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

Jorge De la Roca*†
CEMFI and *IMDEA Social Sciences Institute*

15 November 2011

ABSTRACT: This paper investigates the sorting of more productive workers into denser cities using administrative data for Spain that follow individuals continuously throughout their working lives. Migrants who move to denser cities are positively selected in terms of education, occupational skills, and individual productivity as proxied by pre-migration position in the local earnings distribution. However, not everyone is able to benefit equally from denser cities and this leads to a second round of sorting. Returnees are not only ex-ante less productive than permanent migrants, but are also those who, following the first move, have least boosted up their earnings in denser cities.

Key words: selection, urban migration, return migration, skill sorting
JEL classification: J61, R10, R23

*I thank Diego Puga for excellent supervision throughout this project and Manuel Arellano and Stéphane Bonhomme for helpful conversations. Enzo Cerletti, Pablo Lavado, Anatoli Segura and seminar participants also contributed with valuable comments. Funding from the Bank of Spain Excellence Programme is gratefully acknowledged. This work contains anonymized statistical data from MCVL CDF 2004–2009 which are used with the permission of Spain's Dirección General de Ordenación de la Seguridad Social.

†Centro de Estudios Monetarios y Financieros (CEMFI), Casado del Alisal 5, 28014 Madrid, Spain (e-mail: jorge.delaroca@imdea.org).

1. Introduction

Workers earn substantially more in larger and denser cities (Glaeser and Maré, 2001, Wheaton and Lewis, 2002, Combes, Duranton, Gobillon, and Roux, 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 denser cities facilitate the acquisition of greater skills (Glaeser, 1999, Gould, 2007, Baum-Snow and Pavan, 2010, De la Roca and Puga, 2011). However, it has long been thought that those higher earnings may also partly reflect the sorting of more productive workers into denser 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 denser cities in particular, by regressing individual earnings on worker characteristics and location fixed effects (Glaeser and Maré, 2001, Combes, Duranton, and Gobillon, 2008, Combes *et al.*, 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 (2011) show that the introduction of worker fixed effects, in addition to absorbing unobserved worker heterogeneity, also largely removes the dynamic component of the urban density 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 denser 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, Blum, and Strange, 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 denser city. Using administrative data for Spain that follow individuals continuously over time and across cities throughout their working lives, I show that migrants who move to dense 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 one-half of migrants end up leaving their city of destination within five years. Moreover, around 60% of these second moves involve a return migration to the city of origin. Such return migration is more frequent and happens sooner in high-density 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 high-density 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 high-density cities and the costs of migration. As a result, of all workers in low-density cities, only those with skills above a certain threshold are willing to migrate to high-density cities. Then, of workers who migrate, those with the highest skills remain in high-density 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 denser 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 dense 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. For instance, Borjas, Bronars, and Trejo (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 labour 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 low to high density 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

[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 betwen 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 return migration in European countries.

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 4 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 large 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 labour 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

$$U_i^L = s_i \ . \tag{1}$$

City H is characterized by three differences with respect to 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

$$U_i^H = \alpha(s_i + \delta_i) - R \ . \tag{2}$$

[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, which are briefly discussed below.

[4]Glaeser (1999) develops a learning model where young workers who move into a dense city increase their skills with some probability. De la Roca and Puga (2011) find evidence of substantial skill acquisition by workers in densest cities. On the firm side, Duranton and Puga (2001) develop a model in which dense cities are diversified places that foster innovation and experimentation. Firms can only find their optimal production process in dense 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.

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

$$s_i \leqslant \hat{s} = \frac{R + C - \alpha\delta}{\alpha - 1} \, . \tag{3}$$

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 a further 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. We now solve the model and draw those predictions explicitly.

Solution

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 $\alpha(s_i + \delta_i) - R \leq s_i - C_2$. Thus, a worker returns if ex-post earnings in H are lower than earnings in L minus the return migration cost.[7] 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

[7] 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.

of δ_i. In particular, a worker who migrates to city H returns to city L if and only if

$$\delta_i < \underline{\delta}(s_i) = \begin{cases} 2\delta & \text{if} \quad s_i < \underline{s}\,, \\ \frac{R-C_2-(\alpha-1)s_i}{\alpha} & \text{if} \quad \underline{s} \leqslant s_i < \overline{s}\,, \\ 0 & \text{if} \quad s_i \geqslant \overline{s}\,, \end{cases} \tag{4}$$

where

$$\underline{s} = \frac{R - C_2 - 2\alpha\delta}{\alpha - 1}\,, \tag{5}$$

$$\overline{s} = \frac{R - C_2}{\alpha - 1}\,. \tag{6}$$

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

$$\int_0^{\underline{\delta}(s_i)} \frac{1}{2\delta}(s_i - C_2)\,\mathrm{d}x + \int_{\underline{\delta}(s_i)}^{2\delta} \frac{1}{2\delta}[\alpha(s_i + x) - R]\,\mathrm{d}x - C > s_i \tag{7}$$

For workers with $s_i < \underline{s}$, the condition of equation (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 (4) into (7) and simplifying turns this condition into

$$s_i > \frac{R - C_2 - 2(\alpha\delta + \sqrt{(C + C_2)\,\alpha\delta})}{\alpha - 1}\,. \tag{8}$$

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 (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 (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.[8]

To summarize the results:

- Workers with low initial skills ($s_i \leqslant \frac{R-C_2-2(\alpha\delta+\sqrt{(C+C_2)\,\alpha\delta})}{\alpha-1}$) do not migrate from city L to H.
- Workers with intermediate initial skills ($\frac{R-C_2-2(\alpha\delta+\sqrt{(C+C_2)\,\alpha\delta})}{\alpha-1} < s_i < \frac{R-C_2}{\alpha-1}$) 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.

