CENTRO DE ESTUDIOS MONETARIOS Y FINANCIEROS (CEMFI)
UNIVERSIDAD INTERNACIONAL MENÉNDEZ PELAYO (UIMP)

MIGRATION, SELF-SELECTION AND LEARNING IN CITIES

by

JORGE M. DE LA ROCA MACCHIAVELLO

DOCTORAL DISSERTATION

July 2012

CENTRO DE ESTUDIOS MONETARIOS Y FINANCIEROS (CEMFI)
UNIVERSIDAD INTERNACIONAL MENÉNDEZ PELAYO (UIMP)

MIGRATION, SELF-SELECTION AND LEARNING IN CITIES

by

JORGE M. DE LA ROCA MACCHIAVELLO

SUPERVISOR: DIEGO PUGA PEQUEÑO

DOCTORAL DISSERTATION

July 2012

A Isabel, Lilia, Jorge E. y Nanín

Acknowledgements

This is not an easy section to write. As I think about all the people whose contributions have helped make this project become a reality, some names come readily to mind while others take some time. But when they do, I am reminded that they are no less important. Everybody played a role.

I rather start with those who have been directly involved in the writing process of this dissertation. Diego Puga has not only been an excellent supervisor, teacher and coauthor. He has indeed become a close friend. I am deeply indebted to his patience, worldwide availability and permanent advice and encouragement on my future academic career. I still remember when urban economics was only a curiosity. Thanks to Diego, I have been introduced into this fascinating world where I feel quite comfortable working and meeting people who are attracted by the study of cities.

All faculty in CEMFI should also deserve my kindest regards. Stéphane Bonhomme was highly supportive during all the stages of the PhD program, as a teacher, master thesis director and close supervisor in this dissertation. Claudio Michelacci taught me, always smiling, about the levels of intensity and rigor that the academic career demands. Manuel Arellano showed me how an economist can cover a broad array of fields while grasping the main ideas of all. Samuel Bentolila never hesitated about devoting time to illustrate me on labor economics.

As usual, most of the time at CEMFI was spent in the PhD room. I would like to recognize all my PhD colleagues for making those moments more enjoyable. Of course, there was plenty of support in the development of this dissertation, but more important and what I take, are the tons of illuminating discussions and the spirit of collaboration and sense of humor. All my colleagues are important, thank you Alessandro, Anatoli, Carlos, Cristian, Cristina, David, Elena, Enzo, Felipe, Hernán, Jan, Joan, Kike, Luca, Lucciano, Manu, María Inés, Pablo, Roberto and Serafima.

Last but not least, CEMFI is full of charming people willing to assist at any time. I would like to acknowledge all CEMFI staff for their positive attitude and constant support during all these years. I can not imagine myself going through the process of writing this dissertation while writing labor contracts, arranging seminar meetings, fixing computer problems or photocopying a large array of chapters and papers. I am deeply thanked to all of you Ana, Araceli, Eloy, Eva, Gema, Irene, Isabel, Javi Lido, Javi Palomino, Mari Carmen, Miguel and Ulises.

During my PhD years I also met wonderful people at *Madrid Institute for Advanced Studies (IMDEA) Social Sciences*. The long commuting to Tres Cantos whether using *Cercanías* suburban trains or buses 712/716 was a worthy journey. All researchers and staff have contributed to this dissertation with helpful advice during seminars and personal conversations. I wish all of you the best in your future academic and non-academic endeavors.

Although CEMFI demanded a large part of my time, often more than expected, my experience in day-to-day life in a city like Madrid has been very enriching. I have met very pleasant people from whom I have learned a lot. Thanks Fernando Pascual, Laura Romero, Julián Sotos and Íñigo Picabea for letting me share with you memorable moments. All the CDV volleyball team also deserves special lines. It took us only one season to develop a strong team capable of beating the top-ranked teams. Although volleyball starts to seem part of the past, what I take is the fact that we still think as one team when we meet. Thanks Giulio, Mariano, Pablo, Rubén and Samu. Nicolas Migeot, thank you for being an excellent teammate in the court, in the *marcha madrileña* and in daily life.

Now it is time to start thanking people who were neither at CEMFI nor in Madrid. My parents come on top of the list. Your constant support and encouragement during all these years have been invaluable. I appreciate vey much how you have adapted to the fact of having a son permanently abroad by always being available in good and bad times. After all these years abroad I have realized that, more important than the high-quality education you invested in, are the moral values you have taught me and still continue teaching. I would also like to recognize the constant motivation I have received from my family, specially my siblings Ericka, Jorge Eduardo, Verónica plus Daniel and César. Lastly, and very important, thank you Lilia. I am extremely indebted to you since the first day you dropped me at the nursery and stood there until lunch time.

The final lines are reserved for Nanín for taking the burden of most of the non-enjoyable moments of a PhD program. I can not express how indebted I am for your patience and support during these years. I hope I can return at least only a small part of all this in the future.

Contents

List of Tables

List of Figures

Introduction

Workers in bigger cities earn more than workers in smaller cities and rural areas. In Spain, 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 like the United States and France (Glaeser, 2011, Combes, Duranton, Gobillon, and Roux, 2012). 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.

Three broad reasons can explain 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 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).

The first two chapters in this doctoral dissertation investigate the role and magnitude of these three cited reasons in generating earnings differentials across cities of different sizes. The first chapter, "Selection in initial and return migration: Evidence from moves across Spanish cities", studies how self-selection in migration flows across cities within a country strongly impacts on the sorting of more productive workers into bigger cities. This paper contributes to our understanding of both internal migration as well as the sorting process of workers through initial and return migration. The second chapter, "Learning by working in big cities", provides a quantitative assessment of the relative importance of the three reasons that trigger the observed city-size premium. The main finding reveals that workers in bigger cities are not particularly different in terms of innate ability after accounting for differences in observable skills. It is working in cities of different sizes that makes their earnings diverge. This result is the most important contribution of this dissertation.

In the first chapter I begin by examining migration flows across Spanish cities to determine who migrates on a long-term basis from the city of first employment. I find migrants are positively selected in terms of observable characteristics such as educational attainment and occupational skills. This finding is consistent with previous results in the regional migration literature (Borjas, Bronars, and Trejo, 1992, Bound and Holzer, 2000, Hunt, 2004). However, most studies are not able to go beyond in the analysis mainly due to the difficulties involved in following migrants across locations. As I will explain later, this is not a concern given the rich data set I use. By looking at the direction of migration flows across cities, I show that the positive selection of migrants is largely driven by the group of highly skilled migrants who move from small to big cities.

In addition, I verify there is selection not only on observables but also on unobservables. I proxy for individual productivity more broadly by looking at the relative position of workers in the local pre-migration earnings distribution. Thus, I compare the characteristics of migrants to those of stayers in the city of origin prior to migrating. This is the appropriate way to address selection (Chiquiar and Hanson, 2005, Fernández-Huertas, 2011). I find more productive workers are more likely to migrate and, this remains so, even when looking within given levels of education and occupational skills. However, selection on these unobservables related to productivity while present in the data, is of substantially smaller quantitative importance than selection on observables.

Although workers' skills and education are important predictors of success in the big city, there is still a substantial amount of variation and uncertainty involved in the migration process. I show this leads to a second stage of sorting. About one half of first-time migrants end up leaving their city of destination within five years, and around 70% of these moves involve a return to the city of first employment. I find that returnees from big cities can be characterized as those individuals with initial skills in between those of stayers and those of permanent migrants. Likewise, they are the least successful in boosting their earnings after migrating. This pattern seems to be specific to them as opposed to other repeat migrants. When I examine second-time moves of migrants in big cities to other cities, they are not affected by realized earnings after their first migration episode.

While I document in detail how migration contributes to the sorting of more skilled workers into big cities, it is worth noting that worker' skill sorting can occur through other channels. For instance, both better schools in large urban areas or faster learning associated with working in big cities can widen the skill gap. In the second chapter of this doctoral dissertation I examine in detail the role of static and dynamic advantages of big cities in generating such skill sorting.

The second chapter is a joint-study with my supervisor Diego Puga. In order to measure the magnitude of static advantages of bigger cities, we compare earnings of workers in cities of different sizes using a simple pooled OLS estimation. 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 static advantages. 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 illustrate that it is instead the result of ignoring the dynamic benefits that big cities provide.

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

Lastly, 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 (Glaeser and Resseger, 2010, Bacolod, Blum, and Strange, 2009, Combes *et al.*, 2012). We look at worker-fixed effects because we are interested in capturing time-invariant ability beyond that which may be reflected in observable characteristics. Nevertheless, we show that if we do not estimate worker fixed effects using our full earnings 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. In sum, our results indicate that workers attain a static earnings premium upon arrival in a bigger city and accumulate more valuable experience as they spend more time working there. 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*a*) emphasize.

For all chapters in this dissertation I use data from Spain's Continuous Sample of Employment Histories (*Muestra Continua de Vidas Laborales* or MCVL), a rich and unique data set in many aspects. This administrative data set tracks 4% of Spain's workforce throughout their careers. The data set combines information from social security records, income tax data and the version of the Census that is updated by municipalities. The availability of employee-employer identifiers allows to follow workers as they move between firms across locations. The data processing is by itself a contribution of this dissertation. Given that the unit of observation in the source data is any change in the individual's labor market status or job characteristics, the procedures developed to uncover information have been very time-demanding. After devoting a lot of effort I have been able to transform the data into a simple and usable monthly panel.

The final chapter in this doctoral dissertation is titled "Wage cyclicality: Evidence from Spain using social security data". In contrast to available estimates that use worker-level data for other countries, I find weak procyclicality of real wages in Spain over the period 1988–2010. This finding suggests that for some European countries with high-wage indexation and employment protection –Spain being a good example– wage cyclicality is presumably much lower than the available European estimates.

Spain is a suitable scenario to evaluate real wage cyclicality within a rigid labor market. The

Spanish system of collective bargaining deters firms from adjusting wages along the business cycle. In fact, a large share of collective agreements (more than 60%) include indexation clauses which trigger high inertia in firms wage-setting decisions. In addition, duality in the labor market insulates workers under permanent contracts (around 67–70% of the workforce with high levels of employment protection) from business cycle fluctuations.

The MCVL has strong advantages relative to other data sets that have been used to estimate wage cyclicality. By exploiting the high frequency in the data I identify most labor market transitions, specially those that are of particular interest in the wage cyclicality literature (e.g. estimating cyclicality for job movers or for workers who remain within a match). Moreover, I can calculate net present values of wages in new matches over their duration, which constitutes a key piece of information for the Mortensen-Pissarides search and matching model (Pissarides, 2009).

This chapter contributes to the wage cyclicality literature in several aspects. First, as already mentioned, unlike wage cyclicality estimates available for countries with flexible labor markets, this study provides one estimate for a rigid labor market scenario. Second, the study shows how wage cyclicality responds in a setting with high duality in employment protection. I find cyclicality for temporary workers is four times greater than for permanent workers. Third, I present evidence of wage cyclicality decreasing consistently with the level of tenure, i.e., cyclicality is much higher for newly-hired workers than for job stayers with high levels of tenure.

The estimated difference in wage cyclicality between newly-hired workers and job stayers is relevant for the empirical validity of the Mortensen-Pissarides search and matching model. The model has been challenged on its ability to match the observed cyclicality on vacancies and unemployment. Some studies have in fact suggested wage rigidity as a potential solution to this so called unemployment-volatilty puzzle (Hall, 2005, Shimer, 2005). I estimate cyclicality for the net present value of wages in new matches, a key statistic for job creation in this model, and obtain a similar estimate to the one using current wages of newly-hired workers. This result, the first in the literature using actual data on wages and job duration, does not give support to wage rigidity as a solution to the unemployment volatility puzzle.

Chapter 1

Selection in initial and return migration: Evidence from moves across Spanish cities

1. Introduction

Workers earn substantially more in bigger and denser cities (Glaeser and Maré, 2001, Wheaton and Lewis, 2002, Combes *et al.*, 2010). This may partly reflect the existence of productive advantages in areas where more firms and workers locate nearby (Duranton and Puga, 2004, Rosenthal and Strange, 2004) and also that interactions in bigger cities facilitate the acquisition of greater skills (Glaeser, 1999, Gould, 2007, Baum-Snow and Pavan, 2012*b*, De la Roca and Puga, 2012). However, it has long been thought that those higher earnings may also partly reflect the sorting of more productive workers into bigger cities. Already in 1890, Alfred Marshall wrote "[i]n almost all countries there is a constant migration towards the towns. The large towns and especially London absorb the very best blood from all the rest of England; the most enterprising, the most highly gifted, those with the highest physique and the strongest characters go there to find scope for their abilities." (Marshall, 1890, 5.6).

Existing studies of worker sorting across cities fall in one of two broad categories. Some papers estimate an earnings premium associated with working in cities in general, or in bigger cities in particular, by regressing individual earnings on worker characteristics and location fixed-effects (Glaeser and Maré, 2001, Combes *et al.*, 2008, 2010). A drop in the magnitude of the estimated earnings premium when worker fixed-effects are introduced in such a specification is seen as evidence of positive sorting. However, De la Roca and Puga (2012) show that the introduction of worker fixed-effects, in addition to absorbing unobserved worker heterogeneity, also largely removes the dynamic component of the city size premium. This can be seen as an advantage for the main objective of those papers, which is to estimate the instantaneous boost in earnings from relocating to bigger cities, but also implies that to quantify the importance of sorting one needs to study the issue more directly. A second strand of literature studies worker sorting by looking at differences in observable skills across cities of different size. Workers in larger cities tend to have higher education (Berry and Glaeser, 2005) and greater occupational skills of both cognitive and social type (Bacolod *et al.*, 2009). However, such differences appear to be relatively small in relation to the observed earnings premium.

This paper investigates the contribution of migration to the sorting of workers across cities by studying whether greater skills, both observed and unobserved, increase the likelihood that a worker migrates to a bigger city. Using administrative data for Spain that follow individuals continuously over time and across cities throughout their careers, I show that migrants who move to big cities are positively selected in terms of their level of productivity as proxied by their relative position in the local pre-migration earnings distribution. This remains so even when looking within given levels of education and occupational skills.

In addition, I document a second stage of sorting that happens after a first migration episode. About 45% of migrants end up leaving their city of destination within five years. Moreover, around 70% of these second moves involve a return migration to the city of origin. Such return migration is more frequent and happens sooner in big cities. I find that to understand such return migration, it is important to look not just at initial worker characteristics and relative earnings prior to the first move, but also at the heterogeneous experiences of workers following their first migration episode.

I develop a conceptual framework in which big cities provide workers with a stochastic earnings premium but also involve higher housing costs. Even if faced with the same distribution of the premium, more skilled (and thus higher income) workers are more likely to be able to afford the higher housing costs of big cities and the costs of migration. As a result, of all workers in small cities, only those with skills above a certain threshold are willing to migrate to big cities. Then, of workers who migrate, those with the highest skills remain in big cities while those with intermediate skills end up returning unless the realization of their stochastic earnings premium is sufficiently high. These patterns of return migration are supported by the data. Returnees are not only less productive than permanent migrants prior to their first move. They are also those who, following the first move, have least boosted up their earnings in the big city. This pattern seems to be specific to returnees. When I examine second-time moves of migrants to other cities, they are not affected by realized earnings in the big city.

Studies that analyze selection in initial and return migration have focused mostly on international migration, specially on flows between Mexico and the United States.[1] Besides being of interest per se, studying migration across cities within a country helps overcome two important caveats of international migration studies. First, we can observe migrants' working histories in both the origin and the destination, whereas international studies tend to observe migrants' working histories only in one location, either the origin or the destination country. Second, even if international studies could track individuals across countries, institutional and economic differences between them would make it more difficult to evaluate the performance of migrants and returnees than in the case of internal migration.

Previous studies of regional migration (see Greenwood, 1997, for a survey) find that migrants tend to be more educated, employed in higher skill occupations, and generally more productive.

[1]See Borjas and Bratsberg (1996) for a model on international return migration. See Chiquiar and Hanson (2005), Ibarrarán and Lubotsky (2007), Reinhold (2009), Fernández-Huertas (2011) for selection and return migration flows between Mexico and the United States. See Co, Gang, and Yun (2000), Constant and Massey (2003), Dustmann (2003), DeCoulon and Piracha (2005), Rooth and Saarela (2007), Ambrosini, Mayr, Peri, and Radu (2011) for international return migration in European countries.

Borjas *et al.* (1992) using NLSY data show that more educated and productive workers in the United States are more likely to migrate regardless of their state of origin. In addition, skilled workers in states with low earnings inequality have a higher propensity to outmigrate to states with higher inequality. Bound and Holzer (2000), using US Census data to examine the role of individual characteristics in the sort of labor adjustments to regional shocks studied by Blanchard and Katz (1992), find that workers with low education are less prone to migrate in response to shifts in demand. For Europe, Hunt (2004) examines determinants of migration among federal states in Western Germany. She also finds that migrants are more skilled than stayers. I contribute to this literature by using cities (instead of States or regions) as the units of analysis, and showing that long-term migrants from small to big cities are key to understanding why migrants are positively selected. Moreover, I also allow skills to vary over time by looking at workers' relative position in the local earnings distribution at the time of each migration episode (instead of focusing only on observable skills or a time-invariant worker fixed-effect). This turns out to be particularly important in distinguishing who stays and who returns after a first migration episode. Surprisingly, few studies examine such return migration flows within a country.[2] Considering selection on the basis of characteristics observed in the first as well as in the second location allows me to gain further understanding of the characteristics and experiences of returnees.

The rest of the paper is structured as follows. Section 2 introduces a conceptual framework to help frame the problem. Section 3 presents the econometric framework. Section 3 describes the data. Section 5 presents the results. Finally, section 6 concludes.

2. Conceptual framework

In order to motivate the empirical analysis, I now develop a simple conceptual framework. This considers a pool of heterogeneous workers who are initially located in a small or low-density city (L) to determine the characteristics of those who self-select into migrating to a big or high-density city (H), and also the characteristics of those who, after spending a period of time in city H, self-select into returning to city L.[3]

All workers have identical preferences and are risk neutral but have heterogeneous initial skills. The initial skill (or marginal value product of labor in city L) of worker i is denoted s_i. As in Roback (1982), we wish to consider how differences in earnings and housing costs jointly determine location. Each worker rents a house and spends the rest of her income on a consumption good used as numéraire. I abstract from differences in the characteristics of dwellings, so that everyone rents a house of a standard type. Utility can then be expressed as earnings minus housing costs. Housing costs in city L are normalized to zero, so that utility there is simply

$$U_i^L = s_i \; . \tag{1.1}$$

[2]DaVanzo (1983) and Kennan and Walker (2011) for the US, and Hunt (2004) for Germany are some exceptions. A common feature of these studies is the small sample of return migrants in the survey data they use. Moreover, migration in general is underestimated due to attrition of movers. The large panel of administrative data I use is a great advantage on this respect.

[3]The framework also has implications for migration flows in the opposite direction, from H to L, which are briefly discussed below.

City H is characterized by three differences with respect to city L. First, a worker who works in city H acquires extra skills $\delta_i \sim U[0, 2\delta]$.[4] Second, workers with any given level of skills are α times more productive (and earn α times more) when working in city H.[5] Third, housing in city H involves an extra rental cost R.[6] Thus, utility in city H is

$$U_i^H = \alpha(s_i + \delta_i) - R \; . \tag{1.2}$$

Migrating from city L to city H involves a cost C.

In a simpler framework with irreversible migration and no uncertainty in the realization of skills in city H (e.g., everyone gets $\delta_i = \delta$), the result is straightforward. A worker with initial skill level s_i moves from city L to H if and only if the gain in earnings is enough to at least pay the moving cost C and the extra rent R, i.e., if and only if $\alpha(s_i + \delta) - R - C > s_i$. Thus, in equilibrium, city L would be populated by workers with low skills

$$s_i \leqslant \hat{s} = \frac{R + C - \alpha\delta}{\alpha - 1} \; . \tag{1.3}$$

Anyone with $s_i > \hat{s}$ would migrate to city H. Simply introducing uncertainty in the acquisition of skills in H would not imply any difference for the decision to migrate, since workers are risk neutral and would migrate based on the expected value of additional skills, $\mathbb{E}(\delta_i) = \delta$.

The key ingredient in my framework is the combination of uncertainty in the ex-post realization of skills in H *and* the possibility of return migration after paying an additional moving cost. Together, these imply that some workers with skills low enough that they would be unwilling to undertake irreversible migration ($s_i \leqslant \hat{s}$), given that they can return, are now willing to experiment. If they move to city H and have a good realization of δ_i, great; if not, they can always move back, subject to some cost. Similarly, some workers with higher initial skills ($s_i > \hat{s}$) will now end up returning after migrating from city L to H, if they have a bad realization of δ_i. As a result, city H will exhibit ex-post higher average skills and earnings, but the skill distributions of the two cities will partially overlap because of uncertainty in realization of skills and return moves by unlucky migrants. As we shall see below, this prediction is consistent with what we observe in reality. So are the predictions for initial and return migration flows, the latter being specific to this richer framework.

The intuition for initial migration from city H to L is much simpler. As there is no uncertainty in the ex-post realization of skills in L, return migration can not be optimal.[7] Workers initially located

---

[4]Glaeser (1999) develops a learning model where young workers who move into a big city increase their skills with some probability. De la Roca and Puga (2012) find evidence of substantial skill acquisition by workers in bigger cities. On the firm side, Duranton and Puga (2001) develop a model in which big cities are diversified places that foster innovation and experimentation. Firms can only find their optimal production process in big cities with some probability in every period.

[5]This feature is widely documented in the literature. See Rosenthal and Strange (2004) for a review of the evidence.

[6]This feature is also widely documented in the literature. See Combes, Duranton, and Gobillon (2011*a*) for a recent estimate of the relevant elasticity.

[7]The framework can be extended to incorporate uncertainty in the ex-post realization of skills in L. Now, workers can increase their level of skills in city L, but on average this increase should be lower than in city H. The main advantage of this setting is to allow for return migration from small cities. In the data the incidence of second migration and return migration is more frequent and happens sooner in big cities.

in city H decide to migrate to city L if and only if

$$\alpha(s_i + \delta_i) - R \leqslant s_i - C \; . \tag{1.4}$$

Thus, in equilibrium, workers with skills

$$s_i \leqslant \widetilde{s} = \frac{R - C - \alpha\delta_i}{\alpha - 1} \; , \tag{1.5}$$

migrate to city L. Clearly, the decision to migrate to city L depends on both the level of skill s_i and the realization of extra skills δ_i in H.

We now solve the model and draw all these stated predictions explicitly.

Solution

I first characterize selection in initial and return migration from city L to H. The timing in the framework is the following. In the first stage, based on her initial ability s_i, each worker i decides between staying in city L or migrating to H and paying the migration cost C. In the second stage, workers who have migrated to city H observe their individual realization of δ_i and, with this extra information, decide whether to remain in city H or to return to city L, the latter involving an additional migration cost C_2. Both migration costs, C and C_2 are assumed to be sunk. Furthermore, we assume that $C + C_2 \leqslant \alpha\delta$ (otherwise, as shown below, no migrant ever returns and the framework collapses to the case of irreversible migration discussed above).

