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

Jorge De la Roca*†

University of Southern California

This version, April 2017

Abstract: This paper investigates the contribution of migration to the sorting of more productive workers into big cities using administrative data for Spain that follow individuals over their work lives. While migrants to small cities do not exhibit selection of any type, migrants to big 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 big 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 big cities. Low realized earnings and unemployment affect return decisions of workers who moved to big cities at younger ages in particular, suggesting that older migrants may face less uncertainty upon moving to big cities.

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

*I thank Manuel Arellano, Samuel Bentolila, Laurent Gobillon, Pablo Lavado, Elena Manresa, Diego Puga, Anatoli Segura and Elisabet Viladecans for valuable comments and discussions. Funding from the Bank of Spain Excellence Programme is gratefully acknowledged. This work contains anonymized statistical data from MCVL CDF 2004–2013 which are used with the permission of Spain's Dirección General de Ordenación de la Seguridad Social.

†Sol Price School of Public Policy, University of Southern California, 650 Childs Way RGL 326, Los Angeles, CA 90089, USA (e-mail: `jdelaroc@usc.edu`; website: `http://jorgedelaroca.name`).

1. Introduction

Workers earn substantially more in big cities (Glaeser and Maré, 2001, Wheaton and Lewis, 2002, Combes, Duranton, Gobillon, and Roux, 2010, Moretti, 2012). 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 big cities facilitate the acquisition of greater skills (Glaeser, 1999, De la Roca and Puga, 2017). However, it has long been thought that those higher earnings may also partly reflect the sorting of more able workers into big cities (Combes, Duranton, and Gobillon, 2008). 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).

Exisiting empirical studies of worker sorting on ability across cities of different sizes compare differences in observable skills between big and small cities. Workers in big 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, which hints to an important role for learning advantages in big cities. De la Roca and Puga (2017) 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 innate skills distributions within broad levels of education and occupational groups. Baum-Snow and Pavan (2012) recover measures of ability from a finite-mixture model in a structural estimation setting and find evidence of positive sorting on observed skills to big cities. Again, they find that sorting on unobserved ability within levels of education has a minor contribution to the observed city-size earnings premium.

One common goal of these studies is to examine whether skills vary with city size in a given point in time; however, 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 of different sizes by studying whether greater productivity and skills increase the likelihood that a worker migrates to a big 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 drops substantially after conditioning on observable skills. In contrast, migrants who move to small cities do not exhibit clear selection of any type.

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, 67% of these second moves involve a return migration to the city of origin. Such return migration is more frequent in big cities. I find that to understand return migration, it is important to look not only 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 predicted 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 low realized earnings in the big city. Furthermore, I find different patterns of return migration for workers who moved to big cities at different ages. Workers who moved at later ages are more likely to return, yet, such return decision is not driven by a negative experience in the big city as proxied by low realized earnings or unemployment spells. These findings suggest that older first-time migrants to big cities are more rooted in their city of origin and, hence, more prone to return. At the same time, they seem to face less uncertainty than younger migrants upon moving to big cities.

This study contributes to our understanding of internal or regional migration within countries. Previous studies of regional migration (see Greenwood, 1997, for a survey) find that migrants tend to be more educated, employed in high-skilled 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. Further, 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 and finds that migrants are more skilled than stayers. Bauernschuster, Falck, Heblich, Suedekum, and Lameli (2014) also show that German migrants are more educated and less risk averse than stayers, which makes them less sensitive to cultural differences across regions. I contribute to this literature by using cities (instead of states or regions) as the spatial units of analysis, and showing that migrants from small to big cities are key to understand why migrants are largely positively selected. This positive selection validates 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.[1]

[1]Several recent studies document the strong positive relationship between earnings, inequality and city size (Baum-Snow and Pavan, 2013, Eeckhout, Pinheiro, and Schmidheiny, 2014, De la Roca and Puga, 2017).

The findings in this study also add to the literature in urban economics that examines mobility of individuals across cities with different consumption and production amenities (Rappaport, 2009, Sinha and Cropper, 2013). Chen and Rosenthal (2008) follow a spatial equilibrium framework to rank US cities by quality of life (i.e., cities where real wages are lower as individuals are willing to forgo part of their earnings to enjoy attractive areas) and by quality of business environment (i.e., cities where firms are willing to incur higher rents and wages to access more productive workers). They show that, in the 90s, cities with higher indices of quality of life tended to attract retirees and married couples older than 55, while cities with higher indices of business environment were more appealing to highly-productive individuals aged between 20 and 35. As big cities exhibit higher levels of quality of business environment or agglomeration economies, this latter finding conforms to the positive selection I observe for migrants from small to big cities in Spain.

Finally, the study provides new insights on the degree of selection in return migration from big cities and underscores the role of uncertainty. When examining second migration decisions of individuals, I allow productivity and 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.[2] Considering selection on the basis of productivity and skills observed in the first and 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.

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

[3] 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 (2015) for international return migration in European countries.

2. Conceptual framework

I develop a simple conceptual framework to motivate the empirical analysis. 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.[4]

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 $\delta_i \sim U[0, 2\delta]$.[5] Second, workers with any given level of skills are α times more productive (and earn α times more) when working in city H.[6] Third, housing in city H involves an extra rental cost R.[7] 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, $\mathbb{E}(\delta_i) = \delta$.

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

[5]Glaeser (1999) develops a learning model where young workers who move to a big city increase their skills with some probability. De la Roca and Puga (2017) find evidence of substantial skill acquisition by workers in big 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.

[6]This feature is widely documented in the literature on agglomeration economies. See Duranton and Puga (2004) for microfoundations and Rosenthal and Strange (2004) and Combes and Gobillon (2015) for reviews of the evidence.

[7]This feature is also widely documented in the literature. See Combes, Duranton, and Gobillon (2016) for a recent estimate of urban costs with respect to city population using data for all house and land transactions in France.

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.[8] 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.[9]

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.[10] 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.

I 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

[8] Return migration is one mechanism that can help explain why ability distributions overlap between big and small cities. In Behrens, Duranton, and Robert-Nicoud (2014) more talented individuals sort initially into big 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 workers with high and low skills in big cities.

[9] Following Borjas and Bratsberg (1996), return migration can either be modeled as a correction to an unfavorable draw (or 'mistake') given the uncertainty of economic conditions at destination, or as part of an optimal planned location trajectory over the life cycle. Both models deliver similar testable implications where return migration intensifies the selection driven by the initial migration flow. I do not attempt to distinguish both hypotheses in the empirical section. However, two features of the data suggest that return migration may be generally related to experimentation in big cities or viewed as a correction to an unfavorable outcome. Most returnees are young at the time of their return move from a big city (73% of them under 35) and the great majority of return moves happen a few years after the initial move (77% of individuals return within four years).

[10] 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 this increase has to be on average 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.

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, I 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.[11] 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

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

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 $\underline{s} \leqslant s_i < \overline{s}$, substituting equation (6) into (7) and simplifying turns this condition into

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

[11] 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.

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.[12]

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.[13]

For workers with $s_i < s_*$ (see footnote 13), the condition of equation (4) is always satisfied as $s_* < \tilde{s}$, where $\tilde{s}$ 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 $s_i \leqslant \tilde{s}$ migrate to city L. Lastly, for workers with $s_i \geqslant s*$, the condition of equation (4) is never satified since $\tilde{s} < s^*$. 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.

[12] 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.

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

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

where

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

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:

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

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, $\beta_0(t)$ 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,

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

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.[14] Once again, it is possible to introduce multiple alternatives, such as return migration to the city of origin or move-on migration to a third city.

4. Data

In order to examine selection in initial and return migration I need a data set that follows individuals over time and across locations from the beginning of their work 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 I need the data to record the labor market characteristics of all workers with high frequency since the start of their first job.

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

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 1980. I combine data from ten editions of the MCVL (2004 to 2013), so as to have data on a 4% sample of all individuals who have worked, received benefits or a pension at any point in 2004–2013.

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 remain in the sample while they receive unemployment benefits or a retirement pension, so most exits occur when individuals die or leave 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 type of contract within the same firm). Given that dates of all changes since 1980 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 work life histories since either 1980 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 social security regime. 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 Continuo* or Municipal Register. I construct precise measures of cumulative labor experience and job tenure recording the actual number of work days in each month.

The data include monthly earnings for each job spell, constructed by combining a variety of sources. For the period 2004–2013, uncensored earnings data are available from matched income tax returns for all workers except some self-employed workers and those in the Basque Country and Navarra (where income taxes are not collected by the Central Government). In addition, for the entire period 1980–2013 earnings data from the social security are available for all workers, including all self-employed workers and those in the Basque Country and Navarra, but these are capped for a small fraction of observations.[15]

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 identifiers for each province in which they conduct business. Further, 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. Consequently, location information is at the establishment level.

[15] Appendix A provides details on how earnings are estimated for this small fraction of capped observations (9.9%), following the methodology developed in Card, Heining, and Kline (2013).