[8]If instead $C + C_2 > \alpha\delta$ then $\hat{s} > \overline{s}$ and equation (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.

- 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.

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 migrate to high-density cities. Second, among these migrants, those with the highest initial skills stay in the high-density city while those with intermediate skills return provided they only get an unfavourable earnings boost. Yet, the probability of return is not random, but decreases both with their initial skill/earnings level in the low-density city and with their earnings gain in the high-density 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:

$$h(t) = P[T = t | T \geq t, x(t)] = F[\beta_0(t) + \beta_1' x(t)] \ , \tag{9}$$

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 modelling for an individual working in his 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:

$$L(\beta) = \sum_{i=1}^{N} \left[(1-m_i) \sum_{t=e_i}^{T_i} \log\left(1-h_i(t)\right) + m_i \left(\sum_{t=e_i}^{T_i-1} \log\left(1-h_i(t)\right) + \log h_i(T_i) \right) \right] \tag{10}$$

where i indexes the individual, m_i is an indicator variable which takes value 1 in the last month prior to migration and 0 otherwise, e_i is the month of entry in the sample which usually will correspond to the age of entry in the labour 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,

$$L(\beta) = \sum_{t=1}^{T_i} \left\{ \sum_{i=1}^{N} 1(T_i \geq t \geq e_i) \left[m_i \log h_i(t) + (1 - m_i) \log(1 - h_i(t)) \right] \right\} \tag{11}$$

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 (11) 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 high-density city and initial migration to a low-density 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 low-density city or moving to a high-density city at time t conditional on not having done either before. An alternative possibility is to model conditional hazard rates, i.e. model the probability of moving to a high-density city at time t conditional on not having done so before and on not having moved to a low-density 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 (11) 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.[9] 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.

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 labour 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

[9]One difference between both specifications is how I introduce duration-specific parameters to capture duration dependence. In equation (11), 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.

not, in practice we need the data to record the labour market characteristics of all workers with high frequency since the start of their first job.

The *Muestra Continua de Vidas Laborales* (MCVL), or Longitudinal Sample of Working Lives, 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 labour market status and work characteristics since 1981 are recorded. I combine data from all available editions of the MCVL, from 2004 to 2009, so as to have data on a 4% sample of all individuals who have worked, received benefits or a pension at any point in 2004–2009. 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 labour 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 area deceased). The unit of observation in the source data is any change in the individual's labour 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, and the 3-digit SIC 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 labour market 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 2004–2009, uncensored earnings data is 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 1981–2009 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.[10]

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 Contribution Account 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 firm-establishment level.

[10]Appendix A provides details on how these sources are combined and how earnings for the small fraction (8%) of capped observations is estimated, based on uncensored observations, on the magnitude of Social Security contributions, and individual and job characteristics.

Urban areas

I use official urban area definitions by Spain's Ministry of Housing for 2008. The 85 urban areas in Spain account roughly for 68% of population and 10% of total surface. They represent local labour markets comparable to Metropolitan Statistical Areas (MSAs) in the United States. The mean urban area has a population of 368,659 inhabitants in 2008.

Urban areas enclose 747 municipalities. Given that I know the municipaliy 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.[11]

To measure the scale of an urban area I use the number of people within 10 kilometres of the average resident in the urban area, a measure proposed by De la Roca and Puga (2011). 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.[12] In any case, results are robust to measuring the scale of each urban area by its total population.[13]

Sample restrictions

My initial sample is made up of males born in Spain between 1963 and 1991 (i.e., aged 18–46 during the period 1981–2009) who have been employed or received unemployment benefits at any point over this period. I leave out individuals older than 46 in 2009 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 labour market. Finally, I leave out self-employed workers because their workplace location is not available in MCVL. A total of 290,301 individuals and 34,047,072 monthly observations make up this initial sample.

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

[12] 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 out areas with medium population density adjacent to their border, which makes population per surface area unit artificially high for them. Calculating the number of people within 10 kilometres of the average resident largely gets around both problems.

[13] The correlation between the number of people within 10 kilometres 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.

From this initial sample, I eliminate individuals with low labour 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 2009. This restriction reduces the sample to 266,847 individuals and 33,870,056 observations.

Subsequently, 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 264,407 individuals and 32,174,175 observations.

Next, I exclude observations for which the occupation or workplace location is missing and individuals for whom the educational attainment is missing. This leaves 258,254 individuals and 30,738,787 monthly observations.

Finally, since I wish to focus on urban migration (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 203,902 individuals and 19,631,181 monthly observations.

Identifying migrants

A migration event is defined as a change in workplace location from one urban area to another. In the sample 62,505 individuals can be classified as urban migrants while 141,397 individuals never leave their urban area for working purposes.

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 often not kept up-to-date in the data, thus I detect permanent changes in residence based on the length of the migration episode and the distance between the 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 temporal job. I classify migrants as *short-term* if they never move beyond a 12-month period. Therefore, long-term migrants are movers who experience spells longer than a year in the city of destination. Based on this criterion, I identify 28,990 short-term migrants.

