WORKING PAPER #23 / APRIL 2015

SELECTION IN INITIAL AND RETURN MIGRATION:

Evidence from moves across Spanish cities

+ JORGE DE LA ROCA

ABSTRACT

This paper investigates the contribution of migration to the sorting of more productive workers into bigger cities using administrative data for Spain that follow workers over their lives. Migrants to bigger cities are positively selected in terms of education, occupational skills, and individual productivity as proxied by their pre-migration position in the local earnings distribution. However, not everyone benefits equally from bigger 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 bigger cities.

I thank Manuel Arellano, Pablo Lavado, Elena Manresa, Diego Puga and Anatoli Segura for helpful conversations and Laurent Gobillon for a valuable discussion. Funding from the Bank of Spain Excellence Programme is gratefully acknowledged. This work contains anonymized statistical data from mcvl cdf 2004–2011 which are used with the permission of Spain's Dirección General de Ordenación de la Seguridad Social.

CONTACT

Jorge De la Roca
Furman Center for Real Estate and Urban Policy
New York University
139 MacDougal Street, 2nd Floor, New York, NY 10012
mail: jorge.delaroca@nyu.edu; website: http://jorgedelaroca.name

NYU Marron Institute of Urban Management

1. Introduction

Workers earn substantially more in bigger 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 bigger cities facilitate the acquisition of greater skills (Glaeser, 1999, Gould, 2007, Baum-Snow and Pavan, 2012, De la Roca and Puga, 2012). However, it has long been thought that those higher earnings may also partly reflect the sorting of more able workers into bigger cities. Already in 1890, Alfred Marshall wrote "[i]n almost all countries there is a constant migration towards the towns. The large towns and especially London absorb the very best blood from all the rest of England; the most enterprising, the most highly gifted, those with the highest physique and the strongest characters go there to find scope for their abilities." (Marshall, 1890, 5.6).

Existing empirical studies of worker sorting on ability examine differences in observable skills in big and small cities. 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. De la Roca and Puga (2012) estimate wage regressions including worker fixed-effects and the heterogeneous effects of big-city experience to recover innate distributions of skills in small and big cities. They do not find statistically significant differences in these skills distributions within broad educational and occupational groups. Baum-Snow and Pavan (2012) look at measures of ability from a finite-mixture model in a structural estimation setting and find evidence of positive sorting on observed skills to bigger cities. Again, they find that sorting on unobserved ability within education groups contributes little to the observed city-size earnings premium.

One common goal of these studies is to examine whether skills vary with city size, yet, they tell us little about the dynamic sorting process that may lead to differences in skills between big and small cities. To that end, this study investigates the contribution of internal migration to the sorting of workers across cities by studying whether greater skills, both observed and unobserved, increase the likelihood that a worker migrates to a bigger city. Using rich administrative data for Spain that follow individuals over time and across cities throughout their careers, I show that migrants who move to big cities are positively selected in terms of their level of productivity as proxied by their relative position in the local pre-migration earnings distribution. This remains so even when looking within given levels of educational and occupational groups, yet, the extent of selection is substantially reduced after conditioning on observable skills.

In addition, I document a second stage of sorting that happens after a first migration episode. About 30% of migrants end up leaving their city of destination within five years. Moreover, around 65% of these second moves involve a return migration to the city of origin. Such return migration is more frequent and happens sooner when a migrant moves initially to a big city. I find that to understand such return migration, it is important to look not just at the initial worker characteristics and relative earnings prior to the first move, but also at the heterogeneous experiences of workers following their first migration episode.

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

This study also contributes to our understanding of internal or regional migration in countries. 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. 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 out-migrate to states with higher inequality. Bound and Holzer (2000), using US census data to examine the role of individual characteristics in the sort of labor adjustments to regional shocks studied by Blanchard and Katz (1992), find that workers with low education are less prone to migrate in response to shifts in demand. For Europe, Hunt (2004) examines determinants of migration among federal states in Western Germany. She also finds that migrants are more skilled than stayers. I contribute to this literature by using cities (instead of states or regions) as the spatial units of analysis, and showing that long-term migrants from small to big cities are key to understanding why migrants are largely positively selected. Given the strong positive relationship between the size of a city and its level of earnings and inequality (De la Roca and Puga, 2012, Baum-Snow, Freedman, and Pavan, 2014, Eeckhout, Pinheiro, and Schmidheiny, 2014), this finding confirms the predictions of a standard Roy self-selection model where mismatched high-skilled individuals in small cities move to areas with higher returns to skills, namely big cities.

Furthermore, when examining subsequent migration decisions of individuals, I allow skills to vary over time by looking at the worker's relative position in the local earnings distribution at the time of each migration episode. 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.[1] Considering selection on the basis of skills observed in the first as well as in the second location allows me to gain further understanding of the characteristics and experiences of returnees, as well as characterize the implications that these return moves have for distributions of skills in big and small cities.

[1]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.

An extensive set of studies that analyze selection in initial and return migration have focused mostly on international migration, specially on flows between Mexico and the United States.[2] Besides of the reasons highlighted above, 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 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, as well as high migration costs (both monetary and psychological), would make it more difficult to evaluate the performance of migrants and returnees than in the case of internal migration.

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

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

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

City H is characterized by three differences with respect to city L. First, a worker 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

[2] See Borjas and Bratsberg (1996) for a model on international return migration. See Chiquiar and Hanson (2005), Ibarrarán and Lubotsky (2007), McKenzie and Rapoport (2010), Fernández-Huertas (2011), Kaestner and Malamud (2014) for selection and return migration flows between Mexico and the United States. See Co, Gang, and Yun (2000), Constant and Massey (2003), Dustmann (2003), DeCoulon and Piracha (2005), Rooth and Saarela (2007), Ambrosini, Mayr, Peri, and Radu (2011) for international return migration in European countries.

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

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

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

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

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 an additional moving cost. Together, these imply that some workers with skills low enough that they would be unwilling to undertake irreversible migration ($s_i \leqslant \hat{s}$), given that they can return, are now willing to experiment. If they move to city H and have a good realization of δ_i, great; if not, they can always move back, subject to some cost. Similarly, some workers with higher initial skills ($s_i > \hat{s}$) will now end up returning after migrating from city L to H, if they have a bad realization of δ_i. As a result, city H will exhibit ex-post higher average skills and earnings, but the skill distributions of the two cities will partially overlap because of uncertainty in realization of skills and return moves by unlucky migrants.[7] 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.[8]

[5]This feature is widely documented in the literature on agglomeration economies. See Duranton and Puga (2004) for microfoundations and 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 (2012) for a recent estimate of urban costs with respect to city population using data for all individual land transactions in France.

[7]Return migration is one mechanism that can help explain the overlapping in ability distributions between big and small cities. In Behrens, Duranton, and Robert-Nicoud (2014) more talented individuals sort initially into bigger cities, but once there, a random draw or serendipity opens up the productivity distribution. By assuming prohibitive mobility costs they obtain incomplete sorting on productivity between big and small cities. In Eeckhout, Pinheiro, and Schmidheiny (2014) strong skill complementarities in production between extreme skills generate an overrepresentation of high- and low-skilled workers in bigger cities.

[8]Return migration can either be modeled as a correction to an unfavorable draw (or 'mistake') given the uncertainty of economic conditions in destination or as part of an optimal planned location trajectory between big and small cities over the life cycle. Both models deliver similar testable implications where return migration intensifies the selection driven by the initial migration flow. In the empirical section I do not attempt to distinguish both hypotheses. However, two features of the data suggest that return migration may be generally associated with experimentation in big cities or correction to an unfavorable outcome. Most returnees are young at the time of their return move from a big city (75% of them under 35) and the great majority of return moves happen a few years after the initial move (80% of them within four years).