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 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 the 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 contains four municipalities. I cannot identify three small urban areas in the MCVL because the population of their largest municipality is below the 40,000 population threshold.[16]

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 (2017). 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 (Combes and Gobillon, 2015). 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.[17] In any case, results are robust to measuring the scale of each urban area by its total population.[18]

Sample restrictions

My initial sample is made up of Spanish natives born between 1962 and 1995 (i.e., aged 18–52 during the period 1980–2013) 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 52 in 2013 and foreign-born immigrants since I cannot retrieve complete work histories for them. A total of 626,646 individuals and 98,660,359 monthly observations make up this initial sample.

From this initial sample, I exclude 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 617,425 individuals and 92,926,124 observations.

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

[17]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.

[18]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.

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 cut leaves the sample at 610,452 individuals and 88,868,872 monthly observations. Then, I eliminate individuals with very 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 1980 and 2013. This restriction further reduces the sample to 555,678 individuals and 88,340,673 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 521,226 individuals and 60,562,483 monthly observations.

Identifying migrants

An urban migration event is defined as a change in workplace location between two urban areas. In the sample, when looking at first moves, 123,834 individuals are classified as urban migrants, 27,124 as urban-rural migrants and 20,904 as rural-urban migrants, while 296,696 individuals never leave their urban area for work purposes.[19] Again, I focus only on urban or city-to-city migrations which based on first moves account for 71% of all types of moves.[20]

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. Therefore, I detect permanent changes in residence based on the length of the migration episode and the distance between the cities of origin and destination. Both the conceptual framework of section 2 (the change in housing costs in the framework is associated to a change in residence) and the empirical results presented below suggest that the behavior of short-term or short-distance migrants is rather different from that of long-term and long-distance migrants.

Regarding the length of the migration episode, short-term migrants usually move for brief transfers within a job or to work in a seasonal or temporary job. I classify migrants as *short-term* if they never move beyond a 12-month period. Therefore, long-term migrants are movers who experience spells longer than a year in the city of destination. Based on this criterion, I identify 42,419 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 kms (74.6 miles) driving from the urban area

[19] A total of 1,687 individuals move for the first time across rural areas and later enter the sample when they work in an urban area. To identify moves these moves, I track changes in workplace location between two provinces (52 in total) that lack an urban area identifier, prior to restricting the sample to only urban areas. In addition, 35,882 individuals register occasional moves shorter than a month. These moves are unlikely to involve a change of residence and may appear in the MCVL when individuals work for one or several days in another firm (e.g. one-day service contracts). I exclude these individuals from the group of stayers in table 1, but include their monthly observations in all the estimations, except for the month where they appear linked to a firm in another urban area. Finally, I classify 15,099 individuals as 'special' since they move ten or more times over their lives or record a handful of short job spells (e.g. they appear in the sample with only two short job spells of ten months with many years of inactivity in between). I exclude these special individuals from the estimation sample.

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

where they previously worked. Although urban areas can be conceived 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 large urban areas such as Madrid and Barcelona, which tend to have adjacent smaller urban areas at reasonable commuting distances.[21] Based on this criterion, I identify a total of 37,881 short-distance migrants.

These classification criteria leave 43,534 long-term and long-distance migrants. Of these, 62% move only once in their lives while 30% 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 moves. A total of 32,411 migrants do not return to their city of first employment while 11,123 migrants end up returning.

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 in their second move). In this table, variables are averages over work lives; however, in all estimations I transform the data into a duration analysis format and identify migrants at the time of the migration event, in line with the econometric framework described in the previous section.

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).[22]

This ranking is confirmed by the types of occupations in which individuals tend to work. These occupation categories, assigned by employers, capture the skills required by the job and tend to be related to the level of formal education required for the job. Permanent migrants are twice more likely to work in occupations demanding very-high skills (those typically requiring an engineering or advanced degree) than stayers and short-term/distance migrants. The ranking of permanent migrants, followed by return migrants, stayers, and then short-term/distance migrants

[21] I have collected data on the shortest driving distance between any two urban areas using Google Maps. A threshold of 120 kms is large enough to rule out excessive commutes in Spain. In Madrid and Barcelona, the two urban areas where individuals experience the longest commuting distances, workers can travel to a few nearby urban areas by commuting rail, but none of them is located more than 60 kms away. For instance, traveling from downtown Madrid to Aranjuez (49 kms) and Guadalajara (60 kms) takes 45 and 60 minutes, respectively (excluding the time spent walking or in other transportation modes like local buses). These commuting times exceed by far the mean travel time to work in Spain of 31 minutes according to the Harmonised European Time Use Survey (2007). Other nearby urban areas linked to Madrid through market potential, such as Toledo (81 kms) and Segovia (93 kms), cannot be reached by commuting rail and only recently by high-speed rail at expensive fares. Likewise, in Barcelona, only Manresa (59 kms) is accesible by rail at an average time of 80-90 minutes, whereas Blanes-Lloret de Mar (74 kms), Tarragona (101 kms) and Girona (102 kms), though related in terms of market access, are not connected by commuting rail and require longer commuting times.

[22] Information on educational attainment is that contained in the *Padrón Continuo* or Municipal Register. A large update was held by municipalities in 2001 and further updates relied on voluntary information provided by individuals to municipalities, often when they completed their registration form at a new municipality upon moving (a prerequisite for access to local health and education services). In 2009, the Department of Education started notifying the Spanish Statistical Institute on recent granted diplomas and this information was included in newer versions of the *Padrón Continuo*. Still, for some individuals their levels of education might be underreported, so the fact that the educational ranking across migrant types and stayers holds when considering occupational categories provides further 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				
University	21%	20%	26%	32%
Secondary	34%	31%	35%	35%
Primary	45%	50%	39%	33%
Occupational skills				
Very-highed skilled	8%	7%	11%	15%
High skilled	12%	11%	14%	16%
Medium-high skilled	23%	21%	27%	24%
Medium-low skilled	44%	45%	38%	35%
Low skilled	13%	15%	10%	9%
Earnings				
Mean monthly earnings	1,713	1,604	2,002	2,092
Mean monthly earnings 2^{nd} city			2,067	2,389
Labour market characteristics				
Years of labor experience	8.1	7.1	7.9	7.3
Years of firm tenure	3.7	2.2	2.1	2.4
Self-employed	9%	6%	6%	5%
Public sector employee	7%	9%	8%	11%
Temporary contract	26%	40%	34%	33%
Part-time contract	14%	15%	11%	12%
Unemployed	11%	17%	15%	13%
Male	50%	55%	58%	53%
Age	31.8	31.4	32.5	32.2
Age of entry in labor force	21.6	21.2	21.3	22.0
Individuals	296,696	80,300	11,123	32,411

Notes: Variables are averages over work lives. Only individuals working in urban areas are included. Long-term long-distance migrations are those that exceed 12 months in the city of destination and a distance of 120 kms between city of origin and destination. Earnings are expressed in December 2013 euros. Firm tenure is calculated only for employees.

continues to hold going down to individuals in occupations with high skills. The ranking reverts for occupations demanding low and medium-low skills.[23]

The ranking of monthly earnings across categories again points in the same direction. Permanent migrants exhibit the highest lifetime 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.

[23]Employers assign workers into one of ten social security occupation categories, which I have regrouped into five skill groups. Note that it is the skills required by the job and not those acquired by the worker that dictate the social security category. For instance, someone with a business degree will have social security category 1 (or the 'very-high-skilled occupation' category) when working as a manager, and social security category 7 (included in the 'medium-low-skilled' category) when working as an office assistant.

Other labor market characteristics reveal expected patterns, as stayers are attached to more stable jobs (with permanent instead of temporary 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. Men are also more prone to migrate than women.

In my sample I observe work lives that start in the city of first employment. Although individuals may sort into a few big cities for tertiary education, this pattern is infrequent in a country like Spain where mobility of students is quite low. Until the late 1990s, students wishing to pursue higher education were assigned to a university based on proximity, usually within the same region or *Comunidad Autónoma*.[24] These restrictions were removed by 2003 allowing an open district admission where students could apply to any post-secondary institution in the country. Despite this reform, in 2009, only 12% and 23% of students migrated to another region and province for tertiary education, respectively, and the majority who moved remained within commuting distance and lived with their parents (CRUE, 2010). This incidence of low mobility to pursue higher education is also confirmed by a report by the Department of Education (Ministerio de Educación, Cultura y Deporte, 2013).

A belief about mobility in Spain and other Southern European countries is that it is generally lower than in the United States or Scandinavian countries. However, I do not find evidence of lower mobility rates across urban areas in Spain during the last two decades. Using data from the Internal Revenue Service in the United States, Molloy, Smith, and Wozniak (2011) show a declining trend in annual mobility rates across metropolitan areas between 1985 and 2009 (see panel C in figure 2 in the article where rates fluctuate approximately between 2.8% and 3.7%). Annual mobility rates across urban areas in the MCVL range from 2.6% and 3.9% between 1990 and 2012. Interestingly, as opposed to the downward trend in mobility rates observed in the US, there is a steady increase in inter-metropolitan area migration rates in Spain since the early 90s until the mid 2000s, followed by a moderate decline during the Great Recession.[25]

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 skills and productive characteristics at the time of their first move relative to stayers in the same city. 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. Likewise, explanatory variables such as occupational skills and earnings change over time in order to estimate their effect on the probability of moving at the time of the migration event.

