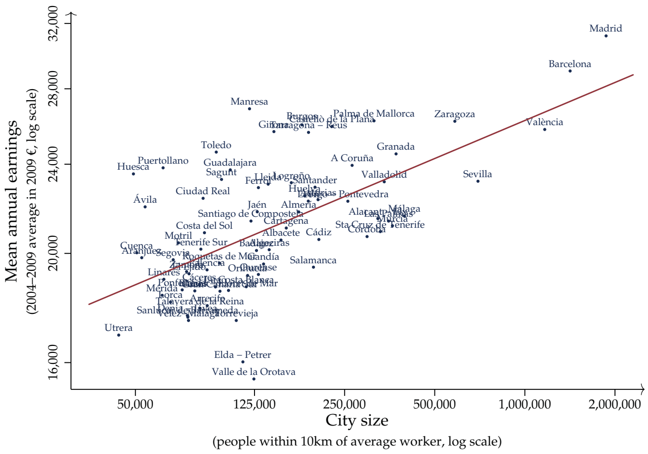
Learning by working in big cities

Jorge De la Roca
University of Southern California
Diego Puga
CEMFI

Published in *Review of Economics Studies* 84(1), January 2017: 106-142.
Replication code available from http://diegopuga.org/data/mcvl/

The earnings premium of bigger cities

- Workers in bigger cities earn significantly more
(Glaeser and Maré, 2001, Wheaton and Lewis, 2002, Gould, 2007, Combes, Duranton, and Gobillon, 2008, Combes, Duranton, Gobillon, and Roux, 2010, Glaeser and Resseger, 2010, Baum-Snow and Pavan, 2012).
- Differences remain large even when we compare workers with the same education and years of experience and in the same industry.



Spatial equilibrium

- From the point of view of workers, higher nominal earnings in bigger cities tend to be offset by differences in the cost of living (housing).
- However, in tradable sectors, if firms are willing to pay higher wages in bigger cities there must be productive advantages.
- The productive advantages that firms experience are confirmed by productivity estimations.

The earnings premium of bigger cities

- Three potential reasons why employers are willing to pay more in bigger cities to similar workers:
  - **Static advantages**, enjoyed while located in bigger cities and lost upon moving away.
  - **Learning advantages**, whereby bigger cities allow workers to acquire more valuable experience.
    Accumulate over time and, if embedded into human capital, beneficial even when a worker moves away.
  - **Sorting** into bigger cities by workers with higher initial ability.
- We use rich administrative (matched social security, tax and census) panel data for Spain to evaluate all three.

Data: Employment histories and earnings

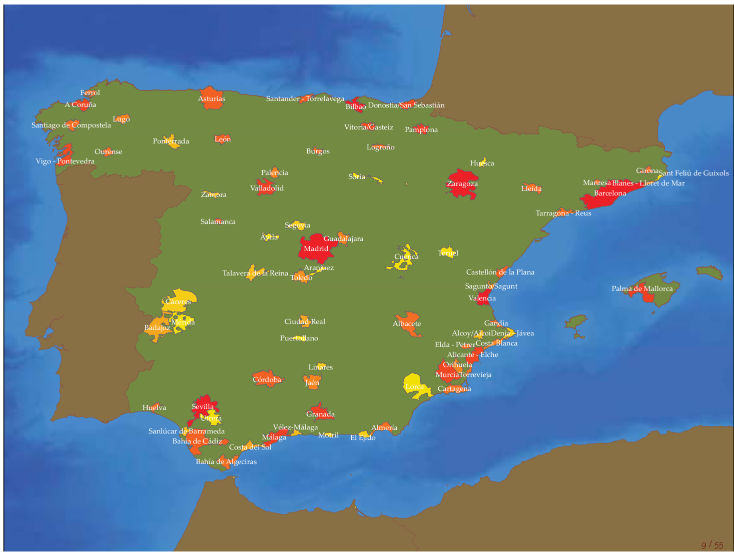
- Main data set is Spain's Continuous Sample of Employment Histories (*Muestra Continua de Vidas Laborales* or MCVL).
- Administrative data with longitudinal information obtained by matching social security, income tax, and census records.
- Tracks a 4% non-stratified random sample of the population 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 unit of observation 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).
- All changes since the date of first employment, or since 1980 for earlier entrants.
- Also personal characteristics from matched Census data (INE)
- Earnings by source from matched income tax returns from 2004 (Agencia Tributaria).

Data: Spatial detail

- Municipality of firm location if municipal population exceeds 40,000 (otherwise, province).
- Firms operating in multiple provinces must have a separate Contribution Account in each province of operation, with its corresponding municipal code.
- Since local labour markets are typically larger than a single municipality, will use urban areas/cities as spatial units.