I proceed backwards, and first concentrate on the second stage. After moving to H the realization of δ_i is revealed to the worker. She decides to return if and only if $\alpha(s_i + \delta_i) - R \leqslant s_i - C_2$. Thus, a worker returns if ex-post earnings in H are lower than earnings in L minus the return migration cost.[8] Given that δ_i takes a minimum value of 0 and a maximum value of 2δ, some workers would always return even in the best-case scenario of $\delta_i = 2\delta$, others will never return even in the worst-case scenario of $\delta_i = 0$, while others return depending on the actual realization of δ_i. In particular, a worker who migrates to city H returns to city L if and only if

$$\delta_i < \underline{\delta}(s_i) = \begin{cases} 2\delta & \text{if} \quad s_i < \underline{s} \; , \\ \frac{R - C_2 - (\alpha-1)s_i}{\alpha} & \text{if} \quad \underline{s} \leqslant s_i < \overline{s} \; , \\ 0 & \text{if} \quad s_i \geqslant \overline{s} \; , \end{cases} \tag{1.6}$$

where

$$\underline{s} = \frac{R - C_2 - 2\alpha\delta}{\alpha - 1} \; ,$$
$$\overline{s} = \frac{R - C_2}{\alpha - 1} \; .$$

I now come back to the first stage. When deciding whether to migrate to city H, workers must take expectations over the possible realizations of δ_i, incorporating the decision of whether to

[8] I assume the realization of δ_i is not portable. None of the qualitative results change if I allow workers to transfer acquired skills back to L.

return or not that they will base on that realization. Thus, a worker will migrate to H if and only if

$$\int_{0}^{\underline{\delta}(s_i)} \frac{1}{2\delta}(s_i - C_2)\,\mathrm{d}x + \int_{\underline{\delta}(s_i)}^{2\delta} \frac{1}{2\delta}[\alpha(s_i + x) - R]\,\mathrm{d}x - C > s_i \tag{1.7}$$

For workers with $s_i < \underline{s}$, the condition of equation (1.7) is never satisfied, so they never migrate. Since they know that they would always find it preferable to return regardless of their realization of δ_i, not migrating to start with and thus saving the migration costs $C + C_2$ must be strictly preferable.

For workers with $\underline{s} \leqslant s_i < \overline{s}$, substituting equation (1.6) into (1.7) and simplifying turns this condition into

$$s_i > \frac{R - C_2 - 2(\alpha\delta + \sqrt{(C + C_2)\,\alpha\delta})}{\alpha - 1} \,. \tag{1.8}$$

For workers with $s_i \geqslant \overline{s}$ (those who know they will never return regardless of their realization of δ_i), the condition of equation (1.7) collapses to that of the simplified framework with irreversible migration, i.e., they will migrate if and only if $s_i > \hat{s}$, where $\hat{s}$ is given by equation (1.3). However, the assumption that $C + C_2 \leqslant \alpha\delta$ ensures that $\hat{s} \leqslant \overline{s}$, so that workers with $s_i \geqslant \overline{s}$ always migrate and never return.[9]

To summarize the results:

- Workers with low initial skills ($s_i < \frac{R-C_2-2(\alpha\delta+\sqrt{(C+C_2)\,\alpha\delta})}{\alpha-1}$) do not migrate from city L to H.
- Workers with intermediate initial skills ($\frac{R-C_2-2(\alpha\delta+\sqrt{(C+C_2)\,\alpha\delta})}{\alpha-1} \leqslant s_i < \frac{R-C_2}{\alpha-1}$) migrate from city L to H. Based on how much they end up gaining from relocating,
  - those who get particularly good outcomes ($\delta_i \geqslant \frac{R-C_2-(\alpha-1)s_i}{\alpha}$) remain in city H,
  - while those who get worse outcomes ($\delta_i < \frac{R-C_2-(\alpha-1)s_i}{\alpha}$) return to city L.
- Workers with high initial skills ($s_i \geqslant \frac{R-C_2}{\alpha-1}$) migrate to city H and do not return, regardless of how much they end up gaining from relocating.

Selection in initial migration from city H to L is straightforward. Workers initially located in H, based on their initial ability s_i and after observing their realization of δ_i, decide between staying in city H or migrating to L. Once again I need to consider cases that vary with realizations of δ_i.[10]

For workers with $s_i < s_*$ (see footnote 10), the condition of equation (1.4) is always satisfied as $s_* < \widetilde{s}$, where $\widetilde{s}$ is given by equation (1.5). These workers always migrate, yet, the least skilled of

[9] If instead $C + C_2 > \alpha\delta$ then $\hat{s} > \overline{s}$ and equation (1.8) is never satisfied for $s_i < \overline{s}$. In this case, we are back to the case of irreversible migration. Only workers with $s_i > \hat{s}$ migrate and no workers ever return.

[10] The values of δ_i for which initial migration from city H to L takes place are the following:

$$\delta_i < \delta_*(s_i) = \begin{cases} 2\delta & \text{if} \quad s_i < s_* \,, \\ \frac{R-C-(\alpha-1)s_i}{\alpha} & \text{if} \quad s_* \leqslant s_i < s^* \,, \\ 0 & \text{if} \quad s_i \geqslant s^* \,, \end{cases} \tag{1.9}$$

where

$$s_* = \frac{R - C - 2\alpha\delta}{\alpha - 1} \,, \quad s^* = \frac{R - C}{\alpha - 1} \,.$$

this group might not be able to afford the moving cost C. For workers with $s_* \leqslant s_i < s^*$, those who get an unfavorable realization of δ_i such that $s_i \leqslant \widetilde{s}$ migrate to city L. Lastly, for workers with $s_i \geqslant s*$, the condition of equation (1.4) is never satisfied since $\widetilde{s} < s^*$. These workers find strictly preferable to remain in city H.

To summarize these latter results:

- Workers with low initial skills ($s_i < \frac{R-C-2\alpha\delta}{\alpha-1}$) migrate from city H to L.
- Workers with intermediate initial skills ($\frac{R-C-2\alpha\delta}{\alpha-1} \leqslant s_i < \frac{R-C}{\alpha-1}$) migrate from city H to L only if they get a bad outcome in city H (i.e., $\delta_i < \frac{R-C-(\alpha-1)s_i}{\alpha}$).
- Workers with high initial skills ($s_i \geqslant \frac{R-C}{\alpha-1}$) do not migrate to city L.

This simple framework delivers some predictions that will be tested in section 5. First, there is sorting in initial migration, whereby workers with sufficiently high initial skills/earnings in small cities migrate to big cities. Likewise, workers with relatively low initial skills/earnings in big cities sort into small cities. Second, among those migrants to big cities, those with the highest initial skills stay in the big city while those with intermediate skills return provided they only get an unfavorable earnings boost. Yet, the probability of return is not random, but decreases both with their initial skill/earnings level in the small city and with their earnings gain in the big city.

3. Econometric framework

I specify a single-exit discrete duration model that can be viewed as a sequence of discrete choice binary models, defined over the population who is at risk of migrating at each period. Thus, in each period, individuals maximize utility by choosing whether to stay in their city or migrate. I focus only on one-way transition events. When focusing on first-time migrants, this implies that an individual can engage in a first migration at most once, and then drops from the population at risk of migrating for the first time.

My unit of analysis is an individual-period pair, where my data is at the monthly level. In each month, I observe different values of individual-level variables (aggregate variables are also captured through location-period indicators) and the migration decision of the individual. I treat each individual-month pair as a distinct observation. I model the hazard rate, i.e. the probability of migrating at time t provided the individual did not migrate up to time t, in the following way:

$$h(t) = P[T = t | T \geq t, x(t)] = F[\beta_0(t) + \beta_1' x(t)] \, , \tag{1.10}$$

where T is the month in which the first migration episode occurs (possibly never), F is a cumulative probability function (always a logistic specification in the study), $x(t)$ is a vector of (possibly time-varying) individual and job characteristics, including the city where the individual is working, $\beta_0(t)$ is a duration-specific parameter that captures duration at t in an additive and unrestricted way and β_1 is a vector of parameters. Therefore, I am modeling for an individual working in her city of first location the probability of migrating, conditioning on observable characteristics.

In the city of first location, the log-likelihood function for a single-exit discrete duration model is the sum of the contributions of N individuals as follows:

$$L(\beta) = \sum_{i=1}^{N} \left[(1-m_i) \sum_{t=e_i}^{T_i} \log\left(1-h_i(t)\right) + m_i \left(\sum_{t=e_i}^{T_i-1} \log\left(1-h_i(t)\right) + \log h_i(T_i) \right) \right] \tag{1.11}$$

where i indexes the individual, m_i is an indicator variable which takes value 1 if a migration is observed and 0 otherwise, e_i is the month of entry in the sample which usually will correspond to the age of entry in the labor force and T_i is the number of months elapsed until first migration.

Alternatively, I can rewrite this function as the log-likelihood of a logit model resulting from the aggregation of the samples surviving at each duration t. With this aim I introduce a sequence of migration indicators at t, such that $Y_t = \mathbf{1}(T = t)$ takes value 1 only in the last month prior to migration and 0 otherwise. Thus,

$$L(\beta) = \sum_{t=1}^{T_i} \left\{ \sum_{i=1}^{N} \mathbf{1}(T_i \geq t \geq e_i) \left[m_i Y_{ti} \log h_i(t) + (1-m_i Y_{ti}) \log(1-h_i(t)) \right] \right\} \tag{1.12}$$

and $\hat{\beta}$ is the maximum likelihood estimator that maximizes $L(\beta)$. Therefore, discrete duration models can be regarded as a sequence of binary models (Jenkins, 1995). I estimate equation (1.12) to examine how the productive characteristics of migrants compare to those of stayers in their city of origin prior to migration.

It will be useful to sometimes split a given risk (e.g., migrating for the first time) into several alternative options (e.g., initial migration to a big city and initial migration to a small city). This requires a multiple-exit discrete duration model. One possibility is to model both transition intensities into such states in a multinomial logit model, i.e. model the probability of either moving to a small city or moving to a big city at time t conditional on not having done either before. An alternative way is to model conditional hazard rates, i.e. model the probability of moving to a big city at time t conditional on not having done so before and on not having moved to a small city either. Bover and Gómez (2004) show that if the transition intensities are multinomial logit, the conditional exit rates are binary logit with the same parameters. Thus, the logit specification is derived from the same model in both cases. Likewise, estimating the model by joint maximum likelihood or conditional maximum likelihood results in consistent and asymptotically normal estimates of the parameters. Although the former approach is generally asymptotically more efficient, this will make little difference in this study as the samples I use are large.

In addition to initial migration episodes, I am also interested in subsequent migration episodes. I can estimate a similar logit specification to that of equation (1.12) to analyze how the productive characteristics of second-time migrants compare to those of workers who engaged in the same initial migration episode but instead remain in the city to which they first moved.[11] Once again, it is possible to introduce multiple alternatives, such as return migration to the city of origin or move-on migration to a third city.

[11]One difference between both specifications is how I introduce duration-specific parameters to capture duration dependence. In equation (1.12), age indicator variables capture time spent in city of origin. In the specification for determinants of second-time moves, I need to include indicator variables for the number of years since the migration event took place.

4. Data

In order to examine selection in initial and return migration I need a data set that follows individuals over time and across locations from the beginning of their working lives. Having data from the start of the first job is important to identify accurately the first migration episode. For migrants, the data should record labor market characteristics both at the origin and at the destination of each migration. However, since we wish to explain migration by comparing migrants both with themselves at times in which they do not migrate and with other workers, whether migrants or not, in practice we need the data to record the labor market characteristics of all workers with high frequency since the start of their first job.

The *Muestra Continua de Vidas Laborales* (MCVL), or Continuous Sample of Employment Histories, satisfies these requirements. This is an administrative data set with information on a 4% non-stratified random draw of the population who on a given year have any relationship with Spain's social security, be it because they are working, receiving unemployment benefits, or receiving a pension. For each of these individuals, all of their changes in labor market status and work characteristics are recorded since 1981. I combine data from all available editions of the MCVL, from 2004 to 2010, so as to have data on a 4% sample of all individuals who have worked, received benefits or a pension at any point in 2004–2010. The criterion for inclusion in the MCVL (based on the individual's social security number) is maintained from year to year, so that the difference across editions is that more recent editions include individuals who enter the labor force for the first time, while they lose those who cease any relationship with the social security (individuals who stop working, continue to be included in the sample while they receive unemployment benefits or a retirement pension, so most exits occur when individuals are deceased). The unit of observation in the source data is any change in the individual's labor market status or job characteristics (including changes in occupation or remuneration within the same firm). Given that all changes since 1981 or the date of first employment are recorded, I am able to construct a panel with day-by-day job characteristics for all individuals in the sample.

I construct for all workers monthly working life histories since either 1981 or entry in social security records, whichever is most recent. For every job spell I know the type of occupation and contract, self-employment status and the 3-digit NACE sector of economic activity. For every unemployment spell I know the amount of monthly unemployment benefits or subsidies. Some individual characteristics like age, gender and province of first affiliation with the social security are also provided. Other individual variables such as level of education and province/country of birth are obtained from the *Padrón* or Municipal Register. I build precise measures of cumulative labor experience and job tenure recording the actual number of working days in each month.

The data includes monthly earnings for each job spell, constructed by combining a variety of sources. For the period 2004–2010, uncensored earnings data are available from matched income tax returns for all workers except those in the Basque Country and Navarre (where income taxes are not collected by the Central Government). In addition, for the entire period 1981–2010 earnings data are available for all workers, including those in the Basque Country and Navarre, from the social security, but these are capped for a small fraction of workers.[12]

A crucial feature of the MCVL is that workers can be tracked across space based on their workplace location. Social security legislation forces employers to keep separate earnings' contribution accounting codes for each province in which they conduct business. Furthermore, within a province, a municipality identification code is provided if the workplace is located in a municipality with population greater than 40,000 inhabitants in 2001. Thus, location information is at the establishment level.

Urban areas

I use official urban area definitions by Spain's Department of Housing for 2008. The 85 urban areas in Spain account roughly for 68% of population and 10% of total surface. They represent local labor markets comparable to Metropolitan Statistical Areas (MSAs) in the United States. The median urban area has a population of 141,498 inhabitants in 2010. From now on, I use the terms cities and urban areas indistinctly to refer to local labor markets.

Urban areas enclose 747 municipalities. Given that I know the municipality of workplace location for each job and unemployment spell in MCVL, I can assign each individual to an urban area in any month, provided the municipality has a population greater than 40,000 inhabitants in 2001. There is large variation in the number of municipalities per urban area. Barcelona is made up of 165 municipalities while 21 urban areas contain a single municipality. The median urban area consists of 4 municipalities. I cannot identify 6 small urban areas in MCVL data because population of their largest municipality in 2001 is below the 40,000 population threshold.[13]

To measure the scale of an urban area I count the number of people within 10 kilometers of the average resident in the urban area, a measure proposed by De la Roca and Puga (2012). This is an index of density, which the literature generally prefers to simple population size as a measure of the potential for interactions that an urban area offers to workers (Puga, 2010, Combes, Duranton, and Gobillon, 2011*b*). At the same time, by considering agglomeration patterns within and around the urban area, it avoids some of the problems derived from the administrative border definitions of urban areas that affect simpler measures of density, like the ratio of total population to total land area.[14] In any case, results are robust to measuring the scale of each urban area by its total population.[15]

[12]Appendix 1.A provides details on how these sources are combined and how earnings for the small fraction (13.4%) of capped observations are estimated, based on uncensored observations, on the magnitude of social security contributions, and individual and job characteristics.

[13]These are in order of population size Denia - Jávea, Valle de la Orotava, Blanes - Lloret de Mar, Sant Feliú de Guixols, Soria and Teruel.

[14]Urban areas are defined as aggregates of municipalities. Several small and medium-sized urban areas (such as Badajoz or Albacete) include in their main municipality large extensions of mostly uninhabited nearby rural land, which makes population per surface area unit artificially low for them. Others instead (such as Burgos) have a municipal border cut with medium-populated suburbs adjacent to their border, which makes population per surface area unit artificially high for them. Calculating the number of people within 10 kilometers of the average resident largely gets around both problems.

[15]The correlation between the number of people within 10 kilometers of the average resident and total population is 0.93. In the context of this paper, the main advantage of the measure I use is that it takes into account the proximity of workers in adjacent urban areas, which are totally excluded when one looks only at total population.

Sample restrictions

My initial sample is made up of males born in Spain between 1963 and 1992 (i.e., aged 18–47 during the period 1981–2010) who have been employed, either as employees or self-employeds, or received unemployment benefits at any point over this period. I leave out individuals older than 47 in 2010 and foreign-born immigrants since I cannot retrieve complete work histories for them. I also leave out women because, particularly in the earlier years of the sample period, their migration decisions may have been more heavily influenced by reasons outside the labor market. A total of 300,202 individuals and 42,628,770 monthly observations make up this initial sample.

From this initial sample, I eliminate individuals with low labor force attachment in their lives, which implies dropping those who have not worked for at least 6 months in at least one calendar year between 1981 and 2010. This restriction reduces the sample to 281,384 individuals and 42,470,128 observations. Next, I drop observations from special social security regimes such as agriculture, fishing and mining. Workers in these regimes tend to self-report earnings and the number of work days recorded is not reliable. Furthermore, these activities are typically rural in nature and linked to natural advantages. At this point, the sample contains 280,031 individuals and 40,489,397 observations.

Subsequently, I exclude observations for which the occupation or workplace location is missing and individuals for whom the educational attainment or the province of entry in the labor force is missing. This leaves 274,605 individuals and 38,102,081 monthly observations. Finally, since I wish to focus on urban migrations (and in any case, only the province is known for rural jobs), I focus on workers located in urban areas. This leaves the final sample at 216,544 individuals and 23,522,113 monthly observations.

Identifying migrants

A migration event is defined as a change in workplace location between two urban areas. In the sample 63,663 individuals can be classified as urban migrants while 152,881 individuals never leave their urban area for working purposes.[16]

The main type of migration I examine necessarily requires a permanent change in home residence. Unlike workplace location (which is precisely measured at any point in time), the residential location of workers (merged from a separate data set) is not kept up-to-date. Thus, I detect permanent changes in residence based on the length of the migration episode and the distance between the cities of origin and destination. Both the conceptual framework of section 2 (the change in housing costs in the framework is associated to a change in residence) and the empirical results presented below suggest that the behavior of short-term or short-distance migrants is rather different from that of long-term and long-distance migrants.

Regarding the length of the migration episode, short-term migrants usually move for brief transfers within a job or to work in a seasonal or temporary job. I classify migrants as *short-term* if they never move beyond a 12-month period. Therefore, long-term migrants are movers who

[16]Based on first-time moves of 96,907 migrants, urban migrants represent 66% of the total, followed by urban-rural (20%), rural-urban (10%) and rural-rural migrants (4%).

experience spells longer than a year in the city of destination. Based on this criterion, I identify 25,502 short-term migrants.

Regarding distance, within the sample of long-term migrants, I label migrants as *short-distance* if they move to an urban area that is less than 120 km. (74.6 miles) driving from the urban area where they previously worked.[17] Although urban areas can be understood as independent local labor markets, in some cases two or more of them may exhibit substantial overlapping in worker flows. This pattern is more prevalent in bigger urban areas such as Madrid and Barcelona, which tend to have smaller urban areas at reasonable commuting distances. Based on this criterion, I identify a total of 17,708 short-distance migrants.

These classification criteria leave a total of 20,453 long-term and long-distance migrants. Of these, 59% move only once in their lives while 31% have moved at most twice. I identify return migrants as those who move back to their city of origin in their second migration. Likewise, move-on migrants are those who do not return to their city of origin after two migrations. Since very few long-term and long-distance migrants register more than two moves, none of the results presented below changes when I restrict the estimation sample to migrants who move at most twice.

Table 1.1 shows summary statistics for non-migrants (stayers), short-term and short-distance migrants, and long-term and long-distance migrants. Within the latter category, I provide separate statistics for permanent migrants (those who never return to their city of first employment) and return migrants (those who eventually return). All variables displayed are individual averages over working lives.

The raw data already shows a clear ranking by educational attainment, where permanent migrants are the most educated, followed by return migrants, and then stayers. Short-term and short-distance migrants exhibit the lowest tertiary and secondary education completion rates (I have grouped short-term and short-distance migrants since both, in general, exhibit similar means in all variables).[18]

This ranking is confirmed by the types of occupations in which individuals tend to work. Permanent migrants are twice more likely to work in occupations demanding very-high skills (those typically requiring an engineering or advanced college degree) than stayers and short-term/distance migrants. The ranking of permanent migrants, then return migrants, then stayers, then short-term or short-distance migrants continues to hold going down to individuals in occupations with high skills.

The ranking of monthly earnings across categories again points in the same direction. Permanent migrants exhibit the highest average earnings and are followed by returnees. Stayers and short-term/distance migrants earn substantially less, the gap being larger for the latter. Among

[17] I have collected data on the shortest driving distance between any two urban areas using Google Maps.

[18] The level of education is that contained in the Municipal Register. A large update to this information was done by municipalities in 2001. Beyond that year any revision takes place only if individuals update their level of education, so individuals who have upgraded their education very recently may have this underreported. However, there is no reason to suspect this affects different categories to different extents. The confirmation of the same ranking using occupational categories provides additional reassurance.

Table 1.1: Summary statistics of stayers and migrant types

	Stayers	Migrants		
		Short-term and short-distance	Long-term long-distance	
			Return	Permanent
Level of education				
Tertiary	13%	12%	16%	22%
Secondary	34%	28%	33%	37%
Primary	53%	60%	51%	41%
Occupational skills				
Very-high skills	7%	6%	9%	14%
High skills	9%	8%	12%	14%
Medium skills	15%	13%	19%	19%
Low skills	51%	53%	47%	42%
Very-low skills	18%	20%	13%	11%
Earnings				
Mean monthly earnings	1,818	1,702	2,063	2,269
Mean monthly earnings 2$^{\text{nd}}$ city			2,086	2,547
Labour market characteristics				
Years of labor experience	7.8	6.6	7.3	7.1
Years of firm tenure	3.4	1.9	1.8	2.2
Self-employed	10%	5%	5%	5%
Public sector employee	4%	5%	4%	6%
Temporary contract	23%	36%	31%	30%
Part-time contract	7%	7%	5%	6%
Unemployed	9%	16%	15%	12%
Age	30.2	30.0	31.2	31.1
Age of entry in labor force	20.6	20.2	20.5	21.3
Individuals	152,881	43,210	5,596	14,857

Notes: Variables are averages for individuals over their lives. Only individuals working in urban areas are included. Long-term long-distance migrations are moves that exceed 12 months in destination and distance of 120 km. Earnings expressed in December 2010 euros.

long-term and long-distance migrants, those who eventually return have lower earnings in their second location than those who do not return.

Other labor market characteristics reveal expected patterns, as stayers are attached to more stable jobs (fixed contracts) and, hence, have accumulated more labor market experience and tenure in the firm. They also have experienced fewer unemployment spells in their lives and are more likely to be self-employed.