[24] The 19 regions and 52 provinces (plus 7 islands) correspond to the NUTS 2 and NUTS level 3 classification, respectively.

[25] Using Current Population Survey microdata, Molloy, Smith, and Wozniak (2011) report interstate mobility rates of 4.5% in the 90s and 3.0% in the 2000s for the age group between 18 and 24. The analogous rates for the age group between 25 and 44 were 3.1% and 2.2%, respectively. I calculate slightly larger inter-metropolitan mobility rates in Spain for both age groups in these two decades. Rates for the age group between 18 and 24 were 4.0% and 5.9% in the 90s and 2000s, whereas for the age group between 25 and 44 they were lower at 2.9% and 3.8%. Although interstate mobility rates are lower than inter-metropolitan rates, the gap is small as shown in figure 2 in Molloy, Smith, and Wozniak (2011).

I focus on long-term long-distance moves, i.e., only those moves that exceed 12 months in the city of destination and 120 kms of distance. The determinants of shorter moves are quite different and discussed in appendix B.

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 five-year periods to confine the analysis of migrants and stayers within an urban area and period. In addition, this allows me to control for unobserved location characteristics that may affect the probability of migration for all individuals in a city.[26] I cluster standard errors by urban area in all specifications.

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 117% relative to having less than secondary education, while working in an occupation that requires very-high skills increases the probability by 47% relative to working in an occupation with low skills.

Other labor market variables have expected signs. An additional year of labor market experience or tenure in the firm decreases the probability of out-migration, 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 or fixed-term contract (with significantly lower job protection) have less to lose from quitting their jobs and thus are 40% more likely to migrate. Self-employed workers and public sector employees appear to be more rooted in 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. More specifically, individuals receiving unemployment benefits are 179% more inclined to migrate than employed individuals. This probability drops to 19% for unemployed individuals receiving unemployment subsidies, a compensation granted to vulnerable income groups with insufficient social security contributions or for whom unemployment benefits have expired. Moreover, by controlling separately for unemployed individuals who have completed their period of entitlement to unemployment benefits or subsidies (with no further extensions), I find they are the ones that particularly drive the large effect of unemployment on the propensity to migrate. Once unemployment benefits or subsidies expire, the probability of migrating jumps by a factor of eight and nineteen, respectively.[27] The disparate

[26] I have combined periods 1980–1984 and 1985–1989 into a single period given that in the earlier years the sample is much smaller, made only by individuals born in the 1960s that enter the sample every year. This regrouping increases the number of moves from urban areas in this initial ten-year period, particularly in small cities. In alternative estimations I have further narrowed periods to three years and results are basically unaltered, however, this increases the number of fixed-effects to be estimated and computation time by one order of magnitude.

[27] I identify unemployed individuals with expired compensations 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. When I exclude controls on expired benefits and subsidies from the specification in column (1) of table 2, I find that the overall effects of receiving unemployment benefits and subsidies raise the odds of migrating by 279% and 115%, respectively.

Table 2: Logit estimation of determinants of first migration

	Dep. variable: long-term long-distance migration			
	(1)	(2)	(3)	(4)
Log mean earnings		1.607 (0.099)***		1.307 (0.057)***
Richest earnings tercile			1.462 (0.082)***	
Poorest earnings tercile			0.950 (0.020)**	
University education	2.170 (0.323)***			2.062 (0.305)***
Secondary education	1.471 (0.085)***			1.437 (0.084)***
Very-high-skilled occupation	1.370 (0.053)***			1.174 (0.044)***
High-skilled occupation	1.158 (0.042)***			1.031 (0.046)
Medium-high-skilled occupation	1.255 (0.044)***			1.194 (0.038)***
Medium-low-skilled occupation	1.058 (0.029)**			1.031 (0.029)
Male	1.244 (0.019)***	1.097 (0.018)***	1.125 (0.017)***	1.195 (0.017)***
Experience	0.925 (0.008)***	0.892 (0.006)***	0.895 (0.006)***	0.919 (0.008)***
Firm tenure	0.922 (0.010)***	0.928 (0.009)***	0.928 (0.009)***	0.920 (0.011)***
Self-employed	0.557 (0.035)***	0.609 (0.037)***	0.575 (0.039)***	0.616 (0.039)***
Public sector employee	0.816 (0.046)***	0.840 (0.050)***	0.866 (0.052)**	0.774 (0.040)***
Temporary contract	1.399 (0.025)***	1.386 (0.027)***	1.377 (0.025)***	1.416 (0.027)***
Part-time contract	1.082 (0.033)**	1.348 (0.069)***	1.175 (0.047)***	1.246 (0.053)***
Receiving unemployment benefits	2.792 (0.095)***	2.443 (0.094)***	2.392 (0.088)***	2.761 (0.098)***
Receiving unemployment subsidy	1.189 (0.063)***	0.981 (0.044)	0.948 (0.043)	1.209 (0.064)***
Expired unemployment benefits	7.964 (0.224)***	8.012 (0.223)***	7.999 (0.222)***	8.028 (0.226)***
Expired unemployment subsidy	19.212 (1.370)***	19.625 (1.363)***	19.600 (1.357)***	19.281 (1.368)***
Urban area $\times$ period indicators	Yes	Yes	Yes	Yes
Age indicators	Yes	Yes	Yes	Yes
Pseudo R^2	0.074	0.071	0.070	0.075

Notes: Odd ratios (exponentiated coefficients) are reported on a sample of 39,384,447 monthly observations and 419,254 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 city of destination and distance of 120 kms. All specifications include month indicator variables. Period is a five-year interval. Log mean earnings are 6-month moving averages, excluding current earnings. Earnings terciles are constructed for all years. Less than secondary education and low-skilled occupations are the omitted categories.

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 suggests 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.

The results so far indicate that workers with greater observable skills have a higher probability of out-migration. To check whether more productive workers, more broadly defined, are also more likely to migrate, I next proxy the productivity of each worker by her relative position in the local earnings distribution of her 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.[28] 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 2.2%.[29] 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-migration by 46%, while being in the lowest one decreases it by 5%.

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 probability of migrating. However, the effect of earnings drops by half relative to column (2)—recall coefficients are odd ratios—whereas observable skills like education remain almost as strong determinants of first-time migration as in column (1). Thus, differences in observable skills are 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.

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.

---

[28] 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, migrants' positive selection attenuates given that point estimates are lower, but only marginally and remain significant. Results are available upon request.

[29] Reported odd ratios are changes in the relative probability of out-migration 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 60.7%. The 2.2% reported in the text is calculated as $10\% \times (1.607 - 1)/e$.

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		1.952		1.407
		(0.129)***		(0.087)***
Richest earnings tercile			1.706	
			(0.063)***	
Poorest earnings tercile			0.966	
			(0.025)	
University education	3.982			3.761
	(0.209)***			(0.207)***
Secondary education	2.044			1.991
	(0.068)***			(0.066)***
Very-high-skilled occupation	1.474			1.228
	(0.109)***			(0.075)***
High-skilled occupation	1.124			0.973
	(0.078)*			(0.052)
Medium-high-skilled occupation	1.511			1.421
	(0.083)***			(0.072)***
Medium-low-skilled occupation	1.088			1.051
	(0.051)*			(0.048)
Male	1.376	1.119	1.164	1.312
	(0.032)***	(0.026)***	(0.027)***	(0.032)***
Experience	0.943	0.886	0.890	0.937
	(0.006)***	(0.006)***	(0.006)***	(0.006)***
Firm tenure	0.908	0.922	0.922	0.905
	(0.006)***	(0.006)***	(0.006)***	(0.006)***
Self-employed	0.697	0.759	0.687	0.790
	(0.060)***	(0.052)***	(0.060)***	(0.064)***
Public sector employee	0.533	0.562	0.598	0.493
	(0.065)***	(0.072)***	(0.070)***	(0.062)***
Temporary contract	1.369	1.329	1.324	1.385
	(0.038)***	(0.036)***	(0.037)***	(0.038)***
Part-time contract	1.177	1.606	1.307	1.411
	(0.048)***	(0.061)***	(0.049)***	(0.053)***
Receiving unemployment benefits	2.945	2.322	2.264	2.895
	(0.179)***	(0.119)***	(0.120)***	(0.177)***
Receiving unemployment subsidy	1.308	0.907	0.866	1.331
	(0.087)***	(0.053)*	(0.050)**	(0.092)***
Expired unemployment benefits	8.107	8.115	8.092	8.189
	(0.347)***	(0.345)***	(0.345)***	(0.350)***
Expired unemployment subsidy	21.406	22.365	22.315	21.504
	(1.852)***	(1.950)***	(1.946)***	(1.857)***
Urban area $\times$ period indicators	Yes	Yes	Yes	Yes
Age indicators	Yes	Yes	Yes	Yes
Pseudo R^2	0.090	0.078	0.077	0.091

Notes: Odd ratios (exponentiated coefficients) are reported on a sample of 29,479,766 monthly observations and 341,001 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 city of destination and distance of 120 kms. Dependent variable takes value one if destination is one of the six biggest cities *and* migrants experience an increase in city size. All specifications include month indicator variables. Period is a five-year interval. Log mean earnings are 6-month moving averages, excluding current earnings. Earnings terciles are constructed for all years. Less than secondary education and low-skilled occupations are the omitted categories.