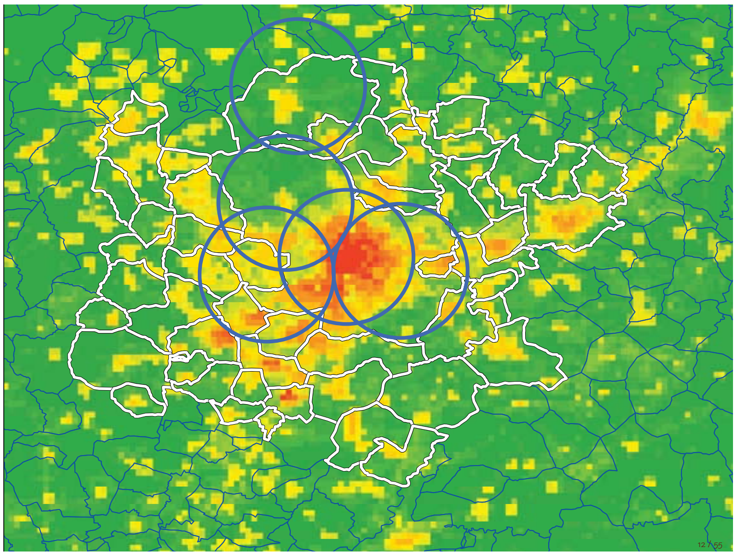
Data: Cities

- City definitions:
  - 2008 urban area/city definitions by Spain's Department of Housing (unchanged).
  - 85 cities comprising 747 municipalities and 30 million people (10% of Spain's surface and 68% of its population).
  - Range: 5,966,067 (Madrid) – 35,396 (Teruel).



Data: City size

- When measuring the scale of each city we wish to capture the potential for interactions between workers.
- We calculate the number of people within 10 kilometres of the average person.
  - Our measure of city size is highly correlated with a simple population count (0.94), but deals more naturally with unusual cities, in particular those that are polycentric.
  - This measure is also less prone to border problems than simple density measures that divide population by area.
- Based on the 1-kilometre-resolution population grid for Spain in 2006 created by Goerlich and Cantarino (2013).
  - They allocate population within each of 35,000 census tracts (áreas censales) covering Spain to 1×1 kilometre cells based on the location of buildings as recorded in high-resolution remote sensing data.
  - We take each 1×1 kilometre cell in the urban area, trace a circle of radius 10 kilometres 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.



The worker sample

- Starting sample are Spanish men aged 18 and over with Spanish citizenship born in Spain since 1962 and employed at any point between January 2004 and December 2009: 246,941 workers and 11,885,511 monthly observations (results for women also reported in the article).
- Data covering 76 out of 85 cities:
  - missing tax data for Basque Country and Navarre (4 cities),
  - location data not available for workers in municipalities with a population below 40,000 (3 cities and rural areas),
  - exclude Ceuta and Melilla (2 enclaves in continental Africa).
- Exclude job spells
  - in the public sector and in education and health services because their salaries are heavily regulated,
  - in agrarian, mining/extractive, fishing and household activites
- Final sample: 157,113 workers and 6,263,446 monthly observations.
- Around 7% of workers relocate across urban areas every year (useful for worker fixed-effects estimation).

A model of wage determination

- Suppose the log wage of worker i in city c at time t is

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

  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,
  - $\mathbf{x}_{it}$ is a vector of time-varying individual and job characteristics,
  - ε_{ict} is an error term.
- This allows for:
  - **Static benefits** of city size: earnings may be higher when working in big cities (σ_c increasing with the size of c).
  - **Learning benefits** of city size: experience may be more valuable when acquired in bigger cities (δ_{jc} increasing with the size of j), and even more valuable when also used in bigger cities (δ_{jc} increasing with the size of c).
  - **Sorting** on unobserved ability (μ_i larger for workers in bigger cities).

Static cross-section or pooled OLS estimation

- For comparison with earlier studies and to highlight the importance of the dynamic benefits of bigger cities, we begin with a simple regression considering only static benefits.
- Relative to the full specification,

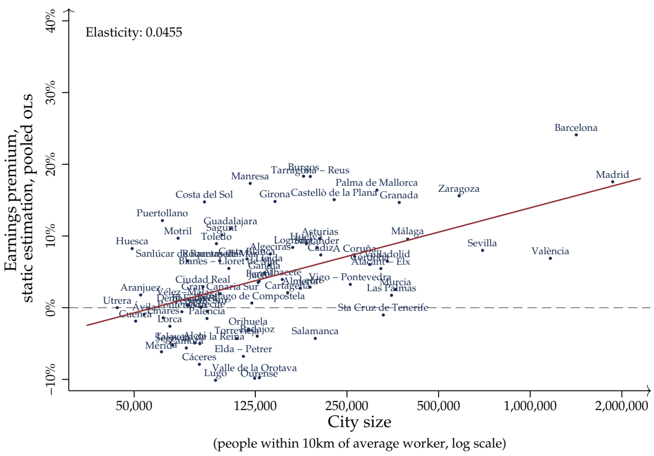
$$w_{ict} = \sigma_c + \mu_i + \sum_{j=1}^{C} \delta_{jc} e_{ijt} + \mathbf{x}'_{it}\beta + \varepsilon_{ict} \ ,$$

  we initially ignore unobserved worker heterogeneity (μ_i) and dynamic benefits of city size (δ_{jc}), taking the log wage to be given by

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

- Can then estimate σ_c with a pooled panel, a cross section, or even aggregate data.
- In a second stage, we regress the coefficients of the city indicators, σ_c, on log city size to estimate the elasticity of earnings with respect to city size.