In my sample I observe working lives that start in the city of first employment. Although individuals may sort into a few big cities for post-secondary education, this pattern is very unlikely in a country like Spain where mobility of students is extremely low. Until 2000, legal restrictions induced individuals to pursue tertiary education in the same region or *Comunidad Autónoma* where they finished high-school. These restrictions were removed by 2003 allowing an open district admission where students could apply to any post-secondary institution in the country. Despite this reform, in 2009, only 12% and 23% of students migrated to another region and province for college education, respectively. Even more, these figures were 7% and 20% in 2001, a more relevant year given the average age of individuals in the study (CRUE, 2002–2010).

5. Results

Selection in initial migration

I begin by studying the determinants of first migration episodes and, in particular, whether migrants are positively selected in terms of productive characteristics at the time of their first move relative to stayers in the same city.[19] In table 1.2 I estimate the probability of outmigration from the individual's first job location using a single-exit discrete duration model as in equation 1.12, where the dependent variable takes value 1 only in the last monthly observation prior to migration. I focus on long-term long-distance moves, i.e., only those moves that exceed 12 months in the city of destination and 120 km. of distance. The determinants of shorter moves are quite different and discussed in appendix 1.C.

Results show migrants are more educated and productive than comparable stayers in their first city. In column (1) I include observable skills, in particular educational attainment and occupational skills. The reported coefficients are odd ratios. Having tertiary education increases the probability of outmigration by 122% relative to having at most primary education, while working in an occupation that requires medium to very-high skills increases the probability by more than 16% relative to working in an occupation with very-low skills.

Other labor market variables reveal expected signs. An additional year of labor market experience or tenure in the firm decreases the probability of outmigrating, conditional on age. Workers in the city of first location who have accumulated less experience and tenure will tend to have lower attachment to their city and their current job. Similarly, those under a temporary contract (with significantly lower job protection) have less to lose from quitting their jobs and thus are over 45% more likely to migrate. Likewise, self-employed workers and public sector employees appear to be more rooted to their city of first employment or may benefit from job amenities that make them less inclined to migrate.

Individuals who are unemployed are also more prone to migrate. However, by controlling separately for unemployeds who have completed their period of entitlement to unemployment benefits, I find they are the ones driving this effect. In fact, workers who are unemployed but receiving unemployment benefits are not any more likely to migrate than those who are employed. Once their unemployment benefits expire, however, the probability of migrating jumps by a factor of nine.[20] The discouraging effect of unemployment benefits for mobility has been previously

[19] In fact, the estimation compares migrants not only with stayers but also with themselves prior to the move, which helps identify the importance of individual characteristics that change over time. Thus, we are trying to explain not only who migrates but also when they migrate.

[20] I identify unemployeds with expired benefits as those unemployeds receiving benefits or subsidies who cease any relationship with the social security immediately after an unemployment spell (as opposed to starting a new job, which is the most common transition for them). When I do not distinguish between both types of unemployment status, I find that the overall effect of unemployment raises the odds of migrating by 79%.

Table 1.2: Logit estimation of determinants of first migration

	Dep. variable: long-term long-distance migration			
	(1)	(2)	(3)	(4)
Log mean earnings		1.678 (0.144)***		1.342 (0.068)***
Richest earnings tercile			1.429 (0.091)***	
Poorest earnings tercile			0.955 (0.026)*	
Tertiary education	2.223 (0.390)***			2.115 (0.367)***
Secondary education	1.506 (0.110)***			1.468 (0.106)***
Very-high skills	1.190 (0.061)***			1.016 (0.049)
High skills	1.165 (0.049)***			1.043 (0.041)
Medium skills	1.280 (0.078)***			1.212 (0.068)***
Low skills	1.031 (0.032)			1.014 (0.030)
Years of experience	0.905 (0.007)***	0.874 (0.004)***	0.875 (0.004)***	0.901 (0.007)***
Years of firm tenure	0.915 (0.011)***	0.920 (0.010)***	0.921 (0.010)***	0.911 (0.012)***
Self-employed	0.456 (0.039)***	0.540 (0.051)***	0.483 (0.040)***	0.517 (0.048)***
Public sector employee	0.757 (0.059)***	0.778 (0.065)***	0.802 (0.072)**	0.720 (0.053)***
Temporary contract	1.487 (0.036)***	1.480 (0.034)***	1.465 (0.036)***	1.516 (0.037)***
Unemployed	1.098 (0.080)	0.771 (0.049)***	0.767 (0.047)***	1.058 (0.076)
Unemployed $\times$ expired benefits	9.681 (0.397)***	10.314 (0.436)***	10.258 (0.426)***	9.818 (0.397)***
Urban area $\times$ period indicators	Yes	Yes	Yes	Yes
Age indicators	Yes	Yes	Yes	Yes
Pseudo R^2	0.073	0.070	0.070	0.074

Notes: Odd ratios (exponentiated coefficients) are reported on a sample of 16,439,937 monthly observations and 215,902 individuals. Standard errors in parentheses are clustered at the urban area level. ***, **, and * indicate significance at the 1, 5, and 10 percent levels. Sample is all individuals who are still in their first city. The reference category is stayers. Long-term long-distance migrations are moves that exceed 12 months in destination and distance of 120 km. All specifications include month indicator variables. Period is a ten-year interval. Log mean earnings are 6-month moving averages, excluding current earnings. Earnings terciles are constructed for all year-month pairs. Primary education and very-low skills are the omitted categories.

noted for Spain by Antolín and Bover (1997) (who proxy for this by looking at registration in Spain's Public Employment Office, INEM) and also for the United States by Goss and Paul (1990). The staggering magnitude of the effect I find indicates that the current design of unemployment insurance in Spain has a detrimental impact on the efficient matching of unemployeds and vacancies across local labor markets.

In all specifications I include indicator variables for age as a way to capture duration dependence in the first city in an additive and flexible way. I also add indicator variables for urban areas interacted with 10-year periods to confine the analysis of migrants and stayers within an urban area and time period.[21] In addition, this allows me to control for unobserved location characteristics that may affect the probability of migration for all individuals in a city.

The above results show that workers with greater observable skills are more likely to migrate. To check whether more productive workers, more broadly defined, are also more likely to migrate, I next proxy the productivity of each worker by their relative position in the local earnings distribution of their city of first employment. Column (2) in table 1.2 repeats the estimation of column (1) but instead of observable measures of skills (educational attainment and occupational skills) it uses average log earnings in the preceding six months to proxy for workers observable and unobservable skills.[22] The inclusion of urban area – period indicators implies that this variable measures the worker's relative position in the local earnings distribution. The corresponding coefficient show that a 10% increase in log mean monthly earnings raises the probability of outmigration by 2.5%.[23] Column (3) looks at this again by splitting the local earnings distribution into terciles. Being in the richest local earnings tercile raises the probability of outmigrating by 43%, while being in the lowest one decreases it by 4%.

I bring in both observable skills and earnings in column (4). Even within given levels of education and occupational skills, higher levels of earnings in the city of first employment increase the probability of migrating. However, the effect of earnings has been reduced by half relative to column (2) — recall coefficients are odd ratios — whereas observable skills like education remain almost as strong determinants of first-time migration as in column (1). Thus, differences in observable skills are essential to characterize the selection of first-time migrants while selection on unobservables, though present in the data, is of smaller quantitative importance.[24]

So far, I have been looking at the probability of migrating in general. However, the selection found in the data is really driven by the particular type of migration modeled in the conceptual framework: from small to big cities. In table 1.3 I estimate the conditional hazard rate of moving to one of the six biggest cities in Spain, i.e., the probability of moving to one of these big cities among those who have not moved before and do not move to small cities. For migrants who move within the six biggest cities I focus only on moves that involve an increment in city size. Other than this, all specifications are identical to those in table 1.2. The results show that the positive selection of

[21] I could further narrow the periods to five years, however, many small cities have very few or no migrants for such short intervals, especially in the early years of the sample.

[22] I construct 6-month moving averages of log employment earnings (excluding current earnings) to lessen the role of temporary fluctuations and to minimize the possible impact of an Ashenfelter (1978)-style dip — a drop in earnings immediately prior to migration. I use only those months of the six most recent where the worker has been employed. This is because productivity is best captured with a measure of earnings that excludes unemployment benefits. Alternatively, I have constructed a log 6-month moving average of income including these benefits. As expected, point estimates are lower, but only marginally and still significant. Results available upon request.

[23] Reported odd ratios are changes in the relative probability of outmigrating when the explanatory variable increases by the value of one. Since earnings are expressed in logs, this implies that when earnings are 2.72 (e) times larger (log earnings 1 unit larger), the probability of migration increases by 67.8%. The 2.5% reported in the text is calculated as $10\% \times (1.678 - 1)/e$.

[24] Individuals working in occupations demanding high- to very-high skills are no longer more inclined to migrate, once I include earnings. I have estimated several alternative specifications for table 1.2 by including as regressors the number of previous short-term migration episodes, their duration, or the number of previous short-term moves to the city of destination. In addition, I have excluded from the sample those migrants who register more than two long-term long-distance moves. Results vary only slightly and are available upon request.

migrants is much stronger when I look only at those who migrate to big cities. In general, the effects of differences in education and pre-move earnings are now two to three times larger than the effects on the probability of migrating in general. In column (1), we can see that long-term long-distance migration to big cities is almost three times more likely to occur for individuals with tertiary education. Having secondary education also raises substantially the odds, making migration to big cities more than twice as likely as for workers with at most primary education. The coefficient on log earnings in column (2) implies that a 10% increase in log mean monthly earnings raises the probability of outmigration by 4.9%. Column (4) shows that, even within given levels of education and occupational skills, more productive workers are more likely to migrate to big cities. In contrast, if I repeat the estimation for migration to small (as opposed to big) cities, I find no significant selection of any type (see appendix 1.B for details).

At the time of first migration, the group of long-term and long-distance migrants is made of permanent migrants —those who never return to their city of first employment — and return migrants — those who eventually return. In table 1.1 the raw data pointed out that permanent migrants have higher earnings and are more educated than returnees. In table 1.4 I narrow down this comparison by examining how productive characteristics differ between these two groups in their first city at the time of migration. Moreover, I investigate whether permanent or return migrants who move to big cities are more skilled or productive than those who move elsewhere. I run pooled OLS regressions where the dependent variable is 6-month moving average of log monthly earnings (excluding current earnings). All specifications include age and urban area interacted with 10-year-period indicator variables. Therefore, I capture the correlation between being a permanent migrant or returnee and earnings in the first city, controlling for other characteristics associated to earnings. I treat observations beyond the migration event as censored.

Overall, permanent and return migrants have higher pre-move earnings than stayers, after controlling for labor market characteristics. In column (1) I divide both migrant categories into those who move to the six biggest cities and those who move elsewhere. Again, for migrants who move within the six biggest cities I keep only those moves that involve an increment in city size. Both permanent and return migrants who move to the biggest cities have higher pre-move earnings than other migrants and stayers. At the time of first migration, earnings of these permanent migrants and eventual returnees are 11% and 4% higher than those of stayers, respectively. Therefore, I confirm the skill ranking found in the raw data (table 1.1) where permanent migrants are the most productive, followed by return migrants, and then stayers. However, both earnings gaps nearly vanish when I include observable skills as controls in column (2). Now, permanent migrants earn only an extra 3% while eventual returnees exhibit similar pre-move earnings as stayers with comparable education and occupational skills. This result verifies that sorting in initial migration of the most productive workers into bigger cities can be be mostly accounted by observable skills.

In columns (3) and (4) I repeat the exercise of columns (1) and (2), but further classify migrants to big cities into those who move to the two biggest cities (Madrid and Barcelona) and those who move to the 3rd - 6th biggest cities (Valencia, Sevilla, Bilbao and Zaragoza). The idea is to investigate whether the degree of sorting in initial migration increases with the size of the city of

Table 1.3: Logit estimation of determinants of first migration to big cities

	Dep. variable: long-term long-distance migration to any of 6 biggest cities			
	(1)	(2)	(3)	(4)
Log mean earnings		2.339 (0.158)***		1.563 (0.106)***
Richest earnings tercile			1.741 (0.081)***	
Poorest earnings tercile			0.979 (0.038)	
Tertiary education	4.229 (0.199)***			3.945 (0.191)***
Secondary education	2.224 (0.098)***			2.143 (0.096)***
Very-high skills	1.311 (0.130)***			1.053 (0.112)
High skills	1.280 (0.119)***			1.094 (0.099)
Medium skills	1.637 (0.097)***			1.513 (0.091)***
Low skills	1.073 (0.068)			1.043 (0.066)
Years of experience	0.924 (0.006)***	0.867 (0.006)***	0.868 (0.006)***	0.918 (0.006)***
Years of firm tenure	0.888 (0.006)***	0.901 (0.006)***	0.906 (0.006)***	0.882 (0.006)***
Self-employed	0.630 (0.068)***	0.784 (0.076)**	0.630 (0.068)***	0.758 (0.077)***
Public sector employee	0.458 (0.088)***	0.484 (0.091)***	0.528 (0.096)***	0.421 (0.080)***
Temporary contract	1.484 (0.053)***	1.448 (0.050)***	1.434 (0.051)***	1.521 (0.054)***
Unemployed	1.350 (0.133)***	0.692 (0.063)***	0.689 (0.064)***	1.281 (0.130)**
Unemployed $\times$ expired benefits	9.297 (0.514)***	10.376 (0.583)***	10.260 (0.570)***	9.481 (0.521)***
Urban area $\times$ period indicators	Yes	Yes	Yes	Yes
Age indicators	Yes	Yes	Yes	Yes
Pseudo R^2	0.093	0.082	0.080	0.094

Notes: Odd ratios (exponentiated coefficients) are reported on a sample of 12,625,095 monthly observations and 178,604 individuals. Standard errors in parentheses are clustered at the urban area level. ***, **, and * indicate significance at the 1, 5, and 10 percent levels. Sample is all individuals who are still in their first city. The reference category is stayers. Long-term long-distance migrations are moves that exceed 12 months in destination and distance of 120 km. Dependent variable takes value 1 if destination is one of six biggest cities *and* migrants experience an increment in city size. All specifications include month indicator variables. Period is a ten-year interval. Log mean earnings are 6-month moving averages, excluding current earnings. Earnings terciles are constructed for all year-month pairs. Primary education and very-low skills are the omitted categories.

destination, e.g., whether among workers who leave Granada those who move to Barcelona are more productive than those who move to Bilbao. I do not find evidence in favor of this argument. Both permanent and return migrants moving to the top 2 and 3rd - 6th biggest cities exhibit similar pre-move earnings. Moreover, when I add in observable skills in column (4), point estimates of earnings of migrants to the top 2 cities are in fact slightly lower than those of migrants to the 3rd -

Table 1.4: Earnings of first-time migrants relative to stayers by city of destination

	Dep. variable: log mean earnings in first city			
	(1)	(2)	(3)	(4)
Permanent migrant to 6 biggest cities	0.112 (0.010)***	0.029 (0.010)***		
Permanent to 1st - 2nd biggest cities			0.116 (0.011)***	0.025 (0.011)**
Permanent to 3rd - 6th biggest cities			0.103 (0.017)***	0.042 (0.016)***
Permanent to other cities	0.048 (0.010)***	0.016 (0.005)***	0.048 (0.010)***	0.016 (0.005)***
Return migrant to 6 biggest cities	0.043 (0.011)***	0.001 (0.009)		
Return to 1st - 2nd biggest cities			0.042 (0.013)***	-0.004 (0.011)
Return to 3rd - 6th biggest cities			0.047 (0.020)**	0.020 (0.017)
Return to other cities	0.015 (0.017)	-0.009 (0.009)	0.015 (0.017)	-0.009 (0.009)
Years of experience	-0.003 (0.003)	0.012 (0.001)***	-0.003 (0.003)	0.012 (0.001)***
Years of firm tenure	0.011 (0.001)***	0.008 (0.000)***	0.011 (0.001)***	0.008 (0.000)***
Self-employed	-0.672 (0.037)***	-0.481 (0.023)***	-0.672 (0.037)***	-0.481 (0.023)***
Public sector employee	0.278 (0.035)***	0.142 (0.022)***	0.278 (0.035)***	0.142 (0.022)***
Unemployed	-0.161 (0.029)***	0.119 (0.011)***	-0.161 (0.029)***	0.119 (0.011)***
Temporary contract	-0.106 (0.017)***	-0.058 (0.008)***	-0.106 (0.017)***	-0.058 (0.008)***
Tertiary education		0.180 (0.019)***		0.180 (0.019)***
Secondary education		0.098 (0.006)***		0.098 (0.006)***
Very-high skills		0.619 (0.022)***		0.619 (0.022)***
High skills		0.459 (0.012)***		0.459 (0.012)***
Medium skills		0.235 (0.006)***		0.235 (0.006)***
Low skills		0.082 (0.007)***		0.082 (0.007)***
Urban area $\times$ period indicators	Yes	Yes	Yes	Yes
Age indicators	Yes	Yes	Yes	Yes
R^2	0.410	0.547	0.410	0.547

Notes: Coefficients reported on a sample of 16,451,300 monthly observations and 216,083 individuals. Standard errors in parentheses are clustered at the urban area level. ***, **, and * indicate significance at the 1, 5, and 10 percent levels. Sample is all individuals who are still in their first city. Dependent variable is 6-month moving average of earnings, excluding current observation. Migrations are moves that exceed 12 months in destination and distance of 120 km. For migrants who move within the six biggest cities, only those moves that involve an increment in city size are considered. All specifications include month indicator variables. Period is a ten-year interval.

6th biggest cities.[25]

[25] I have estimated alternative specifications for columns (3) and (4) like one including increments in city size between the city of origin and destination and its square. None of these specifications indicate that the extent of sorting in initial migration increases with the size of the destination.

Selection in return migration

The conceptual framework of section 2 pointed to a possible second round of sorting after a first migration episode. Some migrants stay in the city to which they have relocated, others return to their city of first employment, and yet others may move on to a third city. The framework suggests that the decision to undertake that second migration or not depends both on individual characteristics that would be observable prior to the first move (such as initial education, occupational skills and relative earnings) and on the extent to which the individual had benefitted from migration.

In table 1.5 I estimate a multiple-exit discrete duration model where the sample is all first-time migrants who are already in the city of destination. As before, I examine one-way transition events. Now, the dependent variable takes value 1 in the last month prior to second migration only if migration is a return move, and it takes value 2 in the last month prior to second migration only if migration is a move to another city.

I first consider all migrants, whatever the size of the city they relocated to. Then, I focus only on migrants who moved to the six biggest cities. All specifications include categories of years elapsed since migration as a way to capture duration dependence in an additive and flexible way. Also, I include as controls labor market characteristics in the city of first employment by computing averages over pre-migration spells (e.g. percent of time spent unemployed, self-employed, in the public sector or under a temporary contract). Given the smaller sample size, instead of including indicator variables for urban areas interacted with 10-year periods, I add urban area indicators for both the first city and the city of destination and year indicators. Therefore, ideally I examine how heterogeneous experiences of migrants who moved to the same destination from the same origin affect the decision to migrate for a second time, controlling for several determinants of this second migration.

Results show, once again, that selection is driven by migrants who initially moved to big cities. In column (3a) a 10% increase in earnings at the first location (i.e., prior to the first migration episode) makes return migration 0.6% less likely[26]. However, even after controlling for initial earnings, earnings at the second location have additional explanatory power. A 10% increase in earnings in the second location (i.e., after the first migration episode) makes return migration 1.4% less likely, after controlling for average earnings in the first city and labor market characteristics in both cities. In column (4a) I consider education and occupational skills as well as earnings in both cities. As for the case of initial migration, observable skills are strong determinants of return migration. Having tertiary education reduces the probability of returning by 27% while having very-high occupational skills decreases the odds of returning by 25%. However, even within education categories and occupational skills, higher realized earnings in big cities make return migration less likely, though the magnitude of the effect attenuates.

The pattern of low realized earnings as a crucial driver of return migration seems to be specific to returnees from big cities. When I look at all returnees who initially moved to any city (not necessarily one of the six biggest), realized earnings in destination do not influence their return decision. In addition, when I examine the migration decision for repeat migrants who do not

[26] $10\% \times (0.831 - 1)/e = -0.6\%$

Table 1.5: Multinomial logit estimation of determinants of second migration

	First move to city of any size				First move to any of 6 biggest cities			
	Return (1a)	Move on (1b)	Return (2a)	Move on (2b)	Return (3a)	Move on (3b)	Return (4a)	Move on (4b)
Log mean earnings $^{\text{2nd loc.}}$	0.885 (0.088)	1.428 (0.084)***	0.989 (0.068)	1.372 (0.084)***	0.623 (0.030)***	1.124 (0.094)	0.732 (0.047)***	1.115 (0.098)
Log mean earnings $^{\text{1st loc.}}$	0.824 (0.036)***	1.110 (0.062)*	0.886 (0.044)**	1.070 (0.066)	0.831 (0.052)***	1.130 (0.110)	0.915 (0.059)	1.124 (0.115)
Self-employed $^{\text{2nd location}}$	0.365 (0.077)***	0.482 (0.070)***	0.391 (0.075)***	0.520 (0.083)***	0.170 (0.045)***	0.385 (0.113)***	0.204 (0.058)***	0.477 (0.171)**
Public sector $^{\text{2nd location}}$	0.900 (0.102)	0.755 (0.081)***	0.973 (0.126)	0.754 (0.083)***	0.841 (0.154)	0.797 (0.190)	0.916 (0.165)	0.805 (0.195)
Temporary contract $^{\text{2nd loc.}}$	1.299 (0.049)***	1.300 (0.079)***	1.282 (0.043)***	1.341 (0.082)***	1.235 (0.073)***	1.246 (0.094)***	1.224 (0.071)***	1.274 (0.097)***
Unemployed $^{\text{2nd location}}$	2.202 (0.209)***	1.446 (0.134)***	2.031 (0.245)***	1.702 (0.212)***	1.343 (0.216)*	1.428 (0.244)**	1.244 (0.224)	1.784 (0.409)**
Unemployed $^{\text{exp.ben. 2nd loc.}}$	6.307 (0.790)***	9.713 (1.097)***	6.368 (0.804)***	9.660 (1.084)***	8.319 (2.079)***	7.573 (2.176)***	8.424 (2.104)***	7.568 (2.177)***
Years of experience	1.030 (0.006)***	0.963 (0.007)***	1.018 (0.005)***	0.969 (0.008)***	1.034 (0.009)***	0.959 (0.014)***	1.014 (0.009)	0.960 (0.014)***
Years of firm tenure	1.041 (0.011)***	1.010 (0.013)	1.046 (0.011)***	1.006 (0.012)	1.074 (0.014)***	1.030 (0.017)*	1.077 (0.014)***	1.028 (0.016)*
Tertiary education			0.879 (0.097)	1.247 (0.068)***			0.725 (0.061)***	1.057 (0.113)
Secondary education			0.830 (0.038)***	1.114 (0.066)*			0.831 (0.047)***	1.017 (0.108)
Very-high skills			0.761 (0.083)**	0.978 (0.120)			0.746 (0.114)*	1.125 (0.235)
High skills			0.879 (0.069)	1.203 (0.121)*			0.903 (0.115)	1.269 (0.226)
Medium skills			1.042 (0.072)	1.136 (0.106)			1.185 (0.119)*	1.358 (0.242)*
Low skills			1.051 (0.045)	1.103 (0.095)			1.120 (0.093)	1.243 (0.197)
Urban area indicators $^{\text{1st loc.}}$	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Urban area indicators $^{\text{2nd loc.}}$	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Labor market controls $^{\text{1st loc.}}$	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Years since migration ind.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age categories	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year indicators	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,058,367		1,058,367		376,376		376,376	
Migrants	19,130		19,130		6,938		6,938	
Pseudo R^2	0.065		0.065		0.058		0.059	

Notes: Relative risk ratios (exponentiated coefficients) reported with standard errors in parentheses clustered at the urban area level of first location. ***, **, and * indicate significance at the 1, 5, and 10 percent levels. The reference category is permanent migrants who remain in the city of destination. In columns (1) – (4), sample is migrants after their first move. In columns (5) – (8), sample is migrants after their first move to one of the six biggest cities, where only increments in city size are considered. All specifications include month indicator variables. The omitted categories are primary education and very-low skills. Labor market controls in first location are averages over pre-migration spells. Log mean earnings in second location are 6-month moving averages, excluding current earnings.

return but move on to a third city, this is not affected by realized earnings (column 4b). In sum, returnees from big cities can be characterized as those individuals with initial skills in between those of stayers and those of permanent migrants, that also are not successful in boosting their earnings after migrating.

