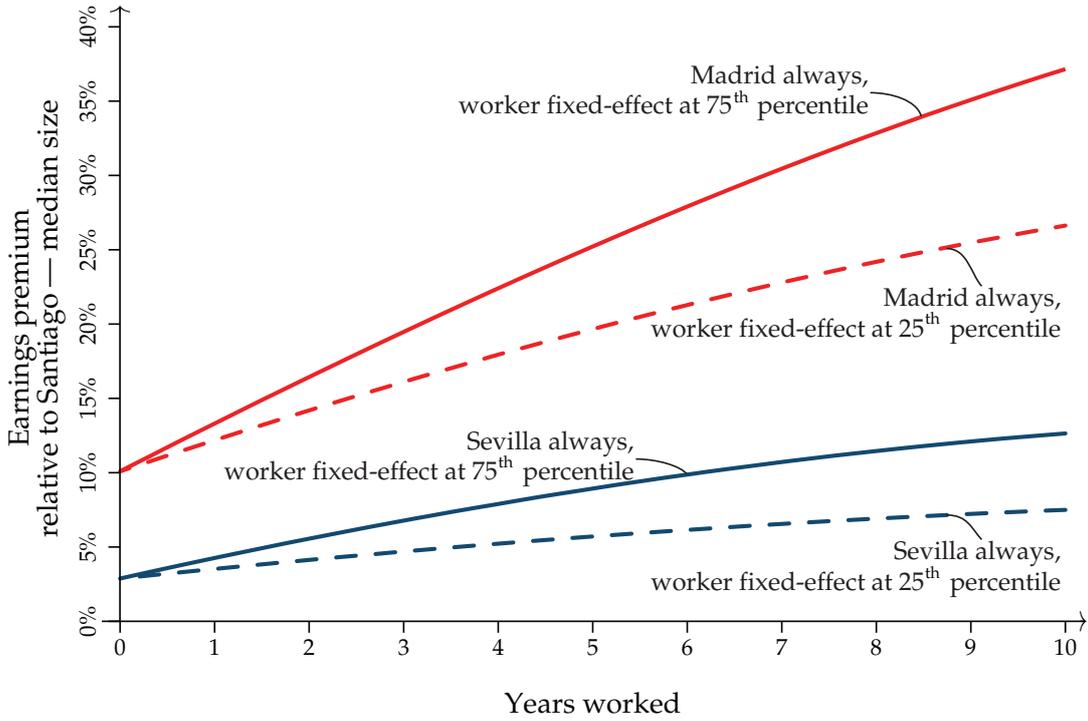
Figure 5: Earnings profile relative to median-sized city, high- and low-ability worker

workers who are idiosyncratically more able choose to locate in bigger cities or whether it is working in bigger cities that makes workers more skilled. Worker fixed-effects allow us to capture time-invariant ability. However, for this to work it is important to estimate worker fixed-effects on the basis of our full specification.

We have seen that a static fixed-effects estimation such as that of column (3) in table 1 gives roughly correct estimates of city fixed-effects. Nevertheless, it yields biased estimates of worker fixed-effects that incorporate not only time-invariant unobserved worker characteristics that affect earnings, but also the time-varying effect of experience in bigger cities as well as interactions between this experience and the time-invariant component of skills. In particular, estimation of μ on the basis of equation (6) if wages are determined as in equation (12) results in a biased estimate of μ:

$$\text{plim } \hat{\mu}_{i\,\text{FE}} = \mu_i(1 + \phi_j \bar{e}_{ijt}) + \sum_{j=1}^{C} \delta_j \bar{e}_{ijt} \,. \tag{14}$$

If we do not take this into account, it could appear from the estimated fixed effects that workers in bigger cities have higher ability on average even if the distribution of μ were identical. Estimation based on equation (12) yields instead plim $\hat{\mu}_i = \mu_i$.