Regarding distance, within the sample of long-term migrants, I label migrants as *short-distance* if they move to an urban that is less than 120 km. (74.6 miles) driving from the urban area where they previously worked. Although urban areas can be understood as independent local labour markets, in some cases two or more of them may exhibit substantial overlapping in worker flows. This pattern is more prevalent in larger 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 14,438 short-distance migrants.

Table 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	12%	10%	16%	20%
Secondary	34%	29%	34%	37%
Primary	54%	61%	50%	43%
Occupational skills				
Very-high skills	6%	5%	8%	12%
High skills	8%	7%	11%	13%
Medium skills	14%	11%	17%	17%
Low skills	53%	53%	48%	44%
Very-low skills	18%	24%	14%	14%
Labour market characteristics				
Mean monthly earnings	1,874	1,632	2,114	2,238
Mean monthly earnings 2^{nd} location			2,094	2,557
Years of labour experience	7.4	5.9	7.1	6.9
Years of firm tenure	3.3	1.8	1.7	2.0
Unemployed	10%	16%	15%	13%
Temporary contract	26%	38%	32%	32%
Part-time contract	7%	7%	5%	6%
Age	29.7	28.8	31.1	30.8
Age of entry in labour force	20.5	20.2	20.7	21.3
Individuals	141,397	43,428	4,029	15,048

Notes: Variables are averages for individuals over their working lives. For stayers, only individuals working in urban areas are included. For migrants, only urban migrations are considered. Long-term long-distance migrations are moves that exceed 12 months in destination and distance of 120 km. Earnings expressed in December 2009 euros.

Table 1 shows summary statistics for non-migrants (stayers), short-term or 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).[14]

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-

[14]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.

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, not very far, by returnees. Stayers and short-term/distance migrants earn substantially less, the gap being larger for the latter. Among long-term and long-distance migrants, those who eventually return have lower earnings in their second location than those who do not return.

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

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.[15] In table 2 I estimate the probability of outmigration from the individual's first job location using a single-exit discrete duration model as in equation 11, 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 below.

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 84% 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 50%.

Other labour market variables reveal expected signs. An additional year of labour 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 50% more likely 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

[15]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.

Table 2: Logit estimation of determinants of first migration

	Dep. variable: long-term long-distance migration			
	(1)	(2)	(3)	(4)
Log mean earnings		1.655 (0.174)***		1.230 (0.080)***
Richest earnings tercile			1.398 (0.117)***	
Poorest earnings tercile			0.881 (0.025)***	
Tertiary education	1.839 (0.380)***			1.776 (0.359)***
Secondary education	1.364 (0.117)***			1.342 (0.113)***
Very-high skills	1.715 (0.138)***			1.511 (0.126)***
High skills	1.505 (0.082)***			1.383 (0.081)***
Medium skills	1.526 (0.099)***			1.467 (0.086)***
Low skills	1.137 (0.033)***			1.120 (0.033)***
Years of experience	0.898 (0.007)***	0.867 (0.004)***	0.868 (0.004)***	0.894 (0.007)***
Years of firm tenure	0.914 (0.013)***	0.917 (0.011)***	0.917 (0.011)***	0.911 (0.013)***
Temporary contract	1.561 (0.037)***	1.521 (0.038)***	1.509 (0.037)***	1.583 (0.037)***
Unemployed	1.062 (0.078)	0.756 (0.053)***	0.761 (0.049)***	1.045 (0.075)
Unemployed, expired benefits	8.975 (0.446)***	9.391 (0.455)***	9.363 (0.452)***	9.044 (0.441)***
Urban area $\times$ period indicators	Yes	Yes	Yes	Yes
Age indicators	Yes	Yes	Yes	Yes
Pseudo R^2	0.068	0.065	0.064	0.068

Notes: Odd ratios (exponentiated coefficients) are reported on a sample of 11,881,792 monthly observations and 169,761 individuals. Standard errors in parentheses clustered at the urban area level. ***, **, and * indicate significance at the 1, 5, and 10 percent levels. The reference category is stayers. Sample is all individuals who are still in their first city. 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 12-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.

benefits, I find they are the ones driving this effect.[16] 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. The discouraging effect of unemployment benefits for mobility has been previously 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

[16] 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).

Table 3: Logit estimation of determinants of first migration to high-density cities

	Dep. variable: long-term long-distance migration to any of 6 densest cities			
	(1)	(2)	(3)	(4)
Log mean earnings		2.358 (0.135)***		1.511 (0.113)***
Richest earnings tercile			1.748 (0.076)***	
Poorest earnings tercile			0.898 (0.052)*	
Tertiary education	3.682 (0.199)***			3.469 (0.198)***
Secondary education	2.027 (0.081)***			1.963 (0.083)***
Very-high skills	1.771 (0.151)***			1.400 (0.120)***
High skills	1.594 (0.137)***			1.354 (0.101)***
Medium skills	1.959 (0.117)***			1.816 (0.109)***
Low skills	1.121 (0.055)**			1.086 (0.053)*
Years of experience	0.916 (0.006)***	0.858 (0.007)***	0.861 (0.007)***	0.910 (0.007)***
Years of firm tenure	0.890 (0.007)***	0.898 (0.007)***	0.902 (0.008)***	0.882 (0.007)***
Temporary contract	1.512 (0.072)***	1.430 (0.065)***	1.427 (0.066)***	1.545 (0.073)***
Unemployed	1.279 (0.117)***	0.681 (0.062)***	0.689 (0.064)***	1.243 (0.117)**
Unemployed, expired benefits	8.576 (0.618)***	9.341 (0.675)***	9.264 (0.663)***	8.695 (0.629)***
Urban area $\times$ period indicators	Yes	Yes	Yes	Yes
Age indicators	Yes	Yes	Yes	Yes
Pseudo R^2	0.090	0.078	0.076	0.091