6. Conclusions

This paper examines selection in initial and in return urban migration. For initial migration, there is a clear selection by observable characteristics. Both higher educational attainment and higher occupational skills increase substantially the probability of migrating. Earlier studies of internal migration also find that migrants are more skilled and educated than stayers (Borjas *et al.*, 1992, Hunt, 2004). By looking at the relative position of migrants in the pre-migration local labor market earnings distribution, I am also able to proxy for individual productivity more broadly. More productive workers are more likely to migrate. This remains so even when looking within given levels of education and occupational skills.

Such selection is largely driven by the group of migrants who moves from small to big cities. The effects of differences in education, occupational skills, or relative earnings on the probability of migrating to big cities are two to three times larger than the effects on the probability of migrating in general. Regarding the role of observables relative to unobservables, I find that the substantial difference in pre-migration earnings between migrants and non-migrants is mostly (but not totally) accounted for by differences in observable characteristics, such as education and occupational skills. Moreover, the marginal effect of relative earnings on the probability of migrating is about half as large once I control for education and occupational skills. This suggests that selection on unobservables, while present in the data, is of smaller quantitative importance.

In addition to selection in initial migration, I also document a second stage of sorting that takes place after a first migration episode. Around one half of migrants move for a second-time within five years of arriving in their city of destination and 70% of these moves involve a return migration. Return moves are more frequent and happen sooner in big cities. I find that returnees from big cities tend to exhibit skills in between those of stayers and those of permanent migrants. They are also typically those who have been least successful in boosting their earnings after migrating to a big city. This pattern seems to be specific to them as opposed to other repeat migrants. When I examine second-time moves of migrants to other cities, they are not affected by realized earnings after their first migration episode.

All of this indicates that positive sorting in cities through migration is important, but that differences in observable characteristics account for much of the observed differences. At the same time, for workers who have already migrated, further sorting is driven not just by workers' initial productivity but also by improvements in that productivity. While I document in detail how initial and return migration contributes to the sorting of more skilled workers into bigger cities, it is worth noting that worker sorting can occur through other channels (Combes *et al.*, 2012). For instance, both faster learning associated with working in big cities (Baum-Snow and Pavan, 2012*b*, De la Roca and Puga, 2012) and better schools in large urban areas can widen the skill gap.

Appendix 1.A. Completing earnings data

For the period 2004–2010, uncensored earnings data are available from matched income tax returns for all workers in the MCVL except those in the Basque Country and Navarre (where income taxes are not collected by the Central Government) and self-employed workers. In addition, for the entire period 1981–2010 earnings data are available from the social security for all workers in the MCVL, including those in the Basque Country and Navarre and self-employeds, but these are capped for some workers. In particular, 10.2% and 3.2% of monthly earnings observations are top- and bottom-coded, respectively. This appendix explains how I estimate earnings for this 13.4% of observations that are capped.

I construct three estimates of earnings using Tobit estimations with varying sets of explanatory regressors and specifications of the error term. The structure of these earnings variables can be briefly summarized as follows:

- *Earnings 1*: individual and job characteristics + i.i.d shock,
- *Earnings 2*: individual and job characteristics + location characteristics + i.i.d shock,
- *Earnings 3*: individual and job characteristics + location characteristics + persistence in shock.

To better take advantage of information on the persistence in earnings over time, for these estimations of capped earnings I enlarge my sample to include all males who were born between 1916 and 1992 and aged 18 to 65 throughout 1981–2010. I restrict observations following the same criteria used to construct the final sample in the study, except for the age restriction. I also exclude self-employed workers because the incidence of censoring is negligible for them. This broader sample contains 493,904 individuals and 85,507,464 monthly observations.

Censoring bounds vary by type of occupation on an annual basis. I use historical information from Spain's *Boletín Oficial del Estado* and plot monthly earnings densities to identify them.

Next, I run 300 occupation-year Tobit regressions (10 occupations $\times$ 30 years). I use as dependent variable log daily earnings expressed in December 2010 euros, adjusted by number of hours worked in the case of part-time jobs. For the Earnings 1 specification, I include as explanatory variables quartics in experience and job tenure, and sets of indicator variables for age, level of education, 3-digit NACE sector, type of contract, public employee status, and month. I also add interactions among many of these variables and level of education. For the Earnings 2 and Earnings 3 specifications, I also include indicator variables for province and urban area of the workplace.

Using the coefficients of these Tobit regressions, I can predict the value of earnings *only* for capped observations under the Earnings 1 and Earnings 2 specifications as follows:

$$\hat{W}_{ijt} = x_{ijt}{}'\hat{\gamma} + z_{ijt}{}'\hat{\theta} + \hat{\sigma}\,\varepsilon_{ijt}, \tag{1.13}$$

where $\hat{W}_{ijt}$ is the value of predicted log earnings for individual i in occupation j at year t, x_{ijt} is a vector of individual and job characteristics, z_{ijt} is a vector of workplace location indicators (not used in the Earnings 1 specification), $\hat{\gamma}$, $\hat{\theta}$ and $\hat{\sigma}$ are estimated parameters, and ε_{ijt} is an i.i.d shock. Finally, since I know whether monthly earnings are originally top- or bottom-coded, I force predicted earnings to be above or below the corresponding bound, respectively.[27]

Persistence in earnings

Both the Earnings 1 and the Earnings 2 specifications include only transitory shocks in income. Yet, earnings exhibit persistent correlation as suggested by studies that use career-long earnings histories. Thus, in the Earnings 3 specification I exploit the panel dimension of the MCVL and introduce persistence in the error term.

To predict earnings in this final specification, I follow the methodology proposed by Haider and Solon (2006). The key assumption is that the joint distribution of uncensored log earnings for an individual is multivariate normal. Therefore, the joint distribution of annual earnings throughout the period 1981–2010 can be fully characterized by the mean and variance of log earnings in each year and the cross-year autocorrelations of log earnings for every pair of years.[28]

The mean and variance of log earnings for each occupation-year pair are estimated in equation (1.13). Haider and Solon (2006) estimate autocorrelations between pairs of years using a bivariate Tobit maximum-likelihood estimator. Instead, I follow a simpler approach based on indirect inference to compute cross-year autocorrelations.

The intuition behind this approach is quite straightforward and can be described in four steps. First, I estimate a regression coefficient for every pair of (standardized) log earnings of the same worker i in two different years. This estimation is carried out only for uncensored earnings observations. I label this regression coefficient $\hat{\lambda}^*$.

Second, I exploit the multivariate normality assumption to generate earnings for worker i in year $t+h$ conditional on his observed earnings in year t, the applicable censoring bounds and a value for the correlation coefficient ρ. Equation (1.14) shows how to generate earnings in year $t+h$ based on a bivariate normal distribution of earnings in years t and $t+h$:

$$\tilde{W}_{i,t+h}\,,\ \tilde{W}_{it} \sim N\left(\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}1 & \rho\\ \rho & 1\end{pmatrix}\right), \tag{1.14}$$

$$\mathbb{E}(\tilde{W}_{i,t+h}\,|\,\tilde{W}_{it},\,\tilde{a}_{t+h}\leqslant\tilde{W}_{i,t+h}\leqslant\tilde{b}_{t+h})=\rho\tilde{W}_{it}+\sqrt{1-\rho^2}\left[\frac{\phi\left(\frac{\tilde{a}_{t+h}-\rho\tilde{W}_{it}}{\sqrt{1-\rho^2}}\right)-\phi\left(\frac{\tilde{b}_{t+h}-\rho\tilde{W}_{it}}{\sqrt{1-\rho^2}}\right)}{\Phi\left(\frac{\tilde{b}_{t+h}-\rho\tilde{W}_{it}}{\sqrt{1-\rho^2}}\right)-\Phi\left(\frac{\tilde{a}_{t+h}-\rho\tilde{W}_{it}}{\sqrt{1-\rho^2}}\right)}\right],$$

where $\tilde{W}_{it}$ are the standardized uncensored log earnings for individual i in year t, ρ is the correlation coefficient, and $\tilde{a}_{t+h}$ and $\tilde{b}_{t+h}$ are the standardized lower and upper bounds applicable in year $t+h$, respectively. Since the only unknown in $\mathbb{E}(\tilde{W}_{i,t+h})$ is the value of ρ, the cross-year

[27]This implies drawing i.i.d shocks from a truncated normal distribution $\varepsilon_{ijt} > (b_{ijt} - x_{ijt}{}'\hat{\gamma} - z_{ijt}{}'\hat{\theta})/\hat{\sigma}$ if earnings are top-coded and $\varepsilon_{ijt} < (a_{ijt} - x_{ijt}{}'\hat{\gamma} - z_{ijt}{}'\hat{\theta})/\hat{\sigma}$ if they are bottom-coded. b_{ijt} and a_{ijt} are earnings levels at which top and bottom censoring occur, respectively. In some exceptional cases, earnings remain capped no matter the size of the shock. After 5,000 iterations only 0.05% and 0.08% of monthly observations remain capped for the Earnings 1 and Earnings 2 specifications, respectively.

[28]Log daily earnings follow a multivariate normal distribution also within each of the ten occupation categories. For simplicity, I omit index j referring to type of occupation.

autocorrelation of interest, based on a grid of 40 values of ρ from 0 to 0.975 on 0.025 intervals, I generate $\mathbb{E}(\tilde{W}_{i,t+h} \,|\, _{\rho=\rho_k})$ where $k = 1, ..., 40$.

Next, I regress each generated $\mathbb{E}(\tilde{W}_{i,t+h} \,|\, _{\rho=\rho_k})$ on $\tilde{W}_{it}$ only for uncensored observations and obtain a regression coefficient $\hat{\lambda}_k$. The optimal value of the cross-year autocorrelation ρ_k^* will be the ρ_k which minimizes the absolute distance between $\hat{\lambda}^*$ and any $\hat{\lambda}_k$. Therefore, if log earnings for individual i follow a multivariate normal distribution, I choose the ρ_k^* that best replicates the observed correlation for uncensored earnings in the data. I construct a variance-covariance matrix (30×30) with the optimal ρ_k^* values calculated for every pair of years throughout 1981–2010.[29]

Finally, I proceed to predict the value of earnings *only* for capped observations under the Earnings 3 specification as follows:

$$\hat{W}_{ijt} = x_{ijt}{'}\hat{\gamma} + z_{ijt}{'}\hat{\theta} + \hat{\sigma} \cdot \hat{p}_{jt}{'}\xi_{ijt}, \tag{1.15}$$

where all variables are the same as in equation (1.13), but now $\hat{p}_{jt}$ is a row vector (1×30) of the Cholesky decomposition of the estimated variance-covariance matrix and ξ_{ijt} is a vector of random shocks.[30] Thus, the main difference between Earnings 3 specification and the previous two is that in the former shocks at $t - j$ are persistent and affect current earnings through the error term.[31]

Fit of estimated earnings

To verify the accuracy of this procedure, I check the predicted values of earnings in top- and bottom-coded observations in social security records against actual uncensored earnings in income tax returns for the same individual and month in those years where both are available (2004–2010). If the fit is satisfactory for 2004–2010, I can be more confident that predicted earnings do a good job in predicting capped earnings for 1981–2003.

The correlation between predicted and actual values for capped month-individual observations in 2004–2010 is high at 0.77 for the three earnings specifications. In addition to predicting with high accuracy individual values, I am also reproducing well the overall shape of the earnings distribution. This can be seen in table 1.6, which shows selected percentiles of the distributions of actual and predicted earnings for all workers and for skilled workers. Overall, the distributions are quite similar. Even for skilled workers, who are top-coded beyond the 59^{th} percentile, predicted earnings approximate salaries quite well in capped percentiles.

Table 1.7 displays estimated order of autocorrelations for salaries and earnings specifications. As expected, the Earnings 3 specification, which accounts for persistence in earnings, is the one

[29] Variance-covariance matrices for each type of occupation are available upon request. Earnings of skilled workers are much more persistent. The first three average estimated order of autocorrelations for an occupation demanding very-high skills are 0.90, 0.85 and 0.80. The same figures for an occupation with very-low skills are 0.75, 0.68 and 0.63.

[30] In four of ten occupation types, the Cholesky decomposition matrix (P) cannot be calculated because the element-by-element estimated variance-covariance matrix ($\hat{\Omega}$) is not positive semidefinite (as it should be). Haider and Solon (2006) face the same problem and impose a non-negativity constraint on the diagonal elements of $\tilde{P}$, where $\tilde{\Omega} = \tilde{P}\tilde{P}'$ and $\tilde{P}$ minimize the distance between $\hat{\Omega}$ and $\tilde{\Omega}$. I consider a less robust but faster solution. I diagonalize $\hat{\Omega}$ and replace negative eigenvalues with zeros –only 10 replacements needed out of 120 eigenvalues. The autocorrelations in this new variance-covariance matrix Ω^* are very similar to the ones in $\hat{\Omega}$. Only 1.2% and 6% of all new autocorrelations differ by more than 0.025 and 0.01 in absolute terms from the originally estimated autocorrelations, respectively.

[31] Because computation of the Earnings 3 specification is extremely time-demanding, I only run 100 iterations for the vector of random shocks. As a result, only 0.91% of earnings observations remain censored.

Table 1.6: Selected percentiles for actual and predicted earnings

	All workers				Skilled workers			
	Actual	*Earnings 1*	*Earnings 2*	*Earnings 3*	Actual	*Earnings 1*	*Earnings 2*	*Earnings 3*
Percentile 5	43.9	43.1	43.1	43.1	40.4	40.6	40.7	40.6
Percentile 10	50.5	50.4	50.4	50.4	48.0	48.4	48.5	48.4
Percentile 25	61.3	61.2	61.3	61.3	63.6	65.1	65.2	65.1
Percentile 50	80.6	80.7	80.8	80.8	84.2	86.2	86.3	86.2
Percentile 75	117.0	120.2	120.4	120.4	119.1	121.2	121.4	121.4
Percentile 90	167.5	170.6	170.6	169.8	166.1	167.7	167.5	167.7
Percentile 95	215.0	216.6	216.3	216.4	207.2	203.4	202.8	203.4

Notes: Monthly earnings expressed as a percentage of the average for the corresponding earnings category. Only individuals working in urban areas are included. Skilled individuals work in the top three out of ten social security occupations demanding high and very-high skills.

Table 1.7: Order of autocorrelations for actual and predicted earnings

Order	Actual	*Earnings 1*	*Earnings 2*	*Earnings 3*
1	0.894	0.869	0.871	0.892
2	0.846	0.827	0.829	0.847
3	0.808	0.794	0.796	0.812
4	0.775	0.769	0.771	0.786
5	0.753	0.750	0.752	0.765

that best matches persistence in salaries. Given this and the accurate fit in reproducing the shape of the earnings distribution, I use estimates from the Earnings 3 specification for the 13.4% of observations that have the value of earnings capped in the MCVL.

Appendix 1.B. Migration to small cities

In table 1.8 I estimate the conditional hazard rate of moving to a small city in Spain, defined as those cities with size below the median-sized city (Santiago de Compostela). Thus, I estimate the probability of moving to one of these small cities among those who have not moved before and do not move elsewhere. For migrants who move within these small cities I keep only those moves that involve a drop in city size. Other than this, all specifications are identical to those in tables 1.2 and 1.3.

The results show there is no evidence of selection of any type for migrants who move to small cities; if anything, and unexpectedly, being in the lowest tercile of the local earnings distribution decreases the probability of migrating to a small city by 19%. Based on the conceptual framework in section 2, this result could hold if individuals with very low earnings in non-small cities can not afford the moving cost. Although the estimated effects are in general non-significant, point estimates suggest that having secondary education or working in occupations demanding medium to very-high skills decrease the probability of moving to a small city. This mild negative selection

Table 1.8: Logit estimation of determinants of first migration to small cities

	Dep. variable: long-term long-distance migration to cities below median city size			
	(1)	(2)	(3)	(4)
Log mean earnings		1.036 (0.184)		1.039 (0.084)
Richest earnings tercile			0.985 (0.145)	
Poorest earnings tercile			0.813 (0.056)***	
Tertiary education	1.392 (0.408)			1.382 (0.399)
Secondary education	0.828 (0.136)			0.825 (0.133)
Very-high skills	0.816 (0.213)			0.799 (0.193)
High skills	0.871 (0.148)			0.858 (0.131)
Medium skills	0.840 (0.149)			0.834 (0.142)
Low skills	1.044 (0.089)			1.041 (0.086)
Years of experience	0.920 (0.012)***	0.916 (0.020)***	0.916 (0.019)***	0.919 (0.013)***
Years of firm tenure	0.913 (0.012)***	0.911 (0.012)***	0.910 (0.012)***	0.912 (0.013)***
Self-employed	0.449 (0.110)***	0.470 (0.111)***	0.532 (0.109)***	0.456 (0.121)***
Public sector employee	1.275 (0.258)	1.253 (0.294)	1.230 (0.289)	1.267 (0.261)
Temporary contract	1.477 (0.118)***	1.505 (0.106)***	1.510 (0.102)***	1.481 (0.121)***
Unemployed	1.226 (0.181)	1.221 (0.130)*	1.223 (0.129)*	1.221 (0.174)
Unemployed $\times$ expired benefits	8.902 (0.665)***	8.944 (0.752)***	9.008 (0.739)***	8.917 (0.657)***
Urban area $\times$ period indicators	Yes	Yes	Yes	Yes
Age indicators	Yes	Yes	Yes	Yes
Pseudo R^2	0.061	0.059	0.060	0.061

Notes: Odd ratios (exponentiated coefficients) are reported on a sample of 15,785,260 monthly observations and 206,033 individuals. Standard errors in parentheses are clustered at the urban area level. ***, **, and * indicate significance at the 1, 5, and 10 percent levels. Sample is all individuals who are still in their first city. The reference category is stayers. Long-term long-distance migrations are moves that exceed 12 months in destination and distance of 120 km. Dependent variable takes value 1 if destination is a city with size below the median-sized city (Santiago de Compostela) *and* migrants experience a decline in city size. All specifications include month indicator variables. Period is a ten-year interval. Log mean earnings are 6-month moving averages, excluding current earnings. Earnings terciles are constructed for all year-month pairs. Primary education and very-low skills are the omitted categories.

in observable skills is in sharp contrast with the strong positive selection of migrants who move to big cities (see table 1.3).

In tables 1.3 and 1.8 I only exploit moving decisions for some long-term long-distance migrants (20,453 individuals). In particular, in the former I use only those migrants who move to one of the six biggest cities and experience an increment in city size (7,217 individuals), while in the latter I use only those who move to cities below the median-sized city and experience a decline in city size

(2,539 individuals). An alternative estimation is to model the probability of migration by splitting it in two comprehensive alternatives: moving to a bigger city (10,726 individuals) or to a smaller one (9,727 individuals). This can be done using a multinomial logit specification or, again, modeling two separate conditional hazard rates.

Results confirm that migrants who move to bigger cities largely drive the positive selection of overall migrants in terms of earnings and observable skills. For these I find that a 10% increase in log mean monthly earnings raises the probability of migrating by 4.2% (odds ratio of 2.13 and standard error of 0.13) while for migrants who move to smaller cities the effect is lower at 1.4% (odds ratio of 1.37 and standard error of 0.10). The difference in earnings between both groups of migrants remains significant when I add in observable skills.

As expected, when I consider these comprehensive migration alternatives, the difference in the effects of observable skills and earnings between both groups of migrants are attenuated relative to the difference shown in the main text. This is not surprising since with comprehensive migration alternatives we count a move from Madrid to Barcelona as one to a smaller city. Likewise, any marginal increase in city size is considered a move to a bigger city. The estimated conditional hazard rates in the main text, i.e., moving to one of the six biggest cities or to a city below the median-sized city, are more suited to the environment of the conceptual framework.

Appendix 1.C. Short-term and short-distance migration

Following the estimations proposed in tables 1.2 and 1.4, where I examined selection of long-term long-distance migrants in terms of productive characteristics at the time of first migration, I now repeat these estimations including short-term and short-distance migrants. In table 1.9, I estimate a multiple-exit discrete duration model (instead of a conditional hazard rate model as in table 1.2) including short-term/distance migration on the one hand and long-term/distance migration on the other as alternative possibilities. The dependent variable takes value 1 if the first migration is a short-term/distance move and value 2 if it is a long-term/distance move. The table shows clearly that the determinants of short-term short-distance migration, which often does not require a permanent change in residence, are very different from the determinants of long-term long-distance migration. The estimates for long-term long-distance migrants are just as in the main text. Regarding short-term /distance migrants, they instead exhibit similar pre-move earnings as stayers. Thus, once I take into account the fact that short-term/distance migrants have accumulated less labor market experience and are in more unstable occupations (e.g., more likely to be under a temporary contract and unemployed), they are no longer less productive than stayers in their first city, as they appeared to be in the raw data. However, they do tend to be special on some dimensions (for instance, occupational skills and education now work in opposite directions, suggesting perhaps that overeducated workers in low-skill occupations are more likely to engage in short-term short-distance moves). In any case, these types of job changes for a very short period or in a nearby urban area are rather different, and best studied separately.