Panel (a) in figure 6 plots the distribution of worker fixed-effects in the five largest cities (solid line) and in cities below the top five (dashed line) based on our full specification with heterogeneous dynamic and static benefits of bigger cities (Table 4, column 1).[21] We can see that both distributions look very similar (we do a more formal comparison below). This suggests that there

---

[21]Each individual is assigned to the city where he was working in May 2007.
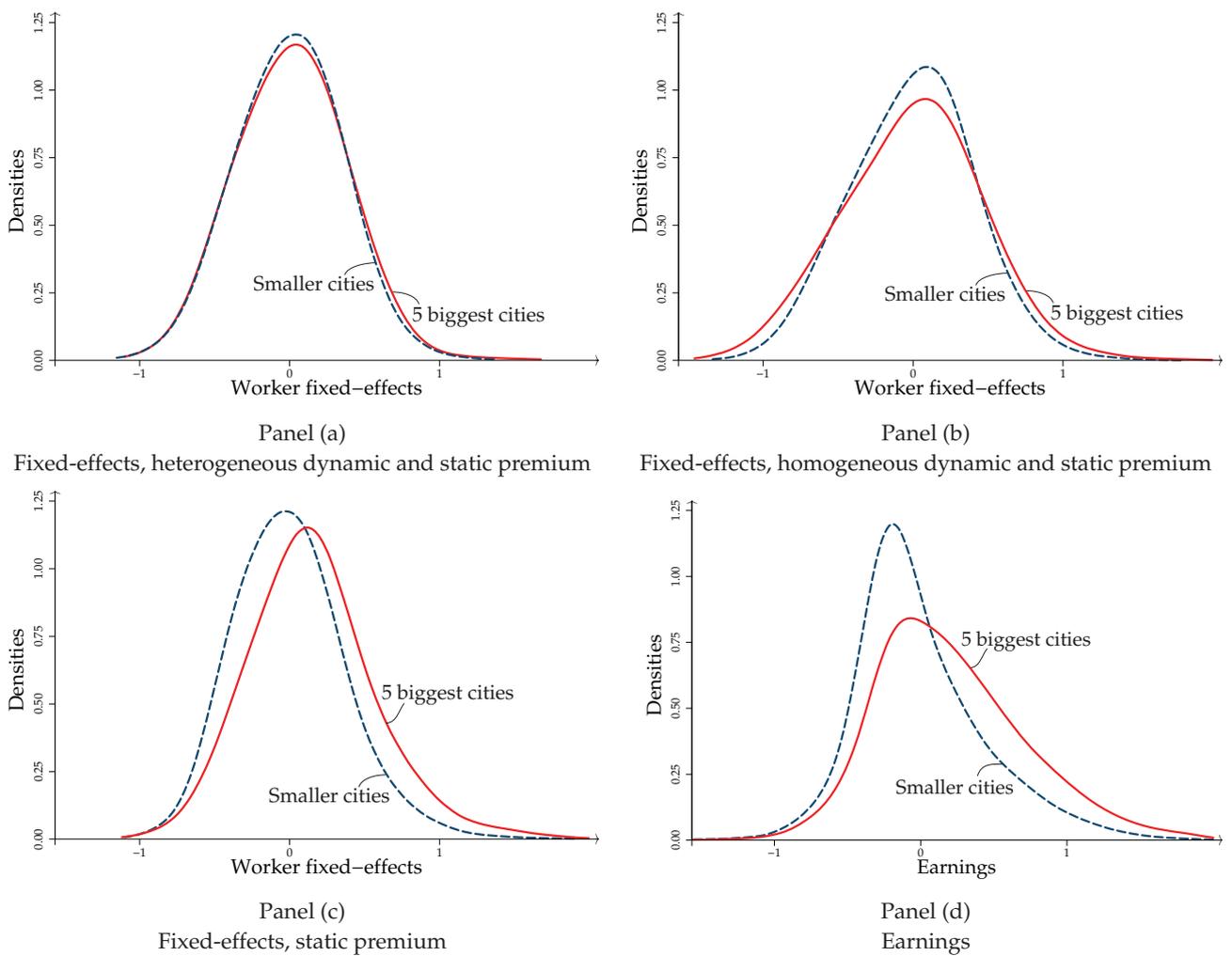
Figure 6: Comparisons of worker fixed-effects distributions across cities

is little sorting: the distribution of workers' innate ability (as measured by their fixed-effects) is very similar across cities of different sizes.