Notes: Odd ratios (exponentiated coefficients) are reported on a sample of 8,952,176 monthly observations and 137,908 individuals. Standard errors in parentheses clustered at the urban area level. ***, **, and * indicate significance at the 1, 5, and 10 percent levels. The reference category is stayers. Sample is all individuals who are still in their first city. 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 densest cities *and* migrants experience an increment in density. All specifications include month indicator variables. Period is a ten-year interval. *Log mean earnings* are 12-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 Spain has a detrimental impact on the efficient matching of unemployeds and vacancies across different locations. If monthly benefits, instead of being almost constant over time, were highest immediately after losing a job and then decreased gradually, the incentives for mobility, and for active job search more generally, could be greatly enhanced.

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 narrow the analysis of migrants and stayers within an urban area and time period. 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 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 year to proxy for workers observable and unobservable skills.[17] 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 2.4%.[18] 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 40%, while being in the lowest one decreases the probability by 12%.

I bring in both observable skills and earnings in column (4). Even within given levels of education and occupational skills, higher levels of earnings increase the probability of migrating. However, the effect of earnings has been reduced by about two-thirds relative to column (2) — recall coefficients are odd ratios — whereas observable skills 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.

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 low-density to high-density cities. In table 3 I estimate the conditional hazard rate of moving to one of the six densest cities in Spain, i.e., the probability of moving to one of these high-density cities among those who have not moved before and do not move to low-density cities. For migrants who move within the six densest cities I restrict the analysis only to moves that involve an increment in density. Other than this, all specifications are identical to those in table 2. The results show that the positive selection of migrants is much stronger when I look only at those who migrate to high-density 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 dense cities is more than three times more likely to occur for individuals with tertiary education. Having secondary education also raises substantially the odds, making migration to high-density cities twice as likely

[17] I construct 12-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 twelve most recent where the worker has been employed, excluding unemployment spells and unemployment benefits from the calculation. This is because productivity is best captured with a measure of earnings that excludes unemployment benefits. Alternatively, I have constructed a log 12-month moving average of income including these benefits. The qualitative results do not change, though as expected, point estimates are lower but still significant. Results available upon request.

[18] 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 65.5%. The 2.4% reported in the text is calculated as $10\% \times (1.655 - 1)/e$.

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 5.0%. Column (4) shows that, even within given levels of education and occupational skills, more productive workers are more likely to migrate to high-density cities. In contrast, if I repeat the estimation for migration to low-density (as opposed to high-density) cities (see appendix C), I find no significant selection (if anything, selection into low-density cities of the workers with low occupational skills).

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 the raw data pointed out that permanent migrants have higher earnings and are more educated than returnees. In table 4 I narrow 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 dense cities are more skilled or productive than those who move elsewhere. I run pooled OLS regressions where the dependent variable is 12-month moving average of log mean 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 labour market characteristics. In column (1) I divide both migrant categories into those who move to the six densest cities and those who move elsewhere. Again, for migrants who move within the six densest cities I keep only those moves that involve an increment in density. Both permanent and return migrants who move to the densest 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 13% and 6% higher than those of stayers, respectively. Therefore, I confirm the skill ranking found in the raw data (table 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 2% while eventual returnees exhibit similar pre-move earnings as stayers with comparable education and occupational skills. This result verifies that sorting of the most productive workers into denser 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 high-density cities into those who move to the two densest cities (Madrid and Barcelona) and those who move to the 3rd - 6th densest cities (Valencia, Seville, Bilbao and Zaragoza). The idea is to investigate whether the degree of sorting increases with the density of the city of 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 favour of this argument. Both permanent and return migrants moving to the top 2 and 3rd - 6th densest cities exhibit similar pre-move earnings. Moreover, when I add in observable skills in column (4), the earnings of migrants to

Table 4: Earnings of first-time migrants relative to stayers by city of destination

	Dependent variable: log mean earnings in first city			
	(1)	(2)	(3)	(4)
Permanent migrant to 6 densest cities	0.133 (0.013)***	0.022 (0.012)*		
Permanent to 1st - 2nd densest cities			0.133 (0.015)***	0.013 (0.015)
Permanent to 3rd - 6th densest cities			0.132 (0.017)***	0.049 (0.015)***
Permanent to other cities	0.048 (0.014)***	0.003 (0.007)	0.048 (0.014)***	0.003 (0.007)
Return migrant to 6 densest cities	0.061 (0.014)***	-0.002 (0.011)		
Return to 1st - 2nd densest cities			0.057 (0.018)***	-0.012 (0.014)
Return to 3rd - 6th densest cities			0.076 (0.017)***	0.034 (0.016)**
Return to other cities	0.028 (0.016)*	-0.014 (0.006)**	0.028 (0.016)*	-0.014 (0.006)**
Years of experience	-0.005 (0.003)	0.016 (0.001)***	-0.005 (0.003)	0.016 (0.001)***
Years of firm tenure	0.014 (0.001)***	0.010 (0.000)***	0.014 (0.001)***	0.010 (0.000)***
Unemployed	-0.177 (0.037)***	0.085 (0.009)***	-0.177 (0.037)***	0.085 (0.009)***
Temporary contract	-0.118 (0.021)***	-0.055 (0.009)***	-0.118 (0.021)***	-0.055 (0.009)***
Tertiary education		0.226 (0.026)***		0.226 (0.026)***
Secondary education		0.119 (0.006)***		0.119 (0.006)***
Very-high skills		0.688 (0.030)***		0.688 (0.030)***
High skills		0.455 (0.010)***		0.455 (0.010)***
Medium skills		0.222 (0.006)***		0.222 (0.006)***
Low skills		0.080 (0.007)***		0.080 (0.007)***
Urban area $\times$ period indicators	Yes	Yes	Yes	Yes
Age indicators	Yes	Yes	Yes	Yes
R^2	0.304	0.499	0.304	0.499