The intuition for initial migration from city H to L is much simpler. As there is no uncertainty in the ex-post realization of skills in L, return migration cannot be optimal.[9] Workers initially located in city H decide to migrate to city L if and only if

$$\alpha(s_i + \delta_i) - R \leqslant s_i - C \, . \tag{4}$$

Thus, in equilibrium, workers with skills

$$s_i \leqslant \widetilde{s} = \frac{R - C - \alpha\delta_i}{\alpha - 1} \, , \tag{5}$$

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

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

Solution

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

I proceed backwards, and first concentrate on the second stage. After moving to H the realization of δ_i is revealed to the worker. She decides to return if and only if $\alpha(s_i + \delta_i) - R \leqslant s_i - C_2$. Thus, a worker returns if ex-post earnings in H are lower than earnings in L minus the return migration cost.[10] Given that δ_i takes a minimum value of 0 and a maximum value of 2δ, some workers always return even in the best-case scenario of $\delta_i = 2\delta$, others never return even in the worst-case scenario of $\delta_i = 0$, while others return depending on the actual realization of δ_i. In particular, a worker who migrates to city H returns to city L if and only if

$$\delta_i < \underline{\delta}(s_i) = \begin{cases} 2\delta & \text{if} \quad s_i < \underline{s} \, , \\ \frac{R - C_2 - (\alpha - 1)s_i}{\alpha} & \text{if} \quad \underline{s} \leqslant s_i < \overline{s} \, , \\ 0 & \text{if} \quad s_i \geqslant \overline{s} \, , \end{cases} \tag{6}$$

where

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

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

[10] 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.

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

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

where the first term of equation (7) refers to the decision to return that takes place under unfavorable realizations of δ_i, while the second term refers to the decision to stay in H under favorable realizations.

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 $\underline{s} \leqslant s_i < \bar{s}$, substituting equation (6) 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 $s_i \geqslant \bar{s}$ (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 $\hat{s} \leqslant \bar{s}$, so that workers with $s_i \geqslant \bar{s}$ always migrate and never return.[11]

To summarize these results:

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

Selection in initial migration from city H to L is straightforward. Workers initially located in H, based on their initial ability s_i and after observing their realization of δ_i, decide between staying in city H or migrating to L. Once again I need to consider cases that vary with realizations of δ_i.[12]

[11] If instead $C + C_2 > \alpha\delta$ then $\hat{s} > \bar{s}$ and equation (8) is never satisfied for $s_i < \bar{s}$. 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.

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

$$\delta_i < \delta_*(s_i) = \begin{cases} 2\delta & \text{if} \quad s_i < s_*\,, \\ \frac{R-C-(\alpha-1)s_i}{\alpha} & \text{if} \quad s_* \leqslant s_i < s^*\,, \\ 0 & \text{if} \quad s_i \geqslant s^*\,, \end{cases} \tag{9}$$

For workers with $s_i < s_*$ (see footnote 12), the condition of equation (4) is always satisfied as $s_* < \widetilde{s}$, where $\widetilde{s}$ is given by equation (5). These workers always migrate, though the least skilled of this group might not be able to afford the moving cost C. For workers with $s_* \leqslant s_i < s^*$, those who get an unfavorable realization of δ_i such that $s_i \leqslant \widetilde{s}$ migrate to city L. Lastly, for workers with $s_i \geqslant s*$, the condition of equation (4) is never satified since $\widetilde{s} < s^*$. These workers find strictly preferable to remain in city H.

To summarize these latter results:

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

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

3. Econometric framework

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

My unit of analysis is an individual-period pair, where my data are at the monthly level. In each month, I observe different values of individual-level variables (aggregate variables are also captured through location 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:

$$h(t) = P[T = t | T \geq t, x(t)] = F[\beta_0(t) + \beta_1' x(t)] \, , \tag{10}$$

where T is the month in which the first migration episode occurs (possibly never), F is a cumulative probability function (always a logistic specification in the study), $x(t)$ is a vector of (possibly time-varying) individual and job characteristics, including the city where the individual is working,

where

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

$\beta_0(t)$ is a duration-specific parameter that captures duration at t in an additive and unrestricted way and β_1 is a vector of parameters. Therefore, I am modeling for an individual working in her city of first location the probability of migrating, conditioning on observable characteristics.

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

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

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

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

$$L(\beta) = \sum_{t=1}^{T_i} \left\{ \sum_{i=1}^{N} \mathbf{1}(T_i \geq t \geq e_i) \left[m_i Y_{ti} \log h_i(t) + (1 - m_i Y_{ti}) \log(1 - h_i(t)) \right] \right\} \tag{12}$$

and $\hat{\beta}$ is the maximum likelihood estimator that maximizes $L(\beta)$. Therefore, discrete duration models can be regarded as a sequence of binary models (Jenkins, 1995). I estimate equation (12) to examine how the productive characteristics of migrants compare to those of non-migrants in their city of origin prior to migration.

It will be useful to sometimes split a given risk (e.g., migrating for the first time) into several alternative options (e.g., initial migration to a big city and initial migration to a small city). This requires a multiple-exit discrete duration model. One possibility is to model both transition intensities into such states in a multinomial logit model, i.e. model the probability of either moving to a small city or moving to a big city at time t conditional on not having done either before. An alternative way is to model conditional hazard rates, i.e. model the probability of moving to a big city at time t conditional on not having done so before and on not having moved to a small city either. Bover and Gómez (2004) show that if the transition intensities are multinomial logit, the conditional exit rates are binary logit with the same parameters. Thus, the logit specification is derived from the same model in both cases. Likewise, estimating the model by joint maximum likelihood or conditional maximum likelihood results in consistent and asymptotically normal estimates of the parameters. Although the former approach is 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 (12) to analyze how the productive characteristics of second-time migrants compare to those of workers who engaged in the same initial migration episode but instead remain in the city to which they first moved.[13] Once again,

[13]One difference between both specifications is how I introduce duration-specific parameters to capture duration dependence. In equation (12), age indicator variables capture time spent in city of origin. In the specification for determinants of second-time moves, I need to include indicator variables for the number of years since first migration took place.

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 labor market characteristics both at the origin and destination of each migration. However, since we wish to explain migration by comparing migrants both with themselves at times in which they do not migrate and with other workers, whether migrants or not, in practice we need the data to record the labor market characteristics of all workers with high frequency since the start of their first job.

The *Muestra Continua de Vidas Laborales* (MCVL), or Continuous Sample of Employment Histories, satisfies these requirements. This is an administrative data set with information on a 4% non-stratified random draw of the population who on a given year have any relationship with Spain's social security, be it because they are working, receiving unemployment benefits, or receiving a pension. For each of these individuals, all of their changes in labor market status and work characteristics are recorded since 1981. I combine data from eight editions of the MCVL (2004 to 2011), so as to have data on a 4% sample of all individuals who have worked, received benefits or a pension at any point in 2004–2011. The requirement for inclusion in the MCVL (based on the individual's social security number) is maintained from year to year, so that the difference across editions is that more recent editions include individuals who enter the labor force for the first time, while they lose those who cease any relationship with the social security (individuals who stop working continue to be included in the sample while they receive unemployment benefits or a retirement pension, so most exits occur when individuals are deceased or left the country). The unit of observation in the source data is any change in the individual's labor market status or job characteristics (including changes in occupation or remuneration within the same firm). Given that all changes since 1981 or the date of first employment are recorded, it is possible to construct a panel with day-by-day job characteristics for all individuals in the sample.

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

The data include monthly earnings for each job spell, constructed by combining a variety of sources. For the period 2004–2011, uncensored earnings data are available from matched income tax returns for all workers except those in the Basque Country and Navarra (where income taxes are not collected by the Central Government). In addition, for the entire period 1981–2011 earnings data from the social security are available for all workers, including those in the Basque Country and Navarra, but these are capped for a small fraction of observations.[14]

A crucial feature of the MCVL is that workers can be tracked across space based on their workplace location. Social security legislation requires employers to keep separate earnings' contribution accounting codes for each province in which they conduct business. Furthermore, within a province, a municipality identification code is provided if the workplace is located in a municipality with population greater than 40,000 inhabitants in 2011. Thus, location information is at the establishment level.

