Discussion Paper Series

IZA DP No. 18526

April 2026

# Where You Arrive Matters: Local Conditions and Migration Duration. Evidence from Italian Registry Data

**Lisa Capretti**
CEIS, University of Rome "Tor Vergata"

**Francesca Centofanti**
DEF, University of Rome "Tor Vergata"

**Alessio Farcomeni**
DEF & CEIS, University of Rome "Tor Vergata"

**Furio Camillo Rosati**
DEF & CEIS, University of Rome "Tor Vergata"
and IZA@LISER



IZA@LISER NETWORK
ISSN: 2365-9793

IZA DP No. 18526 April 2026

# Where You Arrive Matters: Local Conditions and Migration Duration. Evidence from Italian Registry Data


## Abstract

This paper examines the determinants of temporary migration and return decisions among immigrants in Italy, drawing on a novel administrative dataset covering 3.7 million foreign-born individuals between 2011 and 2022. By reconstructing individual migration histories, we estimate migration duration using parametric survival models, quantile regressions for interval-censored data, competing risk models, and a split cure model that separates the determinants of permanent settlement from those shaping the timing of exit, explicitly accounting for the large share of migrants who remain in Italy. Our results show that out-migration is concentrated in the first five years after arrival, while the majority of migrants remain in Italy over the 12-year observation window. Individual characteristics such as age and gender matter, but local conditions within Italy strongly shape migration duration. Higher local incomes are associated with longer stays, while higher rental prices accelerate departures. Regional disparities also matter independently of economic variables: migrants in the South and Islands remain significantly longer than those in the North. These findings highlight how heterogeneity within host countries - rather than national averages alone – shapes migration trajectories, offering novel insights into the geography of return migration and the importance of local labor markets and living conditions.



## JEL classification

F22, J61, C41

## Keywords

migration dynamics, temporary migration, regional disparities, survival analysis



## Corresponding author

Furio Camillo Rosati
f.rosati@economia.uniroma2.it

## 1. Introduction

A significant share of international migrants does not intend to settle permanently in their destination country, but eventually return home or move elsewhere. Understanding the determinants and timing of these return and onward migrations is central to evaluating the long-run effects of immigration on labor markets, public finance, and migrant integration. Despite the importance of the topic and notwithstanding well-developed theoretical analyses, empirical evidence on return and onward migration remains limited, largely because of data constraints. Most existing studies rely on small-scale surveys, retrospective information, or data for specific subgroups, while large-scale administrative evidence is rare.

This paper addresses these gaps by providing the first large-scale study of migration duration and return in Italy, based on administrative registry data covering 3.7 million foreign-born individuals between 2011 and 2022. By reconstructing complete migration histories, we document patterns of temporary and permanent settlement and analyze how individual characteristics, origin-country conditions, and local contexts within Italy shape these outcomes. Italy offers an especially relevant case: it has experienced sustained immigration in recent decades, displays marked regional heterogeneity in labor markets and living costs, and plays a central role in European debates on migration.

Our contribution is fourfold. First, we provide the first large-scale evidence on migration duration and return in Italy, showing that while out-migration is concentrated in the first two years after arrival, most migrants remain in the long term. We also analyze the heterogeneity of responses across different nationalities. Second, we highlight the importance of within-country heterogeneity, showing that local wages, housing costs, and regional disparities (beside relative income, cost of living and political conditions) have independent effects on migration duration. While recent work has emphasized the importance of immigrants' location within destination countries for economic outcomes (Abramitzky et al., 2021), most of the literature on migration duration and return decisions has overlooked how local conditions within the host country affect the timing of migrants' departures. The empirical studies on return migration still treat the host country as spatially homogeneous. As a result, little is known about how subnational variation in wages, housing costs, and regional conditions shapes migrants' decisions to remain, return, or move onward. We demonstrate that migrants' outcomes are significantly affected by where within Italy they reside. Higher local wages prolong stays, while higher rents accelerate departure, and migrants in the South and Islands remain longer than those in the North, even after controlling for economic variables. This highlights the importance of community integration, regional amenities, and institutional differences as determinants of return. Third, we distinguish between return to the country of origin and onward migration to a third country, providing new insights into the geography of migrant trajectories. Fourth, make use of a set of complementary duration models—including parametric survival analysis, quantile regressions for interval-censored data, competing risks models, and a split cure model that accounts for the large fraction of migrants who never leave during the observation period.

Our results indicate that migration in Italy is relatively persistent: around 80% of the 2011 arrival cohort remained after 12 years. Out-migration is more likely among men and older entrants, while women and younger migrants tend to stay longer.

The remainder of the paper is structured as follows. Section 2 reviews the related literature. Before turning to the data, we first situate our analysis within the broader literature. This section reviews the main theoretical and empirical contributions on return and onward migration, highlighting what is known about economic incentives, life-cycle factors, and local conditions, and identifying the gaps that our study seeks to address. Section 3 describes the data and construction of migration histories. Section 4 outlines the conceptual framework and econometric approach. Section 5 presents the results. Section 6 concludes.

## 2. Literature review

Return migration has long been recognized as an important dimension of international mobility, reflecting both planned strategies and responses to shocks. Unlike permanent settlement, return and onward migration often involve sequential decisions taken under uncertainty, shaped by relative economic conditions, life-cycle considerations, and family circumstances.

Understanding these dynamics is crucial not only for assessing the long-run impacts of immigration on labor markets, fiscal systems, and integration processes, but also for evaluating migrants' own welfare and life-cycle trajectories.

Theoretical models generally frame return as an intertemporal optimization problem in which migrants compare expected lifetime utility at home and abroad (Dustmann, 2003; Dustmann & Görlach, 2016). Wage differentials and cost-of-living disparities are key drivers, but non-monetary factors such as amenities, safety, and family networks also matter. Return can thus form part of a target-saving strategy or follow adverse shocks such as unemployment or political instability (Ward, 2017). Anticipated return affects behavior in the host country, influencing investment in host-specific human capital, remittances, and labor supply (Dustmann, 1996; Dustmann & Weiss, 2007).

More recently, Adda et al. (2022) show how strategically timed returns can accelerate human capital accumulation and wage assimilation. Preferences and priorities also evolve with age and family circumstances, leading to shifting valuations of host- versus home-country opportunities (Cobb-Clark & Stillman, 2013).

Empirical research on return migration has long been limited by data availability. Most studies rely on small-scale surveys or group-specific samples. Early work linked return to unemployment, occupational downgrading, or family structure (Vanderkamp, 1971; Ramos, 1992; Schmidt, 1994). More recent contributions emphasize life-cycle considerations (Yahirun, 2014), the role of return in wage dynamics (Adda et al., 2022), and reintegration at origin (Martin & Radu, 2012). Registry-based studies have begun to provide more robust evidence: Bijwaard (2010, 2014, 2023), using Dutch population registers, shows that weak labor market attachment and unemployment spells increase return probabilities.

Economic conditions in host and home countries are strong predictors of return. Dustmann (2003) identifies a non-linear relationship between host-country wages and migration duration: migrants remain longer when wages are low (to accumulate savings) but return earlier when higher wages allow them to meet financial goals. Evidence from West Africa (De Vreyer et al., 2010) shows that accumulated savings and higher purchasing power at origin facilitate earlier return and generate a wage premium upon reintegration. Non-economic factors also matter: origin-country political instability tends to reduce return (Schmidt, 1994), though insecurity may sometimes reinforce intentions to remain abroad (Tezcan, 2019).

More recently, research has expanded to onward or circular migration (Ward, 2017; OECD, 2017, 2020), highlighting that departures from the host country do not necessarily imply return to origin. This also complicates the measurement of return and underscores the need to account for multiple exit pathways.

A smaller but growing body of work examines spatial heterogeneity within host countries. Albert and Monras (2022) show that immigrants cluster in high-productivity, high-price areas as they consume part of their lifetime income in the home country. This connects with a wider literature on local labor market and regional effects of immigration (Dustmann and Preston, 2019), which shows how migrants' outcomes are shaped by subnational disparities in wages, housing, amenities, and institutional quality. Yet the return-migration literature has largely treated destination countries as homogeneous, overlooking how subnational disparities may shape return and onward migration.

Despite these advances, several important gaps remain. Large-scale registry-based studies are limited to the Netherlands, and systematic evidence for USA and Southern Europe is lacking. In Italy, existing research relies mainly on surveys or aggregate statistics (OECD, 2020; INAPP, 2023).

Moreover, the role of local conditions within host countries—wages, rental costs, urban vs. rural contexts, and regional disparities—has received little attention in the return-migration literature, despite descriptive evidence of their importance in Italy (Chiurco et al., 2023) and elsewhere (Saiz, 2007; Ottaviano and Peri, 2007; Sá, 2015).

Our study contributes to filling these gaps by (i) exploiting administrative registry data for the entire immigrant population in Italy; (ii) analyzing how subnational conditions shape duration and out-migration; (iii) distinguishing between return and onward migration; and (iv) applying a set of complementary duration models—including a split cure approach—tailored to the censoring challenges posed by the administrative data.

## 3. The data

Our analysis draws on administrative records of all registrations and cancellations from the municipal registry office, covering both voluntary declarations by individuals and cancellations initiated by the administration.

The construction of this dataset is an innovative contribution, as detailed individual-level migration histories are rarely accessible due to privacy constraints. Focusing on the foreign-born population adds further complexity: registration and cancellation procedures for foreigners are less standardized than for Italians, often fragmented across municipalities, and sometimes subject to delays.

Each registration or cancellation record includes the following information:

- The date of registration or cancellation,
- The country of origin and the municipality of arrival (for registrations from abroad),
- The municipality of origin and destination (for internal movements),
- The country of destination and the municipality of departure (for voluntary cancellations),
- The individual's date of birth, sex, country/municipality of birth, and country of citizenship.