Table 1.9: Multinomial logit estimation of determinants of first migration

	Short-term short-distance (1a)	Long-term long-distance (1b)	Short-term short-distance (2a)	Long-term long-distance (2b)
Log mean earnings	1.031 (0.054)	1.678 (0.144)***	1.061 (0.038)*	1.343 (0.068)***
Tertiary education			1.368 (0.086)***	2.113 (0.367)***
Secondary education			1.063 (0.032)**	1.467 (0.105)***
Very-high skills			0.633 (0.050)***	1.017 (0.049)
High skills			0.693 (0.034)***	1.044 (0.041)
Medium skills			0.761 (0.026)***	1.213 (0.068)***
Low skills			0.852 (0.017)***	1.014 (0.030)
Years of experience	0.936 (0.005)***	0.874 (0.004)***	0.941 (0.004)***	0.901 (0.007)***
Years of firm tenure	0.861 (0.008)***	0.920 (0.010)***	0.860 (0.008)***	0.911 (0.012)***
Self-employed	0.406 (0.035)***	0.540 (0.051)***	0.345 (0.030)***	0.518 (0.048)***
Public sector employee	1.136 (0.186)	0.778 (0.065)***	1.160 (0.198)	0.720 (0.053)***
Temporary contract	1.905 (0.056)***	1.481 (0.034)***	1.869 (0.056)***	1.516 (0.037)***
Unemployed	1.550 (0.083)***	0.771 (0.049)***	1.387 (0.059)***	1.058 (0.076)
Unemployed $\times$ expired benefits	8.240 (0.343)***	10.317 (0.436)***	8.232 (0.323)***	9.815 (0.396)***
Urban area $\times$ period indicators	Yes	Yes	Yes	Yes
Age indicators	Yes	Yes	Yes	Yes
Observations	16,479,146		16,479,146	
Pseudo R^2	0.097		0.099	

Notes: Relative risk ratios (exponentiated coefficients) are reported with standard errors in parentheses clustered at the urban area level. ***, **, and * indicate significance at the 1, 5, and 10 percent levels. Sample is all individuals who are still in their first city. Dependent variable takes value 1 if migration is short-term and short-distance and value 2 if it is long-term and long-distance. Long-term long-distance migrations are moves that exceed 12 months in destination and distance of 120 km. All specifications include month indicator variables. Log mean earnings are 6-month moving averages, excluding current earnings. Primary education and very-low skills are the omitted categories.

Chapter 2

Learning by working in big cities

1. Introduction

Workers in bigger cities earn more than workers in smaller cities and rural areas. Figure 2.1 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 *et al.* (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 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 *et al.*, 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).

[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 *et al.*, 2012).

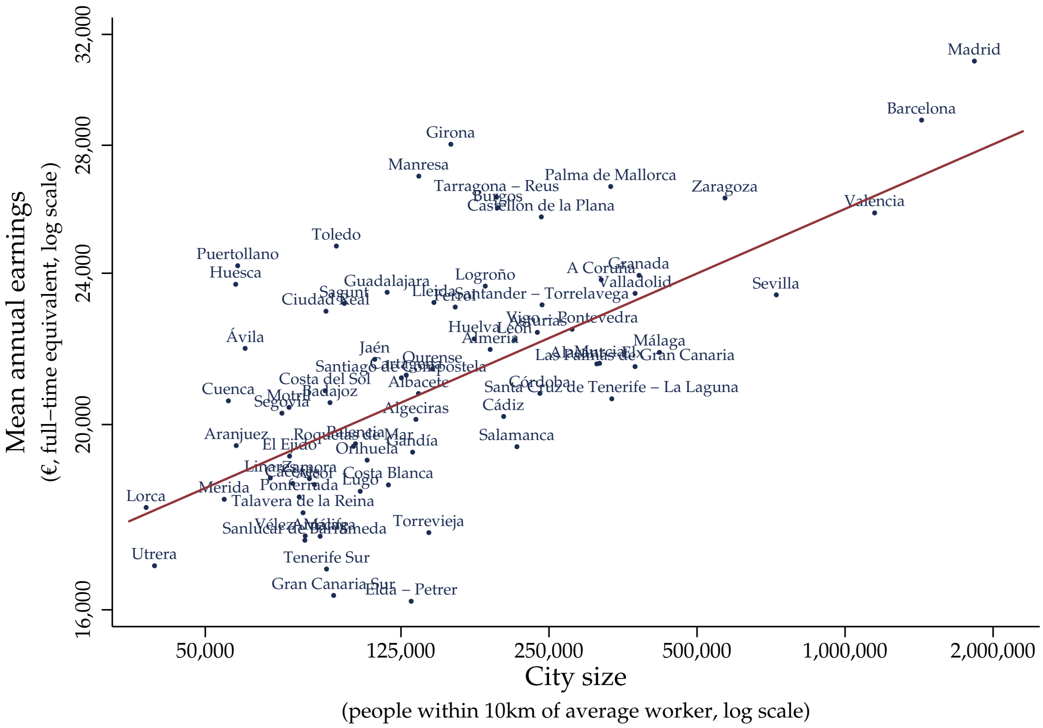


Figure 2.1: Mean earnings and city size

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 various other 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 *et al.*, 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 *et al.* (2010), Combes *et al.* (2012) and Baum-Snow and Pavan (2012*a*) emphasize.

2. Methodology

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

$$w_{ict} = \sigma_c + \mu_i + \sum_{j=1}^{C} \delta_{jc} e_{ijt} + \boldsymbol{x}'_{it}\boldsymbol{\beta} + \varepsilon_{ict} \ , \tag{2.1}$$

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, $\boldsymbol{x}_{it}$ is a vector of time-varying individual and job characteristics, the scalars δ_{jc} and the vector $\boldsymbol{\beta}$ are parameters, and ε_{ict} is an error term uncorrelated with the explanatory variables.[3]

Equation (2.1) 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

[2] This is consistent with the counterfactual simulations of the structural model in Baum-Snow and Pavan (2012*b*), 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*b*) 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.

is that they let workers accumulate more valuable experience, so equation (2.1) 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 (2.1), we ignore both unobserved worker heterogeneity and any dynamic benefits of working in bigger cities, and estimate the following relationship:

$$w_{ict} = \sigma_c + \boldsymbol{x}_{it}'\boldsymbol{\beta} + \eta_{ict} \ . \tag{2.2}$$

Compared with equation (2.1), in equation (2.2) the worker fixed-effect μ_i and the urban experience terms $\sum_{j=1}^{C}\delta_{jc}e_{ijt}$ are missing. Equation (2.2) can be estimated by ordinary least squares with a cross section of workers or a pooled panel.

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

$$\text{Cov}(\iota_{ict}, \eta_{ict}) = 0 \ , \tag{2.3}$$

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 (2.1) holds, the error term of equation (2.2) includes the omitted variables:

$$\eta_{ict} = \mu_i + \sum_{j=1}^{C}\delta_{jc}e_{ijt} + \varepsilon_{ict} \ . \tag{2.4}$$

Hence,

$$\text{Cov}(\iota_{ict}, \eta_{ict}) = \text{Cov}(\iota_{ict}, \mu_i) + \text{Cov}(\iota_{ict}, \sum_{j=1}^{C}\delta_{jc}e_{ijt}) \neq 0 \ . \tag{2.5}$$

Equation (2.5) 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]

[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 $\text{Cov}(\boldsymbol{x}_{it}, \mu_i + \sum_{j=1}^{C}\delta_{jc}e_{ijt}) = \boldsymbol{0}$, 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 $\boldsymbol{x}_{it}$ in equation (2.2). 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.

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 $\operatorname{plim} \hat{\sigma}_{\text{pooled}} = \sigma + \mu + \frac{n+1}{2}\delta$. 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 ($\frac{n+1}{2}\delta$).

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:

$$w_{ict} = \sigma_c + \mu_i + \boldsymbol{x}'_{it}\boldsymbol{\beta} + \zeta_{ict} \ . \tag{2.6}$$

Compared with equation (2.1), the city-specific experience terms $\sum_{j=1}^{C} \delta_{jc} e_{ijt}$ are missing from equation (2.6). Compared with equation (2.2), 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 (2.6) the time average for each worker:

$$(w_{ict} - \bar{w}_i) = \sum_{j=1}^{C} \sigma_c(\iota_{ict} - \bar{\iota}_{ic}) + (\boldsymbol{x}'_{it} - \bar{\boldsymbol{x}}'_i)\boldsymbol{\beta} + (\zeta_{ict} - \bar{\zeta}_i) \ . \tag{2.7}$$

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 $\operatorname{Cov}(\boldsymbol{x}_{it}, \sum_{j=1}^{C} \delta_{jc} e_{ijt}) = \mathbf{0}$, the resulting fixed-effects estimate of σ_c is unbiased if

$$\operatorname{Cov}\big((\iota_{ict} - \bar{\iota}_{ic}), (\zeta_{ict} - \bar{\zeta}_i)\big) = 0 \ . \tag{2.8}$$

However, if the richer wage determination of equation (2.1) holds,

$$(\zeta_{ict} - \bar{\zeta}_i) = \sum_{j=1}^{C} \delta_{jc}(e_{ijt} - \bar{e}_{ij}) + (\varepsilon_{ict} - \bar{\varepsilon}_i) \ , \tag{2.9}$$

and thus

$$\operatorname{Cov}\big((\iota_{ict} - \bar{\iota_{ic}}), (\zeta_{ict} - \bar{\zeta}_i)\big) = \operatorname{Cov}\Big((\iota_{ict} - \bar{\iota_{ic}}), \sum_{j=1}^{C} \delta_{jc}(e_{ijt} - \bar{e}_{ij})\Big) \neq 0 \ . \tag{2.10}$$

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 (2.6) 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 $\hat{\sigma}_{\text{FE}} = \sigma - \frac{m-1}{2}\theta\delta$. 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

[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.

of bigger cities are highly portable, is one that we cannot assess without actually estimating the fully-fledged specification of equation (2.1).

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 (2.2), than when estimated using a specification with worker fixed-effects, such as that of equation (2.6). 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 (2.1) are important, then the estimation of equation (2.6) affects not just one, but two sources of bias present in the estimation of equation (2.2). By including worker fixed-effects, equation (2.6) addresses the bias arising in equation (2.2) 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 (2.1), where the worker fixed-effect μ_i can be again eliminated by subtracting the time average for each worker:

$$(w_{ict} - \bar{w}_i) = \sum_{j=1}^{C} \sigma_c (\iota_{ict} - \bar{\iota}_{ic}) + \sum_{j=1}^{C} \delta_{jc} (e_{ijt} - \bar{e}_{ij}) + (\boldsymbol{x}'_{it} - \bar{\boldsymbol{x}}'_i)\boldsymbol{\beta} + (\varepsilon_{ict} - \bar{\varepsilon}_i) \ . \tag{2.11}$$

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 (2.2) and (2.6) for comparison with earlier studies, we estimate an expression like equation (2.1). 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 severe 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 (2.11), 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 forces 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

[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.

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 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 kilometers 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. (30×30 arc-seconds) showing spatial

[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.

distribution patterns of ambient population (average over 24 hours). Finally, we take each 30×30 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.

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 centers. 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

[9]Nevertheless, we have tried reestimating our results using a simple population count as an alternative measure of city scale and coefficients remain very 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.

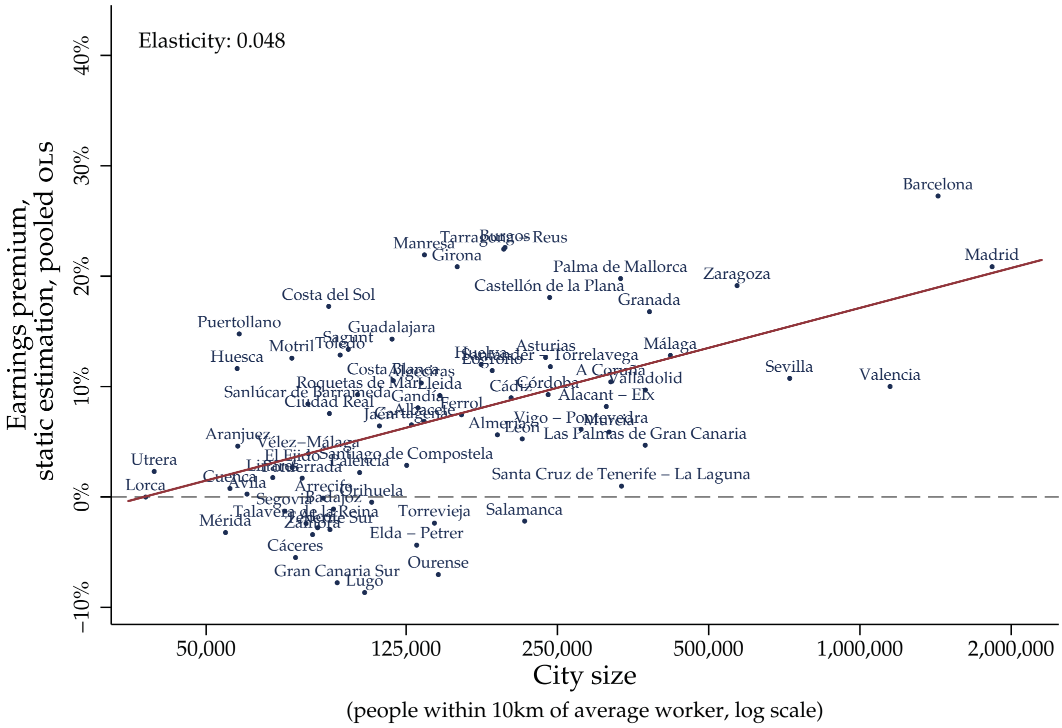


Figure 2.2: Static OLS estimation of the city-size premium

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. Apprenticeships 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.

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 (2.2), 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 2.1 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.

[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 2.1: Estimation of the static city-size earnings premium

	(1)	(2)	(3)	(4)
Dependent variable:	Log earnings	City indicator coefficients column (1)	Log earnings	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.

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 2.2, 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 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 2.1 we estimate equation (2.6) 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 (2.1). 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 2.2 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 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 2.1. 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

[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 2.2: Estimation of the dynamic and static city-size earnings premia

	(1)	(2)	(3)
Dependent variable:	Log earnings	Initial premium (city indicator coefficients column (1))	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 × 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 × now in bigger	0.001 (0.003)		
Experience 3rd-5th biggest cities × 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).

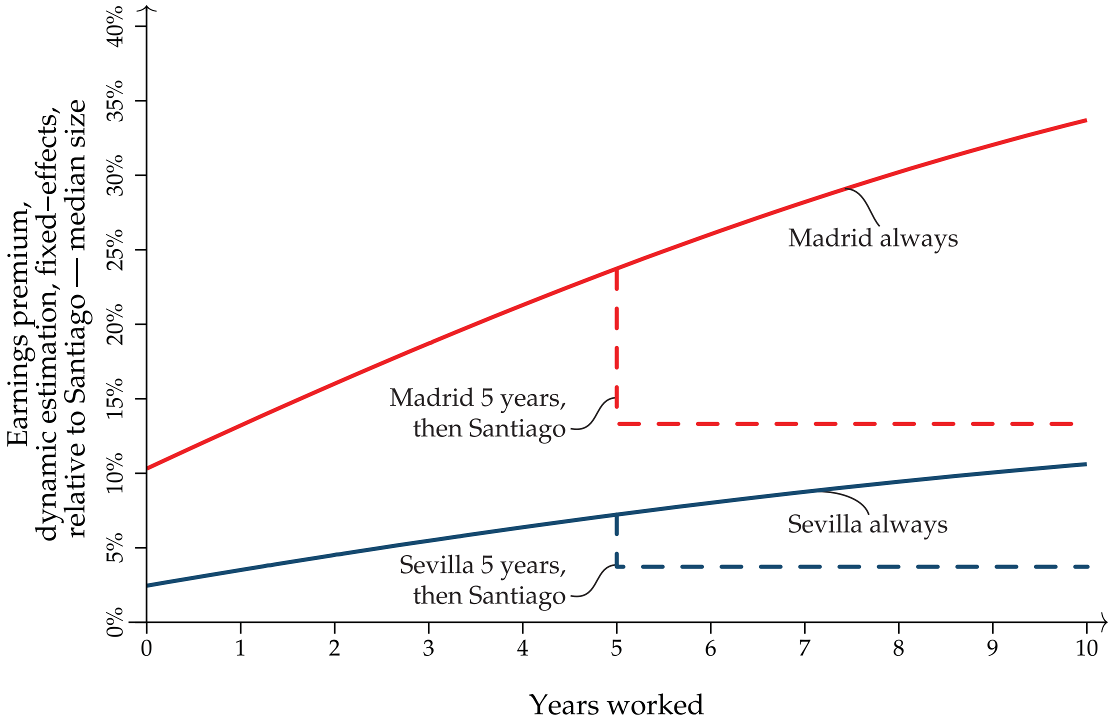
in a city ranked 3rd to 5th raises earnings by 1.1% relative to having worked that same year in a city below the top-five. We have also tried finer groupings of cities by size (not reported), but found no significant differences in the value of experience within the reported groupings (e.g., between 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 3rd to 5th 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.

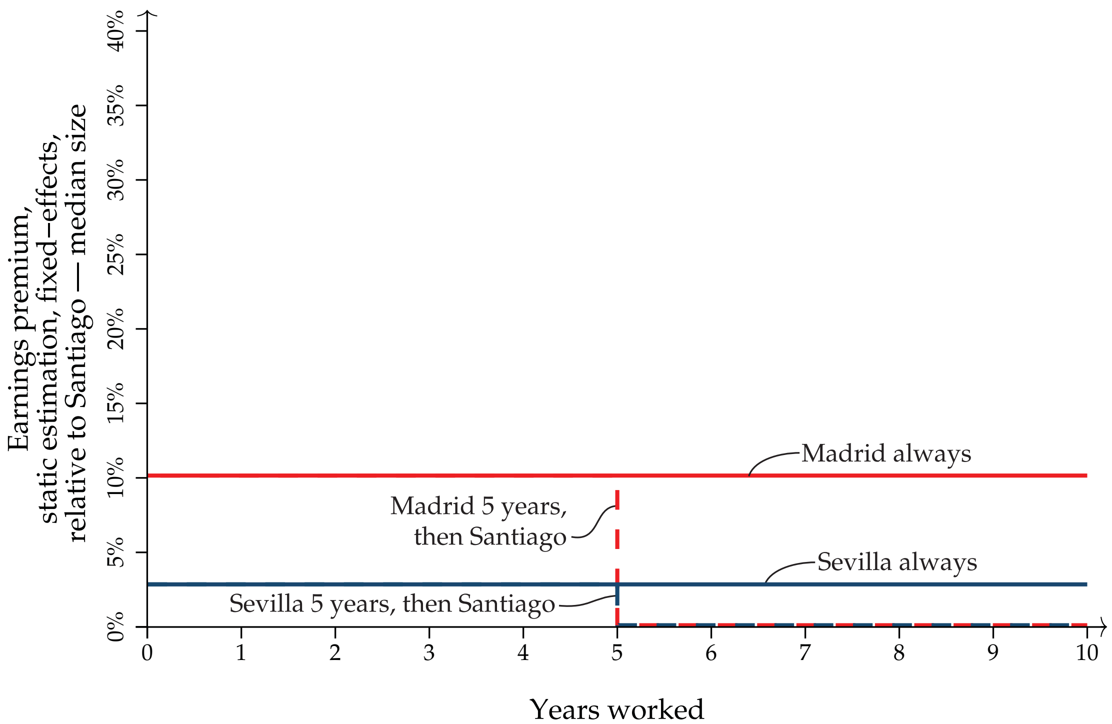
Earnings profiles

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 2.2. In panel (a) of figure 2.3, 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 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

[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 both 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 2.3: Earnings profile relative to median-sized city

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 2.3 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 2.2. 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 2.3 shows that most 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 2.3. 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 2.1 instead of column (1) of table 2.2. 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 2.2 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 2.1. 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

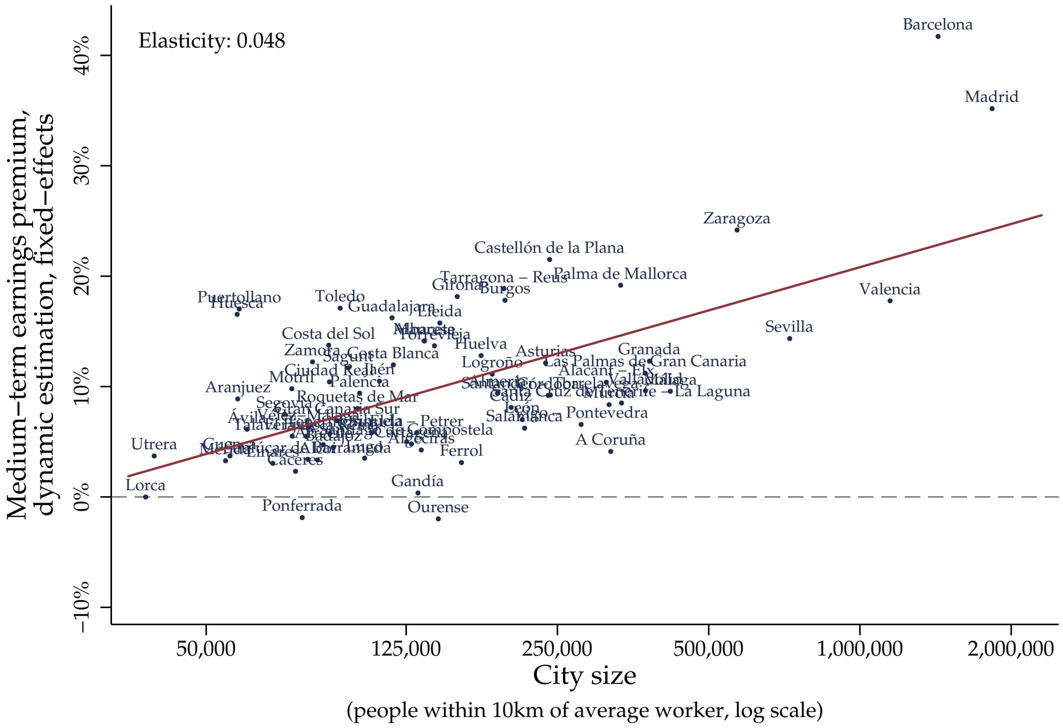


Figure 2.4: Dynamic fixed-effects estimation of the medium-term city-size premium

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 2.2, is 0.049.

Comparison of the 0.049 elasticity of the medium-term premium with respect to city size in column (3) of table 2.2 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 2.2 is almost identical to the static pooled OLS estimate in column (2) of table 2.1. 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 2.2, 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 2.4 plots the estimated medium-term premia against log city size. Compared with the plot for the pooled OLS specification in figure 2.2, 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 2.2, 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 2.2, 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.

Addressing the endogeneity of city sizes

While we have addressed potential sources of bias in the first-stage estimation of column (1) in table 2.2, 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 kilometers 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 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

[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.

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 20$^{\text{th}}$ 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 2.3 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 2.3 shows that the elasticity of the initial premium with respect to city size is not substantially affected by instrumenting (it is 0.023, compared to 0.025 in table 2.2). 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 to 0.049 in table 2.2). In fact, a Hausman test fails to reject that instrumental variables are not required to estimate these elasticities. This is 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).

---

[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 program (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 meters).

[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 2.3: IV estimation of the dynamic city size earnings premium

	(1)	(2)	(3)
Dependent variable:	Log size	Initial premium	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.

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.