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), long-term/distance migration to big cities is more than twice more likely to occur for individuals with tertiary education—the probability increases by 298% relative to an increase of 117% in column (1) of table 2. Having secondary education also raises substantially the odds, making migration to big cities twice more likely (104% in table 3 vs. 47% in table 2), while working in occupations demanding very-high and medium-high 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 3.5% (vs. 2.2% in column 2 of table 2). Column (4) shows that, even within given levels of education and occupational skills, more productive workers are more inclined to migrate to big cities, yet, the probability declines by almost 60% compared to column (2). In contrast, if I estimate the conditional hazard rate of moving 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 skills and 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 instead 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 five-year-period indicator variables. Therefore, I capture the correlation between earnings in the first city and being a permanent migrant or returnee, controlling for other observable 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 consider only those moves with an increase in city size. Permanent migrants to the biggest cities have significantly higher earnings than other permanent migrants and stayers. The same pattern is found when comparing earnings of returnees to biggest cities to those of 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 in 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

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	0.133 (0.011)***	0.040 (0.010)***		
Permanent to cities other than 6 biggest	0.068 (0.012)***	0.025 (0.009)***		
Permanent to 1^{st}–2^{nd} biggest cities			0.135 (0.013)***	0.037 (0.011)***
Permanent to 3^{rd}–6^{th} biggest cities			0.127 (0.016)***	0.052 (0.014)***
Permanent to 7^{th} biggest–median sized city			0.073 (0.010)***	0.027 (0.007)***
Permanent to cities below median size			0.030 (0.036)	0.007 (0.023)
Return migrant to 6 biggest cities	0.074 (0.016)***	0.021 (0.012)*		
Return to cities other than 6 biggest	0.065 (0.018)***	0.036 (0.010)***		
Return to 1^{st}–2^{nd} biggest cities			0.077 (0.019)***	0.020 (0.014)
Return to 3^{rd}–6^{th} biggest cities			0.063 (0.022)***	0.025 (0.020)
Return to 7^{th} biggest–median sized city			0.068 (0.020)***	0.036 (0.011)***
Return to cities below median size			.047 (.017)***	.032 (.011)***
Male	0.140 (0.005)***	0.163 (0.005)***	0.140 (0.005)***	0.163 (0.005)***
Tertiary education		0.207 (0.022)***		0.207 (0.022)***
Secondary education		0.103 (0.007)***		0.103 (0.007)***
Very-high skilled occupation		0.688 (0.022)***		0.688 (0.022)***
High-skilled occupation		0.496 (0.008)***		0.496 (0.008)***
Medium-high-skilled occupation		0.221 (0.004)***		0.221 (0.004)***
Medium-low-skilled occupation		0.109 (0.006)***		0.109 (0.006)***
Labor market characteristics	Yes	Yes	Yes	Yes
Urban area $\times$ period indicators	Yes	Yes	Yes	Yes
Age indicators	Yes	Yes	Yes	Yes
R^2	0.427	0.559	0.427	0.559

Notes: Coefficients reported on a sample of 39,625,875 monthly observations and 421,528 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 city of destination and distance of 120 kms. 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 five-year interval. Less than secondary education and low-skilled occupations are the omitted categories.

controls in column (2). Now, permanent migrants and eventual returnees earn only an extra 4% and 2% more than stayers with comparable levels of education and occupational skills. This sharp decline in the coefficients confirms that sorting in initial migration of the most productive workers into big 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 3rd–6th biggest cities (Valencia, Sevilla, Bilbao and Zaragoza), those who move to any city between the 7th biggest (Málaga) and the median-sized city (Santiago de Compostela), and those who move to small cities below the median-sized city. 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 medium-sized cities (not the six biggest) have earnings that are 8% higher than those of stayers at the time of migration. Further, permanent migrants moving to the 3rd–6th biggest cities earn 14% more than stayers and slightly less than permanent migrants moving to the top two cities. The same pattern in the ranking is found among types of return migrants, although differences across categories are more modest. When I include observable skills in column (4), earnings of permanent migrants to the top two cities are somewhat lower than those of permanent migrants to the 3rd–6th biggest cities and comparable to earnings of permanent migrants to medium-sized 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 match closely 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, 2016, Behrens, Duranton, and Robert-Nicoud, 2014).

In sum, the findings up to this point reveal that the positive selection of long-term and long-distance internal migrants found in extant studies (Borjas, Bronars, and Trejo, 1992, Hunt, 2004, Bauernschuster, Falck, Heblich, Suedekum, and Lameli, 2014) is essentially driven by migrants from small to big cities. This positive selection in skills resembles the findings in Chen and Rosenthal (2008) who show that highly productive individuals (e.g. highly educated) move to productive cities when young in order to take advantage of their skills. However, the finding that migrants to big cities who will eventually return are less skilled and productive than permanent migrants has not been documented in the urban economics literature. We turn to this issue below, after examining some extensions to our models of selection in initial migration.

Extensions: the role of distance and pre-labor market sorting

One natural extension to the analysis is the role that distance plays in the decision to migrate. So far, distance has been treated in a binary way, where any move that exceeds 120 kms is labeled as long distance. Based on the conceptual framework, if individuals could choose to move to several big cities, one could expect that mobility costs (C and C_2) increase with distance of destination. A higher mobility cost would intensify the degree of selection in a small city (L) and make the pool of migrants more skilled as they move to a more distant big city (H). At the same time, uncertainty in the acquisition of skills may amplify with distance, as migrants may be less informed about labor market opportunities in a big city that is farther away (i.e., the distribution of skills from which a workers draws δ_i may exhibit a higher variance in more distant big cities). A higher degree of uncertainty could induce more individuals with intermediate skills to experiment the big city and return in case of a bad realization of δ_i.

In table 5 I explore the sensitivity of the results to the distance of the city of destination. I divide long-term long-distance moves into three categories: below 300 kms, between 300 and 600 kms and over 600 kms. This categorization broadly creates groups of equal size in terms of the number of long-term long-distance migrants. In columns (1a), (1b) and (1c) I estimate a multiple-exit discrete duration model (instead of a conditional hazard rate model as in table 2) including these three categories as alternative possibilities. The dependent variable takes value one if the first migration does not exceed 300 kms, value two if it varies between 300 and 600 kms and value three if it exceeds 600 kms. Results indicate that the degree of positive selection in the pool of migrants does not vary much with distance of the city of destination. If anything, it appears that migrants that move more than 600 kms away are somewhat less selected than the other two types of migrants. For migrants incurring very long distances, the effects of having university or secondary education as well as working in very-high-skilled occupations have a lower incidence on the probability of migrating than for the other two distance categories. On the contrary, the effect of pre-migration earnings on the odds of moving is slightly larger.

In columns (2)–(4) I estimate the conditional hazard rate of moving to one of the six biggest cities for distances of at most 300 kms, between 300 and 600 kms and over 600 kms, respectively. In contrast to the findings on the probability of migrating to any city, migrants that move to big distant cities between 300 and 600 kms appear to be more skilled than migrants who move to big less-distant cities ($\leqslant$ 300 kms). However, this increase in the degree of selection is not observed for moves that exceed 600 kms. Therefore, I find mixed evidence on the effect of distance on the degree of selection of long-term long-distance migrants. It appears that the threshold of 120 kms that divides short-distance and long-distance moves is capturing to a large extent the differences in observable characteristics between both types of migrants as shown in appendix table B.11.[30]

One potential source of concern in the data set is the possibility of pre-labor market sorting by talented individuals as they decide to pursue tertiary education in big cities, where higher-quality universities tend to be located (Ahlin, Andersson, and Thulin, 2016). If this sorting exists then the city of first location is likely the city where students moved to pursue tertiary education and not the city of origin where individuals grew up. Although I have already noted there is low incidence

[30]Although my findings do not closely mirror those of Bauernschuster, Falck, Heblich, Suedekum, and Lameli (2014) who found that conditional on moving, more educated individuals in Germany incur in longer distances, they also show that more educated and risk-friendly migrants are less sensitive to cultural differences between regions of origin and destination. One could also imagine that for migrants from small cities, big cities are also more 'intimidating' or more culturally distant. In line with this pattern, I find that highly-skilled migrants are more able to adapt to this new environment.