A key issue is that many migrants leave Italy without cancelling their registration. In such cases, municipalities eventually apply *administrative cancellations* based on standard checks, third-party information, or other sources. These cancellations provide the same information as voluntary cancellations, but only ex post, meaning they give an upper bound for the length of stay and do not contain information on the country of destination.

As already stated, this study focuses solely on data related to foreign-born individuals, resulting in a dataset of approximately 8.6 million observations (see Table 1).

### 3.1. Data Construction and Tracking Process

The data were accessed via the Laboratory for the Analysis of Elementary Data (ADELE), made available by ISTAT under strict data-protection protocols. For each record we observe: sex, date of birth, country of birth and citizenship[1], year and month of registration ($t$[2]), and the municipality/country of origin ($i$) and destination ($j$). Each observation is assigned a unique identifier $\mathrm{ID}_{t,i,j}$ based on this information.

We reconstruct migration histories by linking sequential records. Starting with an observation at time $t$ (January 2011–December 2022), we search for a subsequent registration involving a move from $j$ to $k$. If a unique match exists ($\mathrm{ID}_{t+1,j,k}$), the movement is attributed to the same individual. Ambiguous cases with multiple matches (i.e., more than one record showing a move from $j$ to $k$ at $t+1$ *for* $\mathrm{ID}_{t,i,j}$), are excluded. This process is repeated iteratively for each time point $t$ and movement $m$ from January 2011 to December 2022. From the initial 8.6 million observations, about 1.8 million are dropped because they cannot be unambiguously attributed to a single individual. A further 2 million observations relating solely to internal movements within Italy are excluded, since our focus is on international migration. We also drop individuals with more than three movements to reduce the risk of misclassification. The resulting dataset includes 4.8 million observations corresponding to 4.4 million individuals. Restricting to those aged 15–69 at entry yields a final sample of 3.7 million individuals. Comparisons with official ISTAT aggregates show that our dataset closely matches national statistics for immigration and emigration, with slightly fewer emigrations due to the exclusion of ambiguous records[3]. Table 1 summarizes the construction of the dataset.

ISTAT registry data are the most reliable and validated official source on foreign residents in Italy. They are standardized across municipalities, regularly checked, and widely used in demographic and migration statistics. However, some important limitations must be acknowledged. Irregular migrants are largely absent, as they cannot register. Similarly, refugees, asylum seekers, and temporary-visa holders are excluded, since they are not legally obliged to register nor the municipalities to accept their registration. Moreover, as the data only include registered movements, some events like short term circular migration or temporary absence might not be observed. Our analysis therefore pertains exclusively to the movements of the regular foreign-born population recorded in municipal registers. While this excludes some groups, the reliance on registry data ensures comparability across municipalities and over time.

[1] In the empirical analysis migrants are identified on the base of the country of birth rather than citizenship. Therefore immigrants that are naturalized remain in our sample.
[2] Deaths are excluded from our data.
[3] In Appendix A, we provide a detailed comparison between our dataset and data obtained from both ISTAT's ADELE laboratory (microdata) and the ISTAT website, serving as a robustness check for the underlying assumptions of our dataset construction. In Appendix B, we show balance tests with the original sample provided by ADELE.

**Table 1. Data set description**

| | Whole sample | Average by year |
|---|---|---|
| Number of cases (Natives and Foreign born) | 24,118,131 | 2,009,844 |
| Number of cases (Foreign born) | 8,610,884 | 717,573.7 |
| Non-unique observations | 1,797,031 | |
| Observations relative to Internal movements only | 1,981,008 | 165,084 |
| Observations in Final sample[4] | 4,832,845 | 406,630.5 |
| Number of individuals[5] | 4,450,256 | 407,286.3 |
| Observations in Final sample (15-69 at time of entry) | 4,044,700 | 338,947 |
| Number of individuals (15-69 at time of entry) | 3,697,693 | 339,085.1 |

## 3.2. Descriptive statistics

### 3.2.1. Temporary migration 2011 - 2022

Table 2 provides basic statistics and offers initial insights into the scale of temporary migration in Italy among individuals aged 15–69 at the time of entry. Over the 12 years covered by the dataset, departures from Italy amounted to approximately half of the arrival flows, with yearly fluctuations partly influenced by administrative cancellations. The total number of out-migrants over this period corresponds to about 33% of the total stock of migrants in 2011.

There is substantial gender parity in the inflow of migrants, but males are more likely to leave Italy. On average, the age at arrival is 33 years, while the age at departure is 41 years.

Both the age of entry and exit, as well as the gender composition, vary significantly across nationalities, as illustrated in Tables 3 for the five largest nationalities present in Italy. Females predominate among arrivals from Romania and Ukraine, as migrants from these countries often specialize in domestic services. The average age at entry does not differ substantially across nationalities, although Romanian and Ukrainian migrants tend to be slightly older at entry. A gender

[4] This refers to those observations that were eliminated as they were duplicates in all the variable available (29,976 obs), or multiple identical ID leaving the same municipality to different places (559,819+24,599= 584,418 obs), IDs that had more than 2 duplicates after the matching procedure - and therefore, more prone to mismatches - (159,937 obs), internal movements (all, not internal only) (2,698,291 obs), or errors related to the combination age of entry/country of birth (children that leave Italy few months/years after birth, result to be born abroad and we do not have info on the arrival in Italy) (60,805 obs), or cancellations for unavailability (these are biased as in 2013 in Italy there was a peak of cancellations due to a verification of those who had not responded to the 2011 Census) (244,612 obs).

[5] This number include people of all ages. Then, from this we will keep only people that move to Italy in the 15-69 age range at entry.

difference is also observed among out-migrants, with males being relatively more likely to leave, except in the case of Romanians and Ukrainians.

The national group with the highest average age at departure is Ukrainians, while the average age of departure is relatively similar across the other groups, at around 42 years. Figures C1 to C4 in Appendix C show the flow of arrivals and departures by province and municipality.

**Table 2. Arrivals in and Departures from Italy**

| Year | Arrivals | Departures | Difference | Stock of Immigrants | Ratio departures to arrivals |
|---|---|---|---|---|---|
| 2011 | 270,182 | 33,079 | 237,103 | 4,100,365 | 0.12 |
| 2012 | 255,516 | 90,069 | 165,447 | 4,318,205 | 0.35 |
| 2013 | 223,379 | 115,348 | 108,031 | 4,609,580 | 0.52 |
| 2014 | 200,887 | 136,231 | 64,656 | 4,786,738 | 0.68 |
| 2015 | 205,003 | 152,412 | 52,591 | 4,834,799 | 0.74 |
| 2016 | 217,963 | 102,190 | 115,773 | 4,830,163 | 0.47 |
| 2017 | 249,729 | 111,608 | 138,121 | 4,817,822 | 0.45 |
| 2018 | 237,616 | 113,662 | 123,954 | 4,883,023 | 0.48 |
| 2019 | 224,153 | 147,565 | 76,588 | 4,995,575 | 0.66 |
| 2020 | 150,139 | 127,441 | 22,698 | 5,039,099 | 0.85 |
| 2021 | 197,998 | 133,999 | 63,999 | 5,171,463 | 0.68 |
| 2022 | 248,728 | 96,320 | 152,408 | 5,030,095 | 0.39 |
| **Total** | **2,681,293** | **1,359,924** | **1,321,369** | | **0.50719** |
| Average age | 33.388 | 41.286 | | | |
| Share of male | .515 | .559 | | | |

**Table 3. Descriptive Statistics – Arrivals and Departures for the 5 largest nationalities**

| | Arrivals | | Departures | |
|---|---|---|---|---|
| | Obs | Mean | Obs | Mean |
| **Country of birth: Romania** | | | | |
| Age at entry | 443,291 | 36.308 | 257,004 | 42.62 |
| Males | 443,291 | .381 | 257,004 | .447 |
| **Country of birth: Albania** | | | | |
| Age at entry | 163,512 | 33.531 | 51,374 | 41.342 |
| Males | 163,512 | .462 | 51,374 | .607 |
| **Country of birth: Morocco** | | | | |
| Age at entry | 138,226 | 32.391 | 96,604 | 42.3 |
| Males | 138,226 | .491 | 96,604 | .638 |
| **Country of birth: China** | | | | |
| Age at entry | 120,908 | 34.161 | 72,224 | 41.239 |
| Males | 120,908 | .479 | 72,224 | .537 |
| **Country of birth: Ukraine** | | | | |
| Age at entry | 86,864 | 37.613 | 32,250 | 48.076 |
| Males | 86,864 | .249 | 32,250 | .233 |

There is also evidence of significant secondary movements, as shown in Table 4. Approximately 36% of out-migrants for which we have information about the country of destination move to a country other than their country of birth (Column 1). This phenomenon is particularly pronounced among Albanians and Moroccans, with around 70% of Moroccan out-migrants relocating to a third country. In contrast, secondary movements are much less frequent (below 15% in all the samples considered, except for Ukrainians) among arrivals (Column 2), supporting the view that Italy is a country of transit for many of the immigrants.