	(1)	(2)	(3)	(4)
Dependent variable:	Log earnings	City indicator coefficients column (1)	Log earnings	City indicator coefficients column (3)
Log city size		0.0455 (0.0080)***		
City indicators	Yes			
Worker fixed-effects	No			
Experience	0.0319 (0.0005)***			
Experience^2	-0.0006 (0.0000)***			
Firm tenure	0.0147 (0.0006)***			
Firm tenure^2	-0.0005 (0.0000)***			
Very-high-skilled occupation	0.7752 (0.0062)***			
High-skilled occupation	0.4976 (0.0046)***			
Medium-high-skilled occupation	0.2261 (0.0031)***			
Medium-low-skilled occupation	0.0542 (0.0021)***			
University education	0.2014 (0.0037)***			
Secondary education	0.1084 (0.0022)***			
Observations	6,263,446	76		
R^2	0.4927	0.2406		



Biases in static cross-section or pooled OLS estimation

- Let $\imath_{ict}$ be an indicator variable for working in city c: $\imath_{ict} = 1$ if worker i is in city c at time t and 0 otherwise. The pooled OLS estimate of σ_c is unbiased if
$$\mathrm{Cov}(\imath_{ict}, \eta_{ict}) = 0 \,.$$
- However, if the more complex wage determination holds,
$$\eta_{ict} = \mu_i + \sum_{j=1}^{C} \delta_{jc} e_{ijt} + \varepsilon_{ict} \,,$$
and thus,
$$\mathrm{Cov}(\imath_i ct, \eta_{ict}) = \mathrm{Cov}(\imath_i ct, \mu_i) + \mathrm{Cov}(\imath_i ct, \sum_{j=1}^{C} \delta_{jc} e_{ijt}) \neq 0 \,.$$

Biases in static cross-section or pooled OLS estimation

- A static cross-section (or pooled) OLS estimation of σ_c suffers from **2 key potential sources of bias**:
  - **Sorting**: the urban premium for city c is overestimated if individuals with high unobserved ability μ_i are more likely to work there:
$$\mathrm{Cov}(\imath_{ict}, \mu_i) > 0$$
  - **Learning effects**: the urban premium for city c is overestimated if individuals with more valuable experience $\sum_{j=1}^{C}(\delta_{jc} - \delta) e_{ijt}$ are more likely to work there:
$$\mathrm{Cov}(\imath_{ict}, \sum_{j=1}^{C} \delta_{jc} e_{ijt}) > 0$$

An example of biases in static cross-section OLS estimation

- 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.
- Experience accumulated in the big city increases log wage by δ per period relative to having worked in the small city.
- n time periods and no migration.
- Pooled OLS estimate of static big city premium (σ) has
$$\operatorname{plim} \hat{\sigma}_{\text{pooled}} = \sigma + \mu + \frac{1+n}{2}\delta \,,$$
which overestimates the actual premium by the value of higher unobserved worker ability (μ) and the higher average value of accumulated experience in the big city ($\frac{1+n}{2}\delta$).

Static fixed-effects estimation (without dynamic premium)

- If we study workers at one isolated point in time, we cannot distinguish the advantages of a location from unobserved characteristics of workers (being creative, hard-working, etc.) that could results in a higher wage.
- A possible solution is to introduce worker fixed-effects.
- We then compare wages for the same worker across different cities as he moves over time (as Glaeser and Maré, 2001, Combes, Duranton, and Gobillon, 2008).
- Suppose we consider unobserved worker heterogeneity but still ignore a dynamic city size premium and take the log wage to be given by

$$w_{ict} = \sigma_c + \mu_i + \mathbf{x}_{it}'\beta + \zeta_{ict} \ ,$$

- To estimate σ_c we now need a panel. μ_i can be eliminated by subtracting from this expression the time average for each worker:

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

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

	(1)	(2)	(3)	(4)
Dependent variable:	Log earnings	City indicator coefficients column (1)	Log earnings	City indicator coefficients column (3)
Log city size		0.0455 (0.0080)***		0.0241 (0.0058)***
City indicators	Yes		Yes	
Worker fixed-effects	No		Yes	
Experience	0.0319 (0.0005)***		0.1072 (0.0018)***	
Experience^2	-0.0006 (0.0000)***		-0.0014 (0.0000)***	
Firm tenure	0.0147 (0.0006)***		0.0042 (0.0004)***	
Firm tenure^2	-0.0005 (0.0000)***		-0.0003 (0.0000)***	
Very-high-skilled occupation	0.7752 (0.0062)***		0.2350 (0.0057)***	
High-skilled occupation	0.4976 (0.0046)***		0.1758 (0.0040)***	
Medium-high-skilled occupation	0.2261 (0.0031)***		0.0873 (0.0029)***	
Medium-low-skilled occupation	0.0542 (0.0021)***		0.0152 (0.0019)***	
University education	0.2014 (0.0037)***			
Secondary education	0.1084 (0.0022)***			
Observations	6,263,446	76	6,263,446	76
R^2	0.4927	0.2406	0.1144	0.1422

Notes: Columns (1) and (3) also include month-year indicators, two-digit sector indicators, and contract-type indicators.