Table 5: Logit estimation of determinants of first migration by distance of destination

	Long-term long-distance migration to any city (kms)			Long-term long-distance migration to 6 biggest cities (kms)		
	$\leqslant$300	>300 & $\leqslant$600	>600	$\leqslant$300	>300 & $\leqslant$600	>600
	(1a)	(1b)	(1c)	(2)	(3)	(4)
Log mean earnings	1.274 (0.086)***	1.205 (0.056)***	1.397 (0.069)***	1.302 (0.092)***	1.292 (0.085)***	1.619 (0.171)***
University education	2.242 (0.299)***	2.388 (0.614)***	1.617 (0.117)***	3.362 (0.460)***	4.557 (0.392)***	3.131 (0.236)***
Secondary education	1.463 (0.103)***	1.613 (0.138)***	1.265 (0.132)**	1.896 (0.109)***	2.143 (0.087)***	1.867 (0.110)***
Very-high-skilled occup	1.186 (0.095)**	1.325 (0.059)***	1.026 (0.074)	1.125 (0.146)	1.274 (0.111)***	1.344 (0.161)**
High-skilled occup	1.019 (0.057)	1.102 (0.068)	0.987 (0.053)	0.841 (0.097)	0.992 (0.064)	1.142 (0.117)
Medium-high-skilled occup	1.165 (0.070)**	1.276 (0.075)***	1.153 (0.049)***	1.268 (0.134)**	1.431 (0.108)***	1.604 (0.136)***
Medium-low-skilled occup	1.038 (0.046)	1.089 (0.042)**	0.961 (0.040)	0.910 (0.090)	1.126 (0.086)	1.090 (0.086)
Male	1.229 (0.035)***	1.254 (0.022)***	1.107 (0.025)***	1.400 (0.049)***	1.359 (0.043)***	1.182 (0.058)***
Experience	0.929 (0.006)***	0.916 (0.013)***	0.924 (0.010)***	0.941 (0.009)***	0.940 (0.009)***	0.941 (0.010)***
Firm tenure	0.931 (0.009)***	0.928 (0.017)***	0.896 (0.009)***	0.937 (0.011)***	0.894 (0.008)***	0.890 (0.011)***
Self-employed	0.568 (0.054)***	0.611 (0.070)***	0.671 (0.083)***	0.782 (0.136)	0.742 (0.092)**	0.873 (0.137)
Public sector employee	1.170 (0.072)**	0.581 (0.060)***	0.488 (0.056)***	0.932 (0.118)	0.350 (0.044)***	0.347 (0.053)***
Temporary contract	1.515 (0.056)***	1.337 (0.034)***	1.351 (0.037)***	1.517 (0.094)***	1.299 (0.057)***	1.345 (0.059)***
Part-time contract	1.211 (0.076)***	1.258 (0.092)***	1.298 (0.073)***	1.256 (0.075)***	1.508 (0.083)***	1.433 (0.083)***
Urban area × period indic	Yes	Yes	Yes	No	No	No
Urban area indicators	No	No	No	Yes	Yes	Yes
Year indicators	No	No	No	Yes	Yes	Yes
Age indicators	Yes	Yes	Yes	Yes	Yes	Yes
Unemployment indicators	Yes	Yes	Yes	Yes	Yes	Yes
Observations		39,443,417		11,641,153	18,683,563	25,983,130
Pseudo R^2		0.088		0.106	0.093	0.087

Notes: Relative risk ratios (exponentiated coefficients) are reported in columns (1a)–(1c) on a sample of 419,648 individuals while odd ratios (also exponentiated coefficients) are reported in columns (2)–(4) on samples of 154,041, 265,754 and 335,598 individuals, respectively. 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 city of destination and distance of 120 kms. In columns (1a)–(1c), dependent variable takes value one if distance to city of destination does not exceed 300 kms, value two if it varies between 300–600 kms and value three if it exceeds 600 kms, respectively. In column (2), dependent variable takes value one if distance to city of destination does not exceed 300 kms. In column (3) it takes value one if distance varies between 300–600 kms and in column (4) it takes value one if it exceeds 600 kms. All specifications include month indicator variables. Period is a five-year interval. Log mean earnings are 6-month moving averages, excluding current earnings. Less than secondary education and low-skilled occupations are the omitted categories.

of regional migration for educational purposes in Spain, I impose two further restrictions to the data to examine whether pre-labor market sorting is driving the results. First, in an analogous estimation to table 2, I exclude those moves in which the province of birth does not match the first observed province in the MCVL. Subsequently, I eliminate those moves in which the province of registration with social security does not match the first observed province in the MCVL. These restrictions reduce the number of moves by 40% and 24%, respectively, compared to the number of long-term long-distance moves examined in table 2.[31]

Individuals that did not change province in these two periods in time are less likely to sort into big urban areas prior to entering the labor market. One should note, however, that these restrictions may also be quite strong. Many individuals may have moved early in life as a result of parental migration. Others may have registered in social security when they started receiving any pension from elder relatives during childhood or when they worked in a seasonal/summer job elsewhere as teenagers. With these caveats in mind, I estimate an analogous version of table 2 in online appendix table D.13 for the two restricted samples I discussed above. For each sample I include specifications with and without log earnings, similar to columns (1) and (4) in table 2. Results show that relative to table 2, the effects of university and secondary education on the odds of out-migration remain as strong, if anything they rise slightly, whereas the effect of higher pre-move earnings falls only marginally. All other coefficients on labor market variables remain significant and similar in magnitude. The only effects that experience a drop in their coefficients are occupational skills: in both subsamples, working in occupations demanding very-high skills no longer has a positive and significant effect on the decision to migrate. In online appendix table D.14 I estimate an analogous version of table 3 for the two restricted samples. Again, results confirm that pre-labor market sorting is not driving the strong positive selection observed for migrants who move from small to big cities.

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 in the city of origin) and on the extent to which the individual had benefitted from migration. Results in table 4 showed that eventual returnees are less skilled and exhibit lower earnings than permanent migrants in their city of first employment. We turn now to examine the experiences of returnees in their city of destination.

In table 6 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

[31] Unfortunately, the municipality of birth or registration with social security is not provided in the MCVL.

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 age categories and labor market characteristics in the city of first employment by computing mean values while working in the first city (e.g. percent of time spent unemployed or self-employed, percent of time working in the public sector or under a temporary). Given the smaller sample size, instead of including indicator variables for urban areas interacted with five-year periods, I add urban area indicators for both the first city and the city of destination. I cluster standard errors at the city of first location, but all results are robust to clustering them at the city of destination or 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 many determinants of this second migration.

Results indicate, once again, that selection is mainly driven by migrants who initially moved to big cities. In column (3a) a 10% increase in earnings at the first location (i.e., prior to the first migration episode) makes return migration 0.6% less likely.[32] 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 1.1% 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. For instance, having university education reduces the probability of returning by 27%. However, even within education categories and occupational skills, higher realized earnings in big cities make return migration less likely.

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 in column (2a) at all returnees who initially moved to any city (not necessarily one of the six biggest), realized earnings in destination do not influence their return decision after controlling for observable skills. In addition, when I examine in column (4b) the migration decision for repeat migrants who did not return but moved on to a third city, this is not negatively affected by realized earnings in the big city. 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.

The intensity and degree of selection in return migration from big cities may depend on the age of individuals at the time of their first move. On the one hand, individuals who move to big cities at later ages may be more aware of their skills, how they compare to those of their peers and how to put them into use in big cities. They may also be better informed about labor market opportunities in big cities. These individuals should experience less uncertainty and, thus, are less likely to return to small cities. On the other hand, when individuals move to big cities at

[32] $10\% \times (0.837 - 1)/e = -0.6\%$.

Table 6: 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 $^{\text{2nd loc.}}$	0.874 (0.072)*	1.435 (0.049)***	0.943 (0.068)	1.392 (0.046)***	0.691 (0.033)***	1.143 (0.073)**	0.756 (0.035)***	1.143 (0.081)*
Log mean earnings $^{\text{1st loc.}}$	0.905 (0.051)*	1.024 (0.068)	0.956 (0.047)	0.995 (0.070)	0.837 (0.037)***	1.064 (0.090)	0.907 (0.046)*	1.064 (0.094)
University education			0.887 (0.084)	1.318 (0.081)***			0.726 (0.045)***	1.109 (0.104)
Secondary education			0.885 (0.034)***	1.207 (0.072)***			0.897 (0.037)***	1.189 (0.118)*
Very-high-skilled occup			0.792 (0.051)***	1.061 (0.120)			0.845 (0.101)	1.313 (0.257)
High-skilled occup			0.895 (0.062)	1.084 (0.110)			0.906 (0.104)	1.240 (0.214)
Medium-high-skilled occup			1.042 (0.044)	1.231 (0.120)**			1.042 (0.087)	1.430 (0.235)**
Medium-low-skilled occup			1.010 (0.042)	1.152 (0.109)			0.970 (0.074)	1.405 (0.222)**
Unemployed $^{\text{2nd location}}$	1.908 (0.123)***	2.229 (0.146)***	1.874 (0.150)***	2.638 (0.298)***	1.580 (0.148)***	2.222 (0.217)***	1.529 (0.187)***	3.064 (0.560)***
Unemployed $^{\text{1st location}}$	2.537 (0.384)***	0.744 (0.080)***	2.385 (0.395)***	0.809 (0.090)*	1.889 (0.247)***	0.778 (0.188)	1.723 (0.237)***	0.815 (0.203)
Unemployed $^{\text{exp.ben. 2nd loc.}}$	8.035 (0.651)***	8.632 (0.798)***	8.036 (0.650)***	8.648 (0.796)***	11.547 (2.085)***	8.597 (1.579)***	11.533 (2.082)***	8.629 (1.580)***
Experience	1.047 (0.004)***	0.954 (0.007)***	1.038 (0.003)***	0.961 (0.007)***	1.049 (0.006)***	0.972 (0.008)***	1.034 (0.006)***	0.974 (0.009)***
Firm tenure	1.060 (0.009)***	1.029 (0.013)**	1.063 (0.009)***	1.025 (0.012)**	1.083 (0.011)***	1.030 (0.012)**	1.087 (0.011)***	1.028 (0.013)**
Urban area indicators $^{\text{1st loc.}}$	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Urban area indicators $^{\text{2nd loc.}}$	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Labor market controls $^{\text{2nd loc.}}$	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Labor market controls $^{\text{1st loc.}}$	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Years since migration indic.	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	2,083,911		2,083,911		833,552		833,552	
Migrants	43,534		43,534		15,611		15,611	
Pseudo R^2	0.071		0.071		0.071		0.072	

Notes: Relative risk ratios (exponentiated coefficients) reported with standard errors in parentheses clustered at the urban area of first location. ***, **, and * indicate significance at the 1, 5, and 10 percent levels. The reference category is permanent (first-time) migrants who remain in the city of destination. In columns (1a)–(2b) the sample is made up of migrants after their first move. In columns (3a)–(4b) the sample is made up of migrants after their first move to one of the six biggest cities. All specifications include a male indicator variable and month indicator variables. Labor market controls in first location are mean values while working in the first city 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. The omitted categories are less than secondary education and low-skilled occupations.