Urban areas

I use official urban area definitions by Spain's Department of Housing for 2008. The 85 urban areas in Spain account roughly for 68% of population and 10% of total surface. They represent local labor markets comparable to Core Based Statistical Areas (CBSAs) in the United States. The median urban area has a population of 140,571 inhabitants in 2008. From now on, I use the terms cities and urban areas indistinctly to refer to local labor markets.

Urban areas enclose 747 municipalities. Given that I know the municipality of workplace location for each job and unemployment spell in MCVL, I can assign each individual to an urban area in any month, provided the municipality has a population larger than 40,000 inhabitants in 2011. 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 three small urban areas in MCVL data because population of their largest municipality is below the 40,000 population threshold.[15]

To measure the scale of an urban area I count the number of people within 10 kilometers of the average resident in the urban area, a measure proposed by De la Roca and Puga (2012). This is an index of density, which the literature generally prefers to simple population size as a measure of the potential for interactions that an urban area offers to workers (Puga, 2010, Combes, Duranton, and Gobillon, 2011). At the same time, by considering agglomeration patterns within and around the urban area, this measure 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.[16] In any case, results are robust to measuring the scale of each urban area by its total population.[17]

[14]Appendix A provides details on how earnings are estimated for the small fraction of capped observations (14.7%), based on uncensored earnings and individual and job characteristics.

[15]These are in order of population size Sant Feliú de Guixols, Soria and Teruel.

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

[17]The correlation between the number of people within 10 kilometers of the average resident and total population is 0.94. In the context of this paper, the main advantage of the measure I use is that it takes into account the proximity of workers in adjacent urban areas, which are totally excluded when one looks only at total population.

Sample restrictions

My initial sample is made up of Spanish natives born between 1963 and 1993 (i.e., aged 18–48 during the period 1981–2011) who have been employed, either as employees or self-employeds, or received unemployment benefits at any point over this period. I leave out individuals older than 48 in 2011 and foreign-born immigrants since I cannot retrieve complete work histories for them. A total of 579,362 individuals and 78,857,078 monthly observations make up this initial sample.

From this initial sample, I eliminate individuals with low labor force attachment in their lives, which implies dropping those who have not worked more than 6 months in at least one calendar year between 1981 and 2011. This restriction reduces the sample to 541,704 individuals and 78,583,304 observations. Next, I drop observations from special social security regimes such as agriculture, fishing and mining. Workers in these regimes tend to self-report earnings and the number of work days recorded is not reliable. Furthermore, these activities are typically rural in nature and linked to natural advantages. At this point, the sample contains 538,403 individuals and 74,642,171 observations.

Subsequently, I exclude observations for which the occupation or workplace location is missing and individuals for whom the educational attainment or the province of entry in the labor force is missing. This leaves 529,739 individuals and 71,260,214 monthly observations. Finally, since I wish to focus on urban migrations (and in any case, only the province is known for rural jobs), I focus on workers located in urban areas. This leaves the final sample at 463,920 individuals and 46,808,842 monthly observations.

Identifying migrants

An urban migration event is defined as a change in workplace location between two urban areas. In the sample, based on first moves, 87,762 individuals can be classified as urban migrants, 30,815 as urban-rural migrants and 20,347 as rural-urban migrants, while 307,400 individuals never leave their urban area for work purposes.[18] Again, I focus only on urban or city-to-city migrations which based on first moves account for 60% of all types of moves.[19]

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

[18] Furthermore, 7,609 individuals move for the first time across rural areas and later enter the sample when they work in an urban area. To identify moves to and from rural areas I track changes in province location (52 in total) that lack an urban area identifier, prior to restricting the sample to urban areas. Also, I classify 9,987 individuals as special since they move ten or more times over their lives or record few or only short job spells in MCVL (e.g. they appear in the sample with only two short job spells of 10 months with many years of inactivity in between). I exclude these special individuals from the estimation sample; results vary marginally when I include them.

[19] This is a lower bound since municipalities with fewer than 40,000 residents are not identified in MCVL and, thus, are coded as rural, although they belong to urban areas. Therefore, many moves between rural and urban areas are likely to be urban moves. Following the econometric framework described above, monthly observations of non-urban migrants while they remain in their city of origin or departure also contribute to the estimation.

results presented below suggest that the behavior of short-term or short-distance migrants is rather different from that of long-term and long-distance migrants.

Regarding the length of the migration episode, short-term migrants usually move for brief transfers within a job or to work in a seasonal or temporary job. I classify migrants as *short-term* if they never move beyond a 12-month period. Therefore, long-term migrants are movers who experience spells longer than a year in the city of destination. Based on this criterion, I identify 27,795 short-term migrants.

Regarding distance, within the sample of long-term migrants, I label migrants as *short-distance* if they move to an urban area that is less than 120 km. (74.6 miles) driving from the urban area where they previously worked.[20] Although urban areas can be understood as independent local labor markets, in some cases two or more of them may exhibit substantial overlapping in worker flows. This pattern is more prevalent in bigger urban areas such as Madrid and Barcelona, which tend to have smaller urban areas at reasonable commuting distances. Based on this criterion, I identify a total of 27,599 short-distance migrants.

These classification criteria leave a total of 32,368 long-term and long-distance migrants. Of these, 64% move only once in their lives while 29% have moved at most twice. I identify return migrants as those who move back to their city of origin in their second migration. Likewise, move-on migrants are those who do not return to their city of origin after two migrations. Since very few long-term and long-distance migrants register more than two moves, none of the results presented below changes when I restrict the estimation sample to migrants who move at most twice.

Table 1 shows summary statistics for non-migrants (stayers), short-term and short-distance migrants, and long-term and long-distance migrants. Within the latter category, I provide separate statistics for permanent migrants (those who never return to their city of first employment) and return migrants (those who return after their second move). All variables displayed are individual averages over working lives.

The raw data already reveal 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).[21]

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

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

[21] The level of education is that contained in the Municipal Register. A large update to this information was done by municipalities in 2001 and, since 2009, the Spanish Statistical Office has been notifying municipalities about any updates it received from the Department of Education. Furthermore, individuals can voluntarily update their level of education when they file any document with the municipality administration. Still, for some individuals levels of education might be underreported, so the fact that the educational ranking across migrants and stayers holds using occupational categories provides additional reassurance.

Table 1: Summary statistics of stayers and migrant types

	Stayers	Migrants		
		Short-term or short-distance	Long-term long-distance	
			Return	Permanent
Level of education				
Tertiary	18%	17%	24%	30%
Secondary	35%	32%	37%	38%
Primary	47%	51%	40%	33%
Occupational skills				
Very-high skills	8%	7%	10%	15%
High skills	11%	10%	14%	16%
Medium skills	22%	21%	27%	25%
Low skills	45%	46%	39%	36%
Very-low skills	14%	17%	10%	9%
Earnings				
Mean monthly earnings	1,715	1,636	2,015	2,134
Mean monthly earnings 2^{nd} city			2,056	2,410
Labour market characteristics				
Years of labor experience	8.1	7.4	8.2	7.6
Years of firm tenure	3.7	2.4	2.2	2.5
Self-employed	8%	5%	5%	4%
Public sector employee	6%	7%	6%	9%
Temporary contract	26%	37%	32%	32%
Part-time contract	12%	12%	9%	10%
Unemployed	10%	14%	13%	11%
Male	53%	57%	60%	55%
Age	30.8	30.4	31.5	31.3
Age of entry in labor force	21.1	20.7	21.0	21.9
Individuals	307,400	55,394	7,527	24,841

Notes: Variables are averages for individuals over their lives. Only individuals working in urban areas are included. Long-term long-distance migrations are moves that exceed 12 months in destination and distance of 120 km. Earnings expressed in December 2011 euros.

stayers, and then short-term/distance migrants continues to hold going down to individuals in occupations with high skills.