**Table 4. Movements from/to Italy by country of destination/origin**

| | (1) From Italy | | (2) To Italy | |
|---|---|---|---|---|
| | Freq. | Percent | Freq. | Percent |
| **Total Sample** | | | | |
| Country of birth | 384,283 | 63.94 | 2,320,752 | 86.55 |
| Other country | 216,768 | 36.06 | 360,541 | 13.55 |
| Total | 601,051 | 100.00 | 2,681,293 | 100.00 |
| **Country of birth: Romania** | | | | |
| Country of birth | 95,343 | 87.08 | 399,077 | 90.03 |
| Other country | 14,149 | 12.92 | 44,214 | 9.97 |
| Total | 109,492 | 100.00 | 443,291 | 100.00 |
| **Country of birth: Albania** | | | | |
| Country of birth | 13,033 | 58.83 | 159,080 | 97.29 |
| Other country | 9,120 | 41.17 | 4,432 | 2.71 |
| Total | 22,153 | 100.00 | 163,512 | 100.00 |
| **Country of birth: Morocco** | | | | |
| Country of birth | 11,843 | 29.85 | 132,100 | 95.57 |
| Other country | 27,827 | 70.15 | 6,126 | 4.43 |
| Total | 39,670 | 100.00 | 138,226 | 100.00 |
| **Country of birth: China** | | | | |
| Country of birth | 11,188 | 90.78 | 117,946 | 97.55 |
| Other country | 1,136 | 9.22 | 2,962 | 2.45 |
| Total | 12,324 | 100.00 | 120,908 | 100.00 |
| **Country of birth: Ukraine** | | | | |
| Country of birth | 10,908 | 85.47 | 66,808 | 76.91 |
| Other country | 1,855 | 14.53 | 20,056 | 23.09 |
| Total | 12,763 | 100.00 | 86,864 | 100.00 |

# 4. Theoretical framework and the econometric model

## 4.1. Conceptual Framework

We build on the model of return migration proposed by Dustmann and Görlach (2016), extending it to account for heterogeneity in regional conditions within the host country. In Italy—as in many destination countries—labor markets, amenities, and living costs vary substantially across regions. Migrants therefore face not only the decision of whether to migrate and when to return, but also where to settle within the host country.

We assume that a potential migrant chooses between staying in the origin country and migrating to one of several regions within Italy[6]. Once in the destination country, the migrant periodically evaluates whether to remain, return to the origin country, or move elsewhere. At each point in time,

[6] In the empirical analysis we will focus on Italian NUTS1 macro region.

the decision is based on a comparison of expected lifetime utility under each alternative[7]. To keep things as simple as possible, we consider only two regions in the destination country, but the conclusions can be easily extended to the case of more than two regions.

In particular the migrant will make their choice comparing the expected lifetime utility

$$\text{Max}\left\{V_i^d, V_j^d, V^o\right\}$$

where $V_i^d$ and $V_j^d$ denote respectively to the lifetime utility in region *i* and *j* of the country of destination and $V^o$ denotes the lifetime utility from remaining in the country of origin. The form of the V's will be discussed below.

Let $w_i$ and $p_i$ denote the wages and the price level in the region *i* of the country of destination, while *w* is the wage and *p* is the cost of living index in the country of origin; $h_i$ indicates a utility shifter reflecting preference for consumption in the home country with respect to consumption in the region *i* of the country of destination or the difference in the level of amenities as perceived by the migrant. $\mu$ is a discount factor. We assume that within the destination country, wages and living costs are positively correlated: regions offering higher wages also tend to have higher prices. This is a mild and empirically supported assumption and does not require the stronger condition of perfect labor market arbitrage, which would imply equalized real wages across regions.

The value of residing in region *i* of the country of destination at time *t* can then be written as:

$$V_i^d(\Theta_t) = \max\ u^d\ (h, c) +\ \mu\ \mathrm{V}(\Theta_{t+1})$$

$$\text{s. t. } A_{t+1} = (1+r)A_t +\ w_i - c_t$$

similarly, the value of out-migrating or returning to the country of origin is given by:

$$V_i^o(\Theta_t) = \max\ u^o\ (h, c) +\ \mu\ \mathrm{V}(\Theta_{t+1})$$

$$\text{s. t. } \frac{p_i}{p} A_{t+1} = \frac{p_i}{p}(1+r)A_t +\ w - c_t$$

At each period *t*, given the optimal consumption path and current assets $A_t$, the migrant compares the value of staying versus out-migrating. The indifference condition equating the utility of staying and out-migrating is:

$$u^d(h, w_i + (1+r)A_t -\ A_{t+1}^d) + \mu\, V\big(t+1, A_{t+1}^d\big) =\ u^o(h, w + (1+r)A_t \frac{p_i}{p} -\ A_{t+1}^o\ \frac{p_i}{p}) +$$

$$\mu\, V(t+1, A_{t+1}^o)$$

Rearranging we have:

[7] We do not consider internal migration movements between NUTS1 macro region as this are very limited, as we will see.

$$u^o(h, w+(1+r)A_t\frac{p_i}{p} - A^o_{t+1}\frac{p_i}{p}) - u^d(w_i+(1+r)A_t - A^d_{t+1}) = \mu\,(V\left(t+1, A^d_{t+1}\right) - V(t+1, A^o_{t+1}))$$

The left-hand side indicates the increase in utility that can be obtained by leaving the country of destination in period *t*, while the right-hand side shows the benefit due to remaining an additional year.

Several well-established factors affect the decision to out-migrate or to return home. Wage differentials play a central role: the smaller the differential, the more likely a migrant is to out-migrate, as the relative benefit of doing so increases. A stronger preference for amenities in the home country similarly raises the probability of an earlier return. Moreover, greater purchasing power in the origin country, relative to the host country, can also incentivize return migration.

Allowing for regional heterogeneity within the destination country adds some interesting considerations. Residing in a high-income, high-cost area has ambiguous effects relative to living in a low-income, low-cost area. On one hand, higher wages increase the opportunity cost of out-migrating, making out-migration less attractive. On the other hand, higher prices in the destination country increase the relative attractiveness of consumption at home or elsewhere, raising the benefit from out-migrating.

This framework implies that the decision to return is not driven by a single national-level trade-off, but rather by region-specific combinations of economic and non-economic factors. It also suggests that individual characteristics (e.g., age, gender, family structure) interact with these local conditions to shape the timing of return.

Our data do not allow us to identify some of the specific mechanisms that might lie behind the observed effects of the covariate considered like family composition, social integration, labour-market matching, access to public services, or informal networks. Our estimates must, therefore, be interpreted as reduced form that subside the specific effects of the various mechanisms. While reduced-form estimates do not allow for a fully structural interpretation, they remain informative about how local conditions shape migration trajectories.

## 4.2. A duration analysis

Our data present several challenges. First, *administrative cancellations* introduce uncertainty about the exact departure date. These events provide an upper bound for the length of stay, resulting in interval-censored data. Second, some migrants arrived before the start of our observation window, generating left-truncated spells where the true exposure to risk is unknown.

Formally, let $Y_i > 0$, $i = 1, \ldots, n$ denote the response variable of interest across $n$ individuals. For the "administrative cancellations" we do not observe $Y_i$, but only ($Y_i^{(l)}$, $Y_i^{(u)}$, $\Delta_i$). The response is of mixed-censoring type, namely, it can be an observed time ($Y_i^{(l)} = Y_i^{(u)}$, $\Delta_i = 1$); a right-censored time ($Y_i^{(l)} = Y_i^{(u)}$, $\Delta_i = 0$) where we only know that the event occurred after, but certainly not before, $Y_i^{(u)}$; or an interval censored time ($Y_i^{(l)} < Y_i^{(u)}$, $\Delta_i = 1$) where we know that the event occurred in the interval between $Y_i^{(l)}$ and $Y_i^{(u)}$.

The problem of left truncation is more difficult to address as we do not have information on the beginning of exposure to risk (date of arrival in Italy). As is well known, there is no well-established methodology to address such cases (Guo et al., 1993). Therefore, we opted to conduct the analysis by cohorts of entry, excluding observations for which the date of entry is not available.

There are additional computational problems linked to the numerosity of the data (more than a quarter million observation for each cohort), for this reason we do not use time varying variable in the estimates but employ them at the date of entry of the cohort. This choice is also supported by the fact that most of the variation is cross sectional.

Moreover, this also helps mitigate concerns about reverse causality between migrants' presence and local conditions, by considering local economic and spatial characteristics at the time of entry, that are predetermined with respect to the arrival of the migrant.

Finally, migrants may leave Italy either for returning to their country of origin or for moving onward to a third country and we take in to consideration the possible difference in behaviour for these different groups.

Of course, we cannot rule out that migrants arrival or departure may affect local labor and housing markets, particularly in smaller areas, and, therefore, some care must be used in interpreting such estimates as causal.

No single estimation method is fully adequate in this setting. We therefore employ a set of complementary approaches that, taken together, provide a robust assessment of the determinants of temporary migration, return, and settlement.

First, we estimate parametric survival models for interval-censored data. These models provide a structured framework to analyze how covariates influence the survival function, and are computationally tractable even with large datasets. We adopt a log-normal specification, which fits the data best based on information criteria, and adapt the likelihood function to account for interval- and right-censored spells.

Second, we use a split cure model (Lambert et al., 2007; Maller & Zhou, 1997). This relaxes the standard assumption that all individuals are at risk of exit and allows for the possibility that a fraction of migrants remain permanently in Italy. The model decomposes the population into two latent subgroups: permanent stayers (“cured”) and temporary migrants (“susceptible”), whose duration of stay follows a parametric survival distribution. Both the cure probability and the timing of exit are modelled as functions of individual and contextual covariates. This approach, widely used in medical survival analysis, has not yet been applied to migration data and provides a natural way to interpret the persistence observed in Kaplan–Meier curves.

Third, to assess heterogeneity in effects across different parts of the duration distribution, we estimate quantile regressions for interval-censored data. Unlike parametric models, which focus on average effects, quantile regression reveals whether covariates influence short-term exits differently from long-term stays, without relying on proportional hazards or other global assumptions.