Notes: Coefficients reported on a sample of 11,881,792 monthly observations and 169,761 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 12-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 densest cities, only those moves that involve an increment in density are considered. All specifications include month indicator variables. Period is a ten-year interval.

the two densest cities are in fact slightly lower than those of migrants to the 3rd - 6th densest cities with comparable education and occupational skills.[19]

[19] I have considered alternative specifications for columns (3) and (4) by including increments in density between the city of origin and destination and its square, city-specific destination indicator variables interacted with permanent and return migrants' categories, among others. None of these specifications indicate an increase in the extent of sorting increases with the density 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 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 transion 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 density of the city they relocated to. Then, I focus only on migrants who moved to the six densest 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 add urban area indicators for both the first city and the city of destination. 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 high-density cities. In column (3a) a 10% increase in earnings at the first location (i.e., prior to the first migration episode) makes return migration 0.7% less likely[20]. 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 1.5% less likely, after controlling for average earnings in the first city and labour 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 first migration, observable skills are strong determinants of return migration. Having tertiary education reduces the probability of returning by 31% while having very-high occupational skills decreases the odds of returning by 29%. However, even within education categories and occupational skills, higher realized earnings in dense 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 high-density cities. When I look at all returnees who initially moved to any city (not necessarily high-density), realized earnings in destination do not influence their return decision once I include observable skills (column 2a). In addition, when I examine the migration decision for repeat migrants who do not return but move on to a third city, this is not affected by realized earnings (column 4b). In sum, returnees from high-density 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.

[20] $10\% \times (0.803 - 1)/e = -0.7\%$

Table 5: Multinomial logit estimation of determinants of second migration

	First move to city of any density				First move to any of 6 densest cities			
	Return	Move on	Return	Move on	Return	Move on	Return	Move on
	(1a)	(1b)	(2a)	(2b)	(3a)	(3b)	(4a)	(4b)
Log mean earnings 2^{nd} loc.	0.840 (0.087)*	1.174 (0.053)***	0.969 (0.060)	1.153 (0.056)***	0.593 (0.033)***	0.974 (0.080)	0.716 (0.057)***	0.935 (0.083)
Log mean earnings 1^{st} loc.	0.775 (0.042)***	1.183 (0.069)***	0.847 (0.051)***	1.176 (0.069)***	0.803 (0.059)***	1.158 (0.119)	0.905 (0.072)	1.155 (0.122)
Unemployed in 2^{nd} location	2.053 (0.184)***	1.825 (0.164)***	1.817 (0.204)***	1.912 (0.172)***	1.689 (0.289)***	1.948 (0.336)***	1.449 (0.241)**	1.987 (0.375)***
Unemployed exp.ben. 2^{nd} loc.	5.487 (0.708)***	8.880 (1.079)***	5.540 (0.711)***	8.927 (1.086)***	7.462 (1.885)***	7.831 (2.257)***	7.600 (1.933)***	7.902 (2.288)***
Unemployed in 1^{st} location	2.169 (0.231)***	1.309 (0.199)*	2.066 (0.227)***	1.333 (0.204)*	1.532 (0.255)**	1.274 (0.288)	1.441 (0.246)**	1.264 (0.297)
Temporary contract 2^{nd} loc.	1.320 (0.061)***	1.385 (0.079)***	1.306 (0.056)***	1.416 (0.083)***	1.298 (0.088)***	1.328 (0.095)***	1.296 (0.085)***	1.345 (0.101)***
Temporary contract 1^{st} loc.	1.148 (0.080)**	1.099 (0.096)	1.157 (0.083)**	1.096 (0.095)	1.141 (0.126)	1.100 (0.157)	1.178 (0.134)	1.105 (0.161)
Years of experience	1.035 (0.009)***	0.986 (0.009)	1.020 (0.007)***	0.988 (0.009)	1.041 (0.009)***	1.000 (0.016)	1.018 (0.009)**	0.999 (0.016)
Years of firm tenure	1.068 (0.014)***	1.014 (0.013)	1.074 (0.014)***	1.012 (0.013)	1.067 (0.017)***	1.027 (0.016)*	1.070 (0.017)***	1.029 (0.016)*
Tertiary education			0.899 (0.129)	1.105 (0.072)			0.689 (0.065)***	0.929 (0.105)
Secondary education			0.823 (0.045)***	0.992 (0.056)			0.857 (0.055)**	0.858 (0.089)
Very-high skills			0.714 (0.099)**	1.077 (0.103)			0.713 (0.125)*	1.391 (0.307)
High skills			0.849 (0.075)*	1.202 (0.107)**			0.837 (0.133)	1.466 (0.274)**
Medium skills			1.033 (0.066)	1.257 (0.103)***			1.171 (0.150)	1.418 (0.259)*
Low skills			1.071 (0.053)	1.183 (0.101)**			1.097 (0.109)	1.294 (0.222)
Years since migration 2 - 3	0.887 (0.036)***	1.059 (0.051)	0.885 (0.036)***	1.061 (0.051)	1.029 (0.082)	1.160 (0.106)	1.034 (0.084)	1.164 (0.107)*
Years since migration 3 - 4	0.581 (0.043)***	0.836 (0.049)***	0.578 (0.043)***	0.838 (0.049)***	0.654 (0.059)***	0.823 (0.087)*	0.658 (0.060)***	0.826 (0.086)*
Years since migration 4 - 5	0.416 (0.033)***	0.701 (0.044)***	0.414 (0.033)***	0.704 (0.045)***	0.564 (0.052)***	0.747 (0.105)**	0.572 (0.054)***	0.749 (0.105)**
Years since migration +5	0.172 (0.018)***	0.507 (0.027)***	0.172 (0.018)***	0.510 (0.027)***	0.223 (0.028)***	0.613 (0.075)***	0.230 (0.029)***	0.615 (0.074)***
Urban area indicators 1^{st} loc.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Urban area indicators 2^{nd} loc.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	718,390		718,390		276,575		276,575	
Migrants	17,143		17,143		6,094		6,094	
Pseudo R^2	0.051		0.051		0.042		0.043	