Comparison with other countries

- Our static pooled OLS estimate of the elasticity of the earnings premium with respect to city size (0.046) is in line with previous estimates:
  - Combes, Duranton, Gobillon, and Roux (2010) find an elasticity of 0.051 for France.
  - Glaeser and Resseger (2010) obtain an elasticity of 0.041 for the US
- When worker fixed-effects are introduced, the elasticity falls to 0.24, similar to previous estimates:
  - Combes, Duranton, Gobillon, and Roux (2010) see the elasticity drop to 0.033 (0.026 when instrumenting).
  - Mion and Naticchioni (2009) find a larger drop for Italy.
- Usual interpretation of the drop: evidence of strong sorting of the more able workers into the biggest cities.
- However, the drop can also be due in part to the fact that the bias from not considering dynamic effects is greatly mitigated when moving from the pooled OLS to the fixed-effects estimation.

Biases in static fixed-effects estimation

- The resulting fixed-effects estimate of σ_c is unbiased if

$$\mathrm{Cov}\left((\imath_{ict} - \bar{\imath}_{ic}), (\zeta_{ict} - \bar{\zeta}_i)\right) = 0 \ .$$

- However, if the more complex wage determination holds,

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

  and thus,

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

- Worker fixed effects take care of unobserved worker heterogeneity.
- Yet, the estimate of σ_c is **biased because dynamic effects are ignored**:
  - The urban premium for city c is overestimated if the value of a worker's experience is particularly high relative to the worker's average when the worker is in city c.
  - It is underestimated when the reverse is true.

An example of biases in static fixed-effects estimation

- Everything as in previous example, but now with some migrants.
- 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.
- Experience accumulated in the big city increases log wage by δ per period relative to having worked in the small city.
- n time periods.
- Assume that a fraction θ of the dynamic component of city premium (value of accumulated experience) is portable.

- Imagine first all migrants relocate from the small to the big city after m of the total n periods. The fixed-effects estimate of the static big city premium σ then has
$$\operatorname{plim} \hat{\sigma}_{\text{FE}} = \sigma + \frac{1+n-m}{2}\delta \,,$$
which overestimates the actual premium by the average extra value of accumulated experience when migrants are in the big city.
- Imagine instead all migrants relocate from the big to the small city after m periods. The fixed-effects estimate of the static big city premium σ then has
$$\operatorname{plim} \hat{\sigma}_{\text{FE}} = \sigma + \left(\frac{1+m}{2} - \theta m\right)\delta \,,$$
which differs from the actual static premium by the difference between the value of the average big-city experience for migrants prior to moving and the (depreciated) value of the big-city experience that migrants take with them after leaving the big city.

- This example shows that estimation with worker fixed-effects deals with the possible sorting of workers across cities on time-invariant characteristics.
- Biases from not considering dynamic city benefits remain, but their sign is no longer clear:
  - Migrants from small to big cities tend to bias the static city size premium upwards.
  - Migrants from big to small cities tend to bias the static city size premium downwards.
- 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 learning benefits are highly portable across cities.

Pending identification issues

- One cannot interpret the difference between the static pooled OLS and fixed-effects estimates as a measure of the importance of sorting (the bias introduced by dynamic effects is also greatly mitigated).
- The static fixed-effects estimation does not give us a measure of dynamic/learning effects, and these may account for much of the benefits of bigger cities.

Are the benefits of agglomeration only static or also dynamic?

- The usual view: the city size premium is associated with a current city,
  - the premium is attained immediately upon arrival in a big city,
  - and lost immediately upon departure.
- Some authors suggest that the key advantages of big cities are dynamic, they facilitate learning, experimentation, and the acquisition of skills (Glaeser, 1999, Glaeser and Maré, 2001, Duranton and Puga, 2001).
- To examine this, we need to relate the city size premium to the entire history of workplace location.
- We estimate:

$$w_{ict} = \underbrace{\sigma_c}_{\substack{\text{Instantaneous}\\ \text{premium of}\\ \text{working in city } c}} + \mu_i + \gamma e_{it} + \nu e_{it}^2 + \underbrace{\sum_{j=1}^{C}(\delta_{jc} e_{ijt} + \lambda_{jc} e_{ijt} e_{it})}_{\substack{\text{Extra value of experience}\\ \text{depending on where}\\ \text{acquired and used}}} + \mathbf{x}'_{it}\beta + \varepsilon_{ict}$$

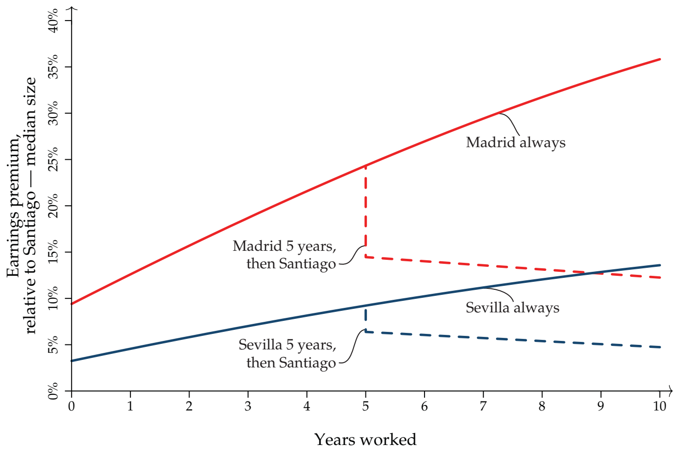
- Note σ_c is still estimated only on the basis of migrants, but δ_{jc} and λ_{jc} use the entire sample.