Table 7: Incidence of second migration by age at fist migration

	First move to any city below median-sized city			First move to any of 6 biggest cities		
	Obs.	Share of returnees	Share of mover-ons	Obs.	Share of returnees	Share of mover-ons
Age at first migration						
18–23	857	34.3%	16.2%	3,055	31.5%	15.8%
24–26	893	33.0%	17.4%	3,389	32.9%	17.1%
27–29	719	27.7%	17.2%	2,748	33.6%	16.7%
30–34	687	31.1%	14.3%	2,243	35.0%	15.0%
35–52	353	29.7%	10.2%	1,040	41.3%	10.8%

Notes: Sample includes first-time migrants that remain in the MCVL at least for five years after their first move.

younger ages, they may benefit more from the valuable experience they accumulate there. Then, individuals who move to big cities at later ages should be less capable of enhancing their skills and, thus, are more likely to return to small cities.[33] Alternatively, as they have spent more time in their city of first employment they may have developed stronger connections, which may increase the odds of return migration from big cities.

In table 7, I present the incidence of return migration for individuals who moved to big and small cities at different ages. I restrict the sample to those migrants who remain in the MCVL at least for five years after their first move to avoid concerns of right truncation in the data. The raw data reveal that the share of first-time migrants who ended up returning from big cities increases with the age of first migration. Individuals who moved to big cities at ages between 18 and 23 are 10 percentage points less likely to return than those individuals who moved after age 35. There is no clear relationship between age at first migration and the share of returnees when looking at first-time migrants that moved to small cities.

Table 8 presents analogous versions of columns (4a) and (4b) in table 6 for the samples of first-time migrants who moved to big cities at ages below and above 30. Results show that the odds of return migration are higher for young first-time migrants who had a negative experience in the big city. For instance, in column (1a) a 10% decrease in earnings raises the odds of returning by 1.1% while being unemployed in the big city increases the probability of returning by 63%. In contrast, in column (2a) the probability of return migration for older first-time migrants is not significantly affected neither by realized earnings in the big city nor by periods of unemployment.

Altogether, the findings in tables 7 and 8 suggest that older first-time migrants to big cities are more rooted in their city of first employment and, hence, more prone to return. Yet, in contrast

[33]See De la Roca, Ottaviano, and Puga (2014) for a dynamic model of urban sorting between big and small cities for individuals with varying levels of ability and self-confidence. Using longitudinal individual data, Baum-Snow and Pavan (2012) and De la Roca and Puga (2017) provide evidence of more valuable experience in large cities. Further, De la Roca and Puga (2017) show that earnings profiles are steeper for individuals who move to big cities early in life than for those who start their careers in small cities and move to big cities some years after.

Table 8: Multinomial logit estimation of determinants of second migration by age at first migration

	First move to any of 6 biggest cities			
	Age at first migration < 30		Age at first migration $\geqslant 30$	
	Return	Move on	Return	Move on
	(1a)	(1b)	(2a)	(2b)
Log mean earnings $^{\text{2nd loc.}}$	0.713	1.098	0.888	1.300
	(0.030)***	(0.091)	(0.094)	(0.230)
Log mean earnings $^{\text{1st loc.}}$	0.904	0.991	0.777	1.295
	(0.063)	(0.099)	(0.055)***	(0.247)
University education	0.710	1.259	0.729	0.642
	(0.042)***	(0.120)**	(0.070)***	(0.127)**
Secondary education	0.858	1.352	0.934	0.667
	(0.043)***	(0.148)***	(0.051)	(0.113)**
Very-high-skilled occupation	0.831	1.146	0.798	2.132
	(0.138)	(0.287)	(0.151)	(1.025)
High-skilled occupation	0.895	1.193	0.873	1.710
	(0.119)	(0.253)	(0.179)	(0.821)
Medium-high-skilled occupation	1.047	1.283	0.983	2.320
	(0.118)	(0.251)	(0.154)	(0.943)**
Medium-low-skilled occupation	1.000	1.383	0.852	1.525
	(0.104)	(0.249)*	(0.113)	(0.622)
Unemployed $^{\text{2nd location}}$	1.628	2.937	1.297	3.442
	(0.237)***	(0.610)***	(0.357)	(1.805)**
Observations	621,971		210,818	
Migrants	10,931		4,680	
Pseudo R^2	0.080		0.066	

Notes: Relative risk ratios (exponentiated coefficients) reported with standard errors in parentheses clustered at the urban area of first location. ***, **, and * indicate significance at the 1, 5, and 10 percent levels. The reference category is permanent (first-time) migrants who remain in the city of destination. Sample is made up of migrants after their first move to one of the six biggest cities. All specifications include the same controls as in table 6. Log mean earnings in second location are 6-month moving averages, excluding current earnings. The omitted categories are less than secondary education and low-skilled occupations.

to return moves of younger first-time migrants, return moves of older first-time migrants are not influenced by low realized earnings or unemployment events, hinting to the possibility that older migrants face less uncertainty upon moving to big cities.

6. Conclusions

This study 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, Bauernschuster, Falck, Heblich, Suedekum, and Lameli, 2014). 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 positive selection is largely driven by the group of migrants who moves from small to big cities. The effects of higher educational attainment, 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. In fact, these variables do not have a statistically significant effect on the probability of migrating to small cities. Regarding the role of observables relative to unobservables, I find that the major 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, differences in pre-move earnings after conditioning on observable skills fall by around two thirds. These findings suggest 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 30% of migrants move for a second time within five years of arriving in their city of destination and 67% of these moves involve a return migration. Return moves are also more frequent in big cities. I find that eventual 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 from big cities to other cities, they are not affected by low realized earnings after their first migration episode. Furthermore, when I analyze return decisions by migrants who move to big cities at different ages, I find that low realized earnings and unemployment episodes increase the odds of returning especially for those who moved to big cities at younger ages. This finding suggests that older migrants face less uncertainty than younger migrants upon moving to big cities.

All of this indicates that positive sorting in big cities through migration is important. However, differences in observable characteristics account for much of the observed differences in skill sorting that results from mobility of workers. At the same time, for workers who have already migrated to big cities, further sorting is driven not only by workers' initial skills and productivity but also by improvements in that productivity, perhaps as a result of the large set of learning advantages and opportunities that big cities provide (De la Roca, Ottaviano, and Puga, 2014). While I document in detail how initial and return migration contributes to the sorting of more skilled workers into big cities, it is worth noting that worker sorting can occur through other channels, some of which appear to have meaningful effects. For instance, both faster learning associated with working in big cities (Baum-Snow and Pavan, 2012, De la Roca and Puga, 2017) and better schools in large urban areas can widen the skill gap.

References

Ahlin, Lina, Martin Andersson, and Per Thulin. 2016. Human capital sorting: The 'when' and 'who' of sorting of talents to urban regions. Processed, Lund University.

Ambrosini, J. William, Karin Mayr, Giovanni Peri, and Drados Radu. 2015. The selection of migrants and returnees: Evidence from Romania and implications. *Economics of Transition* 23(4):753–793.

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. *The Review of Economics and Statistics* 60(1):47–57.

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

Bauernschuster, Stefan, Oliver Falck, Stephan Heblich, Jens Suedekum, and Alfred Lameli. 2014. Why are educated and risk-loving persons more mobile across regions? *Journal of Economic Behavior & Organization* 98:56–69.

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

Baum-Snow, Nathaniel and Ronni Pavan. 2013. Inequality and city size. *Review of Economics and Statistics* 95(5):1535–1548.

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.

Chen, Yong and Stuart S. Rosenthal. 2008. Local amenities and life-cycle migration: Do people move for jobs or fun? *Journal of Urban Economics* 64(3):519–537.

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

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

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

Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon. 2016. The costs of agglomeration: House and 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.

Combes, Pierre-Philippe and Laurent Gobillon. 2015. The empirics of agglomeration economies. In Gilles Duranton, Vernon Henderson, and William Strange (eds.) *Handbook of Regional and Urban Economics*, volume 5. Amsterdam: North-Holland, 247–348.