The ranking of monthly earnings across categories again points in the same direction. Permanent migrants exhibit the highest average earnings and are followed by returnees. Stayers and short-term/distance migrants earn substantially less, the gap being larger for the latter. Among long-term and long-distance migrants, those who eventually return have much lower earnings in their second location than those who do not return.

Other labor market characteristics reveal expected patterns, as stayers are attached to more stable jobs (permanent contracts) and, hence, have accumulated more tenure in a firm. They also have experienced fewer unemployment spells in their lives and are more likely to be self-employed. Males are more prone to migrate on the whole and make up a larger share of the group of return migrants and a smaller one of the group of permanent migrants.

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

5. Results

Selection in initial migration

I begin by studying the determinants of first migration episodes and, in particular, whether migrants are positively selected in terms of productive characteristics at the time of their first move relative to stayers in the same city.[22] In table 2 I estimate the probability of out-migration from the individual's first job location using a single-exit discrete duration model as in equation (12), where the dependent variable takes value one only in the last monthly observation prior to migration. I focus on long-term long-distance moves, i.e., only those moves that exceed 12 months in the city of destination and 120 km. of distance. The determinants of shorter moves are quite different and discussed in appendix B.

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 out-migration by 143% relative to having at most primary education, while working in an occupation that requires medium to very-high skills increases the probability from 20% to 43% relative to working in an occupation with very-low skills.

Other labor market variables reveal expected signs. An additional year of labor market experience or tenure in the firm decreases the probability of out-migrating, 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 36% more likely to migrate. Likewise, self-employed workers and public sector employees appear to be more rooted to their city of first employment or may benefit from job amenities that make them less inclined to migrate. Men are more likely to migrate, as suggested by the raw data, even after controlling for skills and labor market characteristics.

Individuals who are unemployed are also more prone to migrate. However, by controlling separately for unemployed individuals who have completed their period of entitlement to unem-

[22] In fact, the estimation compares urban 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		1.754 (0.144)***		1.346 (0.081)***
Richest earnings tercile			1.463 (0.066)***	
Poorest earnings tercile			0.963 (0.019)*	
Tertiary education	2.431 (0.344)***			2.290 (0.313)***
Secondary education	1.614 (0.096)***			1.566 (0.092)***
Very-high skills	1.429 (0.074)***			1.208 (0.054)***
High skills	1.202 (0.044)***			1.059 (0.042)
Medium skills	1.286 (0.075)***			1.223 (0.063)***
Low skills	1.077 (0.034)**			1.054 (0.031)*
Male	1.254 (0.020)***	1.080 (0.022)***	1.115 (0.018)***	1.199 (0.015)***
Years of experience	0.938 (0.006)***	0.908 (0.006)***	0.911 (0.006)***	0.934 (0.006)***
Years of firm tenure	0.920 (0.009)***	0.919 (0.010)***	0.923 (0.009)***	0.914 (0.010)***
Self-employed	0.539 (0.044)***	0.601 (0.048)***	0.546 (0.041)***	0.602 (0.054)***
Public sector employee	0.823 (0.049)***	0.838 (0.050)***	0.901 (0.061)	0.770 (0.045)***
Temporary contract	1.357 (0.024)***	1.354 (0.025)***	1.342 (0.025)***	1.379 (0.025)***
Unemployed	2.106 (0.078)***	1.690 (0.102)***	1.681 (0.098)***	2.049 (0.082)***
Unemployed × expired benefits	9.067 (0.338)***	9.295 (0.363)***	9.277 (0.362)***	9.128 (0.338)***
Urban area × period indicators	Yes	Yes	Yes	Yes
Age indicators	Yes	Yes	Yes	Yes
Pseudo R^2	0.066	0.061	0.060	0.067

Notes: Odd ratios (exponentiated coefficients) are reported on a sample of 31,989,746 monthly observations and 408,596 individuals. Standard errors in parentheses are clustered at the urban area level. ***, **, and * indicate significance at the 1, 5, and 10 percent levels. Sample is all individuals who are still in their first city. The reference category is stayers. Long-term long-distance moves are those that exceed 12 months in destination and distance of 120 km. All specifications include month indicator variables. Period is a ten-year interval. Log mean earnings are 6-month moving averages, excluding current earnings. Earnings terciles are constructed for all year-province pairs. Primary education and very-low skills are the omitted categories.

ployment benefits, I find they are the ones that particularly drive this large effect. Workers who are unemployed but receiving unemployment benefits are over 110% more likely to migrate than those who are employed. Once their unemployment benefits expire, however, the probability of migrating jumps by a factor of seven.[23] The disparate effect of unemployment for mobility of individuals with and without unemployment benefits has been previously highlighted 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 gap in the magnitude of the effects I find indicates that the current design of unemployment insurance in Spain has a detrimental impact on the efficient matching of unemployeds and vacancies across local labor markets.

In all specifications I include age indicator variables 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 ten-year periods to confine the analysis of migrants and stayers within an urban area and period.[24] 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 indicate 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 six months to proxy for workers observable and unobservable skills.[25] The inclusion of urban area–period indicators implies that the earnings variable measures the worker's relative position in the local earnings distribution. The corresponding coefficient shows that a 10% increase in log mean earnings raises the probability of out-migration by 2.8%.[26] 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 out-migrating by 46%, while being in the lowest one decreases it by 4%.

I bring in both observable skills and earnings in column (4). Even within given levels of education and occupational skills, higher earnings in the city of first employment increase the

[23]I identify unemployed individuals with expired benefits as those receiving benefits or subsidies who cease any relationship with the social security immediately after an unemployment spell—as opposed to starting a new job, which is the most common transition for them. When I do not distinguish between both types of unemployment status, I find that the overall effect of unemployment raises the odds of migrating by 226%.

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

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

[26]Reported odd ratios are changes in the relative probability of out-migrating 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 75.4%. The 2.8% reported in the text is calculated as $10\% \times (1.754 - 1)/e$.

probability of migrating. However, the effect of earnings has been reduced by more than half relative to column (2)—recall coefficients are odd ratios—whereas observable skills like education remain almost as strong determinants of first-time migration as in column (1). Thus, differences in observable skills are key to characterize the selection of first-time migrants while selection on unobservables, though present in the data, appears to be of less quantitative importance.[27]

So far, I have been looking at the probability of migrating in general. However, the selection found in the data is really driven by the particular type of migration modeled in the conceptual framework: from small to big cities. In table 3 I estimate the conditional hazard rate of moving to one of the six biggest cities in Spain, i.e., the probability of moving to one of these big cities among those who have not moved before and do not move to small cities. For migrants who move within the six biggest cities I focus only on moves with an increase in city size. Other than this, all specifications are identical to those in table 2.

Results reveal that the positive selection of migrants is much stronger when I look only at those who migrate to big cities. In general, the effects of differences in education and pre-move earnings are now much larger than the effects on the probability of migrating in general. In column (1), we can see that long-term/distance migration to big cities is more than twice more likely to occur for individuals with tertiary education—the probability increases by 329% relative to an increase of 143% in table 2. Having secondary education also raises substantially the odds, making migration to big cities twice as likely as for workers with at most primary education, while working in occupations demanding very-high and medium skills are now stronger determinants of migration to big cities. The coefficient on log earnings in column (2) implies that a 10% increase in log mean earnings raises the probability of out-migration by 4.6%. Column (4) shows that, even within given levels of education and occupational skills, more productive workers are more likely to migrate to big cities, yet, the probability declines by 60% compared to column (2). In contrast, if I repeat the estimation for migration to small (as opposed to big) cities, I find no clear evidence of selection of any type (see appendix C for details).