Panel (b) repeats the plot, but now constrains the dynamic benefits of bigger cities to be homogenous across workers (worker fixed-effects in this panel come from table 2, column 1). While both distributions have almost the same mean, the distribution in bigger cities exhibitsa higher variance. This is the result of forcing experience acquired in bigger cities to be equally valuable for everyone, so the ability of workers at the top of the distribution appears larger than it is (this estimation mixes the extra value that big-city experience has for them with their innate ability) while the ability of workers at the bottom of the distribution appears smaller than it is. Hence, by ignoring the heterogeneity of the dynamic benefits of bigger cities we can get the erroneous impression that there is greater dispersion of innate ability in bigger cities.

Panel (c) leaves out any dynamic benefits of bigger cities and plots worker fixed-effects from a purely static specification. This corresponds to the same comparison of fixed-effects carried out by Combes *et al.* (2012). They find a higher mean and greater dispersion of worker fixed-effects in bigger cities for France, which is also what this panel shows for Spain. This higher mean and

Table 5: Comparison of earnings and worker fixed-effects distributions, 5 biggest vs. other cities

Worker fixed-effects estimation	Shift ($\hat{A}$)	Dilation ($\hat{D}$)	Mean square quantile diff.	R^2	Obs.
Worker fixed-effects, heterogeneous dynamic and static premium (Table 4, column (1))	0.011 (0.003) ***	1.040 (0.007) ***	6.6e-04	0.919	84,662
Worker fixed-effects, homogenous dynamic and static premium (Table 2, column (1))	-0.004 (0.006)	1.147 (0.008) ***	7.0e-03	0.994	84,662
Worker fixed-effects, static premium (Combes et al., 2012)	0.150 (0.006) ***	1.106 (0.005) ***	4.9e-02	0.981	84,662
Log earnings	0.216 (0.003) ***	1.211 (0.008) ***	.11	0.982	84,662

Notes: The table applies the methodology of Combes, Duranton, Gobillon, Puga, and Roux (2011) to approximate the distribution of worker fixed-effects in the five biggest cities, $F_B(\mu_i)$, by taking the distribution of worker fixed-effects in smaller cities, $F_S(\mu_i)$, shifting it by an amount A, and dilating it by a factor D. $\hat{A}$ and $\hat{D}$ are estimated to minimize the mean quantile difference between the actual big-city distribution $F_B(\mu_i)$ and the shifted and dilated small-city distribution $F_S((\mu_i - A)/D)$. $M(0,1)$ is the total mean quantile difference between $F_B(\mu_i)$ and $F_S(\mu_i)$. $R^2 = 1 - M(\hat{A}, \hat{D})/M(0,1)$ is the fraction of this difference that can be explained by shifting and dilating $F_S(\mu_i)$. Coefficients are reported with bootstrapped standard errors in parenthesis (re-estimating worker fixed-effects in each of the 100 bootstrap iterations). ***, **, and * indicate significance at the 1, 5, and 10 percent levels.

variance carries over with greater strength to the distribution of log earnings, plotted in panel (d). Combes *et al.* (2012) carefully acknowledge that their estimated fixed-effects capture 'average skills' over a worker's lifetime. In contrast, panel (a) separates initial innate ability from the cumulative effect of the experience acquired in different cities, showing that differences arise as a result of the greater value of experience acquired in bigger cities, which is amplified for more able workers.

Table 5 performs a more formal comparison of the plotted distributions, using the methodology developed by Combes *et al.* (2011) to approximate two distributions. In particular we approximate the distribution of worker fixed-effects in the five biggest cities, $F_B(\mu_i)$, by taking the distribution of worker fixed-effects in smaller cities, $F_S(\mu_i)$, shifting it by an amount A, and dilating it by a factor D. $\hat{A}$ and $\hat{D}$ are estimated to minimize the mean quantile difference between the actual big-city distribution $F_B(\mu_i)$ and the shifted and dilated small-city distribution $F_S((\mu_i - A)/D)$.[22]

The top row compares the distributions of worker fixed-effects from our full specification with heterogeneous dynamic and static benefits of bigger cities (Table 4, column 1). The second row forces these benefits to be homogenous across workers. The third row constrains the benefits of bigger cities to be purely static. The bottom row compares log earnings. The table confirms what was visually apparent from figure 6.