	(1)	(2)	(3)
Dependent variable:	Log earnings	Initial premium (city indicator coefficients column (1))	Medium-term premium (initial + 7.7 years local experience)
Log city size		0.0223 (0.0058)***	0.0510 (0.0109)***
City indicators	Yes		
Worker fixed-effects	Yes		
Experience 1st-2nd biggest cities	0.0309 (0.0029)***		
Experience 1st-2nd biggest cities × experience	-0.0008 (0.0001)***		
Experience 3rd-5th biggest cities	0.0155 (0.0045)***		
Experience 3rd-5th biggest cities × experience	-0.0006 (0.0002)**		
Experience	0.0912 (0.0019)***		
Experience^2	-0.0011 (0.0000)***		
Experience 1st-2nd biggest × now in 5 biggest	-0.0014 (0.0028)		
Experience 1st-2nd biggest × experience × now in 5 biggest	0.0000 (0.0001)		
Experience 3rd-5th biggest × now in 5 biggest	-0.0025 (0.0043)		
Experience 3rd-5th biggest × experience × now in 5 biggest	0.0003 (0.0002)		
Experience outside 5 biggest × now in 5 biggest	0.0064 (0.0024)***		
Experience outside 5 biggest × experience × now in 5 biggest	-0.0002 (0.0001)*		
Observations	6,263,446	76	76
R^2	0.1165	0.1282	0.3732

Notes: Column (1) also includes tenure and its square, occupational skill indicators, month-year indicators, two-digit sector indicators, and contract-type indicators. City medium-term premium calculated for workers' average experience in one city (7.72 years).

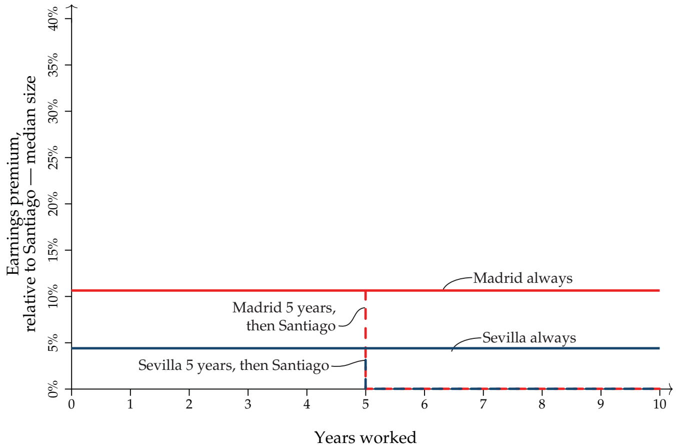
Gradual accumulation of the city size premium

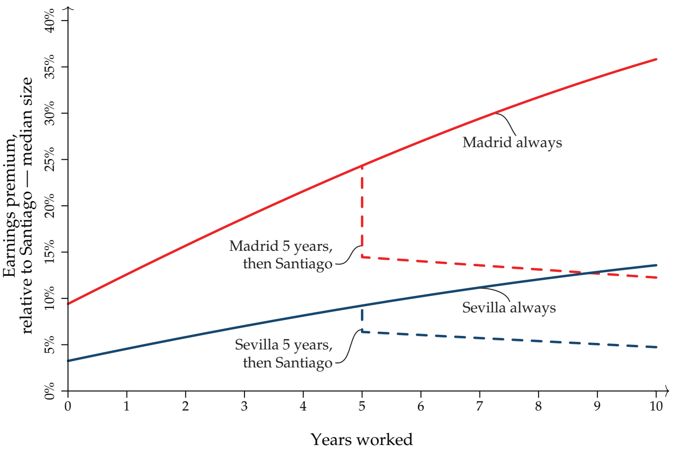
- These estimations show that
  - a premium is attained immediately upon arrival in a big city,
  - but much of the gains accumulate gradually over time,
  - and workers who leave take most of the accumulated premium upon departure.
- To visualize the last two points more clearly, based on the above estimation, we can calculate how differences between the earnings of workers with particular location histories evolve over time.