At the time of first migration, the group of long-term and long-distance migrants is made of permanent migrants—those who never return to their city of first employment—and return migrants—those who eventually return. In table 1 the raw data pointed out that permanent migrants have higher lifetime earnings and are more educated than returnees. In table 4 I narrow down this comparison by examining how productive characteristics differ between these two groups in their first city at the time of migration. Moreover, I investigate whether permanent or return migrants who move to big cities are more skilled or productive than migrants from the same city who move elsewhere. I run pooled OLS regressions where the dependent variable is a 6-month moving average of log earnings (excluding current earnings). All specifications include age and urban area interacted with ten-year-period indicator variables. Therefore, I capture the correlation

[27]Individuals working in occupations demanding high skills are no longer more inclined to migrate, once I condition on pre-move earnings. I have estimated several alternative specifications for table 2 by including as regressors the number of previous short-term migration episodes, their duration, or the number of previous short-term moves to the city of destination. In addition, I have excluded from the sample those migrants who register more than two long-term/distance moves. Results remain consistent and are available upon request.

Table 3: Logit estimation of determinants of first migration to big cities

	Dep. variable: long-term long-distance migration to any of 6 biggest cities			
	(1)	(2)	(3)	(4)
Log mean earnings		2.245 (0.123)***		1.505 (0.076)***
Richest earnings tercile			1.699 (0.053)***	
Poorest earnings tercile			0.997 (0.030)	
Tertiary education	4.292 (0.161)***			3.986 (0.160)***
Secondary education	2.245 (0.070)***			2.160 (0.071)***
Very-high skills	1.647 (0.136)***			1.326 (0.097)***
High skills	1.249 (0.093)***			1.057 (0.070)
Medium skills	1.656 (0.101)***			1.551 (0.090)***
Low skills	1.159 (0.062)***			1.123 (0.059)**
Male	1.358 (0.033)***	1.065 (0.027)**	1.119 (0.030)***	1.279 (0.031)***
Years of experience	0.942 (0.005)***	0.897 (0.006)***	0.900 (0.006)***	0.937 (0.005)***
Years of firm tenure	0.909 (0.006)***	0.910 (0.007)***	0.918 (0.007)***	0.900 (0.006)***
Self-employed	0.727 (0.064)***	0.789 (0.059)***	0.678 (0.058)***	0.841 (0.072)**
Public sector employee	0.518 (0.071)***	0.525 (0.074)***	0.603 (0.083)***	0.467 (0.065)***
Temporary contract	1.340 (0.036)***	1.319 (0.035)***	1.312 (0.036)***	1.363 (0.036)***
Unemployed	2.226 (0.166)***	1.486 (0.085)***	1.485 (0.086)***	2.132 (0.160)***
Unemployed $\times$ expired benefits	9.537 (0.388)***	9.935 (0.404)***	9.900 (0.406)***	9.623 (0.389)***
Urban area $\times$ period indicators	Yes	Yes	Yes	Yes
Age indicators	Yes	Yes	Yes	Yes
Pseudo R^2	0.083	0.071	0.069	0.084

Notes: Odd ratios (exponentiated coefficients) are reported on a sample of 24,353,675 monthly observations and 337,166 individuals. Standard errors in parentheses are clustered at the urban area level. ***, **, and * indicate significance at the 1, 5, and 10 percent levels. Sample is all individuals who are still in their first city. The reference category is stayers. Long-term long-distance moves are those that exceed 12 months in destination and distance of 120 km. Dependent variable takes value one if destination is one of six biggest cities *and* migrants experience an increase in city size. All specifications include month indicator variables. Period is a ten-year interval. Log mean earnings are 6-month moving averages, excluding current earnings. Earnings terciles are constructed for all year-province pairs. Primary education and very-low skills are the omitted categories.

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.

Results show that permanent and return migrants have higher pre-move earnings than stayers, after controlling for labor market characteristics. In column (1) I divide both migrant categories into those who move to the six biggest cities and those who move elsewhere. Again, for migrants who move within the six biggest cities I keep only those moves with an increase in city size. Permanent migrants who move to the biggest cities have significantly higher earnings than other permanent migrants and stayers. The same pattern is found between returnees to biggest cities and returnees to other cities and stayers, although the difference in earnings between both types of returnees is not significant at the 10% level. At the time of first migration, earnings of permanent migrants and eventual returnees who move to big cities are 14% and 8% higher than those of stayers, respectively. Therefore, I confirm the skill ranking found in the raw data (see table 1) where permanent migrants are the most productive, followed by return migrants, and then stayers. However, both earnings gaps fall substantially when I include observable skills as controls in column (2). Now, permanent migrants and eventual returnees earn only an extra 4% and 3% more than stayers with comparable education and occupational skills. This finding confirms that sorting in initial migration of the most productive workers into bigger cities can be to a large extent explained by observable skills.

In columns (3) and (4) I repeat the exercise of columns (1) and (2), but further classify migrants into those who move to the two biggest cities (Madrid and Barcelona), those who move to the 3$^{\text{rd}}$–6$^{\text{th}}$ biggest cities (Valencia, Sevilla, Bilbao and Zaragoza), those who move to small cities below the median-sized city (Santiago de Compostela), and those who move elsewhere. The idea is to examine whether the degree of skill sorting in initial migration increases with the size of the city of destination, i.e., whether among those workers who leave a medium-sized city like Granada, those who move to Barcelona are more productive than those who move to Sevilla, and much more productive than those who move to a small city like Huelva.

In column (3) I find some evidence in favor of skill sorting in mobility patterns. Permanent migrants to small cities exhibit similar earnings as stayers, whereas permanent migrants to cities outside of the six biggest have earnings that are around 8% higher than those of stayers at the time of migration. However, earnings for permanent migrants moving to the top 2 cities, while higher than for other permanent migrants, do not exceed those of permanent migrants moving to the 3$^{\text{rd}}$–6$^{\text{th}}$ biggest cities. The same patterns in the ranking are found among types of return migrants, although some of the differences in earnings across categories are not significant at the 10% level. When I add in observable skills in column (4), earnings of migrants to the top 2 cities are in fact lower than those of migrants to the 3$^{\text{rd}}$–6$^{\text{th}}$ biggest cities (the difference being significant for returnees at the 5% level) and similar to earnings of migrants to other cities. While these findings suggest there is broad skill sorting in initial migration based on the size of the city of destination, they do not fully match the predictions of theoretical models of urban sorting with a continuum of cities of different sizes, in which the most talented or skilled individuals move to the cities in the top of the urban hierarchy (Davis and Dingel, 2012, Behrens, Duranton, and Robert-Nicoud, 2014).

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

	Dep. variable: log mean earnings in first city			
	(1)	(2)	(3)	(4)
Permanent migrant to 6 biggest cities	0.136 (0.010)***	0.042 (0.008)***		
Permanent to cities other than 6 biggest	0.070 (0.012)***	0.026 (0.009)***		
Permanent to 1st–2nd biggest cities			0.135 (0.011)***	0.037 (0.010)***
Permanent to 3rd–6th biggest cities			0.141 (0.017)***	0.061 (0.014)***
Permanent to cities below median size			0.022 (0.029)	0.001 (0.018)
Permanent to other cities			0.077 (0.011)***	0.029 (0.008)***
Return migrant to 6 biggest cities	0.077 (0.010)***	0.027 (0.007)***		
Return to cities other than 6 biggest	0.052 (0.013)***	0.026 (0.007)***		
Return to 1st–2nd biggest cities			0.073 (0.012)***	0.020 (0.009)**
Return to 3rd–6th biggest cities			0.096 (0.015)***	0.054 (0.012)***
Return to cities below median size			0.025 (0.022)	0.009 (0.014)
Return to other cities			0.056 (0.016)***	0.028 (0.008)***
Male	0.152 (0.005)***	0.173 (0.006)***	0.152 (0.005)***	0.173 (0.006)***
Tertiary education		0.217 (0.021)***		0.217 (0.021)***
Secondary education		0.115 (0.007)***		0.115 (0.007)***
Very-high skills		0.662 (0.017)***		0.662 (0.017)***
High skills		0.492 (0.008)***		0.492 (0.008)***
Medium skills		0.203 (0.005)***		0.203 (0.005)***
Low skills		0.092 (0.005)***		0.092 (0.005)***
Labor market characteristics	Yes	Yes	Yes	Yes
Urban area × period indicators	Yes	Yes	Yes	Yes
Age indicators	Yes	Yes	Yes	Yes
R^2	0.346	0.507	0.346	0.507

Notes: Coefficients reported on a sample of 32,013,502 monthly observations and 409,832 individuals. Standard errors in parentheses are clustered at the urban area level. ***, **, and * indicate significance at the 1, 5, and 10 percent levels. Sample is all individuals who are still in their first city. Dependent variable is 6-month moving average of earnings, excluding current observation. Migrations are moves that exceed 12 months in destination and distance of 120 km. For migrants who move within the six biggest cities, only moves with an increase in city size are considered. For migrants who move below the median-sized city, only moves with a decline in city size are considered. Labor market variables are the same as those in tables 2 and 3. All specifications include month indicator variables. Period is a ten-year interval. Primary education and very-low skills are the omitted categories.