Notes: Relative risk ratios (exponentiated coefficients) reported with standard errors in parentheses clustered at the urban area level of 1^{st} location. ***, **, and * indicate significance at the 1, 5, and 10 percent levels. The reference category is permanent migrants who stay 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 six densest cities, where only increments in density are considered. All specifications include indicator variables for year and month and age categories. The omitted categories are *Primary education*, *Very-low skills* and *Years since migration 0 - 2*. Variables *in 1^{st} location* are averages over pre-migration spells. *Log mean earnings 2^{nd} location* are 6-month moving averages, excluding current earnings.

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 regional 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 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 low- to high-density cities. The effects of differences in education, occupational skills, or relative earnings on the probability of migrating to high-density 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 only about one-third 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 60% of these moves involve a return migration. Return moves are more frequent and happen sooner in high-density cities. I find that returnees from high-density 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 high-density 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 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 migration contributes to the sorting of more skilled workers into denser cities, it is worth noting that worker sorting can occur through other channels (Combes, Duranton, Gobillon, and Roux, 2011*c*). For instance, both faster learning associated with working in dense cities (Baum-Snow and Pavan, 2010, De la Roca and Puga, 2011) and better schools in dense urban areas can widen the skill gap.

References

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.

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

Baum-Snow, Nathaniel and Ronni Pavan. 2010. Understanding the city size wage gap. *Review of Economic Studies* (forthcoming).

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

Blanchard, Olivier Jean and Lawrence F. Katz. 1992. Regional evolutions. *Brooking Papers on Economic Activity* (1):1–61.

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.

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.

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, 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 Sèbastien Roux. 2011*c*. 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.

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. 2011. Learning by working in dense cities. Processed, IMDEA Social Sciences Institute.

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.

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.

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

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

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

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.

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

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.

Jenkins, Stephen P. 1995. Easy estimation methods for discrete-time duration models. *Oxford Bulletin of Economics and Statistics* 57(1):129–38.

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

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

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.

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

Appendix A. Completing earnings data

For the period 2004–2009, uncensored earnings data is 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). In addition, for the entire period 1981–2009 earnings data is available for all workers in the MCVL, including those in the Basque Country and Navarre, from the Social Security, but these are capped for some workers. In particular, 5.7% and 2.3% of monthly earnings observations are top- and bottom-coded, respectively. This Appendix explains how I estimate earnings for this 8% 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 1991 and aged 18 to 65 throughout 1981–2009. I restrict observations following the same criteria used to construct the final sample in the study, except for the age restriction. This broader sample contains 473,979 individuals and 70,709,561 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 290 occupation-year Tobit regressions (10 occupations×29 years). I use as dependent variable log daily earnings expressed in December 2009 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 SIC sector, type of contract, public worker 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:

$$\hat{W}_{ijt} = x_{ijt}{}'\hat{\gamma} + z_{ijt}{}'\hat{\theta} + \hat{\sigma}\,\varepsilon_{ijt}, \tag{A.1}$$

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. Last, 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.[21]

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 1981–2009 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.[22]

The mean and variance of log earnings for each occupation-year pair are estimated in equation (A.1). 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

[21]This implies drawing i.i.d shocks from a truncated normal distribution $\varepsilon_{ijt} > (b_{ijt} - x_{ijt}{}'\hat{\beta})/\hat{\sigma}$ if earnings are top-coded and $\varepsilon_{ijt} < (a_{ijt} - x_{ijt}{}'\hat{\beta})/\hat{\sigma}$ 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 2,000 iterations only 0.08% and 0.11% of observations remain capped for the Earnings 1 and Earnings 2 specifications, respectively.

[22]Log daily earnings follow a multivariate normal distribution also within each of the 10 occupation categories. For simplicity, I omit the j index referring to type of occupation.