Finally, because exits from Italy occur through distinct pathways, we estimate competing risks models (Fine & Gray, 1999) that distinguish between (i) return to the country of origin, (ii) onward migration to a third country, and (iii) exits with unknown destination (*administrative cancellations*). This allows us to link the timing of departure to the type of migration outcome, a distinction that standard survival models cannot capture.

Taken together, these approaches enable us to address censoring, account for heterogeneity in stay lengths, and differentiate exit pathways. In Section 5, we present the results from each model in turn, showing how they complement each other and provide a coherent picture of migration duration in Italy. In that section, we will present the analysis relative to the 2011 cohort, the one for which we can observe the individuals for the longest period. We run, however, estimates also for other cohorts to assess if there are relevant differences across cohorts, and the related estimates are reported in Appendix D.

### 4.2.1 Parametric survival model for temporary migration

We begin by estimating a fully parametric survival model to analyze the duration of stay in Italy. This type of approach provides a structured framework to assess how covariates influence the survival function over time. However, given the complexity introduced by the high proportion of right and interval-censored observations in our data, we have to keep in mind that parametric approaches may not fully capture the heterogeneity of migration duration. We discuss this limitation and how to deal with it below.

As conventionally defined in the standard literature of the field (e.g., Kalbfleisch & Prentice, 2002), let $S(t|X)$ denote the survival function, which gives the probability that the event has not yet occurred by time $t$:

$$S(t|X) = P(T > t|X) \tag{1}$$

Following Lawless (2003), the corresponding hazard function, which represents the instantaneous risk of event occurrence at time $t$, is given by:

$$h(t|X) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t | T \geq t, X)}{\Delta t} \tag{2}$$

The likelihood function for interval-censored data is constructed as follows: for exact observations, the contribution to the likelihood is $f(Y_i|X)$, where $f(\cdot)$ is the density function of the chosen survival distribution. For right-censored observations, the contribution is $S(Y_i|X)$. For interval-censored observations, the contribution is $S\left(Y_i^{(l)}\middle|X\right) - S\left(Y_i^{(u)}\middle|X\right)$, capturing the probability that the event occurs within the given interval.

The model parameters are estimated via maximum likelihood, ensuring consistency and efficiency in estimation. We chose a log-normal distribution, since it ensures the best model fit based on BIC and AIC[8].

Parametric models directly estimate baseline hazards and provide interpretable effects of covariates, but rely on distributional assumptions that may not hold in all subpopulations. We therefore complement this approach with a semiparametric alternative.

### 4.2.2. Split cure model for interval-censored data

A standard assumption of survival models is that all individuals are eventually at risk of experiencing the event—in our case, leaving Italy. This assumption is often violated in migration settings, where a substantial share of individuals remains permanently. In fact, the evidence that will be presented in Section 5.1 suggests that a large fraction of migrants remain in the country over the entire 12-year observation window, consistent with the notion of "permanent stayers." Classical proportional hazards or fully parametric models may therefore overstate the eventual probability of exit.

To address this issue, we estimate a split cure model (Lambert et al., 2007, Maller & Zhou, 1997), which assumes that the population consists of two latent subgroups: a cured group (permanent stayers) who will never experience the event, and a susceptible group who may eventually leave, with the timing modeled through a parametric survival distribution. This approach has been widely developed in the statistical literature (Peng & Taylor, 2014; Latouche et al., 2007; Yu & Peng, 2008) and has recently begun to attract attention in economics, though it has not yet been applied to migration studies.

Formally, the survival function for individual *i* can be written as

$$S_i(t) = \pi_i + (1 - \pi_i)\, S^*(t)$$

[8] We run the parametric regression using the command *stintreg* by StataCorp.

where $\pi_i = \Pr(cured_i = 1 \mid x_i)$ is the probability of permanent stay, modeled via a logit link, and $S^*(t)$ is the survival function for the susceptible group, assumed to follow a log-normal distribution as in Section 4.2.1. Both the cure and latency components are modeled as functions of the same covariates used in the baseline survival model, allowing us to assess whether the determinants of permanent settlement differ from those affecting the timing of exit.

In particular, we observe durations $T_i$ that may be: either exact (event observed at t), interval-censored (event occurs in $(L_i,R_i]$), or right-censored (no event by $R_i$).

A share of the population is assumed to be immune (cured) and will never experience the event. The rest are "susceptible" with a parametric survival distribution. For individual *i*, the probability of being cured is modeled with a logit:

$$\pi_i = \Pr(cured_i = 1 \mid x_i) = \frac{1}{1 + \exp(-x_i \beta_{cure})}$$

Conditional on being susceptible, survival follows a Weibull distribution:

$$S_i(t|uncured) = \exp\left[-\left(\frac{t}{S_i}\right)^{k_i}\right],$$

For each observation the contribution to the likelihood is given by: (i) exact failure at t: $T_i = t_i$ ; (ii) interval censored at (L,R]$T_i \in (L_i, R_i]$; and (iii) right censored at R: $T_i > C_i$.

The overall likelihood is given by the sum $\sum_i l_i$, where $l_i$ is respectively: (i) for exact failure at t:

$$l_i = \ln(1 - \pi_i) + \ln f_i(t \mid uncured) \;;$$

(ii) for interval-censored at (L, R]:

$$l_i = \ln(1 - \pi_i) + \ln(S_i(L \mid uncured) - S_i(R|uncured);$$

(iii) for right-censored at R:

$$l_i = \ln(\pi_i + (1 - \pi_i)\ S_i(R|uncured))$$

Estimation is carried out by maximum likelihood, adapting the likelihood contributions to account for right- and interval-censored spells as shown.

### 4.2.3. Quantile regression for interval censored data for temporary migration

To capture heterogeneity across the distribution of stay lengths, we estimate quantile regressions for interval-censored data. Unlike survival models, which provide average effects, quantile regression identifies whether covariates affect short-term and long-term migrants differently. In our setting, the 10th and 25th percentiles correspond roughly to short-term stays (2–5 years), while the 75th and 95th percentiles represent long-term stays (10–12 years). This mapping allows us to interpret how covariates influence different parts of the duration distribution.

Quantile regression is distribution-free and does not require proportional hazards, making it well suited to our data with high censoring and low event rates. Estimates are obtained through a two-step procedure: first, a nonparametric estimate of the conditional CDF via piecewise exponential models (Friedman, 1982), and second, optimization of the quantile-specific residuals. When the task involves studying the association of duration outcomes with covariates, classical methods involve Cox regression models. As already mentioned, in our case there are computational and technical issues that hinder the possibility of using classical methods. The most crucial ones are linked to the very high proportion of right and interval censored observations, which makes classical methodologies instable (even when two-part models involving a cured fraction is considered). Notably, the assumption of proportionality of hazards is not tenable in the presence of a low event rate. To overcome these issues, we use a quantile regression approach (see e.g. Portnoy (2003), Farcomeni and Geraci (2020) and references therein). This approach has been recently extended to the case of mixed-type censored data (Frumento, 2022). This method is particularly valuable in our context, as both economic and demographic variables may operate differently at various stages of a migrant's stay. For example, the impact of political instability or relative prices may be muted for early exits but more pronounced for long-term residents. Quantile regression thus provides a richer understanding of the migration process and complements the log-normal model by relaxing distributional assumptions and identifying distributional shifts.

Quantile regression involves separately modeling different quantiles $\tau \in (0, 1)$ of the conditional response variable. The method is distribution-free (that is, it does not require any parametric assumption on the actual conditional distribution) and does not involve any assumption of proportionality of hazards. A possible limitation is the fact that a full model is not available: models are quantile-specific.

Our approach proceeds as follows. Let $X_i$ denote a $p$-dimensional column vector of covariates, $i = 1, \ldots, n$. First of all, denote the conditional Cumulative Distribution Function (CDF) as $F(t|X) = \Pr(Y \leq t|X)$.

For each of different quantiles $\tau_j$, $j = 1, . . . , q$, we model

$$F(X_i'\beta_j|X) = \tau_j \tag{3}$$

where $\beta_j$ is a vector of quantile-specific coefficients to be estimated, and by construction the estimating equation leads to predict conditional quantiles at level $\tau_j$.

For completeness, we note that the estimates $\hat{\beta}_j$ are based on a two-step approach. At the first step we obtain a non-parametric estimate of $F(t|X)$ based on piecewise exponential models as in Friedman (1982); and at the second step we optimize residuals involved in (3) after plug in of the estimated conditional CDF.

As customary in quantile regression, $\hat{\beta}_{jh}$ can be interpreted directly as the variation in the conditional quantile of $Y$ for each unit of the $h$-th variable, $h = 1, . . . , p$, when all other predictors are held fixed.

### 4.2.4. Competing Risk model

Out-migration from Italy occurs through distinct pathways: (i) return to the country of origin, (ii) onward migration to a third country, and (iii) exit without a specified destination (i.e., administrative cancellations). To capture these alternatives, we estimate competing risks models using Fine and Gray's (1999) proportional subhazards framework, separately for each nationality group.

This framework is particularly suited to our setting for three reasons. First, it explicitly links the timing of departures with their type, which is essential for distinguishing return from onward migration. Second, it properly accounts for right censoring, which is pervasive in registry data given that a large share of migrants remains in Italy at the end of the observation period. Third, unlike multinomial logit models, it accommodates time-to-event information and can thus trace how risks evolve over time, rather than reducing migration decisions to static outcomes.

The Fine and Gray model estimates the subdistribution hazard of a given event type. In our case, the event of interest is return migration (destination equals origin country). Onward migration and administrative cancellations are treated as competing risks. Grouping administrative cancellations with onward moves is necessary because cancellations do not provide information on the actual destination; this strategy allows us to retain departures in the analysis while avoiding artificial distinctions unsupported by the data.