Starting from the bottom row, earnings are higher on average in bigger cities. The shift parameter is $\hat{A} = 0.216$, indicating that average earnings are 24% ($e^{0.216} - 1$) higher in the five biggest cities. Earnings are also more dispersed in bigger cities. The dilation parameter is $\hat{D} = 1.211$ indicating that the distribution of earnings in the five biggest cities is amplified by that factor relative to smaller cities.

---

[22]Combes *et al.* (2011) also allow for truncation of one distribution to approximate the other. We find no significant truncation when comparing our two distributions, and so in table 5 we restrict ourselves to shift and dilation.

Moving one row up, the distribution of worker fixed-effects from a static specification also exhibits a higher mean and greater dispersion in bigger cities. However, both the shift and the dilation parameter are smaller than for earnings, and the distributions are more similar (the mean squared quantile difference is one order of magnitude smaller, $4.9e-02$ instead of .11). This implies that observables, such as employment in different sectors, account for a significant fraction of the differences.

The next row up introduces dynamic effects. This brings the distributions even closer (the mean squared quantile difference is reduced by another order of magnitude). The estimated shift parameter is not statistically significantly different from zero, indicating both distributions are centered on the same mean. However, the distribution of worker fixed-effects is still more dispersed in the five biggest cities ($\hat{D} = 1.147$).

The top row corresponds to our full specification. Once we allow experience in bigger cities to be more valuable and workers with higher innate ability to take greater advantage of this, worker fixed-effects exhibit extremely similar distributions in big and small cities (the mean squared quantile difference is reduced by yet another order of magnitude). The shift and dilation parameters, while statistically significant, are very close to 0 and to 1, respectively.

Several recent studies (Eeckhout *et al.*, 2010, Combes *et al.*, 2012, Baum-Snow and Pavan, 2012*b*) emphasize that earnings are higher on average and also exhibit greater dispersion in bigger cities. Our results in this section indicate that this is not the result of sorting. In fact, differences in the distribution of innate ability are quite similar in big and small cities. Instead, workers in big cities attain higher earnings on average precisely thanks to working there, which provides them with static advantages and also allows them to accumulate more valuable experience. Because more able workers benefit the most and less able workers benefit the least from working in bigger cities, a similar distribution of underlying ability translates into greater dispersion of earnings in bigger cities. In sum, workers in big and small cities are not particularly different to start with, it is working in cities of different sizes that makes their earnings diverge.

8. Conclusions

We have examined three reasons why firms may be willing to pay more to workers in bigger cities. First, there may be some static advantages associated with bigger cities. Second, bigger cities may allow workers to accumulate more valuable experience. Third, workers who are inherently more productive may choose to locate in bigger cities. Using a large and rich panel data set for workers in Spain, we provide a quantitative assessment of the importance of each of these three mechanisms in generating earnings differentials across cities of different sizes.

We find that there are substantial static and dynamic advantages from working in bigger cities. The medium-term elasticity of earnings (after seven years) with respect to city size is close to 0.05. About one-half of these gains are static and tied to currently working in a bigger city. About another half accrue over time as workers accumulate more valuable experience in bigger cities. Furthermore, workers are able to take these dynamic gains with them when they relocate, which

we interpret as evidence that learning in bigger cities is important. Sorting of more able workers into bigger cities plays at best a minor role in explaining earnings differentials.

In the process of deriving our results, we also make some methodological progress. We confirm that estimations of the static city-size premium that use worker fixed-effects to address sorting but ignore the learning advantages of bigger cities provide an accurate estimate of the purely static gains. However, besides not capturing learning, they overestimate the importance of sorting because they mix innate ability with the extra value of big-city experience. Once we disentangle ability and the value of accumulated experience, cities of different sizes have quite similar distributions of worker ability.

Overall, we conclude that workers in bigger cities are not particularly different in terms of innate ability. It is working in cities of different sizes that makes their earnings diverge. The combination of static gains and learning advantages together with the fact that higher-ability workers benefit more from bigger cities explain why the distribution of earnings in bigger cities has higher mean and higher variance.

References

Bacolod, Marigee, Bernardo Blum, and William C. Strange. 2009. Skills and the city. *Journal of Urban Economics* 65(2):127–135.

Baker, Michael. 1997. Growth-rate heterogeneity and the covariance structure of life-cycle earnings. *Journal of Labor Economics* 15(2):338–375.