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, applicable censoring bounds and a value for the correlation coefficient ρ. Equation (A.2) explains how to generate earnings in year $t+h$ based on a bivariate normal distribution of earnings in years t and $t+h$:

$$\tilde{W}_{i,t+h}\,,\ \tilde{W}_{it} \sim N\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & \rho\\ \rho & 1\end{pmatrix}\right), \tag{A.2}$$

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

where $\tilde{W}_{it}$ are the standardised 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 standardised 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 $k = 1,...,40$.

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^* 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^* that best replicates the observed correlation for uncensored earnings in the data. I construct a variance-covariance matrix (29×29) with the optimal ρ^* values calculated for every pair of years throughout 1981–2009.[23]

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

$$\hat{W}_{ijt} = x_{ijt}{}'\hat{\gamma} + z_{ijt}{}'\hat{\theta} + \hat{\sigma}\cdot\ \hat{p}_{jt}{}'\zeta_{ijt}, \tag{A.3}$$

where all variables are the same as in equation A.1, but now $\hat{p}_{jt}$ is a row vector (1×29) of the Cholesky decomposition of the estimated variance-covariance matrix and ζ_{ijt} is a vector of random shocks.[24] Thus, the main difference between the 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.[25]

[23]Variance-covariance matrices for each type of occupation are available upon request. Earnings for skilled workers are much more persistent. The first three average estimated order of autocorrelations for the highest skill oocupation are 0.90, 0.85 and 0.80. The same figures for the lowest-skill occupation are 0.75, 0.68 and 0.63.

[24]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 $\tilde{P}$ minimizes 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 8 replacements needed out of out of 116 eigenvalues. The autocorrelations in this new variance-covariance matrix Ω^* are very similar to the ones in $\hat{\Omega}$. Only 1.1% and 5.3% of all new autocorrelations differ by more than 0.025 and 0.01 in absolute terms from the originally estimated autocorrelations, respectively.

[25]Because computation of the Earnings 3 specification is extremely time-demanding, I only run 20 iterations for the vector of random shocks. As a result, only 1.56% of earnings observations remain censored.

Fit of estimated earnings

Table A.1: Correlations for actual and predicted earnings

	Actual	*Earnings 1*	*Earnings 2*	*Earnings 3*
Actual	1.000			
Earnings 1	0.912	1.000		
Earnings 2	0.914	0.954	1.000	
Earnings 3	0.937	0.953	0.955	1.000

Table A.2: 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	42.6	43.2	43.3	43.4	37.7	40.3	40.4	40.4
Percentile 10	48.9	50.6	50.7	50.7	44.8	48.0	48.2	48.1
Percentile 25	59.1	61.4	61.5	61.5	59.5	64.6	64.8	64.8
Percentile 50	77.5	80.5	80.6	80.7	78.8	85.5	85.8	85.7
Percentile 75	112.4	119.8	120.0	120.1	112.1	120.6	120.9	120.7
Percentile 90	161.6	170.0	170.0	167.5	157.6	167.5	167.1	167.6
Percentile 95	208.3	216.3	215.8	215.8	201.6	203.9	202.6	203.6

Notes: Monthly earnings expressed as a percentage of the average for the corresponding category. Only individuals working in urban areas included. Skilled individuals work in the top three out of ten Social Security occupations.

Table A.3: Order of autocorrelations for actual and predicted earnings

Order	Actual	*Earnings 1*	*Earnings 2*	*Earnings 3*
1	0.891	0.866	0.867	0.889
2	0.843	0.823	0.824	0.843
3	0.806	0.791	0.792	0.809
4	0.778	0.766	0.768	0.784
5	0.754	0.748	0.750	0.764

To check 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 (2004–2009). If the fit is satisfactory for 2004–2009, I can be more confident that predicted earnings do a good job in predicting capped earnings for 1981–2003.

As shown in table A.1 the correlation between predicted and actual values for capped month-individual observations in 2004–2009 is 0.91 for the Earnings 1 and Earnings 2 specifications and 0.94 for the Earnings 3 specification. 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 A.2, 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.

Table B.1: Sample of migrants based on minimum length of migration episodes

Minimum length of migration episode	Number of migrants	Share of individuals	Monthly observations	Mean # of migrations	Max # of migrations
One month	96,494	37.4%	11,754,992	3.5	147
Over 3 months	76,637	29.7%	9,773,368	2.5	34
Over 6 months	63,905	24.7%	8,419,267	2.0	13
Over 12 months	47,822	18.5%	6,660,654	1.6	9
Over 24 months	28,122	10.9%	4,343,699	1.3	6

Notes: For any cut-off migration period specified (e.g. 6 months, 1 year), a migrant must experience spells of such duration in both urban areas, origin and destination.

Even for high-skilled workers, who are top-coded beyond the 62$^{\text{nd}}$ percentile, predicted earnings approximate salaries quite well in capped percentiles.

Table A.3 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 that better matches persistence in salaries. Given this and the higher correlation of individual observations with uncensored earnings, I use estimates from the Earnings 3 for the 8% of observations that have the value of earnings capped in the MCVL.

Appendix B. Identifying migrant types

The first distinction I consider is between short- and long-term migration events. Short-term migrations take place if an individual moves to another urban area for a short-term transfer within a job or to work in a seasonal or temporal job. Table B.1 shows the number of migrants in the sample for different cut-offs for what constitutes a long-term migration. For any cut-off (e.g. 6 months, 1 year) an individual must experience spells of such duration in both urban areas, origin and destination. As expected, the longer the cut-off the smaller the number of migrants. One out of five migrants in the sample has only made very short-term moves (under 3 months) in his life. I choose a one year cut-off because it is a reasonable period of time for the migrant to get familiar with the city of destination. From now on, I distinguish between short-term migrants (those who have never moved beyond a 12-month period) and long-term migrants (50% of all migrants).