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 transition events. Now, the dependent variable takes value one in the last month prior to second migration only if migration is a return move, and it takes value two in the last month prior to second migration only if migration is a move to another city.

I first consider all migrants, whatever the size of the city they relocated to. Then, I focus only on migrants who moved to the six biggest cities. All specifications include categories of years elapsed since migration as a way to capture duration dependence in an additive and flexible way. Also, I include as controls labor market characteristics in the city of first employment by computing averages over pre-migration spells (e.g. percent of time spent unemployed, self-employed, in the public sector or under a temporary contract). Given the smaller sample size, instead of including indicator variables for urban areas interacted with 10-year periods, I add urban area indicators for both the first city and the city of destination and cluster the standard errors at both locations. 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 indicate, once again, that selection is driven by migrants who initially moved to big cities. In column (3a) a 10% increase in earnings at the first location (i.e., prior to the first migration episode) makes return migration 0.8% less likely.[28] 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 1.2% less likely, after controlling for average earnings in the first city and labor market characteristics in both cities. In column (4a) I include education and occupational skills as well as earnings in both cities. As for the case of initial migration, observable skills are strong determinants of return migration. Having tertiary education reduces the probability of returning by 31% while working in an occupation demanding very-high skills decreases the odds of returning by 17%, this latter effect being close to significance at the 10% level. However, even within education categories and occupational skills, higher realized earnings in big cities make return migration less likely, although the magnitude of the effect attenuates.

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

[28] $10\% \times (0.79 - 1)/e = -0.8\%$

Table 5: Multinomial logit estimation of determinants of second migration

	First move to city of any size				First move to any of 6 biggest cities			
	Return	Move on	Return	Move on	Return	Move on	Return	Move on
	(1a)	(1b)	(2a)	(2b)	(3a)	(3b)	(4a)	(4b)
Log mean earnings 2nd loc.	0.866	1.382	0.937	1.353	0.664	1.056	0.746	1.071
	(0.099)	(0.135)***	(0.085)	(0.106)***	(0.040)***	(0.091)	(0.049)***	(0.099)
Log mean earnings 1st loc.	0.812	1.091	0.855	1.068	0.790	1.117	0.862	1.115
	(0.033)***	(0.078)	(0.034)***	(0.077)	(0.047)***	(0.120)	(0.059)**	(0.121)
Self-employed 2nd location	0.341	0.242	0.346	0.283	0.212	0.197	0.225	0.297
	(0.049)***	(0.056)***	(0.046)***	(0.061)***	(0.036)***	(0.085)***	(0.042)***	(0.139)***
Public sector 2nd location	0.819	0.821	0.860	0.821	0.759	0.699	0.808	0.701
	(0.072)**	(0.059)***	(0.081)	(0.061)***	(0.128)	(0.136)*	(0.135)	(0.141)*
Temporary contract 2nd loc.	1.148	1.113	1.146	1.142	1.186	1.108	1.179	1.123
	(0.033)***	(0.064)*	(0.031)***	(0.061)**	(0.056)***	(0.098)	(0.057)***	(0.100)
Unemployed 2nd location	1.711	2.211	1.649	2.690	1.396	2.004	1.340	3.055
	(0.175)***	(0.160)***	(0.159)***	(0.390)***	(0.164)***	(0.280)***	(0.208)*	(0.732)***
Unemployed exp.ben. 2nd loc.	7.501	8.053	7.488	8.088	9.469	10.471	9.433	10.512
	(0.714)***	(1.015)***	(0.713)***	(1.014)***	(1.834)***	(1.900)***	(1.817)***	(1.903)***
Years of experience	1.047	0.965	1.037	0.970	1.046	0.965	1.026	0.966
	(0.006)***	(0.009)***	(0.005)***	(0.009)***	(0.008)***	(0.011)***	(0.008)***	(0.012)***
Years of firm tenure	1.050	1.042	1.053	1.038	1.071	1.056	1.076	1.052
	(0.012)***	(0.013)***	(0.012)***	(0.013)***	(0.014)***	(0.017)***	(0.014)***	(0.017)***
Tertiary education			0.892	1.240			0.688	1.119
			(0.118)	(0.122)**			(0.051)***	(0.116)
Secondary education			0.865	1.173			0.840	1.170
			(0.036)***	(0.062)***			(0.042)***	(0.114)
Very-high skills			0.797	1.112			0.829	1.360
			(0.031)***	(0.124)			(0.104)	(0.346)
High skills			0.878	1.110			0.882	1.403
			(0.046)**	(0.124)			(0.099)	(0.283)*
Medium skills			1.013	1.260			1.046	1.547
			(0.028)	(0.148)**			(0.100)	(0.283)**
Low skills			0.997	1.202			0.988	1.611
			(0.028)	(0.146)			(0.093)	(0.306)**
Urban area indicators 1st loc.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Urban area indicators 2nd loc.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Labor market controls 1st loc.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Years since migration ind.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age categories	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year indicators	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,495,790		1,495,790		594,388		594,388	
Migrants	32,368		32,368		11,650		11,650	
Pseudo R^2	0.069		0.069		0.066		0.068	

Notes: Relative risk ratios (exponentiated coefficients) reported with standard errors in parentheses clustered at the urban areas of first and second location. ***, **, and * indicate significance at the 1, 5, and 10 percent levels. The reference category is permanent migrants who remain in the first city of destination. In columns (1)–(4), sample is migrants after their first move. In columns (5)–(8), sample is migrants after their first move to one of the six biggest cities, where only an increase in city size is considered. All specifications include a male indicator and month indicator variables. The omitted categories are primary education and very-low skills. Labor market controls in first location are averages over pre-migration spells and are the same as those in tables 2 and 3. Log mean earnings in second location are 6-month moving averages, excluding current earnings.

return decision. 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 big cities can be characterized as those individuals with initial skills in between those of stayers and those of permanent migrants, that are also not successful in boosting their earnings after migrating.

6. Conclusions

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

Such selection is largely driven by the group of migrants who moves from small to big cities. The effects of differences in education, occupational skills, or relative earnings on the probability of migrating to big cities are much 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 to a large extent (but not totally) accounted for by differences in observable characteristics, such as education and occupational skills. In fact, the drop in pre-move earnings after conditioning on observables averages 65%. Moreover, the marginal effect of relative earnings on the probability of migrating declines by more than half once I control for education and occupational skills. This suggests that selection on unobservables, while present in the data, appears to be of less 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 a third of migrants move for a second time within five years of arriving in their city of destination and 65% of these moves involve a return migration. Return moves are more frequent and happen sooner in big cities. I find that returnees from big cities tend to exhibit skills in between those of stayers and those of permanent migrants. They are also typically those who have been least successful in boosting their earnings after migrating to a big city. This pattern seems to be specific to them as opposed to other repeat migrants from big cities. When I examine second-time moves of migrants to other cities, they are not affected by realized earnings after their first migration episode.