We construct a categorical event variable with four values: 0 if the individual remains in Italy in 2022 (right-censored); 1 if the destination is unknown (administrative cancellation); 2 if the destination coincides with the origin country (return migration); and 3 if the destination is a third country (onward migration). For administrative cancellations, we take the reported cancellation date as the actual departure date. As explained in Section 3, these cancellations reflect ex post adjustments when migrants leave without formally deregistering. Ideally, one would model these as interval-censored events, with the true departure occurring at some point between the last recorded presence and the cancellation date. However, incorporating interval censoring into a competing risks framework with time-varying covariates is computationally demanding and, given the size of our dataset, infeasible. We therefore adopt the cancellation date as the departure date, which offers a tractable solution while preserving consistency across groups. This assumption may slightly shift the estimated timing of some departures but does not affect their classification as exits, meaning that the relative risks of return versus onward migration remain robust.

The analysis relies on the dataset with time-varying covariates measured at the origin-country level. This choice ensures consistency across nationality groups and reduces the risk of collinearity, thereby improving the robustness of the estimates.

### 4.3. Covariates included in the analysis

Our analysis includes three broad sets of covariates in line with the conceptual framework: individual characteristics, conditions in Italy with respect to the country of origin, and local conditions within Italy.

At the individual level, the administrative data provide information on sex and age at entry, while educational attainment is unfortunately not recorded for foreign-born individuals in our dataset. We therefore use these variables to control for individual characteristics.

To account for the relative conditions in Italy with respect to the country of origin we use several variables. In particular, we use the (log) difference between our proxies for return to work in Italy and in the home country using the GDP per capita adjusted for purchasing power parity (PPP), drawn from the World Development Indicators. While GDP is not an ideal substitute for wages, it offers broad country coverage and correlates well with wage data in those countries where the latter are available[9].

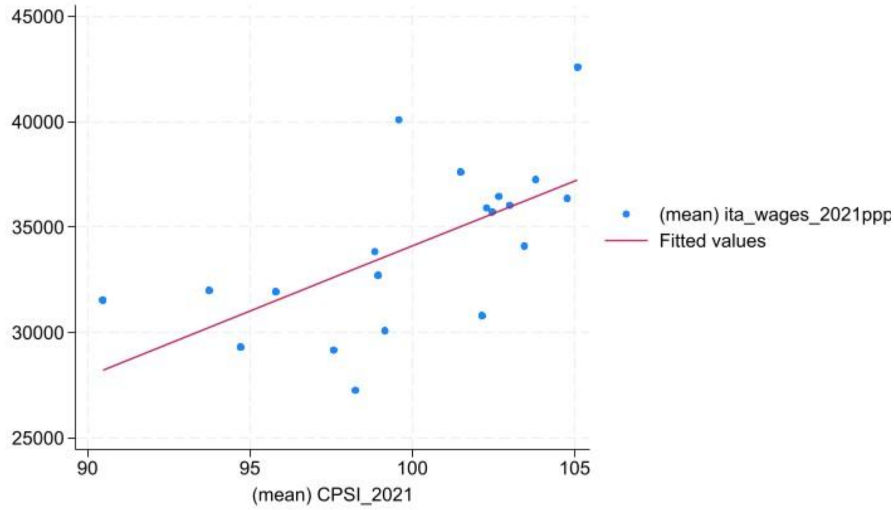
The relative cost of living between Italy and the origin country is captured through a purchasing power parity index, calculated as the ratio between Italy's and the origin country's PPP conversion factors. This variable reflects the incentive to return when purchasing power in the origin country is relatively high. In addition, we include a measure of political instability in the country of origin, using

[9] Complementary regressions show that for the subset of countries for which we have good wage data the correlation between wages and GDP per capita is high.

the Global Peace Index produced by the Institute for Economics and Peace. This index, which ranges from 1 (most peaceful) to 5 (most violent), reflects not only the presence of armed conflict but also broader societal safety and perceptions of violence, which can significantly affect migrants' willingness or ability to return.

An important contribution of our study is the incorporation of detailed subnational information on conditions within Italy. To capture local labor market opportunities, we use income per worker obtained from the tax returns published by the Ministry of Finance. at the Local Labor Market (LLM) level. We preferred to use the income tax returns rather than the hourly gross wages, as the former are available with a much larger geographical granularity (tax returns are available at municipality level and can be aggregated at LLM level, while the hourly wages are only available at provincial level). In any case the correlation between the two is very high at the level of province (about 0.9). We proxy cost of living with the average rental prices, calculated using data from the *Osservatorio Immobiliare* of the Italian Revenue Agency reflecting housing affordability. The Agency reports minimum and maximum rental prices per square meter at the municipality level; we compute the midpoint and aggregate these values to the LLM level to maintain consistency with other spatial variables. Differences in cost of living and wages may reflect the significant disparities among Italy's major socio-economic regions (NUTS1). In fact, as illustrated in Figure 1, there is a strong positive correlation between the cost of living (measured by the Consumer Price Spatial Index, CPSI) and wages across Italian regions.[10]

**Figure 1. Income per worker in Italy and Consumer Price Spatial Index (CPSI) in 2021 – Regions**




*Note: CPSI for the Italian regions based on Italy=100*

[10] Data on the difference on cost of living across the country are available from ISTAT on an experimental base only at the level of Region (NUTS2) and for a limited number of years. In Figure 1 we show the correlation using variable values observed in 2021.

However, we think that these two variables alone do not fully capture the regional differences. The cost of rentals is only a part, albeit relevant, of the cost of living. For these reasons, we include a dummy variable identifying urban areas, where living costs, job density, and mobility patterns tend to differ from rural settings. Broader regional effects are captured through a set of indicator variables for the four major socio-economic areas of Italy, defined at the NUTS1 level: North, Center, South, and Islands. These regional dummies proxy for institutional differences, labor market characteristics, and non-economic factors such as discrimination, community cohesion, and service provision. For example, recent survey evidence (Chiurco et al., 2023) suggests that migrants in Southern regions are relatively less educated than one residing in other areas, they are more likely to work in agriculture, they show a bigger sense of belonging to the local communities and feel less discriminated, despite weaker economic indicators.

To verify robustness, we estimated alternative specifications. These included replacing GDP with wage data for a subset of countries, estimating models with and without socio-economic region dummies, distinguishing internal movements across municipalities or regions and adding the share of foreign workers employed in agriculture by region. In all cases, the main results remained unchanged. Table 5 reports some descriptive statistics for the main variables used in the survival analysis.

**Table 5. Descriptive statistics of the main covariates used in the analysis**

| | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|
| Males | .455 | .498 | 0 | 1 |
| Age | 33.641 | 11.527 | 15 | 69 |
| Average rental price (€/MQ monthly) | 4.66 | 1.73 | 0 | 19.69 |
| Relative cost of living | .396 | .462 | 0 | 4.419 |
| Urban | .843 | .363 | 0 | 1 |
| Average income per worker in Italy (2021 PPP $) | 36,860.05 | 6,306.3 | 14,889.6 | 47,466.35 |
| Global Peace Index (Political instability) | 2.026 | .381 | 1.084 | 3.361 |
| GDP per capita (2021 PPP $) | 17,602.27 | 12,010.4 | 951.189 | 145,591 |

*Note: Average rental price and income per worker in Italy are at the LLM level*

## 5. Results

In this section we present the results of the empirical analysis, focusing on the 2011 arrival cohort to allow for a sufficiently long follow-up period. We proceed in stages, following the sequence of models introduced in Section 4.2. We begin with descriptive evidence, followed by Kaplan–Meier estimates, which provide a non-parametric benchmark of survival patterns. We then report estimates from parametric survival models, before turning to the split cure model, quantile regressions, and competing risks analysis. We further explored heterogeneity across migrant groups by interacting local variables with migrants' characteristics (gender and age). The results do not change substantially and are available on request.

## 5.1. Duration of stay: descriptive statistics for the 2011 cohort

We begin with descriptive evidence from the 2011 cohort, the earliest in our dataset and the one observed for the longest period. Over the 12-year observation period, approximately 20% of individuals of this cohort left Italy (see Table 6). Most departures occurred in the early years: after five years, 14% of them had departed. Out-migration differs markedly by nationality. Ukrainians and Albanians have the lowest exit rates (11–12% over 12 years), whereas Chinese and Romanians are much more likely to leave (30% and 22%, respectively).

Gender differences also emerge: males are consistently more likely to leave than females. By age, differences are small, though younger migrants tend to remain slightly longer. Regional disparities are more pronounced (last panel of Table 6): migrants entering in Northern Italy are more likely to leave (22% over 12 years) than those entering in the South or Islands (14%).