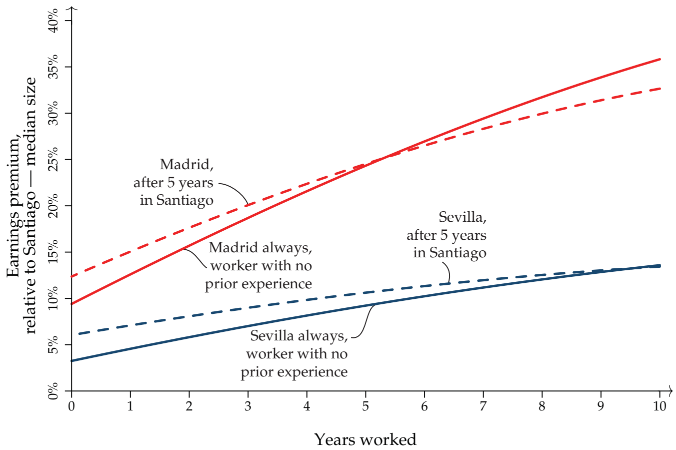


Comparison with static estimation

- Note how this differs from the usual static estimation. The dynamic estimation shows that
  - the earnings premium associated with working in a big city is not a one-off instantaneous gain, instead there is an initial jump but then the premium grows over time;
  - the accumulated premium is not lost upon departure.







Initial city size premium

- By regressing the instantaneous premium

$$\underbrace{\sigma_c}_{\substack{\text{Instantaneous}\\ \text{premium of}\\ \text{working in city } c}}$$

  on log size, we can calculate the elasticity with respect to city size of the initial earnings premium a worker gets right upon moving into a city.
- We can think of this as capturing the instantaneous benefits of working in big cities.
- The elasticity of the initial (instantaneous) earnings premium with respect to size in the dynamic fixed-effects estimation (0.022) is similar to the static fixed-effects estimation.

Medium-term city size premium

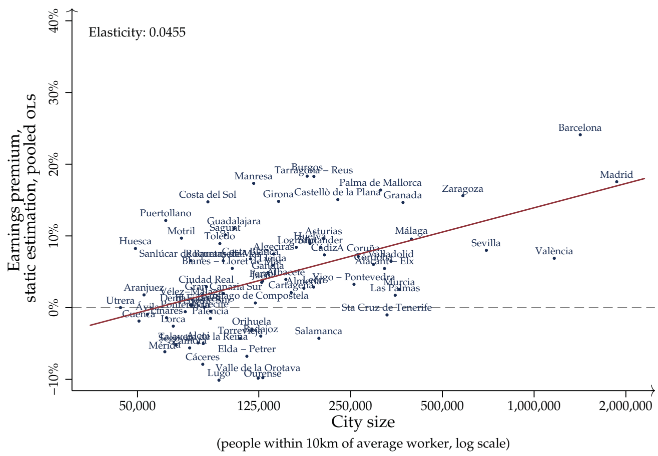
- Let $\bar{e}$ be mean experience accumulated in one city: 7.72 years.
- By regressing the medium-term premium

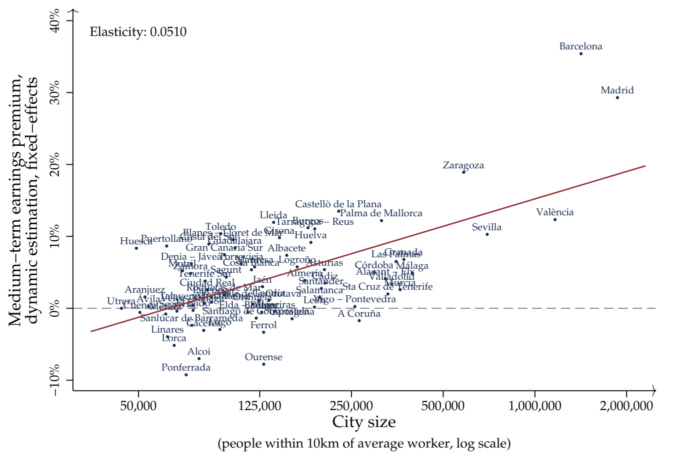
$$\underbrace{\sigma_c}_{\substack{\text{Instantaneous}\\ \text{premium of}\\ \text{working in city } c}} + \underbrace{\delta_c \bar{e} + \lambda_c \bar{e}^2}_{\substack{\text{Extra value}\\ \text{of experience}\\ \text{acquired in city } c}}$$

  on log size, we can calculate the size elasticity of the medium-term earnings premium a worker gets after $\bar{e}$ years working in a city.
- We can think of this as capturing the medium-term benefits of learning by working in big cities.

Learning accounts for lower elasticity with worker fixed-effect

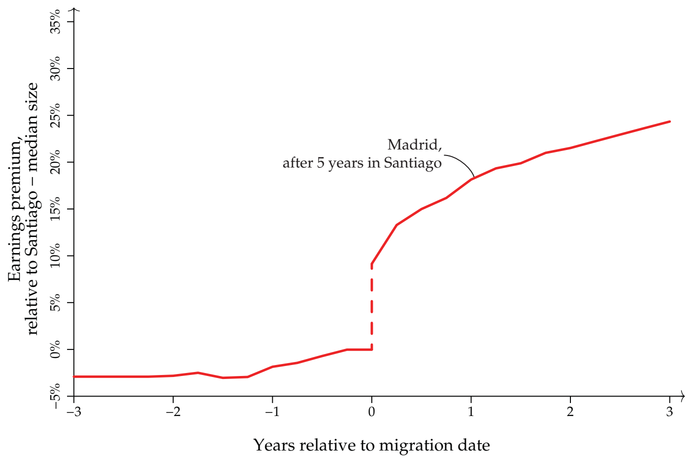
- Elasticity of medium-term earnings premium w.r.t. city size: 0.0510.
- Elasticity of short-term earnings premium w.r.t. city size: 0.0223.
- About half of the gains from working in bigger cities are static and about half are dynamic.
- Elasticity of the medium-term earnings premium w.r.t. city size is not significantly different from the standard static pooled OLS estimate.
  - In this estimation, we have brought dynamic effects 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 estimate to the magnitude of the static pooled OLS estimate indicates that learning effects can fully account for the difference.