Most of long-term migrants have moved only once in their lives (58%), while 31% of them have moved twice. Very few long-term migrants register more than three moves (3%), so to facilitate tracking across locations, I drop migrants with more than three migration events. Furthermore, since I exclude non-urban migrations, the final sample of long-term urban migrants is made of 33,515 individuals and 4,617,408 monthly observations.

The second distinction I consider is between short- and long-term distance. Although urban areas can be understood as independent local labour markets, in some cases two or more of them may exhibit substantial overlapping in worker flows. This pattern is more prevalent in dense urban areas such as Madrid and Barcelona which tend to have smaller (satellite) urban areas at reasonable commuting distances. Because the type of migration I analyse requires a change in residence, I need to distinguish potential commuters from long-distance migrants.

In order to identify short-distance moves, I have collected data on the shortest driving distance between any two urban areas using Google Maps. I classify migrations as long-distance if they exceed a driving distance threshold of 120 km. (74.6 miles). Once restricted to long-distance moves, the sample of long-term urban migrants is made up of 19,077 individuals and 2,609,908 monthly observations.

Appendix C. Migration to low-density cities

Table C.1: Logit estimation of determinants of first migration to low-density cities

	Dep. variable: long-term long-distance migration to cities below median density			
	(1)	(2)	(3)	(4)
Log mean earnings		0.944 (0.195)		0.860 (0.082)
Richest earnings tercile			0.932 (0.189)	
Poorest earnings tercile			0.874 (0.056)**	
Tertiary education	1.173 (0.406)			1.205 (0.405)
Secondary education	0.764 (0.146)			0.773 (0.146)
Very-high skills	1.321 (0.320)			1.454 (0.334)
High skills	1.078 (0.157)			1.150 (0.164)
Medium skills	1.013 (0.141)			1.043 (0.136)
Low skills	1.216 (0.080)***			1.229 (0.080)***
Years of experience	0.916 (0.010)***	0.914 (0.022)***	0.914 (0.021)***	0.918 (0.011)***
Years of firm tenure	0.913 (0.013)***	0.913 (0.012)***	0.911 (0.013)***	0.916 (0.013)***
Temporary contract	1.692 (0.105)***	1.676 (0.095)***	1.688 (0.095)***	1.676 (0.105)***
Unemployed	1.281 (0.156)**	1.249 (0.193)	1.255 (0.185)	1.297 (0.155)**
Unemployed, expired benefits	7.826 (0.567)***	7.789 (0.552)***	7.829 (0.552)***	7.787 (0.552)***
Urban area $\times$ period indicators	Yes	Yes	Yes	Yes
Age indicators	Yes	Yes	Yes	Yes
Pseudo R^2	0.057	0.055	0.055	0.057

Notes: Odd ratios (exponentiated coefficients) are reported on a sample of 11,443,407 monthly observations and 162,726 individuals. Standard errors in parentheses clustered at the urban area level. ***, **, and * indicate significance at the 1, 5, and 10 percent levels. The reference category is stayers. Sample is all individuals who are still in their first city. 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 density below the median density city *and* migrants experience a decline in density. The median density city is Santiago de Compostela. All specifications include month indicator variables. Period is a ten-year interval. *Log mean earnings* are 12-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.

Appendix D. Short-term and short-distance migration

Table D.1: 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	1.089 (0.059)	1.654 (0.174)***	1.052 (0.038)	1.230 (0.080)***
Tertiary education			1.283 (0.089)***	1.774 (0.358)***
Secondary education			1.027 (0.031)	1.341 (0.113)***
Very-high skills			0.900 (0.069)	1.510 (0.126)***
High skills			0.958 (0.056)	1.382 (0.081)***
Medium skills			0.964 (0.037)	1.466 (0.086)***
Low skills			0.953 (0.021)**	1.120 (0.033)***
Years of experience	0.942 (0.006)***	0.867 (0.004)***	0.947 (0.004)***	0.894 (0.007)***
Years of firm tenure	0.856 (0.008)***	0.917 (0.011)***	0.855 (0.008)***	0.911 (0.013)***
Temporary contract	2.002 (0.062)***	1.521 (0.038)***	2.000 (0.064)***	1.584 (0.037)***
Unemployed	1.846 (0.122)***	0.756 (0.053)***	1.926 (0.104)***	1.044 (0.075)
Unemployed, exp. benefits	7.532 (0.291)***	9.394 (0.454)***	7.457 (0.276)***	9.037 (0.440)***
Urban area $\times$ period ind.	Yes	Yes	Yes	Yes
Age indicators	Yes	Yes	Yes	Yes
Observations	11,902,686		11,902,686	
Pseudo R^2	0.093		0.094	

Notes: Relative risk ratios (exponentiated coefficients) reported on a sample of 11,902,686 monthly observations. 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 short-distance and value 2 if it is long-term 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 12-month moving averages, excluding current earnings. *Primary education* and *Very-low skills* are the omitted categories.

Following the estimations proposed in tables 2, 3 and 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 D.1, I estimate a multiple-exit discrete duration model (instead of a conditional hazard rate model as in tables 2 and 3) 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 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.