**Table 6. Foreign born: 2011 cohort departures and arrivals**

| | After 1 year (%) | After 5 years (%) | After 7 years (%) | By 2022 (%) | Cumulated departures by 2022 | Arrivals |
|---|---|---|---|---|---|---|
| Total | 1 | 14 | 15 | 20 | 53,626 | 270,182 |
| **By nationality** | | | | | | |
| Romania | 2 | 14 | 16 | 22 | 14,659 | 66,701 |
| Albania | 0 | 8 | 9 | 12 | 1,573 | 13,282 |
| Morocco | 1 | 12 | 13 | 17 | 2,664 | 15,625 |
| China | 1 | 25 | 26 | 30 | 4,260 | 14,138 |
| Ukraine | 1 | 9 | 9 | 11 | 1,681 | 14,823 |
| **By age category** | | | | | | |
| Age 15-25 | 1 | 13 | 14 | 19 | 14,004 | 72,738 |
| Age 26-50 | 1 | 14 | 15 | 19 | 32,970 | 169,241 |
| Age 51-69 | 2 | 16 | 18 | 23 | 6,652 | 28,424 |
| **By sex** | | | | | | |
| Female | 1 | 12 | 13 | 18 | 26,384 | 147,488 |
| Male | 1 | 16 | 17 | 22 | 27,242 | 122,915 |
| **By major socio-economic region of arrival** | | | | | | |
| North | 2 | 16 | 17 | 22 | 32,160 | 148,181 |
| Center | 1 | 13 | 14 | 21 | 13,783 | 67,223 |
| South | 0.8 | 9 | 10 | 14 | 5,676 | 40,004 |
| Islands | 0.6 | 8 | 9 | 14 | 2,007 | 14,774 |

Table 7 displays the share of arrivals by major socio-economic region (NUTS1) in 2011 (Panel A) and from 2011 to 2022 (Panel B) for the main five nationalities considered. Shares are calculated with respect to the total number of migrants from each nationality. North-west is the main region of arrival for all the nationalities, particularly for Moroccans (38.5%), Chinese (31.1%) and Ukrainians (31.8%). Albanians are distributed among North-west (35.7%), Center (26.9%) and North-east (24.6%). Romanians' main destination in Italy is Central Italy (27.8%), North-west (25.2%) and Southern Italy (18.5%). The Ukrainians stand out for their strong presence also in the South (30.3%), unlike other nationalities. The Islands host a very low share of immigrants from all the nationalities considered, and in particular the Albanians and Ukrainians. In Appendix C, we report the shares by Italian region (NUTS-2), that are very similar to the ones in Table 7. In the same Appendix we also show the share of arrival by major socio-economic region by sex and age.

**Table 7. Share of arrivals by major socio-economic region (NUTS1)**

| | Total | Albanian | Romanian | Moroccan | Chinese | Ukrainian | Other nat. |
|---|---|---|---|---|---|---|---|
| **Panel A. Arrivals in 2011** | | | | | | | |
| North-east | 0.22 | 0.25 | 0.20 | 0.26 | 0.27 | 0.19 | 0.22 |
| North-west | 0.33 | 0.36 | 0.25 | 0.39 | 0.31 | 0.32 | 0.36 |
| Center | 0.25 | 0.27 | 0.28 | 0.14 | 0.28 | 0.18 | 0.25 |
| South | 0.15 | 0.11 | 0.19 | 0.17 | 0.10 | 0.30 | 0.12 |
| Islands | 0.06 | 0.02 | 0.08 | 0.05 | 0.03 | 0.02 | 0.05 |
| **Panel B. Arrivals from 2011 to 2022** | | | | | | | |
| North-east | 0.22 | 0.26 | 0.23 | 0.26 | 0.23 | 0.23 | 0.21 |
| North-west | 0.31 | 0.33 | 0.25 | 0.37 | 0.30 | 0.31 | 0.32 |
| Center | 0.24 | 0.25 | 0.26 | 0.14 | 0.34 | 0.19 | 0.24 |
| South | 0.17 | 0.13 | 0.18 | 0.18 | 0.11 | 0.25 | 0.17 |
| Islands | 0.07 | 0.04 | 0.08 | 0.05 | 0.02 | 0.03 | 0.07 |

As a complement to the descriptive evidence in Table 7, we estimate a multinomial logit for the initial macro-area of settlement. The specification includes nationality fixed effects and a nationality-specific responsiveness to co-national networks. Co-national networks are measured using pre-period migrant stocks (2002), mitigating simultaneity concerns. All continuous covariates are standardized (z-scores). The model is intended as a descriptive mapping of "who settles where" upon arrival, controlling for individual and local conditions. Full coefficients are provided in Appendix C (Table C7). Moreover, in the same Appendix (Table C8) we also report two complementary summaries (Panels A and B).

Table 8 shows the number and the related percentage of migrants moving from the municipality within the same region (at LAU[11] level), between municipalities in different regions (at NUTS2 level) or between major socio-economic regions of arrival (at NUTS1 level) throughout the period considered. In the total cohort of arrivals in 2011, 25.8% of individuals moves from a municipality to another in the same region, while only 7.6% and 4.5% moves towards a different region or major socio-economic region, respectively. Moroccans are the group of immigrants that moves most frequently between municipalities, while the Chinese are those who move more often between regions and major socio-economic region.

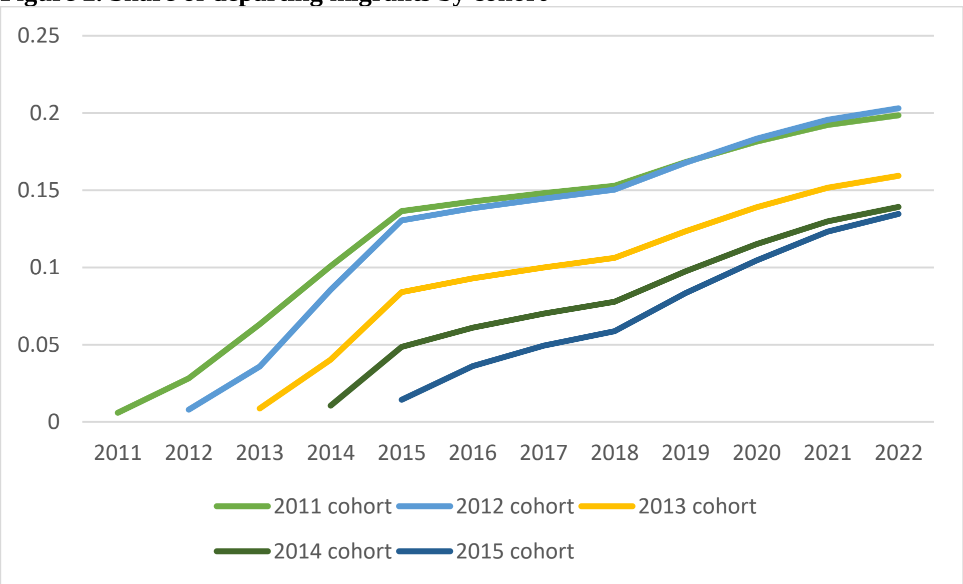
[11] The Nomenclature of territorial units for statistics by the European Commission defines municipalities as Local Administrative Units (LAUs).

**Table 8. Internal movements between municipalities**

| | No. | % |
|---|---|---|
| **Total** | | |
| In the same region – LAU | 69,684 | 25.79 |
| In different regions – NUTS2 | 20,505 | 7.59 |
| In major socioeconomic regions – NUTS1 | 12,257 | 4.54 |
| **Romanian** | | |
| In the same region – LAU | 16,960 | 25.43 |
| In different regions – NUTS2 | 4,240 | 6.36 |
| In major socioeconomic regions – NUTS1 | 2,551 | 3.82 |
| **Albanian** | | |
| In the same region – LAU | 3,789 | 28.53 |
| In different regions – NUTS2 | 1,188 | 8.94 |
| In major socioeconomic regions – NUTS1 | 729 | 5.49 |
| **Moroccan** | | |
| In the same region – LAU | 5,320 | 34.04 |
| In different regions – NUTS2 | 1,202 | 7.69 |
| In major socioeconomic regions – NUTS1 | 718 | 4.59 |
| **Chinese** | | |
| In the same region – LAU | 4,648 | 32.87 |
| In different regions – NUTS2 | 2,486 | 17.58 |
| In major socioeconomic regions – NUTS1 | 1,250 | 8.84 |
| **Ukrainian** | | |
| In the same region – LAU | 3,931 | 26.52 |
| In different regions – NUTS2 | 900 | 6.07 |
| In major socioeconomic regions – NUTS1 | 622 | 4.20 |

Figure 2 plots the cumulative out-migration rate for a subset of the observed cohorts of immigrants.

**Figure 2. Share of departing migrants by cohort**

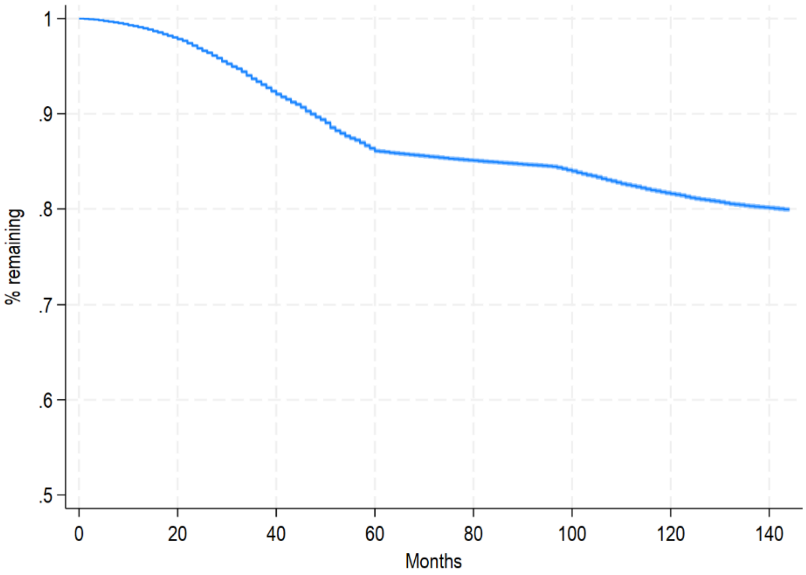



As it is easy to see, there does not appear to be relevant differences in terms of out-migration behavior across the different cohorts.