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. 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. 2016. A spatial knowledge economy. Processed, Chicago Booth, University of Chicago.

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, Gianmarco I.P. Ottaviano, and Diego Puga. 2014. City of dreams. Processed, CEMFI.

De la Roca, Jorge and Diego Puga. 2017. Learning by working in big cities. *Review of Economic Studies* 84(1):106–142.

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.

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.

Harmonised European Time Use Survey. 2007. HETUS. [online database version 2.0], Statistics Finland and Statistics Sweden. https://www.h5.scb.se/tus/tus/default.htm.

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.

Ministerio de Educación, Cultura y Deporte. 2013. *Datos Básicos del Sistema Universitario Español. Curso 2013/2014*. Madrid, Spain: Ministerio de Educación, Cultura y Deporte.

Molloy, Raven, Christopher L. Smith, and Abigail Wozniak. 2011. Internal migration in the United States. *Journal of Economic Perspectives* 25(3):173–196.

Moretti, Enrico. 2012. *The New Geography of Jobs*. New York, NY: Houghton Mifflin Harcourt.

Rappaport, Jordan. 2009. The increasing importance of quality of life. *Journal of Economic Geography* 9(6):779–804.

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.

Sinha, Paramita and Maureen L. Cropper. 2013. The value of climate amenities: Evidence from US migration decisions. Working Paper 18756, National Bureau of Economic Research.

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 2004–2013, uncensored earnings data are available from matched income tax returns for all workers in the MCVL except for some 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 1980–2013, earnings data are available from the social security for all workers in the MCVL, including all self-employed workers and those in the Basque Country and Navarra, but these are capped for some observations. In particular, 11.1% and 9.6% of monthly earnings observations are top- and bottom-coded between 1980 and 2013. However, since I have earnings data after 2004, only 6.8% and 3.1% of all observations are top- and bottom-coded for the period 1980–2003. This appendix explains how I estimate earnings for this 9.9% of observations that are capped. Further extensions are provided in De la Roca (2014).

I construct estimates of earnings using Tobit models with individual, job and location characteristics, as well as variables that use the panel dimension of the MCVL. For these estimations of capped earnings I enlarge the sample to include all individuals who were born between 1916 and 1995, with ages ranging from 18 to 65 throughout 1980–2013. Except for this age restriction, I follow the same criteria used to construct the final sample in the study. 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 earnings that exceed (are below) daily censoring bounds are flagged as top-coded (bottom-coded).

Next, I estimate 2,720 Tobit regressions by groups of age, occupation and year (8 age groups $\times$ 10 occupations $\times$ 34 years).[34] The dependent variable is log daily earnings expressed in December 2013 euros and as explanatory variables I include age and sets of indicator variables for gender, level of education, temporary contract, part-time contract 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 earnings imputation for German social security data. More specifically, I include the worker's mean of log daily earnings over her career (excluding current earnings) and both fractions of top and bottom censored earnings observations over her career (again excluding the current censoring status). Furthermore, I treat cities as they treat firms in their analysis, so instead of including the annual mean of earnings in the firm and firm size as regressors, I include the annual mean of earnings in the city and the measure of city size described

[34]The eight age groups are the following: 18–24, 25–29, 30–34, 35–39, 40–44, 45–49, 50–54 and 55–65.

Table A.9: Order of autocorrelations for actual and simulated earnings

Order	Actual	Simulated
1	0.910	0.921
2	0.877	0.892
3	0.848	0.865
4	0.823	0.842
5	0.805	0.823
6	0.791	0.807
7	0.777	0.794
8	0.768	0.781
9	0.757	0.771

in section 4. Lastly, I add as a control the worker's mean of log daily earnings from tax returns and an indicator variable that takes value one in the years that these data are available.

Using the coefficients of these Tobit regressions, I can simulate 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 earnings, the fraction of censored earnings observations in all other periods, city size and mean earnings in the city, $\hat{\gamma}$ 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.[35]

Fit of simulated earnings

To verify the accuracy of estimated earnings, I compare simulated values of earnings in top- and bottom-coded observations in social security records relative to actual uncensored earnings in income tax returns for the same individual and month in those years where both are available (2004–2013). If the fit is satisfactory for 2004–2013, I can be more confident that simulated earnings do a good job in approximating capped earnings for 1980–2003.

The correlation between simulated and actual values for capped month-individual observations in 2004–2013 is very high at 0.93. Table A.9 shows that the estimated order of annual autocorrelations for actual and simulated earnings looks very similar with simulated earnings exhibiting slightly higher levels of autocorrelation. Simulated earnings also reproduce well the shape of the earnings distribution. This can be seen in table A.10, which presents selected percentiles of the distributions of actual and simulated earnings for all workers and for skilled workers. Overall,

[35] 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 earnings 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 earnings level at which bottom censoring occurs, then I define $\varepsilon = \Phi^{-1}[k_a \times u]$.

Table A.10: Selected percentiles for actual and simulated earnings

	All workers		Skilled workers	
	Actual	Simulated	Actual	Simulated
Percentile 1	33.6	32.8	29.6	29.0
Percentile 5	43.4	42.9	42.1	41.5
Percentile 10	49.7	49.2	50.2	49.6
Percentile 25	61.8	60.6	66.4	66.1
Percentile 50	82.4	81.4	86.3	84.5
Percentile 75	119.4	118.8	118.4	122.0
Percentile 90	168.0	172.0	165.0	173.3
Percentile 95	212.3	227.4	202.2	206.3
Percentile 99	336.1	338.3	307.4	278.1

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

the distributions are quite similar. Even for skilled workers, who are top-coded beyond the 70th percentile, simulated earnings approximate actual earnings quite well in capped percentiles.

Appendix B. Short-term and short-distance migration

Following the estimations proposed in table 2, where I examined selection of long-term long-distance migrants in terms of skills and productive characteristics at the time of first migration, I now repeat these estimations including short-term and short-distance migrants. In table B.11, 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. Because of the large computation requirements for a multinomial logit with multiple urban area-period fixed-effects, I sample 50% of observations in non-migration episodes.

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 exhibit similar pre-move earnings as stayers (column 1a). Therefore, once I take into account that short-term/distance migrants have accumulated less labor market experience and are in more unstable jobs (e.g., more likely to be under a temporary contract and unemployed), they are no longer less productive than stayers in their first city, as they appeared to be in the raw data. However, they do tend to be special on some dimensions. For instance, occupational skills and education now work in opposite directions (column 2a), suggesting that over-educated workers in low-skilled 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.

Table B.11: 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	1.041 (0.044)	1.606 (0.100)***	1.073 (0.029)***	1.306 (0.057)***
University education			1.252 (0.082)***	2.053 (0.303)***
Secondary education			1.022 (0.038)	1.433 (0.083)***
Very-high-skilled occupation			0.654 (0.032)***	1.179 (0.044)***
High-skilled occupation			0.677 (0.025)***	1.037 (0.046)
Medium-high-skilled occupation			0.843 (0.018)***	1.199 (0.039)***
Medium-low-skilled occupation			0.888 (0.016)***	1.034 (0.029)
Male	1.250 (0.035)***	1.098 (0.018)***	1.248 (0.032)***	1.196 (0.018)***
Experience	0.954 (0.003)***	0.892 (0.006)***	0.956 (0.002)***	0.919 (0.008)***
Firm tenure	0.877 (0.008)***	0.928 (0.009)***	0.878 (0.008)***	0.920 (0.011)***
Self-employed	0.561 (0.047)***	0.609 (0.037)***	0.490 (0.040)***	0.618 (0.039)***
Public sector employee	0.980 (0.066)	0.839 (0.050)***	1.030 (0.068)	0.772 (0.040)***
Temporary contract	2.004 (0.046)***	1.386 (0.027)***	1.987 (0.042)***	1.416 (0.027)***
Part-time contract	1.244 (0.030)***	1.345 (0.069)***	1.249 (0.026)***	1.244 (0.053)***
Receiving unemployment benefits	3.149 (0.084)***	2.443 (0.094)***	2.759 (0.079)***	2.767 (0.097)***
Receiving unemployment subsidy	1.685 (0.062)***	0.981 (0.044)	1.516 (0.051)***	1.211 (0.064)***
Expired unemployment benefits	8.904 (0.253)***	7.986 (0.225)***	8.941 (0.253)***	7.985 (0.226)***
Expired unemployment subsidy	18.239 (1.086)***	19.637 (1.407)***	18.177 (1.095)***	19.291 (1.408)***
Urban area × period indicators	Yes	Yes	Yes	Yes
Age indicators	Yes	Yes	Yes	Yes
Observations	19,804,146		19,804,146	
Pseudo R^2	0.105		0.107	

Notes: Relative risk ratios (exponentiated coefficients) are reported on a sample of 442,033 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. The reference category is stayers. 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 city of destination and distance of 120 kms. All specifications include month indicator variables. Period is a five-year interval. Log mean earnings are 6-month moving averages, excluding current earnings. Less than secondary education and low-skilled occupations are the omitted categories.

Appendix C. Migration to small cities

In table C.12 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 consider only moves that involve a drop in city size. Other than this, all specifications are identical to those in tables 2 and 3.