All of this indicates that positive sorting in big cities through migration is important, but that differences in observable characteristics account for much of the observed differences in skill sorting due to mobility. At the same time, for workers who have already migrated to big cities, further sorting is driven not just by workers' initial productivity but also by improvements in that productivity, perhaps as a result of the large set of opportunities and learning advantages that big cities provide. While I document in detail how initial and return migration contributes to the sorting of more skilled workers into bigger cities, it is worth noting that worker sorting can occur through other channels (Combes, Duranton, and Gobillon, 2012). For instance, both faster learning associated with working in big cities (Baum-Snow and Pavan, 2012, De la Roca and Puga, 2012) and better schools in large 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, Matthew Freedman, and Ronni Pavan. 2014. Why has urban inequality increased? Processed, Brown University.

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

Behrens, Kristian, Gilles Duranton, and Frédéric Robert-Nicoud. 2014. Productive cities: Sorting, selection, and agglomeration. *Journal of Political Economy* 122(3):507–553.

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* 23(1):1–76.

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

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

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

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

Card, David, Jörg Heining, and Patrick Kline. 2013. Workplace heterogeneity and the rise of West German wage inequality. *Quarterly Journal of Economics* 128(3):967–1015.

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. 2011. The identification of agglomeration economies. *Journal of Economic Geography* 11(2):253–266.

Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon. 2012. The costs of agglomeration: Land prices in French cities. Processed, Wharton School, University of Pennsylvania.

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.

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

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

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.

Davis, Donald R. and Jonathan I. Dingel. 2012. A spatial knowledge economy. Working Paper 18188, National Bureau of Economic Research.

De la Roca, Jorge. 2014. Wage cyclicality: Evidence from Spain using social security data. SERIEs *Journal of the Spanish Economic Association* 5(2–3):173–195.

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

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

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.

Eeckhout, Jan, Roberto Pinheiro, and Kurt Schmidheiny. 2014. Spatial sorting. *Journal of Political Economy* 122(3):554–620.

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.

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):129–138.

Kaestner, Robert and Ofer Malamud. 2014. Self-selection and international migration: New evidence from Mexico. *Review of Economics and Statistics* 96(1):78–91.

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.

McKenzie, David and Hillel Rapoport. 2010. Self-selection patterns in Mexico-U.S. migration: The role of migration networks. *Review of Economics and Statistics* 92(4):811–821.

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

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 2004–2011, uncensored earnings data are available from matched income tax returns for all workers in the MCVL except for self-employed workers and for those in the Basque Country and Navarra (where income taxes are not collected by the Central Government). In addition, for the entire period 1981–2011 earnings data are available from the social security for all workers in the MCVL, including self-employeds and those in the Basque Country and Navarra, but these are capped for some workers. In particular, 8.1% and 6.6% of monthly earnings observations are top-and bottom-coded, respectively. This appendix explains how I estimate earnings for this 14.7% of observations that are capped. A more detailed description of the methodology and further extensions are provided in De la Roca (2014).

I construct estimates of earnings using Tobit estimations with individual, job and location characteristics, as well as individual-specific variables that use the longitudinal dimension of the MCVL. For these estimations of capped earnings I enlarge the sample to include all individuals who were born between 1916 and 1993 and aged 18 to 65 throughout 1981–2011. I restrict observations following the same criteria used to construct the final sample in the study, except for the age restriction. I also exclude self-employed workers because the incidence of censoring for them is negligible.

Censoring bounds vary by type of occupation in the social security on an annual basis. In order to identify these bounds, I use historical information from Spain's *Boletín Oficial del Estado* and plot monthly earnings densities. I calculate daily top and bottom censoring bounds by dividing monthly bounds by the number of calendar days in each month. Daily wages that exceed daily censoring bounds are flagged as top-coded and viceversa.

Next, I estimate 1,240 Tobit regressions by groups of age, occupation and year (4 age groups × 10 occupations × 31 years).[29] The dependent variable is log daily earnings expressed in December 2011 euros and as explanatory variables I include age and sets of indicator variables for gender, level of education, temporary contract, part-time contract, province of workplace and month. To account for the high persistence in earnings found in studies that use career-long earnings histories, I follow Card, Heining, and Kline (2013) on their wage imputation for German social security data. More specifically, I include the worker's mean of log daily wages over her career (excluding the current wage) and the fractions of top or bottom censored wage observations over her career (again excluding the current censoring status).

Using the coefficients of these Tobit regressions, I can predict the value of earnings *only* for capped observations as follows:

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

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, job and location characteristics, including the mean of log daily wages and the fraction of censored wage observations in all other periods, $\hat{\gamma}$ and $\hat{\sigma}$ are estimated parameters, and ε_{ijt} is an i.i.d shock. Finally, since I know whether daily earnings are top- or bottom-coded, I force predicted earnings to be above or below the corresponding bound, respectively.[30]

Fit of predicted earnings

To verify the accuracy of estimated earnings, I compare predicted values of earnings in top- and bottom-coded observations in social security records relative to actual uncensored earnings in

[29] The four age groups are the following: 18–29, 30–39, 40–49 and 50–65.

[30] This implies drawing i.i.d shocks from a truncated normal distribution. In particular, if $k_b = \Phi[(b_{ijt} - x_{ijt}{}'\hat{\gamma})/\hat{\sigma}]$, where Φ represents the standard normal density, b_{ijt} is the daily wage level at which top censoring occurs and $u \sim U[0,1]$ is a uniform random variable, then I define $\varepsilon = \Phi^{-1}[k_b + u \times (1 - k_b)]$. Likewise, if $k_a = \Phi[(a_{ijt} - x_{ijt}{}'\hat{\gamma})/\hat{\sigma}]$, where a_{ijt} is the daily wage level at which bottom censoring occurs, then I define $\varepsilon = \Phi^{-1}[k_a \times u]$.

Table A.6: Order of autocorrelations for actual and predicted earnings

Order	Actual	Predicted
1	0.918	0.913
2	0.879	0.882
3	0.848	0.855
4	0.820	0.832
5	0.797	0.815
6	0.780	0.800
7	0.762	0.786

Table A.7: Selected percentiles for actual and predicted earnings

	All workers		Skilled workers	
	Actual	Predicted	Actual	Predicted
Percentile 5	43.2	42.2	42.9	42.3
Percentile 10	50.0	49.2	51.0	50.5
Percentile 25	62.4	61.0	67.3	67.2
Percentile 50	82.9	81.7	86.4	85.6
Percentile 75	119.2	119.7	117.7	121.4
Percentile 90	166.7	170.5	163.7	171.4
Percentile 95	210.8	224.6	201.4	203.3

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

income tax returns for the same individual and month in those years where both are available (2004–2011). If the fit is satisfactory for 2004–2011, I can be more confident that predicted earnings do a good job in approximating capped earnings for 1981–2003.

The correlation between predicted and actual values for capped month-individual observations in 2004–2011 is high at 0.81. Table A.6 shows that the estimated order of autocorrelations for actual and predicted earnings look very similar with predicted earnings exhibiting slightly higher levels of autocorrelation. In addition to predicting with high accuracy individual values, predicted earnings also reproduce well the shape of the earnings distribution. This can be seen in table A.7, which presents selected percentiles of the distributions of actual and predicted earnings for all workers and for skilled workers. Overall, the distributions are quite similar. Even for skilled workers, who are top-coded beyond the 58^{th} percentile, predicted earnings approximate salaries quite well in capped percentiles.

Appendix B. Short-term and short-distance migration

Following the estimations proposed in tables 2, 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 B.8, I estimate a multiple-exit discrete duration model (instead of a conditional hazard rate model as in table 2) including short-term/distance migration on the one hand and long-term/distance migration on the other as alternative possibilities. The dependent variable takes value one if the first migration is a short-term/distance move and value two if it is a long-term/distance move.