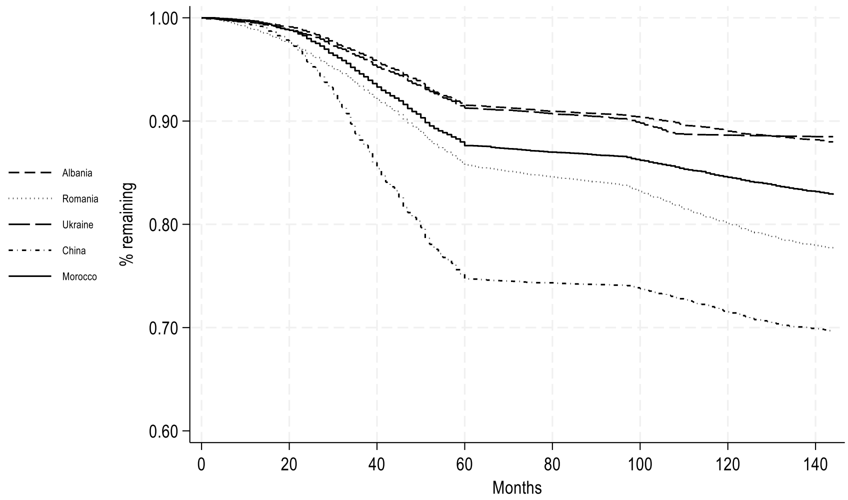
An understanding of migration duration can be initially gained through the Kaplan-Meier survival function. The Kaplan-Meier product-limit estimator presents a non-parametric estimate of the marginal survival curve, showing the proportion of individuals who remain in Italy over the study period and providing insight into out-migration probabilities based on migration duration.

Figure 3 displays the Kaplan-Meier survival function for the whole sample of individual entered in Italy in 2011 and for the five main countries of origin, respectively. In particular, Figure 3 (Panel A) shows that around 80% of the individuals remain in the country at the end of our period of analysis. The largest number of out-migrations seems to happen during the first 20 months after the arrival, even if the majority of individuals remains in Italy at the end of the period. Figure 3 (Panel B) shows how out-migration behavior is very similar for individuals born in Albania and Ukraine. Across all five nationalities, the most significant reduction in the percentage of individuals remaining in Italy occurs within the first 5 years. After this period, the decline becomes more gradual, albeit at varying rates depending on the nationality group. The largest out-migration rate is observed among individuals from China, with approximately 75% of the initial group still residing in Italy after 60 months, dropping to less than 70% by 144 months. Panel C shows the Kaplan- Meier by major socio-economic region (NUTS1). Individuals who arrived in Northern Italy are those most likely to leave the country: after 60 months, 16% had left, compared to just 8% of those who arrived in the Islands over the same period.

**Figure 3. Kaplan-Meier survival function**
**Panel A – Whole sample**




**Panel B - Five largest nationalities**




**Panel C. by major socio-economic region (NUTS1) of arrival**

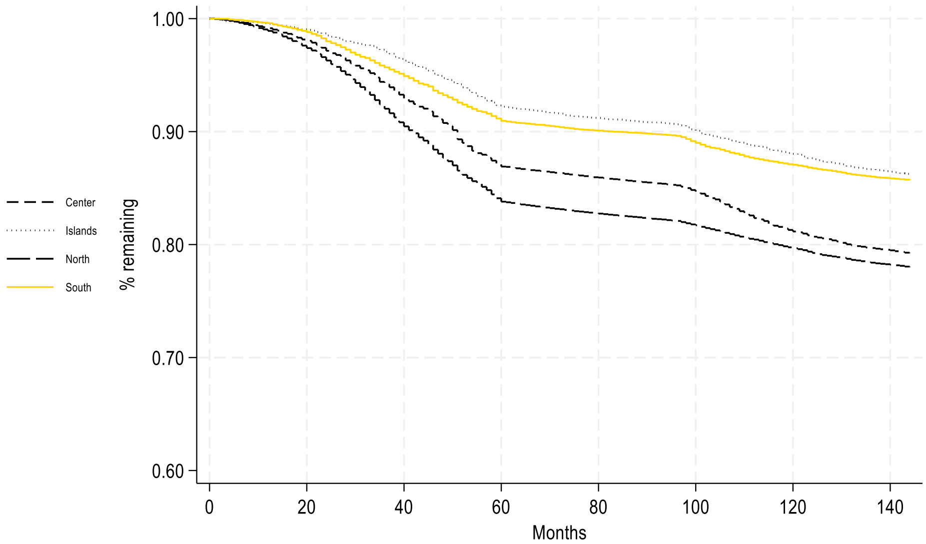


Table 9 shows the number of out-migrants, the time at risk and the incidence rates for out-migration of the total sample considered and by country of origin. Time at risk indicates the sum of months that individuals stay in the sample while at risk of the event, namely, leaving Italy. Consistently with figure 3 (Panel B), the incidence rate, i.e. the probability of out-migration, is higher for Chinese immigrants. The ones for Romanian and Moroccan immigrants follows at 0.00182 and 0.00137, respectively.

**Table 9. Incidence rates of out-migration**

| | Total | Romania | Albania | Morocco | China | Ukraine |
|---|---|---|---|---|---|---|
| #observations | 270,182 | 66,701 | 13,282 | 15,628 | 14,141 | 14,823 |
| #out-migrants | 53,626 | 14,659 | 1,573 | 2,664 | 4,260 | 1,681 |
| Time at risk | 32,879,124 | 8,047,215 | 1,708,598 | 1,943,465 | 1,560,170 | 1,909,097 |
| Incidence rate | .001631 | .001822 | .000921 | .001371 | .002731 | .000881 |

*Notes: Incidence rate is the probability of out-migrating, namely the number of out-migrants divided by the time at risk.*

## 5.2. Out-migration

### 5.2.1. Parametric survival model for interval-censored data

Table 10 reports estimates from a log-normal parametric survival model, using the 2011 cohort to ensure maximum follow-up duration. Several individual and contextual factors emerge as significant predictors of migration duration.

First, individual characteristics matter. Male migrants are more likely to leave earlier than females, possibly reflecting greater labor market mobility or transnational job ties, consistent with IOM evidence (Bauloz et al., 2024). Age at entry is negatively associated with duration: each additional year of age reduces the expected stay length by about 1.6%, suggesting that younger migrants are more likely to integrate or remain longer in the host country.

Economic conditions at origin and destination play an important role in determining the duration of stay. A higher relative cost of living in Italy, compared to the country of origin, significantly reduces migration duration. A 10% increase in the relative cost of living index reduces expected stay length by about 1%. In contrast, a higher income differential in favor of Italy increases expected duration: a 10% increase in the log income ratio extends the median stay by 3.5%.

Political instability in the country of origin, proxied by the Global Peace Index, is not statistically significant in the baseline model. However, as shown below, its effect varies substantially across the distribution of migration durations.

Local economic and social conditions within Italy strongly affect out-migration. Migrants in areas with higher average income levels tend to stay longer, while those in high-rent areas exit earlier.

**Table 10. Estimation results – Parametric survival analysis**

| | Total sample |
|---|---|
| Female | 0.852*** |
| | (0.0229) |
| Age of entry | -0.0160*** |
| | (0.000957) |
| Political Instability | -0.0525 |
| | (0.0391) |
| Relative costs of living | -0.244*** |
| | (0.0271) |
| Log differential between Italian and origin country's GDP | 0.358*** |
| | (0.0214) |
| Income per worker in Italian LLMs (log) | 2.476*** |
| | (0.0814) |
| Average rent in Italian LLMs | -0.0236*** |
| | (0.00740) |
| Urban | -0.0972*** |
| | (0.0319) |
| Major socio-economic region (NUTS1) of arrival: Center | 0.272*** |
| | (0.0282) |
| Major socio-economic region (NUTS1) of arrival: Islands | 2.012*** |
| | (0.0620) |
| Major socio-economic region (NUTS1) of arrival: South | 1.914*** |
| | (0.0423) |
| ***Observations*** | ***254,418*** |

*Notes: Standard errors in parentheses, * p<0.10 ** p<0.05 *** p<0.01*

These findings are consistent with several channels. Higher local incomes might reflect also better labor-market matching and employment stability, increasing the opportunity cost of leaving, whereas higher housing costs tighten budget constraints and accelerate exit, particularly among migrants who do not intend to settle permanently.

Because higher average income and living costs are positively correlated, as we have seen, the net effect of exit from Italy of living in a high-income high-living cost area depends on the specifics of the area and cannot be determined in general.

Regional disparities within Italy are also relevant. Migrants residing in Southern Italy or the Islands tend to remain significantly longer than those settling in the North. These differences persist even after controlling for wage levels and housing costs, suggesting that broader regional characteristics—such as community networks, perceived quality of life, or discrimination—may play a role. These findings are consistent with survey evidence showing greater social integration and life satisfaction among migrants in Southern regions (INAPP, 2023).

While the log-normal survival model provides useful insights into the determinants of migration duration, it assumes that all migrants are eventually at risk of leaving Italy. Yet, the descriptive evidence in Section 5.1 indicates that a substantial fraction never exits during the observation window. To address this, we estimate a split cure model that explicitly distinguishes between permanent stayers and those at risk of departure.

### 5.2.2. Split cure model results

Table 11 reports the results of the split cure model for the 2011 cohort. Panel A presents the determinants of the cure probability (permanent settlement), while Panel B shows the determinants of the timing of out-migration among the susceptible.