Baum-Snow, Nathaniel and Ronni Pavan. 2012*a*. Understanding the city size wage gap. *Review of Economic Studies* 79(1):88–127.

Baum-Snow, Nathaniel and Ronni Pavan. 2012*b*. Inequality and city size. Processed, Brown University.

Black, Duncan and Vernon Henderson. 2003. Urban evolution in the usa. *Journal of Economic Geography* 3(4):343–372.

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

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

Combes, Pierre-Philippe, Gilles Duranton, Laurent Gobillon, Diego Puga, and Sébastien Roux. 2011. The productivity advantages of large cities: Distinguishing agglomeration from firm selection. Processed, IMDEA Social Sciences.

Combes, Pierre-Philippe, Gilles Duranton, Laurent Gobillon, and Sébastien Roux. 2010. Estimating agglomeration effects with history, geology, and worker fixed-effects. In Edward L. Glaeser (ed.) *Agglomeration Economics*. Chicago, il: Chicago University Press, 15–65.

Combes, Pierre-Philippe, Gilles Duranton, Laurent Gobillon, and Sébastien Roux. 2012. Sorting and local wage and skill distributions in France. Processed, University of Toronto.

Duranton, Gilles and Diego Puga. 2001. Nursery cities: Urban diversity, process innovation, and the life cycle of products. *American Economic Review* 91(5):1454–1477.

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

Eaton, Jonathan and Zvi Eckstein. 1997. Cities and growth: Theory and evidence from France and Japan. *Regional Science and Urban Economics* 27(4–5):443–474.

Eeckhout, Jan, Roberto Pinheiro, and Kurt Schmidheiny. 2010. Spatial sorting: Why New York, Los Angeles and Detroit attract the greatest minds as well as the unskilled. Discussion Paper 8151, Centre for Economic Policy Research.

European Environment Agency. 1990. CORINE *Land Quality Project*. Copenhagen: European Environment Agency.

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

Glaeser, Edward L. 2011. *Triumph of the City: How Our Greatest Invention Makes Us Richer, Smarter, Greener, Healthier, and Happier*. London: MacMillan.

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

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

Goerlich, Francisco J. and Matilde Mas. 2009. Drivers of agglomeration: Geography vs. history. Processed, Instituto Valenciano de Investigaciones Económicas.

Goerlich, Francisco J., Matilde Mas, Joaquín Azagra, and Pilar Chorén. 2006. *La localización de la poblacion española sobre el territorio. Un siglo de cambios: un estudio basado en series homogéneas*. Bilbao, Spain: BBVA Foundation.

Gould, Eric D. 2007. Cities, workers, and wages: A structural analysis of the urban wage premium. *Review of Economic Studies* 74(2):477–506.

Holmes, Thomas J. 2010. Structural, experimentalist, and descriptive approaches to empirical work in regional economics. *Journal of Regional Science* 50(1):5–22.

Jarvis, Andrew, Hannes Isaak Reuter, Andrew Nelson, and Edward Guevara. 2008. *Hole-filled Seamless* SRTM *Data, Version 4.1*. Cali, Colombia: International Centre for Tropical Agriculture.

Mion, Giordano and Paolo Naticchioni. 2009. The spatial sorting and matching of skills and firms. *Canadian Journal of Economics* 42(1):28–55.

Oak Ridge National Laboratory. 2009. *LandScan Global Population Database 2008*. Oak Ridge, TN: Oak Ridge National Laboratory.

Rosenthal, Stuart S. and William Strange. 2004. Evidence on the nature and sources of agglomeration economies. In Vernon Henderson and Jacques-François Thisse (eds.) *Handbook of Regional and Urban Economics*, volume 4. Amsterdam: North-Holland, 2119–2171.

Saiz, Albert. 2010. The geographic determinants of housing supply. *Quarterly Journal of Economics* 125(3):1253–1296.

Stock, James H. and Motohiro Yogo. 2005. Testing for weak instruments in linear iv regression. In James H. Stock and Donald W. K. Andrews (eds.) *Identification and Inference for Econometric Models: Essays in Honor of Thomas J. Rothenberg*. Cambridge: Cambridge University Press, 109–120.