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

$$w_{ict} = \sigma_c + \mu_i + \sum_{j=1}^{C} (\delta_j + \phi_j \mu_i) e_{ijt} + \boldsymbol{x}'_{it}\boldsymbol{\beta} + \varepsilon_{ict} \; . \qquad (2.12)$$

Table 2.4: Estimation of the heterogeneous dynamic and static city-size earnings premia

	(1)	(2)	(3)
Dependent variable:	Log earnings net of worker fixed-effect	Initial premium (city indicator coefficients column (1))	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).

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 (2.12) recursively. Given a set of fixed-effects (for instance, those coming from estimating equation (2.11), which corresponds to $\phi_j = 0$), we can estimate equation (2.12) by ordinary least squares, then obtain a new set of

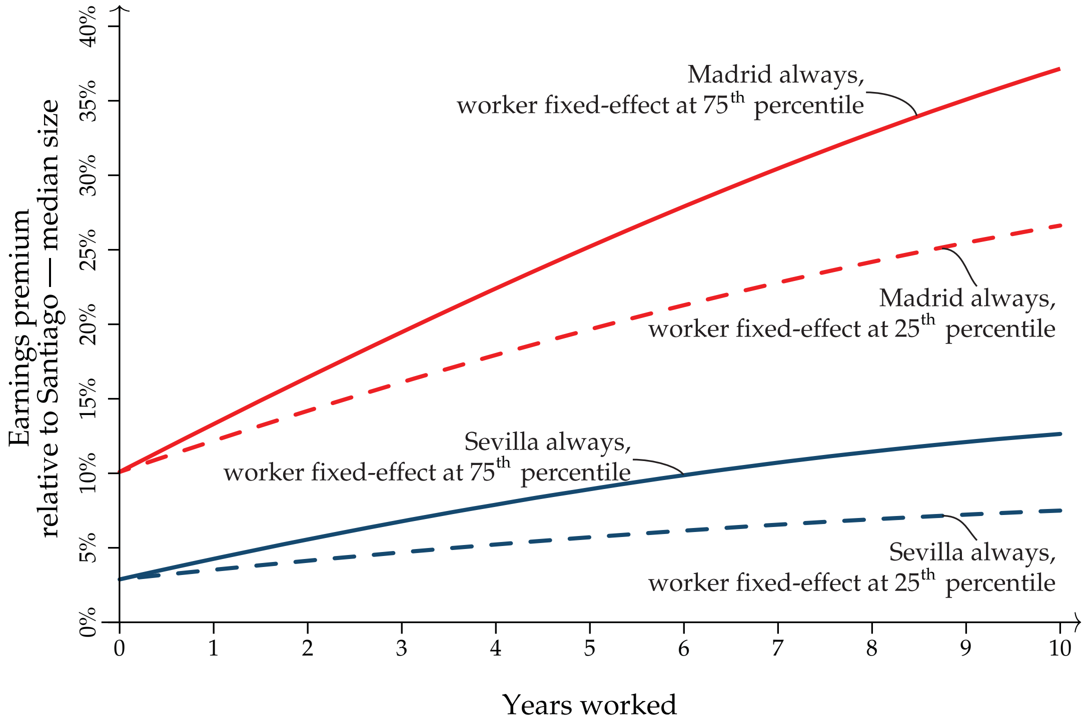


Figure 2.5: Earnings profile relative to median-sized city, high- and low-ability worker

estimates of worker fixed-effects as

$$\hat{\mu}_i = \frac{w_{ict} - \hat{\sigma}_c - \sum_{j=1}^{C} \hat{\delta}_j e_{ijt} - \boldsymbol{x}'_{it}\hat{\boldsymbol{\beta}}}{1 + \sum_{j=1}^{C} \hat{\phi}_j e_{ijt}} , \tag{2.13}$$

then, given these new worker fixed-effects estimate again equation (2.12), and so on until convergence is achieved.[20]

Table 2.4 shows the results of our iterative estimation. Relative to column (1) of table 2.2 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 2.4, figure 2.5 replicates the earnings profiles of figure 2.3 for workers of different ability based on this table. We consider two different workers, a high-ability one (in the 75$^{\text{th}}$ percentile of the estimated overall worker fixed-effects distribution) and a low-ability one (in the 25$^{\text{th}}$ 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.

[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.

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 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 2.1 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 (2.6) if wages are determined as in equation (2.12) results in a biased estimate of μ:

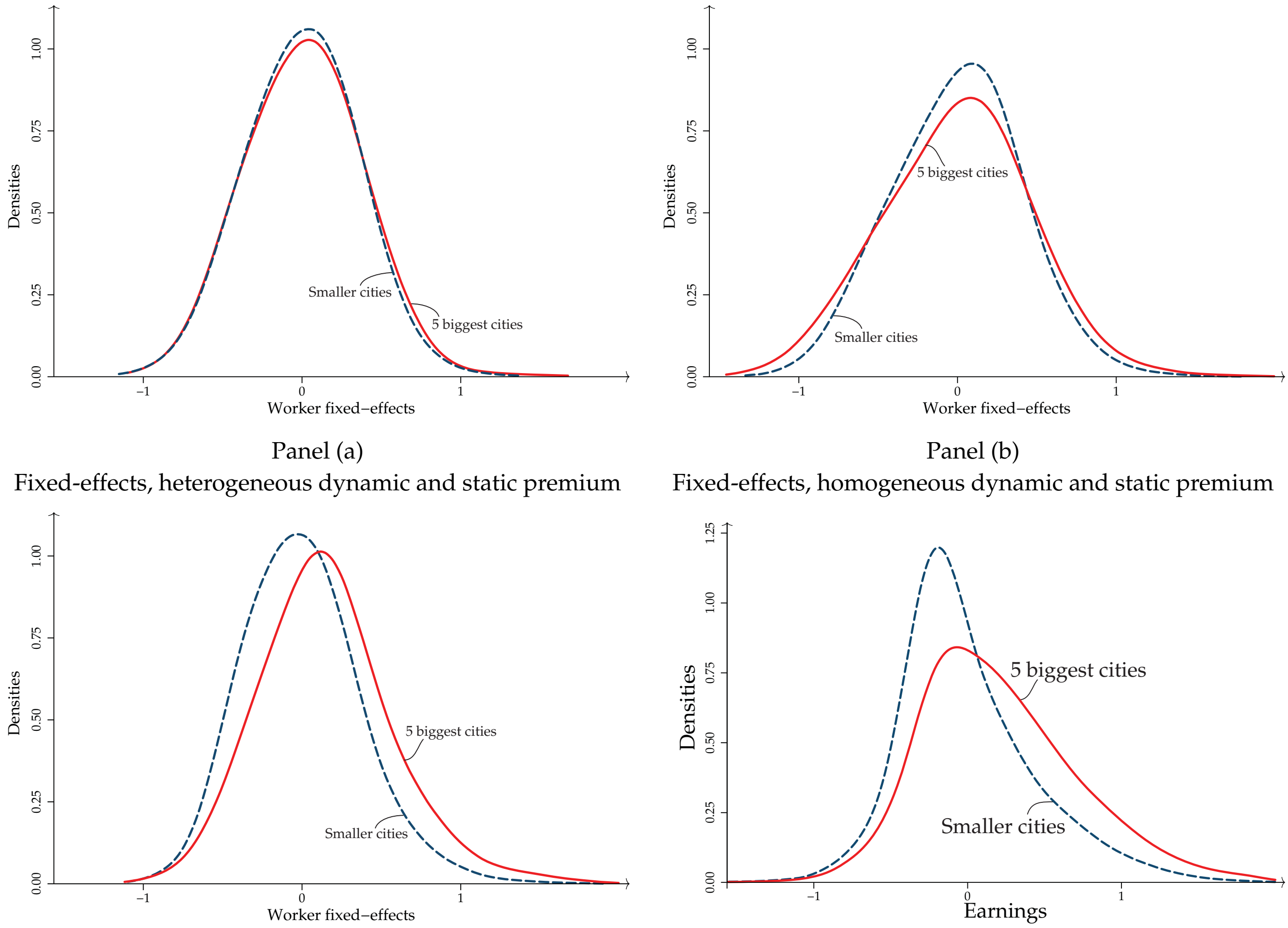
$$\text{plim } \hat{\mu}_{i\,\text{FE}} = \mu_i(1 + \phi_j \bar{e}_{ijt}) + \sum_{j=1}^{C} \delta_j \bar{e}_{ijt} \,. \tag{2.14}$$

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 (2.12) yields instead plim $\hat{\mu}_i = \mu_i$.

Panel (a) in figure 2.6 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 2.4, column 1).[21] We can see that both distributions look very similar (we do a more formal comparison below). This suggests that there 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 2.2, column 1). While both distributions have almost the same mean, the distribution in bigger cities exhibits a higher variance. This is the result of forcing experience acquired in bigger cities to be equally valuable

[21] Each individual is assigned to the city where he was working in May 2007.



Panel (a)
Fixed-effects, heterogeneous dynamic and static premium

Panel (b)
Fixed-effects, homogeneous dynamic and static premium

Panel (c)
Fixed-effects, static premium

Panel (d)
Earnings

Figure 2.6: Comparisons of worker fixed-effects distributions across cities

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 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 2.5 performs a more formal comparison of the plotted distributions, using the method-

Table 2.5: 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 2.4, 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 2.2, 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*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.

ology developed by Combes *et al.* (2011*c*) 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 2.4, 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 2.6.

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.

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

[22]Combes *et al.* (2011*c*) 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 2.5 we restrict ourselves to shift and dilation.

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*a*) 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, these 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 accrues 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.

Chapter 3

Wage cyclicality: Evidence from Spain using social security data

1. Introduction

Recent evidence on real wage cyclicality using worker-level data shows wages are much more cyclical than previously thought. Following the lead of Bils (1985), many studies have found that since the early 1970s individual wages respond to changes in the unemployment rate (see Pissarides (2009) for a summary of the evidence). However, most of this evidence has been obtained for countries with flexible labor markets, mainly the United States. Even more, in contrast to many countries in Continental Europe, the European countries for which estimations are available (UK, Germany, Portugal and Italy) do not exhibit a large incidence of wage indexation policies or centralized collective agreements.

Spain is a suitable scenario to evaluate real wage cyclicality within a much rigid labor market. The Spanish system of collective bargaining is based on two principles that deter firms from adjusting wages along the business cycle. The first principle automatically extends any collective agreement beyond the scope of a firm to all workers in the same sector and province, even if they had not participated in the bargaining process. The second principle secures the validity of collective agreements after their expiration. Likewise, a large share of agreements (more than 60%) include indexation clauses which trigger high inertia in firms wage-setting decisions. Lastly, duality in the labor market insulates workers under permanent contracts (around 67–70% of the workforce with high levels of employment protection) from business cycle fluctuations.

I find weak procyclicality of real wages in Spain over the period 1988–2010. My baseline estimate is a 0.4% increase in wages in response to a 1% decline in the (lagged) unemployment rate. This estimate is much lower than the range of available estimates, which usually vary between 1.3%–1,5% increase in wages in the United States –again for a 1% drop in the unemployment rate– and an even larger increment in wages between 2.0%–2.2% in European countries. When I use total salaries instead of base salaries for a restricted period in which the former are available, I still obtain a low level of procyclicality at 0.6%. As expected, I find real wage cyclicality is lower in a country with institutions that hinder firms to respond to business cycle fluctuations.

In this line, this finding indicates that for some European countries with high-wage indexation and employment protection (e.g., Belgium, France and Scandinavian countries), wage cyclicality is presumably much lower than the available European estimates.

To obtain these cyclicality estimates I use a rich social security data set –*Muestra Continua de Vidas Laborales* (MCVL). This is an administrative data set that tracks career histories for a 4% of individuals who in a calendar year have any type of relation with social security. For each individual, all employment and most unemployment spells are available at the daily level since 1981 or entry in social security, whichever is more recent. Thus, I can construct a monthly panel recording labor market status, some individual traits, job characteristics and wages.

This unique data set has strong advantages relative to other data sets that have been used to estimate wage cyclicality. By exploiting the high frequency in the data I can identify most labor market transitions, specially those that are of particular interest in the wage cyclicality literature (e.g. estimating cyclicality for job movers, for workers who start jobs from periods of inactivity, or for workers who remain within an employer-employee match). Such transitions are not available in surveys with high attrition or are difficult to detect in administrative data sets with higher periodicity. Moreover, I can estimate cyclicality of net present values of wages in new matches over their duration, which constitutes a key piece of information for the Mortensen-Pissarides search and matching model.

One drawback in the data set is the intermediate level of censoring in wages. I propose a simple approach to simulate wages using information on individual and job characteristics, uncensored wage observations and wage persistence estimated by exploiting the longitudinal dimension in MCVL. This is by itself one empirical contribution of the paper. In the line of Haider and Solon (2006), I assume uncensored wages for a worker follow a multivariate log-normal distribution. I estimate using Tobit regressions the mean and variance of wages in each period. To approximate wage correlation coefficients between any two periods I develop an indirect inference approach. Lastly, I simulate wages only for censored observations and evaluate the fit of the simulation by comparing these simulated wages to total salaries available from income tax return data for a restricted period. Overall, the fit of the simulation is quite satisfactory.

This study contributes to the wage cyclicality literature in several aspects. First, as already mentioned, unlike wage cyclicality estimates available for countries with flexible labor markets, this study provides one estimate for a rigid labor market scenario. Second, the study shows how wage cyclicality responds in a setting with high duality in employment protection. I find cyclicality for temporary workers is four to five times greater than for permanent workers. Hence, temporary workers carry most of the burden of wage adjustments along the cycle, while permanent workers are much less affected. Third, I present evidence of wage cyclicality decreasing consistently with the level of job tenure. I find cyclicality is much higher for newly-hired workers (those who enter jobs from periods of unemployment or inactivity) than for job stayers with high levels of tenure. The availability of employer identifiers allows to calculate level of tenure with high precision and to estimate cyclicality within an employer-employee match as in Devereux (2001).

The estimated difference in wage cyclicality between newly-hired workers and job stayers is relevant for the empirical validity of the Mortensen-Pissarides search and matching model

(Mortensen and Pissarides, 1994, Pissarides, 2000). The model has been challenged on its ability to match the observed cyclicality on vacancies and unemployment. Some studies have suggested wage rigidity as a potential solution to this so called unemployment-volatilty puzzle (Hall, 2005, Shimer, 2005). In this model, the cyclicality of the net present value of wages in new matches is a key statistic to determine job creation (Pissarides, 2009). I estimate such cyclicality for the net present value of wages in new matches and obtain a similar estimate to the one using current wages of newly-hired workers. This result, the first in the literature using actual data on monthly wages and job duration, is encouraging since net present values of wages are rarely observed and difficult to calculate in most data sets. Overall, this finding does not give support to wage rigidity as a solution to the unemployment volatility puzzle.

The rest of the paper is structured as follows. Section 2 reviews some institutional aspects of the Spanish labor market. Section 3 describes the data and the approach developed to simulate wages for censored observations. Section 4 explains the estimation methodology. Section 5 presents the results. Finally, section 6 concludes.

2. Institutional aspects of the Spanish labor market

In this section I highlight some features of the Spanish labor market that influence the response of wages to changes in economic conditions. In particular, I focus on the system of collective bargaining and the duality in labor market contracts. The combination of these two factors makes of Spain an interesting scenario to examine wage cyclicality.[1]

The Spanish system of collective bargaining follows the principles established in the 1980 Workers' Statute, which has experienced only minor changes since its adoption. Two main principles govern the Statute. First, the principle of automatic effectiveness states that any collective agreement of a higher level than a firm agreement is immediately extended to all firms and workers in the same sector and province. Workers need not be to be affiliated in a union or participate in the bargaining process. Second, the ultra-activity principle guarantees the permanent validity of collective agreements after their expiration. Even more, if the terms on a new agreement are less beneficial to workers than the earlier one, then the new agreement can not be endorsed.

The system of collective bargaining in Spain can be characterized by its large scope, intermediate degree of centralization, substantial inertia in wage indexation and homogeneity in wage setting decisions. The rate of coverage reaches more than 80% of private sector workers despite a low rate of unionization below 15%. This disparity in rates is sustained, of course, by the principle of automatic effectiveness. The bargaining process takes place mainly at the sectoral level within a provincial scope –only less than 15% of workers are subject to a firm agreement, which are frequent in large firms and negligible in firms with fewer than 200 workers. The high rate of inertia in wages is reflected in the large share of agreements (between 60% and 70%) that incorporate indexation clauses. Collective agreements last around 2.5 years –a long duration similar to those

[1] See Estrada and Izquierdo (2005) for a review of institutional aspects of the labor market. Bentolila, Izquierdo, and Jimeno (2010) provide a recent description of the Spanish system of collective agreement. Dolado, García-Serrano, and Jimeno (2002) analyze the causes and characteristics of labor market duality in Spain.

of Scandinavian countries– and include annual wage setting policies and protection clauses in case of deviations from the inflation rate of reference.

All these features of the Spanish system of collective bargaining do not contribute to the adjustment of wages and employment levels to macroeconomic conditions or the evolution of labor demand and supply. In fact, several studies find a high incidence of nominal and real wage rigidities in Spain.[2] Most of them exploit micro-data from the Wage Dynamics Network (WDN) survey –a project sponsored by the European Central Bank that records information on the determinants of price and wage setting decisions by European firms. In this survey the share of Spanish firms that have frozen wages in the past five years is only 2.4%, a share four times lower than the European average at 9.6%. Likewise, a substantial share of 55% of firms apply an automatic indexation mechanism, a rate three times larger than the European average at 17%. Cuadrado, Hernándes de Cos, and Izquierdo (2011), using the same data set as in this study but a different methodology, find evidence of a high level of real wage rigidity in Spain. In fact, their findings indicate Spain ranks fifth among 17 OECD countries only after Belgium, Sweden, Finland and France.

The duality in the Spanish labor market is the result of two hiring mechanisms with different firing costs. On one hand, workers under permanent contracts benefit from a high level of employment protection through generous severance payments and legal defense in case of firing. On the other hand, workers under temporary contracts have much lower severance payments and do not face legal proceedings when the contract expires. As a result, in this dual labor market, workers in permanent contracts (around 67%–70%) enjoy high protection and bargaining power, while workers in temporary contracts earn lower wages and suffer from high turnover rates and low levels of job tenure.[3]

This dual labor market and the restrictions imposed by a stringent system of collective bargaining hinder firms to react to changes in economic conditions by adjusting wages. Workers under permanent contracts benefit from wage indexation policies and high firing costs. Therefore, it is not surprising that, temporary workers take most of the burden of labor market adjustments. According to the WDN survey, when Spanish firms were asked about ways to reduce costs in response to potential demand shocks, a sizable share of 58% responded they would adjust by reducing temporary employment while only 10% by lowering wages (Izquierdo and Cuadrado, 2009). The corresponding averages across European countries show that only 22% of firms would react by reducing employment and 13% by lowering wages (Bentolila *et al.*, 2010).

[2]Nominal (downward) wage rigidities arise when there is a low incidence of wage cuts, while real (downward) wage rigidities are induced by institutional mechanisms that generate wage increments larger than inflation. Holden and Wulfsberg (2008) find no significant evidence of nominal wage cuts in Spain using industry-level wage data for 19 OECD countries during the period 1973–99. Likewise, Babecký, Caju, Kosma, Lawless, Messina, and Room (2010) show wages in Spain exhibit both nominal and real (downward) rigidities using data from the Wage Dynamics Network survey in the years 2007–2008

[3]The rise in the share of temporary workers induced the government to implement reforms to mitigate the 1984 labor market liberalization policy. In 1994, the government toughened some rules for the use of temporary contracts and expanded the array of reasons for job dismissal. This reform had no effect in the share of temporary workers. Furthermore, in 1997, another reform introduced a new permanent contract with lower firing costs and social security contribution (such contract was extended and amplified in a later reform in 2001). According to Dolado *et al.* (2002) the net effect was a small decline in the share of temporary workers.

3. Data

The main data set used in the study is Spain's Continuous Sample of Employment Histories (*Muestra Continua de Vidas Laborales* or MCVL). This administrative data set is a 4% non-stratified random draw of the population of individuals related with the social security system in a calendar year. Individuals can either be working as employees or self-employeds, receiving unemployment benefits or receiving a pension. The MCVL records all changes in labor market status and job characteristics for each individual in the sample since 1981 or entry in social security, whichever is more recent. I combine all six available editions of the MCVL, so I can obtain a 4% sample of all individuals who appear in social security records at any time throughout 2004–2010.

Individuals enter the sample based on their anonymized social security number and remain in subsequent editions. Each new MCVL wave adds individuals who first enter the labor market and losses those who were deceased or left the country in the previous calendar year (those who stopped working remain in the sample while they receive unemployment benefits, disability benefits or a retirement pension). The unit of observation in the source data records the starting and ending date for any change in the individual's labor market status or job characteristics (including changes in occupation or remuneration within the same firm). Therefore, as long as an individual registers one day of activity with social security in any calendar year between 2004–2010, her complete working life history can be recovered up to 1981.

I construct for all workers monthly working life histories including labor market status, individual and main job characteristics.[4] For every job spell I know the type of occupation and contract (permanent or temporary), the 3-digit NACE sector of economic activity and whether the individual is self-employed or a public sector worker. Some individual characteristics like age and gender are provided while other individual variables such as level of education and country of birth are obtained from the *Padrón* or Municipal Register.

Monthly wages are available for all workers but some observations are censored. These wages correspond to base salaries and do not include overtime, commissions or bonuses.[5] Later in this section, I explain in detail how I simulate wages only for capped observations exploiting the panel dimension in the MCVL and using data on uncensored wage observations and individual and job characteristics. I build detailed measures of cumulative labor market experience by adding up the actual number of work days in each month. Similarly, I can construct a measure of job tenure at the establishment location. This is possible since legislation forces employers to keep separate earnings' contribution accounting codes for each province in which they conduct business. Unfortunately, employers need not keep a common firm identifier across provinces; hence, I cannot build a second measure of job tenure at the firm level.

This rich administrative data set has unique advantages over other data sets that have been used to estimate wage cyclicality. First, its large coverage makes it a representative sample of the

[4] In months when individuals become unemployed, I classify them as such if the amount of daily unemployment benefits exceed her daily wage in the same month. Likewise, I identify the main job as the one with the highest wage (and highest number of work days in case of a tie).

[5] According to García-Pérez (2008) and Cuadrado *et al.* (2011) base salaries in MCVL are a good proxy of total salaries for a large share of the population of workers.

Spanish labor force, as opposed to surveys such as the PSID or the NLSY for the United States.[6] Second, its daily frequency allows to accurately identify most labor market transitions such as new jobs of workers who come from unemployment or inactivity spells, job moves between firms and long-term separations. These transitions are not easy to detect in annual data sets or surveys with high levels of attrition. Third, the presence of an establishment identifier makes it possible to construct precise measures of job tenure at this level and to estimate wage cyclicality within a job or match. Finally, the last two cited advantages allow to calculate the present value of wages in a new match, a key statistic for job creation in the canonical search and matching model. In fact, this is the first study that calculates such present values using actual data on job durations and wages.[7]

Sample restrictions

The initial sample is a monthly data set of men born between 1924 and 1992 who have worked at any time between January 1988 and December 2010 (i.e., aged 18–64 during this period).[8] I also compute wage cyclicality estimates for women and for all workers. Yet, all other results include only males to make them comparable to estimates from other countries. A total of 609,765 individuals and 67,591,582 monthly observations make up this initial sample.[9]

From this initial sample, I drop observations with non-contributory occupations, with missing values of occupation and individuals for whom educational attainment is missing. These restrictions decrease the sample to 559,690 individuals and 63,992,898 monthly observations. Then, I eliminate those years in which individuals show low labor force attachment, i.e. those calendar years when individuals work less than a month. After this constraint the sample contains 547,427 individuals and 63,548,335 monthly observations.