The table shows clearly that the determinants of short-term/distance migration, which often does not require a permanent change in residence, are very different from the determinants of long-term/distance migration. The estimates for long-term/distance migrants are basically identical to those in the main text. Regarding short-term /distance migrants, they instead exhibit only slightly higher pre-move earnings than stayers (see column 1a). Thus, once I take into account the fact that short-term/distance migrants have accumulated less labor market experience and are in more unstable occupations (e.g., more likely to be under a temporary contract and unemployed), they are no longer less productive than stayers in their first city, as they appeared to be in the raw data. Furthermore, they do tend to be special on some dimensions. For instance, occupational skills and education now work in opposite directions (see column 2a), suggesting perhaps that over-educated workers in low-skill occupations are more likely to engage in short-term/distance moves. In any case, these types of job changes for a very short period or to a nearby urban area are rather different and best studied separately.

Appendix C. Migration to small cities

In table C.9 I estimate the conditional hazard rate of moving to a small city in Spain, defined as those cities with size below the median-sized city (Santiago de Compostela). Thus, I estimate the probability of moving to one of these small cities among those who have not moved before and do not move elsewhere. For migrants who move within these small cities, I keep only those moves with a decline in city size. Other than this, all specifications are identical to those in tables 2 and 3.

Results show there is no clear evidence of selection of any type for migrants who move to small cities. While individuals with a tertiary education have a 65% higher probability of moving to small cities, workers in occupations demanding medium to very-high skills are not more likely to follow such move (see column 1). In fact, estimates (though not significant) suggest that working in such occupations reduces the probability of moving to a small city. Column (2) confirms the absence of any type of selection given that average pre-migration earnings do not influence the odds of moving to a small city. Further, when I include earnings and observable skills in column (4), individuals with tertiary education continue to have a higher probability of moving to small cities, but now the effects of working in high-skilled occupations become stronger and deter such type of move. These findings suggest that college-educated individuals who are mismatched in their city of first employment are more likely to migrate to small cities. Overall, the absence of clear selection in observable skills is in sharp contrast with the strong positive selection of migrants who move to big cities (see table 3).

Table B.8: 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	1.108 (0.069)*	1.711 (0.130)***	1.075 (0.044)*	1.336 (0.071)***
Tertiary education			1.502 (0.121)***	2.170 (0.293)***
Secondary education			1.088 (0.054)*	1.510 (0.088)***
Very-high skills			0.785 (0.064)***	1.173 (0.049)***
High skills			0.696 (0.034)***	1.037 (0.038)
Medium skills			0.849 (0.030)***	1.177 (0.059)***
Low skills			0.886 (0.020)***	1.028 (0.029)
Male	1.014 (0.029)	1.146 (0.027)***	1.032 (0.034)	1.262 (0.015)***
Years of experience	0.964 (0.005)***	0.914 (0.005)***	0.972 (0.003)***	0.938 (0.005)***
Years of firm tenure	0.924 (0.007)***	0.917 (0.009)***	0.924 (0.007)***	0.912 (0.010)***
Self-employed	0.293 (0.024)***	0.630 (0.050)***	0.254 (0.020)***	0.618 (0.053)***
Public sector employee	1.076 (0.109)	0.804 (0.049)***	1.074 (0.110)	0.742 (0.043)***
Temporary contract	1.483 (0.038)***	1.378 (0.023)***	1.469 (0.038)***	1.400 (0.022)***
Unemployed	1.895 (0.080)***	1.716 (0.099)***	1.719 (0.077)***	2.011 (0.081)***
Unemployed × expired benefits	10.868 (0.398)***	9.856 (0.310)***	10.838 (0.391)***	9.690 (0.294)***
Urban area × period indicators	Yes	Yes	Yes	Yes
Age indicators	Yes	Yes	Yes	Yes
Observations	32,035,795		32,035,795	
Pseudo R^2	0.091		0.094	

Notes: Relative risk ratios (exponentiated coefficients) are reported on a sample of 409,186 individuals with standard errors in parentheses clustered at the urban area level. ***, **, and * indicate significance at the 1, 5, and 10 percent levels. Sample is all individuals who are still in their first city. Dependent variable takes value one if migration is short-term and short-distance and value two if it is long-term and long-distance. Long-term long-distance moves are those that exceed 12 months in destination and distance of 120 km. All specifications include month indicator variables. Log mean earnings are 6-month moving averages, excluding current earnings. Primary education and very-low skills are the omitted categories.

Table C.9: Logit estimation of determinants of first migration to small cities

	Dep. variable: long-term long-distance migration to cities below median city size			
	(1)	(2)	(3)	(4)
Log mean earnings		1.167 (0.190)		1.050 (0.127)
Richest earnings tercile			1.114 (0.114)	
Poorest earnings tercile			0.954 (0.051)	
Tertiary education	1.651 (0.337)**			1.634 (0.306)***
Secondary education	1.068 (0.122)			1.063 (0.114)
Very-high skills	0.892 (0.141)			0.867 (0.094)
High skills	0.931 (0.063)			0.911 (0.055)
Medium skills	0.817 (0.113)			0.811 (0.098)*
Low skills	0.984 (0.060)			0.980 (0.057)
Male	1.207 (0.037)***	1.148 (0.047)***	1.152 (0.042)***	1.198 (0.037)***
Years of experience	0.937 (0.008)***	0.925 (0.012)***	0.925 (0.012)***	0.936 (0.009)***
Years of firm tenure	0.917 (0.016)***	0.916 (0.017)***	0.916 (0.016)***	0.916 (0.018)***
Self-employed	0.511 (0.057)***	0.584 (0.071)***	0.584 (0.065)***	0.520 (0.068)***
Public sector employee	1.428 (0.193)***	1.467 (0.211)***	1.481 (0.218)***	1.412 (0.195)**
Temporary contract	1.476 (0.079)***	1.498 (0.079)***	1.498 (0.080)***	1.479 (0.080)***
Unemployed	2.044 (0.153)***	2.087 (0.187)***	2.090 (0.188)***	2.034 (0.153)***
Unemployed $\times$ expired benefits	9.584 (0.872)***	9.646 (0.916)***	9.648 (0.914)***	9.595 (0.886)***
Urban area $\times$ period indicators	Yes	Yes	Yes	Yes
Age indicators	Yes	Yes	Yes	Yes
Pseudo R^2	0.063	0.061	0.061	0.063

Notes: Odd ratios (exponentiated coefficients) are reported on a sample of 31,020,861 monthly observations and 398,473 individuals. Standard errors in parentheses are clustered at the urban area level. ***, **, and * indicate significance at the 1, 5, and 10 percent levels. Sample is all individuals who are still in their first city. The reference category is stayers. Long-term long-distance moves are those that exceed 12 months in destination and distance of 120 km. Dependent variable takes value one if destination is a city with size below the median-sized city (Santiago de Compostela) *and* migrants experience a decline in city size. All specifications include month indicator variables. Period is a ten-year interval. Log mean earnings are 6-month moving averages, excluding current earnings. Earnings terciles are constructed for all year-province pairs. Primary education and very-low skills are the omitted categories.

In tables 3 and C.9 I only exploit moving decisions for some long-term long-distance migrants. In particular, in the former I use only those migrants who move to one of the six biggest cities and experience an increase in city size, while in the latter I use only those who move to cities below the median-sized city and experience a decline in city size. An alternative estimation is to model the probability of migration by splitting it in two comprehensive alternatives: moving to a bigger city or to a smaller one. This can be done using a multinomial logit specification or, again, modeling two separate conditional hazard rates.

Results in these estimations (available upon request) confirm that migrants who move to bigger cities largely drive the positive selection of all migrants in terms of earnings and observable skills. For these migrants I find that a 10% increase in log mean monthly earnings raises the probability of migrating by 4.2% (odds ratio of 2.139 and standard error of 0.103) while for migrants who move to smaller cities the effect is lower at 1.8% (odds ratio of 1.485 and standard error of 0.122). The difference in the role of earnings on mobility between both groups of migrants remains significant when I add in observable skills.

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