Dependent variable:	(1) Log size	(2) Short-term premium	(3) Medium-term premium
Instrumented log city size		0.0203 (0.0079)***	0.0530 (0.0143)***
Log city size 1900	0.6489 (0.0810)***		
% high-quality land within 25km of city centre	0.0151 (0.0065)**		
% water within 25km of city centre	0.0059 (0.0029)**		
% steep terrain within 25km of city centre	-0.0134 (0.0057)**		
Log mean elevation within 25km of city centre	0.2893 (0.0834)***		
Roman road rays 25km from city centre	0.0674 (0.0372)*		
Observations	76	76	76
R^2	0.6503	0.1271	0.3726
F-test weak ident. (H_0: instruments jointly insignificant)		25.2482	25.2482
P-value LM test (H_0: model underidentified)		0.0236	0.0236
P-value J test (H_0: instruments uncorr. with error term)		0.3025	0.2051
P-value endog. test (H_0: exogeneity of instrumented var.)		0.5757	0.5998

No ‘Ashenfelter dip’ in earnings pre-migration



Heterogeneous dynamic advantages of bigger cities

- We have seen that an important part of the advantages associated with bigger cities is that they provide steeper earnings profiles.
- Following Baker (1997), a large literature emphasizes heterogeneity in earnings profiles across workers.
- This suggests, not just allowing for heterogeneous profiles across workers, but also exploring whether there are complementarities between bigger cities and greater individual skills in terms of the value of experience.
- Is the additional value of experience in bigger cities even greater for more able workers?

Estimation with heterogeneous dynamic effects

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

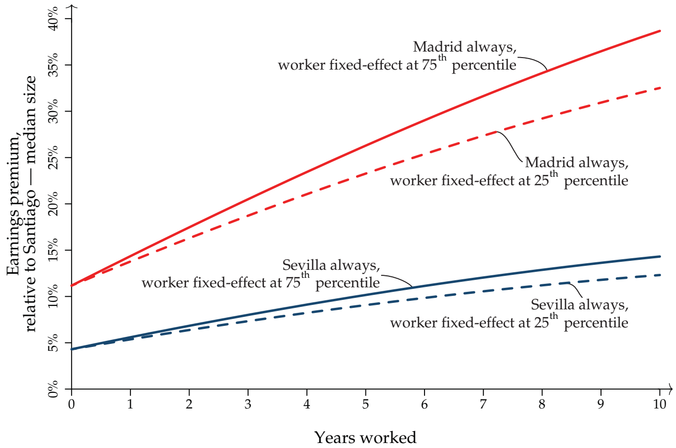
$$w_{ict} = \sigma_c + \mu_i + \sum_{j=1}^{C} (\delta_j + \phi_j \mu_i) e_{ijt} + \mathbf{x}'_{it}\beta + \varepsilon_{ict} \, .$$

(the change is the interaction term between the worker fixed-effect and the value of experience in different cities, μ_i)

- We can estimate this recursively. Given a set of fixed-effects (for instance, those coming from our previous estimation), we can estimate the equation by OLS, then obtain a new set of estimates of worker fixed-effects as

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

then, given these new worker fixed effects estimate again the wage equation above, and so on until convergence is achieved.



Sorting

- Our estimations simultaneously consider
  - static advantages associated with workers' current location,
  - learning by working in bigger cities
  - spatial sorting.
- 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 of different sizes.

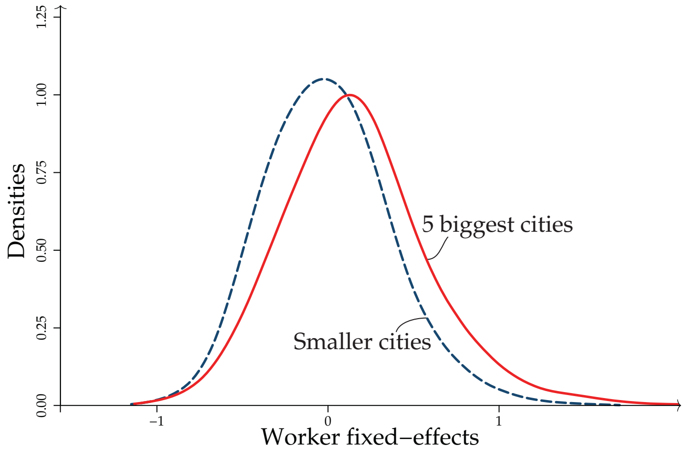
	Occupational groups				
	Very-high-skilled	High-skilled	Medium-high skilled	Medium-low skilled	Low-skilled
1^{st}-2^{nd} biggest cities	10.9%	13.8%	24.2%	41.7%	9.4%
3^{rd}-5^{th} biggest cities	6.3%	10.9%	21.0%	48.2%	13.8%
Other cities	3.5%	7.9%	18.4%	54.0%	16.1%

Notes: Employers assign workers into one of ten social security categories which we regroup into five occupational-skill categories. Shares are averages of monthly observations in the sample.