I further restrict the sample to full-time job spells in the private sector. Earnings in the public sector are heavily regulated by the national and regional governments and only a small subset of workers are under a part-time contract. This leaves the final sample at 528,205 individuals and 56,474,295 monthly observations.

Job tenure categories

I consider two job tenure definitions. In the first one I count the number of days accumulated in an establishment throughout a worker's career. When the worker switches jobs the level of tenure resets to zero. If he later returns to an earlier job, the tenure count starts from its former level, i.e., the one when he left that job. In the second definition a return to a previous job is considered a

[6] García-Pérez (2008) tabulates worker characteristics for MCVL and for the Spanish Labor Force Survey EPA and finds similar population magnitudes.

[7] In addition, the MCVL longitudinal dimension allows to control for compositional bias in the (unobserved) characteristics of workers along the cycle. Lastly, the fact that wages are not self-reported also increases data reliability.

[8] I set 1988 as the initial year because prior to it job tenure is heavily left-censored – recall it can be measured at most since 1981. Still, as I show later in the results, when I estimate wage cyclicality for long-term stayers I will group workers with more than six years of tenure.

[9] I exclude spells workers spent as self-employed because earnings and the number of work days are self-reported and, hence, less reliable. I also eliminate spells in agriculture, fishing, mining and households since non-pecuniary payments are more prevalent in these activities and the number of work days is again self-reported. These latter spells account only for 7% of the sample at this stage.

Table 3.1: Tenure categories

	Monthly observations	Share in sample	Share under temporary contract	Mean monthly wages (€)
Definition 1				
Newly-hireds	6,572,764	11.8%	82.6%	1,334
Job-movers	8,162,033	14.6%	66.4%	1,690
1 year - 2 years	7,712,658	13.8%	42.2%	1,757
2 years - 4 years	9,503,472	17.0%	26.3%	1,940
4 years - 6 years	6,004,060	10.7%	16.1%	2,120
More than 6 years	17,985,815	32.2%	5.6%	2,533
Definition 2				
Newly-hireds	8,061,622	14.5%	83.2%	1,342
Job-movers	8,330,305	15.0%	65.1%	1,710
1 year - 2 years	7,689,087	13.9%	37.9%	1,795
2 years - 4 years	9,142,458	16.5%	21.3%	1,994
4 years - 6 years	5,644,631	10.2%	11.4%	2,183
More than 6 years	16,608,914	29.9%	3.7%	2,571

Notes: All variables are sample means over 1988–2010, except for the share of workers under temporary contract which is restricted to 1997–2010. Monthly wages are expressed in December 2010 euros.

new spell. Thus, by construction job tenure levels can not be higher in the latter definition. More important, the former definition allows to estimate cyclicality of wages for new workers within a match, i.e., those who start a job *and* have not worked in the firm before.

For each job tenure definition I construct six categories. I divide workers who start a new job into newly-hireds –those who come from periods of unemployment or inactivity– and job-movers –those who change jobs in two consecutive months. Workers remain under these categories during their first year of tenure.[10] The other tenure categories are the following: 1 year – 2 years, 2 years – 4 years, 4 years – 6 years and more than 6 years.

Column (2) in table 3.1 shows the share of observations for each tenure category and definition. Newly-hireds and job-movers account for more than 26% of observations while workers with more than four years of tenure comprise a large share greater than 40%. The share of observations in the first year of tenure is larger than for other countries. For instance, in Portugal it accounts for 17% (Carneiro, Guimaraes, and Portugal, 2012) while the proportion of newly-hired workers in the United States is much lower at 8% (Haefke, Sonntag, and van Rens, 2012). This polarization in the level of tenure confirms the duality in the labor market mentioned in the previous section. Column (3) reveals that the use of temporary contracts is predominant among workers who start a job

[10] I have been extremely careful in identifying these two categories. Given the high frequency of the data, it is common to find many jobs with very low durations (e.g. less than a week), so I exclude all jobs that last less than one month. Not all job entries can be classified. In some cases a worker starts a job, leaves it for a while (e.g. 3 months) and returns after. Thus, I prefer not labeling him as newly-hired or job-mover in the first year of tenure, but I do classify future levels of tenure in this job into other categories. These cases account for a trivial share of new jobs.

–greater than 80% for newly-hireds and somewhat lower for job-movers (65%–66%).[11] This high incidence of temporary work increases turnover rates in the labor market and generates dispersion in job durations of new matches –in the sample, using the first tenure definition, the 25th percentile in the distribution of job duration of new matches is quite low at 4 months, while the mean and 75th percentile are much larger at 24 and 27 months, respectively. Column (4) also highlights large raw wage differentials among tenure categories.

Simulating wages for censored observations

The MCVL reports data on wages for the entire period 1988–2010, but these are capped for some workers. In particular, 15.8% and 2.3% of monthly wage observations are top- and bottom-coded, respectively. In this subsection I explain how I simulate wages for the 18.1% of observations that are capped.

Censoring bounds vary by type of occupation on an annual basis. Since these bounds are not available in the MCVL I have gathered annual bulletins from Spain's *Boletín Oficial del Estado* BOE, the official newspaper, to identify them. Then, I have checked consistency by plotting monthly wage densities for each year-month-occupation combination. In general, censored observations can be easily detected using the information in the BOE bulletins.

Next, I run 230 occupation-year Tobit regressions (10 occupations $\times$ 23 years) in which the dependent variable is log daily wages expressed in December 2010 euros. As explanatory variables I include quartics in experience and job tenure, and sets of indicator variables for age, level of education, 3-digit NACE sector, type of contract, province and urban area of workplace, and month. I also add interactions among many of these variables and level of education.

After these estimations, I can simulate wages for capped observations as follows:

$$\hat{W}_{ijt} = x_{ijt}{}'\hat{\gamma} + \hat{\sigma}\,\varepsilon_{ijt}, \tag{3.1}$$

where $\hat{W}_{ijt}$ is the simulated log wage for individual i in occupation j at year t, x_{ijt} is a vector of individual and job characteristics, $\hat{\gamma}$ and $\hat{\sigma}$ are estimated parameters, and ε_{ijt} is an i.i.d shock. However, as shown by studies that exploit data on career-long earnings histories, earnings exhibit great persistence (Bjorklund, 1993, Haider and Solon, 2006). Thus, I exploit the panel dimension in the MCVL to introduce persistence in wage shocks.

To this end I follow the methodology proposed by Haider and Solon (2006). The main assumption is that the joint distribution of uncensored log wages for an individual is multivariate normal. Hence, annual wages throughout the period of interest 1988–2010 can be fully characterized by the mean and variance of log wages in each year –estimated in equation (3.1)– and the cross-year autocorrelations of log wages for every pair of years.[12]

Haider and Solon (2006) estimate autocorrelations between pairs of years using a bivariate Tobit maximum-likelihood estimator. Instead, I use a more simple approach based on indirect inference to compute cross-year autocorrelations.

[11] Although the sample records information on type of contract in all years, the incidence of temporary work matches aggregate figures after 1996. In column (3) of table 3.1 I restrict the period to 1997–2010.

[12] Log daily wages follow a multivariate normal distribution also within each of the ten occupation categories. For simplicity and from now on I omit index j referring to type of occupation.

The approach can be summarized in four steps. First, I estimate a regression coefficient for every pair of (standardized) log wages of the same worker i in two different years. This estimation is carried out only for uncensored wage observations. I label this regression coefficient $\hat{\lambda}^*$.

Second, I exploit the multivariate normality assumption to generate wages for worker i in year $t+h$ conditional on his observed wage in year t, the relevant censoring bounds and a value for the correlation coefficient ρ. Equation (3.2) shows how to generate wages in year $t+h$ based on a bivariate normal distribution of wages in years t and $t+h$:

$$\tilde{W}_{i,t+h}\,,\,\tilde{W}_{it} \sim N\left(\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}1 & \rho\\ \rho & 1\end{pmatrix}\right), \tag{3.2}$$

$$\mathbb{E}(\tilde{W}_{i,t+h}\,|\,\tilde{W}_{it},\,\tilde{a}_{t+h} \leqslant \tilde{W}_{i,t+h} \leqslant \tilde{b}_{t+h}) = \rho\tilde{W}_{it} + \sqrt{1-\rho^2}\left[\frac{\phi\left(\frac{\tilde{a}_{t+h}-\rho\tilde{W}_{it}}{\sqrt{1-\rho^2}}\right)-\phi\left(\frac{\tilde{b}_{t+h}-\rho\tilde{W}_{it}}{\sqrt{1-\rho^2}}\right)}{\Phi\left(\frac{\tilde{b}_{t+h}-\rho\tilde{W}_{it}}{\sqrt{1-\rho^2}}\right)-\Phi\left(\frac{\tilde{a}_{t+h}-\rho\tilde{W}_{it}}{\sqrt{1-\rho^2}}\right)}\right],$$

where $\tilde{W}_{it}$ are the standardized uncensored log wages for individual i in year t, ρ is the correlation coefficient, and $\tilde{a}_{t+h}$ and $\tilde{b}_{t+h}$ are the standardized lower and upper bounds applicable in year $t+h$, respectively. Since the only unknown in $\mathbb{E}(\tilde{W}_{i,t+h})$ is the value of ρ –the cross-year autocorrelation of interest– based on a grid of 40 values of ρ from 0 to 0.975 on 0.025 intervals, I generate $\mathbb{E}(\tilde{W}_{i,t+h}\,|\,_{\rho=\rho_k})$ where $k = 1,...,40$.

Third, I regress each generated $\mathbb{E}(\tilde{W}_{i,t+h}\,|\,_{\rho=\rho_k})$ on $\tilde{W}_{it}$ only for uncensored observations and obtain a regression coefficient $\hat{\lambda}_k$. The optimal value of the cross-year autocorrelation ρ_k^* will be the ρ_k which minimizes the absolute distance between $\hat{\lambda}^*$ and any $\hat{\lambda}_k$. Thus, if log wages for individual i follow a multivariate normal distribution, I choose the ρ_k^* that best replicates the observed correlation for uncensored wages in the data. I construct variance-covariance matrices (23 $\times$ 23) for each occupation with the optimal ρ_k^* values calculated for every pair of years throughout 1988–2010.

Lastly, I simulate wages *only* for capped observations as follows:

$$\hat{W}_{ijt} = x_{ijt}{}'\hat{\gamma} + \hat{\sigma}\cdot\,\hat{p}_{jt}{}'\xi_{ijt}, \tag{3.3}$$

where all variables are the same as in equation (3.1), but now $\hat{p}_{jt}$ is a row vector (1 $\times$ 23) of the Cholesky decomposition of the estimated variance-covariance matrix and ξ_{ijt} is a vector of random shocks.[13] Given that I know whether monthly wages are originally top- or bottom-coded, I force simulated wages to be above or below the corresponding bound, respectively.[14]

[13]In one type of occupation the Cholesky decomposition matrix (P) cannot be calculated because the element-by-element estimated variance-covariance matrix ($\hat{\Omega}$) is not positive semidefinite (as it should be). Haider and Solon (2006) face the same problem and impose a non-negativity constraint on the diagonal elements of $\tilde{P}$, where $\tilde{\Omega} = \tilde{P}\tilde{P}'$ and $\tilde{P}$ minimize the distance between $\hat{\Omega}$ and $\tilde{\Omega}$. I consider a less robust but faster solution. I diagonalize $\hat{\Omega}$ and replace negative eigenvalues (only 2) with zeros. The autocorrelations in this new variance-covariance matrix Ω^* are very similar to the ones in $\hat{\Omega}$.

[14]This implies drawing i.i.d shocks from a truncated normal distribution $\varepsilon_{ijt} > (b_{ijt} - x_{ijt}{}'\hat{\gamma})/\hat{\sigma}$ if wages are top-coded and $\varepsilon_{ijt} < (a_{ijt} - x_{ijt}{}'\hat{\gamma})/\hat{\sigma}$ if they are bottom-coded. b_{ijt} and a_{ijt} are wage levels at which top and bottom censoring occur, respectively. Because computation is extremely time-demanding due to the large number of observations, I only run 25 iterations for the vector of random shocks. Only 1.5% of observations remain censored and I do not use them in the estimations.

The fit of the simulation can be tested for the period 2004–2010 in which uncensored earnings from income tax data have been merged to MCVL. These earnings include all labor income (salaries plus overtime and bonuses) and are matched to MCVL based on both employee and employer identifiers separately for each job in a calendar year. Appendix 3.A presents details on the fit of the simulation. In general, I am able to match the shape of the upper tail of the earnings distribution with high precision, even for high-skilled workers who are top-coded beyond the 51st percentile. Moreover, the degree of persistence observed in labor income is quite similar to the one I obtain in the MCVL after simulating censored observations. These results increase the odds that simulated wages can accurately predict actual labor income also for the period 1988–2003.

4. Estimation methodology

In order to estimate wage cyclicality I use a level wage equation of the following form:

$$\ln w_{it} = \alpha_i + x_{it}{}'\beta + \delta_1 U_t + \delta_2 T + \epsilon_{it}, \tag{3.4}$$

where $\ln w_{it}$ is the log real wage of worker i in period t (a year-month pair), α_i is a worker fixed-effect, x_{it} are individual and job characteristics, U_t is the national civilian unemployment rate (used as a cyclical variable), T is a linear time-trend and ϵ_{it} is the error term with zero mean and constant variance. The previous literature has emphasized the need to address composition bias resulting from various types of individuals working in different phases of the business cycle (Solon, Barsky, and Parker, 1994). To this end, I introduce worker fixed-effects to capture time-invariant unobserved individual heterogeneity.

The most common solution used to address such composition bias has been to estimate a wage regression in first differences (Bils, 1985, Solon *et al.*, 1994, Shin, 1994, Devereux, 2001, Devereux and Hart, 2006). However, this strategy restricts the sample to individuals who work in two consecutive periods –or years given that most studies are based on annual or bi-annual surveys. For this reason, studies that estimate wage cyclicality for newly-hired workers –or any group of workers with weak labor force attachment– run a wage equation in levels (Kudlyak, 2011, Carneiro *et al.*, 2012, Haefke *et al.*, 2012). By following this approach, I can exploit the full sample of workers including job stayers, newly-hired workers and job-movers. Also, given the high frequency of the data, most individuals appear more than once so I lose very few observations when introducing worker fixed-effects.[15]

In a regression of equation (3.4) the standard error of the estimated coefficient of interest, δ_1, is substantially underestimated in the presence of a year specific error since all workers in period t face the same level of unemployment (Moulton, 1986). For this reason, like previous studies (Solon *et al.*, 1994, Shin, 1994, Devereux, 2001), I adopt a two-stage estimation method and transform the specification in equation (3.4) in the following two equations:

[15]Note that given the monthly frequency of the data I could still estimate cyclicality for newly-hired workers using first differences. In this setup, I would be restricting the sample to those who are currently working, were employed in the same month during the past year, but faced periods of unemployment or inactivity before entering the current job. Still, I rather prefer to estimate wage cyclicality in levels to use the full sample while losing vey few observations. In my baseline estimation I only lose 0.8% of individuals and 1.5% of monthly observations.

$$\ln w_{it} = \alpha_i + x_{it}{}'\beta + \sum_{t=1}^{T} \eta_t D_t + \psi_{it}, \tag{3.5}$$

$$\hat{\eta}_t = \theta_1 U_t + \theta_2 T + v_t. \tag{3.6}$$

Equation (3.5) now includes a set of indicator variables η_t for each year-quarter pair –wages are observed monthly but the unemployment rate is at the quarterly level– that capture in an unrestricted way all temporal variation in wages net of observed and (time-invariant) unobserved individual characteristics. In the second stage, equation (3.6), I regress the estimated year-quarter indicators $\hat{\eta}_t$ on the unemployment rate and a linear time trend. Therefore, the standard error of the new coefficient of interest, θ_1, is now free from the aggregate bias present in equation (3.4).[16]

The cyclicality coefficient θ_1 measures the semi-elasticity of wages with respect to the unemployment rate. Most studies using micro-data on wages have used the unemployment rate as the preferred proxy of the economic cycle, following the lead of Bils (1985). A negative estimated value for θ_1 would indicate wages are procyclical. I multiply log wages by 100 so that θ_1 approximates the percentage change in wages in response to a one percentage point increase in the unemployment rate.

5. Results

Baseline estimates of wage cyclicality

I begin by estimating the two-stage method described in equations (3.5) and (3.6) for different groups. In the first stage I regress log daily wages –deflated using the Consumer Price Index and expressed in December 2010 euros– on worker fixed-effects, quartics of experience and job tenure, occupation and contract-type indicators and a set of year-quarter indicators.[17] In the second stage I regress the estimated year-quarter indicator coefficients on the yearly-lagged quarterly unemployment rate, a linear time trend and quarter indicators. I include the lagged unemployment rate given that most wages are set in advance.[18] I report only results for the coefficient of interest, θ_1, the semi-elasticity of real wages with respect to the (lagged) unemployment rate.

Row (1) in table 3.2 shows evidence of weak real wage procyclicality for men in Spain. A one percentage point decline in the unemployment rate is associated with a small increment in real wages of 0.36%. Similar to previous studies, estimates in row (3) indicate that the degree of cyclicality for women is even lower at 0.26% (the difference being significant at the 5% level). In row (4) I estimate wage cyclicality for men but using first-differences in both stages, as in Solon *et al.* (1994) and Devereux (2001). Cyclicality increases only marginally and is not statistically different from my baseline estimate in row (1). I have also carried out alternative second stage

[16]One alternative is to estimate equation (3.4) using robust-clustered standard errors by year-quarter. However, when worker fixed-effects are introduced this is not plausible as workers are observed in different periods. Carneiro *et al.* (2012) provide a simple method to obtain standard errors in a one-step estimation.

[17]I use this set of regressors in all other estimations. I have also included 3-digit sector indicators and province and urban area of workplace indicators with little effect in the second stage results. Unless stated otherwise, I always use the first job tenure definition.

[18]As stated in section 2, collective agreements may last several years but include annual wage setting policies. According to the WDN survey Spain has the second highest share of firms that change wages only once a year.

Table 3.2: Baseline estimates of wage cyclicality

	Cyclical indicator: Unemployment rate (levels)	Cyclical indicator: Unemployment rate (first-differences)	Observations (first stage)
Sample: Male workers	-0.357 (0.052)***		55,601,982
Sample: All workers	-0.329 (0.052)***		81,935,931
Sample: Female workers	-0.266 (0.055)***		26,333,949
Sample: Male workers		-0.414 (0.072)***	45,099,386

Notes: Each coefficient is a separate second stage regression of the estimated year-quarter coefficients on the yearly-lagged quarterly unemployment rate. All second stage regressions have 92 quarterly observations (1988:1 to 2010:4) and include a constant term, a linear time trend and quarter indicators. In row (4) the first stage regression is in first-differences and the second stage regresses estimated year-quarter coefficients on the lagged annual change in the quarterly unemployment rate (88 observations). ***, **, and * indicate significance at the 1, 5, and 10 percent levels.

estimations by using unemployment rates for men or other lags of unemployment. In general, the semi-elasticity increases slightly to 0.40.

The level of real wage cyclicality I find is not far from the one obtained by Bentolila *et al.* (2010). Using micro-data on collective agreements over the period 1990–2007 they estimate a regression of nominal wage increments on the lagged annual changes of the regional unemployment rate and sectoral productivity, and the inflation rate (considering positive and negative deviations from the rate of reference). They also include some characteristics of collective agreements as controls. They find that a one percentage point decline in the unemployment rate is related to an increase of wages of 0.24%, but only for newly-signed agreements.

The level of real wage cyclicality I estimate, 0.4%, is the lowest among studies that use worker-level data for developed countries. Pissarides (2009) summarizes results of most available studies. For the United States a drop of one percentage point in the unemployment rate is correlated with a real wage increment of 1.3%–1.5%. For European countries (UK, Germany, Italy and Portugal) the estimated real wage cyclicality is even greater at 2.0%–2.2%. However, based on the results of the International Wage Flexibility Project and the Bank of Spain –see Cuadrado *et al.* (2011)–, none of these European countries exhibits a high degree of real wage rigidity. Spain ranks fifth in such ranking after countries with predominant wage indexation (Belgium, Sweden, Finland and France), while countries such as Germany and Italy have low levels of real wage rigidity. In sum, the fact that estimated levels of real wage cyclicality in Europe are higher than expected may be driven by the selected pool of countries for which estimations are available.

Wage cyclicality for selected samples

One potential reason for the low level of real wage cyclicality I find is the use of base salaries. If firms compensate workers during expansions by offering bonuses aside from the base salary (De-

Table 3.3: Wage cyclicality for selected samples

	Cyclical indicator: Unemployment rate (levels)	Estimation period	Observations (first stage)
Sample: Total salaries	-0.546 (0.050)***	2004:1–2010:4	18,614,684
Sample: Base salaries	-0.325 (0.040)***	2004:1–2010:4	21,319,792
Sample: Censored base salaries	-0.217 (0.045)***	1988:1–2010:4	46,243,253
Sample: Permanent contract	-0.228 (0.055)***	1988:1–2010:4	41,695,370
Sample: Temporary contract	-1.069 (0.136)***	1988:1–2010:4	13,906,612

Notes: Each coefficient is a separate second stage regression of the estimated year-quarter coefficients on the yearly-lagged quarterly unemployment rate. All second stage regressions include a constant term, a linear time trend and quarter indicators. Rows (1) and (2) contain 28 quarterly observations (2004:1–2010:4) while rows (3), (4) and (5) have 92 quarterly observations (1988:1 to 2010:4). ***, **, and * indicate significance at the 1, 5, and 10 percent levels.

vereux, 2001), then my estimate is a lower bound.[19] For this reason I estimate real wage cyclicality using total salaries (all labor income) from income tax data for the period 2004–2010. Although the period is relatively short, it exhibits enough quarterly variation in wages and unemployment as it coincides with a period of economic expansion and the great recession. Rows (1) and (2) in table 3.3 indicate that the semi-elasticity of wages with respect to the unemployment rate is larger when using total salaries (0.55 vs. 0.33 in absolute terms). If I extrapolate my baseline estimate of 0.36 over the period 1988–2010, then the new estimate of real wage cyclicality is at most 0.6. Still, it reflects much lower cyclicality than for other countries.[20]

Row (3) in table 3.3 estimates real wage cyclicality using only uncensored wages in MCVL and, hence, excludes mainly those in the upper tail of the wage distribution. The lower estimate obtained suggests that real wages are more cyclical for high-skilled workers. Both low- and high-skilled workers benefit from wage setting policies in collective agreements, but the latter may profit more during expansionary periods given their greater bargaining power. Overall, the cyclicality difference between the baseline estimate and the one using only uncensored observations confirms the need to address the censoring issue in MCVL.