Results indicate there is no clear evidence of selection of any type for migrants who move to small cities. Observable skills in column (1) and average pre-migration earnings in column (2) do not influence the odds of moving to a small city. Unexpectedly, being in the lowest tercile of the local earnings distribution in column (3) decreases the probability of migrating to a small city by 10%. Based on the conceptual framework in section 2, it might be that individuals with low earnings in non-small cities cannot afford the moving cost. When I include earnings and observable skills in column (4), point estimates show that individuals with university education have a higher probability of moving to small cities, yet working in medium-high skilled occupations reduces the probability by approximately 20%. These findings hint that university-educated individuals who are mismatched in their city of first employment (i.e., they tend to work in low-skilled occupations despite their diplomas) are more likely to migrate to small cities. Overall, the absence of clear selection in observable skills and pre-migration earnings is in sharp contrast with the strong positive selection of migrants who move to big cities.

In tables 3 and C.12 I only exploit moving decisions for some long-term long-distance migrants. In particular, in table 3 I use only those migrants who move to one of the six biggest cities and experience an increase in city size, while in table C.12 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 out-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 3.2% (odds ratio of 1.878 and standard error of 0.105) while for migrants who move to smaller cities the effect is much lower at 1.5% (odds ratio of 1.409 and standard error of 0.090). 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 attenuate relative to the differences presented 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, define better the environment of the conceptual framework.

Table C.12: 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		1.223 (0.202)		1.140 (0.134)
Richest earnings tercile			1.106 (0.149)	
Poorest earnings tercile			0.896 (0.049)**	
University education	1.408 (0.311)			1.372 (0.281)
Secondary education	0.967 (0.116)			0.956 (0.107)
Very-high-skilled occupation	1.078 (0.153)			0.997 (0.094)
High-skilled occupation	0.990 (0.072)			0.934 (0.056)
Medium-high-skilled occupation	0.848 (0.095)			0.827 (0.077)**
Medium-low-skilled occupation	1.051 (0.054)			1.038 (0.053)
Male	1.162 (0.031)***	1.111 (0.049)**	1.116 (0.045)***	1.138 (0.033)***
Experience	0.941 (0.009)***	0.928 (0.014)***	0.929 (0.013)***	0.939 (0.009)***
Firm tenure	0.917 (0.019)***	0.916 (0.018)***	0.917 (0.018)***	0.916 (0.019)***
Self-employed	0.687 (0.069)***	0.761 (0.092)**	0.772 (0.092)**	0.720 (0.089)***
Public sector employee	1.310 (0.184)*	1.311 (0.196)*	1.336 (0.204)*	1.278 (0.171)*
Temporary contract	1.663 (0.068)***	1.693 (0.071)***	1.684 (0.071)***	1.672 (0.067)***
Part-time contract	1.008 (0.058)	1.109 (0.128)	1.072 (0.090)	1.079 (0.105)
Receiving unemployment benefits	2.923 (0.245)***	2.949 (0.177)***	2.922 (0.175)***	2.908 (0.237)***
Receiving unemployment subsidy	1.594 (0.176)***	1.593 (0.161)***	1.577 (0.156)***	1.608 (0.182)***
Expired unemployment benefits	9.402 (0.703)***	9.421 (0.699)***	9.420 (0.697)***	9.439 (0.699)***
Expired unemployment subsidy	14.896 (1.992)***	14.966 (1.977)***	14.982 (1.977)***	14.918 (1.991)***
Urban area × period indicators	Yes	Yes	Yes	Yes
Age indicators	Yes	Yes	Yes	Yes
Pseudo R^2	0.071	0.070	0.070	0.071

Notes: Odd ratios (exponentiated coefficients) are reported on a sample of 37,258,745 monthly observations and 323,472 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 city of destination and distance of 120 kms. 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 five-year interval. Log mean earnings are 6-month moving averages, excluding current earnings. Earnings terciles are constructed for all years. Less than secondary education and low-skilled occupations are the omitted categories.

Appendix D. Online appendix

Table D.13: Logit estimation of determinants of first migration for restricted samples

	Dep. variable: long-term long-distance migration			
	Province of first location is the same as province of birth		Province of first location is the same as province of social security registration	
	(1)	(2)	(3)	(4)
Log mean earnings		1.240 (0.038)***		1.268 (0.034)***
University education	2.390 (0.282)***	2.295 (0.281)***	2.333 (0.309)***	2.229 (0.307)***
Secondary education	1.587 (0.081)***	1.559 (0.082)***	1.561 (0.089)***	1.530 (0.090)***
Very-high-skilled occupation	0.931 (0.056)	0.823 (0.050)***	1.086 (0.047)*	0.946 (0.041)
High-skilled occupation	0.902 (0.032)***	0.822 (0.024)***	0.966 (0.033)	0.870 (0.026)***
Medium-high-skilled occupation	1.172 (0.046)***	1.127 (0.044)***	1.186 (0.051)***	1.135 (0.049)***
Medium-low-skilled occupation	1.002 (0.027)	0.982 (0.026)	1.019 (0.031)	0.996 (0.030)
Male	1.267 (0.028)***	1.224 (0.028)***	1.225 (0.026)***	1.181 (0.026)***
Experience	0.980 (0.005)***	0.976 (0.004)***	0.958 (0.005)***	0.953 (0.005)***
Firm tenure	0.894 (0.006)***	0.892 (0.006)***	0.907 (0.008)***	0.906 (0.008)***
Self-employed	0.623 (0.037)***	0.676 (0.039)***	0.566 (0.035)***	0.619 (0.035)***
Public sector employee	0.702 (0.053)***	0.672 (0.050)***	0.751 (0.045)***	0.717 (0.041)***
Temporary contract	1.510 (0.028)***	1.525 (0.031)***	1.443 (0.025)***	1.459 (0.025)***
Part-time contract	1.265 (0.041)***	1.417 (0.050)***	1.210 (0.039)***	1.372 (0.047)***
Urban area × period indicators	Yes	Yes	No	No
Age indicators	Yes	Yes	Yes	Yes
Unemployment indicators	Yes	Yes	Yes	Yes
Observations	39,267,390	39,267,390	39,328,307	39,328,307
Pseudo R^2	0.081	0.081	0.076	0.077

Notes: Odd ratios (exponentiated coefficients) are reported on a sample of 417,887 individuals in columns (1) and (2) and 418,545 individuals in columns (3) and (4). 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 city of destination and distance of 120 kms. All specifications include month indicator variables. Period is a five-year interval. Log mean earnings are 6-month moving averages, excluding current earnings. Less than secondary education and low-skilled occupations are the omitted categories.

Table D.14: Logit estimation of determinants of first migration to big cities for restricted samples

	Dep. variable: long-term long-distance migration to any of 6 biggest cities			
	Province of first location is the same as province of birth		Province of first location is the same as province of social security registration	
	(1)	(2)	(3)	(4)
Log mean earnings		1.265 (0.087)***		1.295 (0.085)***
University education	3.997 (0.206)***	3.846 (0.215)***	4.150 (0.218)***	3.977 (0.221)***
Secondary education	2.142 (0.073)***	2.105 (0.068)***	2.136 (0.075)***	2.095 (0.072)***
Very-high-skilled occupation	1.111 (0.098)	0.981 (0.075)	1.192 (0.099)**	1.038 (0.072)
High-skilled occupation	0.958 (0.082)	0.868 (0.056)**	0.985 (0.082)	0.883 (0.057)*
Medium-high-skilled occupation	1.441 (0.088)***	1.383 (0.080)***	1.455 (0.087)***	1.390 (0.079)***
Medium-low-skilled occupation	1.085 (0.053)*	1.059 (0.051)	1.078 (0.056)	1.050 (0.054)
Male	1.418 (0.032)***	1.369 (0.034)***	1.372 (0.033)***	1.321 (0.033)***
Experience	0.975 (0.007)***	0.971 (0.006)***	0.963 (0.006)***	0.959 (0.006)***
Firm tenure	0.891 (0.007)***	0.889 (0.007)***	0.897 (0.006)***	0.895 (0.006)***
Self-employed	0.755 (0.073)***	0.824 (0.075)**	0.684 (0.055)***	0.751 (0.056)***
Public sector employee	0.459 (0.064)***	0.435 (0.062)***	0.480 (0.059)***	0.453 (0.057)***
Temporary contract	1.466 (0.046)***	1.479 (0.047)***	1.423 (0.042)***	1.436 (0.043)***
Part-time contract	1.311 (0.061)***	1.488 (0.062)***	1.292 (0.058)***	1.483 (0.057)***
Urban area $\times$ period indicators	Yes	Yes	No	No
Age indicators	Yes	Yes	Yes	Yes
Unemployment indicators	Yes	Yes	Yes	Yes
Observations	29,167,787	29,167,787	29,353,297	29,353,297
Pseudo R^2	0.091	0.092	0.091	0.092

Notes: Odd ratios (exponentiated coefficients) are reported on a sample of 337,879 individuals in columns (1) and (2) and 339,679 individuals in columns (3) and (4). 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 city of destination and distance of 120 kms. All specifications include month indicator variables. Period is a five-year interval. Log mean earnings are 6-month moving averages, excluding current earnings. Less than secondary education and low-skilled occupations are the omitted categories.