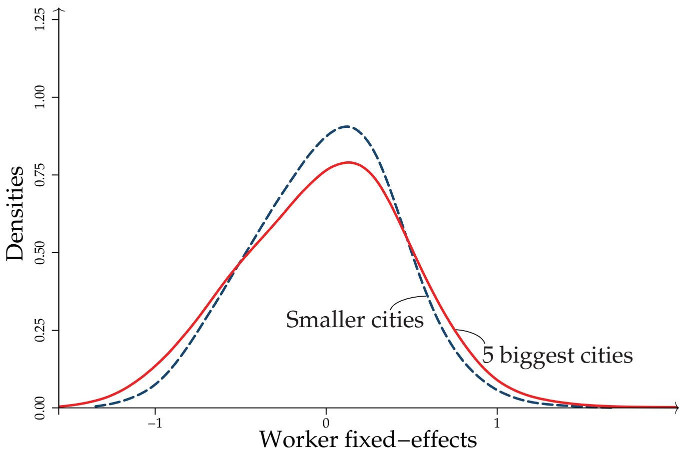
Not considering learning biases worker fixed-effects

- If the heterogeneous dynamic fixed-effects estimation is correct,
$$w_{ict} = \sigma_c + \mu_i + \sum_{j=1}^{C}(\delta_j + \phi_j\mu_i)e_{ijt} + \mathbf{x}'_{it}\beta + \varepsilon_{ict}\ ,$$
then estimated worker fixed-effects capture initial unobserved ability.
- In the static fixed-effects specification,
$$w_{ict} = \sigma_c + \mu_i + \mathbf{x}'_{it}\beta + \zeta_{ict}\ ,$$
even assuming city fixed-effects (σ_c) and other coefficients (β) are unbiased, worker fixed-effects are biased:
$$\text{plim}\ \hat{\mu}_{i\,\text{FE}} = \mu_i(1 + \phi_j\bar{e}_{ijt}) + \sum_{j=1}^{C}\delta_j\bar{e}_{ijt}\ .$$

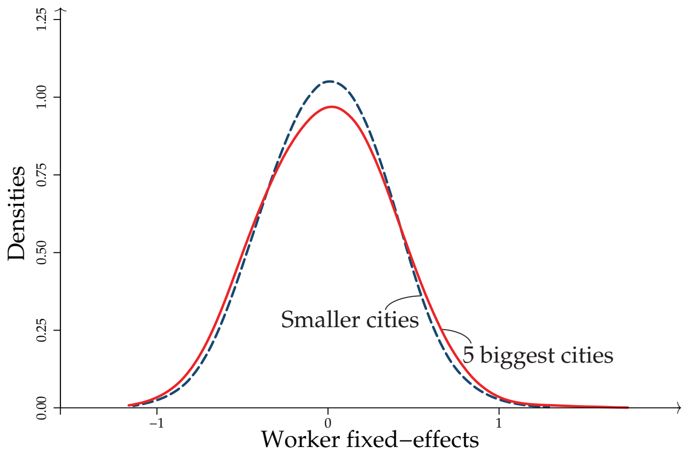
Fixed-effects, static premium only



Fixed-effects, homogeneous dynamic and static premium



Fixed-effects, heterogeneous dynamic and static premium



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)	0.0009 (0.0026)	1.0854 (0.0090) ***	1.7e-03	0.9738	90,628
Worker fixed-effects, homogenous dynamic and static premium (table 2, column 1)	-0.0039 (0.0071)	1.1633 (0.0078) ***	8.8e-03	0.9974	90,628
Worker fixed-effects, static premium (Combes et al., 2012*b*)	0.1571 (0.0050) ***	1.1670 (0.0066) ***	5.6e-02	0.9908	90,628
Log earnings	0.2210 (0.0031) ***	1.2153 (0.0073) ***	.11	0.9825	90,628

Notes: The table applies the methodology of Combes, Duranton, Gobillon, Puga, and Roux (2012*c*) 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.

Sorting

- Clear sorting by observables (education, broad occupation).
- But no sorting by innate ability within broad occupation.
- It is the differential value of big-city experience by ability that increases the variance of earnings in big cities as workers gain experience.

Conclusions

- By tracking not only workers' current job location, but also their entire workplace location histories, we show that
  - an earnings premium is attained upon arrival in a big city,
  - workers accumulate more valuable experience in a big city,
  - and take most of the accumulated premium when they relocate.
- Furthermore, differences in worker skills across cities
  - appear not to be the result of sorting (workers in big and small cities appear initially very similar),
  - but the result of workers accumulating more valuable experience in bigger cities (which increases mean skills),
  - and this benefits more able workers (which increases the variance).
- We interpret this as evidence that an important part of the benefits of big cities are related to learning.