Finally, I look at real wage cyclicality for workers under different contract types. Rows (4) and (5) in table 3.3 show that wage cyclicality of temporary workers exceeds that of permanent workers by a factor of four. Moreover, when estimating the second stage for temporary workers using the current or yearly-lagged quarterly unemployment rate makes little difference (-0.95% vs.

[19]Workers and firms may agree to also increase the number of hours of work and, hence, the relevant wage measure becomes the hourly wage rate including overtime. The MCVL records the number of work hours only for workers in part-time contracts who account for a small share of the labor force (less than 10% among males).

[20]Few studies estimate the cyclicality of base salaries. One exception is Devereux (2001) who obtains also a low estimate for salaried stayers of -0.3% using PSID data. When he looks at total salaries the cyclicality is greater at -0.51%.

Table 3.4: Wage cyclicality for tenure categories

	Job tenure (definition 1)	Job tenure (definition 2)	Job tenure (definition 1, excl. migrants)
Newly-hired workers	-0.800	-0.662	-0.826
	(0.069)***	(0.064)***	(0.069)***
Job-movers	-0.556	-0.529	-0.557
	(0.064)***	(0.064)***	(0.064)***
1 year – 2 years	-0.532	-0.512	-0.532
	(0.068)***	(0.068)***	(0.068)***
2 years – 4 years	-0.336	-0.322	-0.337
	(0.060)***	(0.062)***	(0.061)***
4 years – 6 years	-0.233	-0.217	-0.238
	(0.053)***	(0.052)***	(0.053)***
More than 6 years	-0.135	-0.122	-0.132
	(0.041)***	(0.042)***	(0.041)***
Observations (first stage)	44,084,170	43,719,077	40,745,734
Unemployment rate (levels)	Yes	Yes	Yes

Notes: Each coefficient is a separate second stage regression of the estimated interaction between tenure category and year-quarter coefficients on the yearly-lagged quarterly unemployment rate. All second stage regressions have 92 quarterly observations (1988:1 to 2010:4) and include a constant term, a linear time trend and quarter indicators. I randomly drop 20% of observations in the first stage given the large number of coefficients to be estimated. ***, **, and * indicate significance at the 1, 5, and 10 percent levels.

-1.07%, respectively). This pattern is specific to them. When I estimate the second stage regression for permanent workers using current unemployment, the elasticity falls almost by half to -0.13. This reduction also occurs for other selected samples. Therefore, temporary workers face a much higher level of real wage cyclicality similar to that in countries with much more flexible labor markets. In addition, given that collective agreements apply mainly to permanent workers as they are continuously employed with a higher probability, temporary workers are the only group affected by current economic conditions.

Wage cyclicality for tenure categories

The employer-employee setup in MCVL allows to estimate wage cyclicality for stayers within a match. Also, in contrast to other data sets, newly-hired workers and job-movers can be clearly detected. In table 3.4 I estimate real wage cyclicality for the tenure categories described in section 3. For this purpose, I modify equation (3.5) by including indicators for tenure categories and interactions between them and year-quarter indicators. Now, the second stage (equation 3.6) regresses estimated year-quarter indicator coefficients for each tenure category on the yearly-lagged quarterly unemployment rate.

Column (1) presents cyclicality measures using the first tenure definition, i.e., a worker accumulates tenure in a job and keeps that level if he eventually returns in the future. Real wage cyclicality declines consistently with level of tenure. For instance, a drop of one percentage point in the unemployment rate is associated to a 0.80% increase in wages for newly-hired workers, a smaller increase of 0.34% for workers with tenure between two and four years, and a meager increase of 0.14% for workers with more than six years of tenure. In column (2) I use the alternative tenure definition in which all returns to previous jobs are labeled as new spells. The cyclicality estimates decline also with the level of tenure but are slightly lower than those in column (1), particularly for newly-hired workers. Therefore, when comparing both estimates, real wage cyclicality is, as expected, greater for newly-hired workers who start a new job in which they have not acquired tenure in the past.

I have performed several alternative estimations. In column (3) I restrict the sample of newly-hireds and job-movers to those who do not migrate across provinces –recall establishment identifiers are unique within a province but a main firm identifier is not provided. In this way, I eliminate newly-hired and job-mover spells that may actually correspond to within-firm relocations. The estimated wage cyclicality for newly-hired workers is marginally greater while the one for job-movers remains unaffected. In another estimation (not shown), I have limited duration of newly-hired and job-mover spells to three months while including a separate tenure category between 4 months and one year. The estimated cyclicality levels for newly-hired workers and job-movers are virtually identical.

Lastly, as in the estimation for temporary workers, results in the second stage regression for newly-hired workers and job-movers vary only slightly if I replace the yearly-lagged quarterly unemployment rate by the current unemployment rate. Of course, this is due to the high correlation between low levels of tenure and incidence of temporary contract (see table 3.1). The estimated real wage cyclicality falls only by one fifth when we include the current unemployment rate for either newly-hireds or job-movers, while it declines by more than 60% for workers with more than six years of tenure.

Overall, workers who enter new jobs or matches –specially newly-hireds or those who come from periods of unemployment or inactivity– are the ones more exposed to the economic cycle. This is consistent with the evidence summarized in Pissarides (2009) and the findings of Haefke *et al.* (2012). However, the magnitude of the response again does not lie in the range of earlier estimates. A consensus estimate is that a percentage point decline in the unemployment rate is related to a real wage increment in new matches of 3%. Hence, the cyclicality estimate I find for newly-hired workers is at best half of the level found for other countries, though it is still substantially larger than the wage cyclicality of stayers.

Cyclicality of net present values of wages

The available estimates of wages in new matches being more cyclical than wages of job stayers have contributed to the discussion of wage rigidities and fluctuations of unemployment and vacancies along the cycle. Hall (2005) and Shimer (2005) argue that the Mortensen-Pissarides canonical search and matching model can not account for the observed cyclical volatility of unemployment and vacancies. They propose wage rigidity as a potential solution to this so called unemployment-volatility puzzle (Pissarides, 2009).

Pissarides (2009) shows that the wage flexibility that is relevant for the model to amplify unemployment fluctuations is the one in new matches. In fact, in this model, the job creation condition

Table 3.5: Cyclicality of net present value of wages

	Estimation period (1988:1–2010:4)	Estimation period (1988:1–2003:4)
NPV of wages: equivalent monthly wage	-0.793 (0.122)***	-0.744 (0.120)***
NPV of wages: NPV / match duration	-0.750 (0.096)***	-0.755 (0.111)***

Notes: Each coefficient is a separate second stage regression of the estimated year-quarter coefficients on the yearly-lagged quarterly unemployment rate. All second stage regressions include a constant term, a linear time trend and quarter indicators. Row (1) and row(2) have 92 (1988:1–2010:4) and 64 (1988:1–2003:4) quarterly observations, respectively, while rows (3), (4) and (5) have 92 quarterly observations (1988:1 to 2010:4). The number of observations in the first stage is 1,748,756 in column (1) and in column (2). ***, **, and * indicate significance at the 1, 5, and 10 percent levels.

depends on the difference between the expected productivity and cost of labor in new matches, while the level of wage flexibility in ongoing jobs is irrelevant. Using available estimates of the cyclicality in new matches, he concludes that wages in new matches are as cyclical as productivity. However, given that job creation is a forward-looking decision, the statistic that is of real interest is not necessarily the wage cyclicality in new matches but the one in the net present of value of wages over the duration of these new matches.

To this end, I calculate the net present value (NPV) of wages for each new match in period t by adding up discounted wages along the duration of the match. I use an annual discount rate of 5%. However, the NPV of wages for matches of different duration can not be compared directly. For this reason, I calculate both the equivalent annuity, i.e., the equivalent monthly wage along the duration of the match, and the NPV of wages divided by its duration. Then, I proceed to estimate the cyclicality of these measures using the two-stage method described in equations (3.5) and (3.6), but replacing log wages in the first stage by any of the two measures. In this setup I still address compositional effects along the cycle by including worker fixed-effects, yet, now the estimation drops those workers who do not switch jobs throughout 1988–2010.[21]

Table 3.5 presents the results of the estimation. The NPV of equivalent wages in new matches is as cyclical as current wages of newly-hired workers and somewhat higher than current wages of job-movers. In general, a decline of one percentage point in the unemployment rate raises the NPV of equivalent wages in a new match by 0.75%–0.79%. When I restrict the sample in column (2) to matches created before 2004 to address right censoring in durations, I find a similar semi-elasticity.[22] This finding resembles that of Haefke *et al.* (2012) who obtain using CPS data a virtually identical response of wages of newly-hireds and NPV of wages to changes in productivity. In contrast, Kudlyak (2011) finds evidence of greater cyclicality of NPV of wages relative to wages of newly-hired workers using NLSY data.

[21]Of the 528,205 workers in the initial sample, 72,541 (13.7%) do not switch jobs during the estimation period.

[22]The 90$^{\text{th}}$ percentile in the distribution of job durations between 1988–2003 (and further observed until 2010) is 91 months. Thus, I mostly get rid of right censoring in durations when applying such restriction.

The main difference between my findings and the estimates in these studies is that I exploit actual match-specific data on wages and durations and, hence, there is no need to simulate matches. This is possible given the monthly frequency of the wage data and the availability of employee and employer identifiers. Finally, given that the NPV of wages in new matches is rarely available or difficult to construct in available data sets, it is encouraging that its estimated cyclicality is quite close to that of current wages of newly-hired workers.

6. Conclusions

This paper provides estimates of real wage cyclicality in Spain during the period 1988–2010. I find wages are weakly procyclical with a small increment of 0.4% in response to a decline of one percentage point in the unemployment rate. Earlier studies that estimate wage cyclicality using worker-level data do so for countries with more flexible labor markets, mainly the United States and UK. Therefore, this result confirms that in countries with institutions that deter firms to adjust wages to business cycle conditions –Spain being one good example–, real wage cyclicality is weaker than previously thought.

The results I find on different wage cyclicality estimates for workers under permanent and temporary contracts validate the duality in the Spanish labor market. The estimated real wage cyclicality for temporary workers is four to five times larger than for permanent workers. Also, their wages react to the current unemployment rate suggesting a more immediate response to the economic cycle. Although I do not examine differences in extensive margin cyclical responses between workers in permanent and temporary contracts, it is documented that the latter also face higher turnover rates (Dolado *et al.*, 2002). Therefore, policies in favor of reducing disparities between both types of contract should generate a more balanced response of wages to business cycles conditions for both types of workers.

Finally, I provide evidence that wage cyclicality decreases consistently with the level of tenure. My finding that wages of newly-hired workers are much more volatile than wages of job stayers does not support wage rigidity as a solution to the unemployment volatility puzzle. In the Mortensen-Pissarides search and matching model the key piece of information to determine the number of jobs created is the cyclicality of the net present value of wages in new matches. I obtain a cyclicality level for this net present value of wages in new matches of the same order as the one I find for wages of newly-hired workers.

Appendix 3.A. Fit of simulated earnings

To examine the fit of the simulation, I compare simulated wages in top- and bottom-coded observations from MCVL and actual uncensored total salaries from income tax returns for the same individual and month in those years where both are available (2004–2010). If the fit is reasonably close for 2004–2010, I can be more confident that simulated wages accurately approximate capped earnings for 1981–2003.

The correlation between simulated wages and actual salaries for capped month-individual observations in 2004–2010 is high at 0.75. Simulated wages can reproduce well the overall shape

Table 3.6: Selected percentiles for actual and simulated wages

	All workers		Skilled workers	
	Total salaries	Simulated wages	Total salaries	Simulated wages
Percentile 5	47.1	46.9	37.2	38.0
Percentile 10	52.9	53.0	43.6	44.7
Percentile 25	62.2	62.2	58.4	60.2
Percentile 50	78.1	78.1	83.1	85.8
Percentile 75	112.1	113.9	121.6	123.5
Percentile 90	168.9	171.2	172.9	172.6
Percentile 95	222.9	225.5	220.1	210.9

Notes: Monthly wages expressed as a percentage of the average in each category. Skilled individuals work in the top three out of ten social security occupations demanding high and very-high skills.

Table 3.7: Order of autocorrelations for actual and simulated wages

Order	Total salaries	Simulated wages
1	0.937	0.922
2	0.906	0.887
3	0.882	0.861
4	0.861	0.842
5	0.846	0.826
6	0.831	0.811

of the salaries distribution. This can be seen in table 3.6, which presents selected percentiles of the distributions of salaries and simulated wages for all workers and for skilled workers (those in the top three occupation categories in social security). Overall, the distributions are quite similar. More important, for skilled workers, who are top-coded beyond the 51st percentile, simulated wages approximate salaries quite well in capped percentiles.

Table 3.7 displays estimated order of autocorrelations for total salaries and simulated wages. Although salaries exhibit a higher degree of persistence, the level of persistence observed in MCVL after including simulated wages provides a good approximation.

Concluding remarks

The core of this doctoral dissertation examines 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, these cities may allow workers to accumulate more valuable experience. Third, workers who are inherently more productive may choose to locate there. Using a large and rich panel data set for workers in Spain, I provide a quantitative assessment of the importance of each of these three mechanisms in generating earnings differentials across cities of different sizes.

The results show there are substantial static and dynamic advantages from working in bigger cities. About one-half of these gains are static and tied to currently working in a big city. The other half accrues over time as workers accumulate more valuable experience in big cities. Furthermore, workers are able to take these dynamic gains with them when they relocate, which suggests that learning in big cities is important.

Although the highly-skilled and productive workers in small cities migrate to the big cities, most of this selection can be accounted by observables such as educational attainment or occupational skills. When I compare the distributions of innate ability across cities of different sizes, i.e., the level of ability prior to any accrued dynamic benefits in big cities, the distributions are remarkably similar. Sorting of more able workers into big cities plays at best a minor role in explaining earnings differentials.

I 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 working in bigger cities, explain why the observed distribution of earnings in bigger cities has a higher mean and variance.

The evidence provided on the relative importance of these three mechanisms should guide further research to focus on micro-founded theories that can explain how static and dynamic advantages generate higher earnings in big cities. These two mechanisms are the main drivers of the observed city-size earnings premium.

Bibliography

Ambrosini, J. William, Karin Mayr, Giovanni Peri, and Drados Radu. 2011. The selection of migrants and returnees: Evidence from Romania and implications. Working Paper 16912, National Bureau of Economic Research.

Antolín, Pablo and Olimpia Bover. 1997. Regional migration in Spain: The effect of personal characteristics and of unemployment, wage and house price differentials using pooled cross-sections. *Oxford Bulletin of Economics and Statistics* 59(2):215–235.

Ashenfelter, Orley. 1978. Estimating the effect of training programs on earnings. *Review of Economics and Statistics* 60(1):47–50.

Babecký, Jan, Philip Du Caju, Theodora Kosma, Martina Lawless, Julián Messina, and Tairi Room. 2010. Downward nominal and real wage rigidity: Survey evidence from european firms. *Scandinavian Journal of Economics* 112(4):884–910.

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*. Inequality and city size. Processed, Brown University.

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

Bentolila, Samuel, Mario Izquierdo, and Juan F. Jimeno. 2010. Negociación colectiva: La gran reforma pendiente. *Papeles de Economía Española* 124:176–192.

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

Bils, Mark J. 1985. Real wages over the business cycle: Evidence from panel data. *Journal of Political Economy* 93(4):666–689.

Bjorklund, Anders. 1993. A comparison between actual distributions of annual and lifetime income: Sweden 1951–89. *Review of Income and Wealth* 39(4):377–386.

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

Blanchard, Olivier Jean and Lawrence F. Katz. 1992. Regional evolutions. *Brooking Papers on Economic Activity* 23(1):1–76.

Borjas, George J. and Bernt Bratsberg. 1996. Who leaves? The outmigration of the foreign-born. *Review of Economics and Statistics* 78(1):165–76.

Borjas, George J., Stephen G. Bronars, and Stephen J. Trejo. 1992. Self-selection and internal migration in the United States. *Journal of Urban Economics* 32(2):159–185.

Bound, John and John J. Holzer. 2000. Demand shifts, population adjustments, and labor market outcomes during the 1980s. *Journal of Labor Economics* 18(1):20–54.

Bover, Olympia and Ramón Gómez. 2004. Another look at unemployment duration: exit to a permanent vs. a temporary job. *Investigaciones Económicas* 28(2):285–314.

Carneiro, Anabela, Paulo Guimaraes, and Pedro Portugal. 2012. Real wages and the business cycle: Accounting for worker and firm heterogeneity. *American Economic Journal: Macroeconomics* 4(2):133–152.

Chiquiar, Daniel and Gordon H. Hanson. 2005. International migration, self-selection, and the distribution of wages: Evidence from Mexico and the United States. *Journal of Political Economy* 113(2):239–281.

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

Co, Catherine Y., Ira N. Gang, and Myeong-Su Yun. 2000. Returns to returning. *Journal of Population Economics* 13(1):57–79.

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

Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon. 2011*a*. The costs of agglomeration: Land prices in French cities. Processed, University of Toronto.

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

Combes, Pierre-Philippe, Gilles Duranton, Laurent Gobillon, Diego Puga, and Sébastien Roux. 2011*c*. 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.

Constant, Amelie and Douglas S. Massey. 2003. Self-selection, earnings, and out-migration: A longitudinal study of immigrants to Germany. *Journal of Population Economics* 16(4):631–653.

CRUE. 2002–2010. *La Universidad Española en Cifras*. Madrid: Conferencia de Rectores de las Universidades Españolas.

Cuadrado, Pilar, Pablo Hernándes de Cos, and Mario Izquierdo. 2011. El ajuste de los salarios frente a las perturbaciones de España. Economic Bulletin 2, Banco de España.

DaVanzo, Julie. 1983. Repeat migration in the United States: Who moves back and who moves on? *Review of Economics and Statistics* 65(4):552–59.

De la Roca, Jorge and Diego Puga. 2012. Learning by working in big cities. Processed, CEMFI.

DeCoulon, Augustin and Matloob Piracha. 2005. Self-selection and the performance of return migrants: The source country perspective. *Journal of Population Economics* 18(4):779–807.

Devereux, Paul J. 2001. The cyclicality of real wages within employer-employee matches. *Industrial & Labor Relations Review* 54(4):835–850.

Devereux, Paul J. and Robert A. Hart. 2006. Real wage cyclicality of job stayers, within-company job movers, and between-company job movers. *Industrial & Labor Relations Review* 60(1):105–119.

Dolado, Juan J., Carlos García-Serrano, and Juan F. Jimeno. 2002. Drawing lessons from the boom of temporary jobs in spain. *Economic Journal* 112(721):270–295.

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.

Dustmann, Christian. 2003. Return migration, wage differentials, and the optimal migration duration. *European Economic Review* 47(2):353–369.

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.

Estrada, Ángel and Mario Izquierdo. 2005. La producción y el mercado de trabajo. In Servicio de Estudios del Banco de España (ed.) *El análisis de la economía española*. Madrid: Alianza Editorial, 347–377.

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

Fernández-Huertas, Jesús. 2011. New evidence on emigrant selection. *Review of Economics and Statistics* 93(1):72–96.

García-Pérez, José Ignacio. 2008. La muestra continua de vidas laborales: una guía de uso para el análisis de transiciones. *Revista de Economía Aplicada* 26(1):5–28.

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.

Goss, Ernst P. and Chris Paul. 1990. The impact of unemployment insurance benefits on the probability of migration of the unemployed. *Journal of Regional Science* 30(3):349–358.

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

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

Haefke, Christian, Marcus Sonntag, and Thijs van Rens. 2012. Wage rigidity and job creation. Discussion Paper 8968, Centre for Economic Policy Research.

Haider, Steven and Gary Solon. 2006. Life-cycle variation in the association between current and lifetime earnings. *American Economic Review* 96(4):1308–1320.

Hall, Robert E. 2005. Employment fluctuations with equilibrium wage stickiness. *American Economic Review* 95(1):50–65.

Holden, Steinar and Fredrik Wulfsberg. 2008. Downward nominal wage rigidity in the OECD. *The B.E. Journal of Macroeconomics* 8(1):15.

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

Hunt, Jennifer. 2004. Are migrants more skilled than non-migrants? Repeat, return, and same-employer migrants. *Canadian Journal of Economics* 37(4):830–849.

Ibarrarán, Pablo and Darren Lubotsky. 2007. Mexican immigration and self-selection: New evidence from the 2000 Mexican census. In *Mexican Immigration to the United States*. Cambridge, MA: NBER, 159–192.

Izquierdo, Mario and Pilar Cuadrado. 2009. La encuesta sobre formación de salarios de las empresas españolas: nueva evidencia sobre la relación entre precios y salarios, y la respuesta de las empresas a perturbaciones económicas. Economic Bulletin 3, Banco de España.

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.

Jenkins, Stephen P. 1995. Easy estimation methods for discrete-time duration models. *Oxford Bulletin of Economics and Statistics* 57(1):129–138.

Kennan, John and James R. Walker. 2011. The effect of expected income on individual migration decisions. *Econometrica* 79(1):211–251.

Kudlyak, Marianna. 2011. The cyclicality of the user cost of labor with search and matching. Working Paper 09-12, Federal Reserve Bank of Richmond.

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

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

Mortensen, Dale T. and Christopher A. Pissarides. 1994. Job creation and job destruction in the theory of unemployment. *Review of Economic Studies* 61(3):397–415.

Moulton, Brent R. 1986. Random group effects and the precision of regression estimates. *Journal of Econometrics* 32(3):385–397.

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

Pissarides, Christopher A. 2000. *Equilibrium Unemployment Theory, 2nd Edition*. Cambridge, MA: MIT Press.

Pissarides, Christopher A. 2009. The unemployment volatility puzzle: Is wage stickiness the answer? *Econometrica* 77(5):1339–1369.

Puga, Diego. 2010. The magnitude and causes of agglomeration economies. *Journal of Regional Science* 50(1):203–219.

Reinhold, Steffen. 2009. Temporary migration and skill upgrading: Evidence from Mexican migrants. Discussion Paper 09182, Mannheim Research Institute for the Economics of Aging.

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

Rooth, Dan-Olof and Jan Saarela. 2007. Selection in migration and return migration: Evidence from micro data. *Economics Letters* 94(1):90–95.

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.

Shimer, Robert. 2005. The cyclical behavior of equilibrium unemployment and vacancies. *American Economic Review* 95(1):25–49.

Shin, Donggyun. 1994. Cyclicality of real wages among young men. *Economics Letters* 46(2):137–142.

Solon, Gary, Robert Barsky, and Jonathan A. Parker. 1994. Measuring the cyclicality of real wages: How important is composition bias? *The Quarterly Journal of Economics* 109(1):1–25.

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.

Wheaton, William C. and Mark J. Lewis. 2002. Urban wages and labor market agglomeration. *Journal of Urban Economics* 51(3):542–562.