**Table 11. Estimation results – Split cure model**

| | (1)<br>Cure equation | (2)<br>Latency equation |
|---|---|---|
| Female | 0.334*** | 0.299*** |
| | (0.0105) | (0.0149) |
| Age of entry | -0.00728*** | -0.00139** |
| | (0.000449) | (0.000575) |
| Political Instability | 0.0168 | -0.172*** |
| | (0.0179) | (0.0274) |
| Relative costs of living | -0.100*** | -0.0558*** |
| | (0.0111) | (0.0214) |
| Log differential between Italian and origin country's GDP | 0.218*** | -0.117*** |
| | (0.0100) | (0.0141) |
| Income per worker in Italian LLMs (log) | 1.056*** | 0.206*** |
| | (0.0356) | (0.0625) |
| Average rent in Italian LLMs | -0.00176 | -0.0356*** |
| | (0.00337) | (0.00482) |
| Urban | 0.0141 | -0.180*** |
| | (0.0149) | (0.0206) |
| Major socio-economic region (NUTS1) of arrival: Center | 0.126*** | 0.0205 |
| | (0.0130) | (0.0184) |
| Major socio-economic region (NUTS1) of arrival: | 0.852*** | 0.248*** |
| South and Islands | (0.0185) | (0.0291) |
| Constant | -10.45*** | -3.300*** |
| | (0.369) | (0.652) |
| lnshape | -0.114*** | |
| | (0.00474) | |
| ***Observations*** | ***254,418*** | ***254,418*** |

*Notes: Standard errors in parentheses, * p<0.10 ** p<0.05 *** p<0.01*
Panel A reports logit coefficients for the probability of permanent settlement ("cured fraction").
Panel B reports coefficients from the log-normal latency model for the susceptible group.

The estimates confirm the presence of a substantial cured fraction: the average predicted probability of remaining permanently in Italy is around 0.79, closely aligned with the Kaplan–Meier survival curve in Figure 3. This validates the interpretation that most right-censored spells reflect genuine long-term settlement rather than delayed reporting of exits.

Female and age at entry exert similar effects on cure probability and duration of stay (conditional on not being permanent settlers): women are more likely to never leave and, if leaving, they do so later than males. On the contrary, older individuals are less likely to settle permanently and tend to leave earlier than younger individuals. Cross country income differential increases the probability of permanent settling, but shorten the length of stay for those who do leave. This is consistent with theoretical findings that individual who plan to temporary migrate might aim to accumulate a target amount of resources and higher relative income allows then to reach the target quicker. Higher cost of living reduces both the probability of permanent settlement and the duration of stay. Local level characteristics exert a differentiated effect. Individuals residing in area with higher wages are more likely to settle permanently, while the duration of stay is not affected by local labor market conditions (for those who plan to leave). On the contrary, higher cost of housing do not affect the probability of settling permanently, but reduce the duration of for those leaving. Finally, residing in the South strongly increases both the share of permanent settlers with respect to the other area of the country (especially with respect to the North), and the duration of stay of those leaving.

Taken together, these results enrich our analysis by showing that determinants of long-term settlement differ partly from those of temporary stays. The split cure model reconciles the high persistence observed in the Kaplan–Meier curve with the dynamics captured by the parametric and quantile regressions, and highlights that regional disparities within Italy operate not only through differences in duration but also through the probability of permanent settlement. From a policy perspective, distinguishing permanent settlers from temporary migrants is crucial: local conditions not only influence the length of temporary stays but also whether migrants become part of the long-term resident population. This distinction has direct implications for integration policies and local public finance.

### 5.2.3. Quantile regression for interval-censored data

The survival model described above provides an overview of the determinants of migration duration, however its assumptions and possible limitations call for a more flexible approach, also to validate its robustness. Given the challenges posed by the data, particularly the high proportion of right-censored observations and the lack of proportional hazards, the quantile regression framework adopted below allows for a more detailed exploration of the different factors shaping migration duration across various points in the distribution. This approach not only relaxes strong parametric assumptions but also provides a richer understanding of how migration patterns vary among different subgroups of the migrant population. We estimate quantile regressions for interval-censored data at the 10th, 25th, 50th, 75th, and 95th percentiles (Table 12). The findings provide valuable insights into the factors influencing onward or return migration across different points of the duration distribution. As already mentioned in Section 4.2, this framework aligns well with the interpretation of lower quantiles as reflecting short-term stays, while higher quantiles capture long-term stays. Therefore, estimating models at the quantiles considered allows us to identify possible heterogeneity in the determinants of migration durations.

The results of the quantile regressions show a substantial consistency with the parametric estimates, but also highlights non-negligible heterogeneity across deciles and indicate some lack of robustness in the estimates for a couple of variables. Several variables exhibit strong variation across quantiles. Gender effects are highly non-linear: men are more likely to exit early (positive coefficients for women at lower quantiles), but also more likely to remain among the long-stayers (negative coefficients for women at upper quantiles). At the 95th percentile (long-term stays), women stay almost three years less than men, while the effect reverses at the 10th percentile (very short-term stays). This suggests a bifurcated pattern in male migration—between early exits and long-term settlers.

While men are underrepresented among early exits, they are more likely to persist among long-term stayers, revealing a nuanced gender dynamic in migration durations.

Age at entry follows a similar pattern to that of male migrants: older migrants are more likely to leave early, but those who stay longer tend to remain longer still. This likely reflects differing migration strategies by age group, with older migrants possibly engaging in shorter-term labor migration.

Political instability has no significant effect at short durations, but becomes increasingly important for long-term stays. At the 95th percentile, a 10% increase in the Global Peace Index score is associated with a four-month increase in expected stay. This suggests that migrants from conflict-affected countries are more likely to remain in Italy in the long-term, possibly due to constrained return options.

The relative cost of living consistently reduces stay length across all quantiles, with the effect more pronounced for long-term stays (higher percentiles). Similarly, the income gap between Italy and the origin country becomes a stronger predictor of longer durations. Possibly migrants who have stayed longer and accumulated some wealth become more sensitive to relative prices, as they have the possibility to consume the accumulated wealth in a low cost environment. The (log) differential between Italian GDP and that of the country of origin has a non-significant effect at the 10th quantile, but a strong and significant positive effect from the 25th quantile onward. At the 95th quantile, a 10%

increase in the average log differential between income in Italy and in the country of origin increases the expected duration by 0.214 years[12] (approximately 2.5 months).

Local income levels in Italian labor markets are strongly associated with longer stays, particularly at the upper end of the distribution. This finding indicates that higher income areas correlate with longer stays, highlighting the importance of local economic opportunity for migrant retention and reinforcing the role of economic incentives and job opportunities. The average rent in the LLM area shows a distinct pattern. The coefficient at the 10th quantile is negative and non-significant, and it remains negligible at the 25th, so it seems to be not so relevant for short-term stays. However, at higher quantiles, the effect becomes positive and significant.

Urban residence continues to be associated with shorter stays across the distribution, reinforcing the earlier finding that urban settings foster greater circulation or onward migration.

**Table 12. Estimation results – Quantile regression analysis (total sample)**

| | (1)<br>10% | (2)<br>25% | (3)<br>50% | (4)<br>75% | (5)<br>95% |
|---|---|---|---|---|---|
| Female | 0.404*** | 0.283*** | -0.300*** | -1.066*** | -2.861*** |
| | (0.040) | (0.016) | (0.036) | (0.020) | (0.007) |
| Age of entry | -0.005*** | -0.006*** | 0.004** | 0.016*** | 0.042*** |
| | (0.001) | (0.001) | (0.002) | (0.001) | (0.0004) |
| Political instability | -0.092 | 0.049** | 0.260*** | 0.640*** | 1.624*** |
| | (0.062) | (0.017) | (0.058) | (0.033) | (0.013) |
| Relative costs of living | -0.105** | -0.154*** | -0.269*** | -0.573*** | -1.162*** |
| | (0.038) | (0.014) | (0.045) | (0.033) | (0.022) |
| Log differential between Italian and | -0.009 | 0.287*** | 0.494*** | 1.016*** | 2.243*** |
| origin country's GDP | (0.036) | (0.011) | (0.032) | (0.018) | (0.007) |
| Income per worker in Italian LLMs (log) | 1.084*** | 1.115*** | 1.371*** | 2.481*** | 4.641*** |
| | (0.126) | (0.043) | (0.126) | (0.079) | (0.045) |
| Average rent in Italian LLMs | -0.017 | 0.001 | 0.070*** | 0.182*** | 0.433*** |
| | (0.012) | (0.005) | (0.012) | (0.007) | (0.002) |
| Urban | -0.151** | -0.046** | -0.303*** | -0.734*** | -1.637*** |
| | (0.048) | (0.018) | (0.053) | (0.029) | (0.013) |
| Major socio-economic region (NUTS1) | -0.062 | 0.246*** | 0.555*** | 1.266*** | 3.038*** |
| of arrival: Centre | (0.060) | (0.018) | (0.046) | (0.025) | (0.007) |
| Major socio-economic region (NUTS1) | 1.326*** | 0.561*** | 0.137** | -0.197*** | -0.799*** |
| of arrival: South | (0.085) | (0.018) | (0.055) | (0.031) | (0.013) |
| Major socio-economic region (NUTS1) | 2.031*** | 0.769*** | 0.754*** | 1.142*** | 1.936*** |
| of arrival: Islands | (0.160) | (0.023) | (0.076) | (0.042) | (0.017) |
| Constant | -9.99*** | -0.360 | -1.758 | -11.820*** | -30.960*** |
| | (1.303) | (0.451) | (1.308) | (0.823) | (0.464) |
| ***Observations: 254,418*** | | | | | |

*Notes: Standard errors in parentheses, * p<0.10 ** p<0.05 *** p<0.01. The number of observations remains the same across the five columns as this type of regressions takes into consideration censored data exploiting information from all observations (censored and uncensored) to obtain a more robust estimate.*

[12] Again, for the log-transformed variables we have: log(1.10)≈0.0953. So, we approximate the result as the product between this value and the estimated coefficient. For the (log) differential between Italian and origin country's GDP we have: 2.243×0.0953≈0.214.

Finally, dummies for the major socio-economic region (NUTS1) of residence confirm substantial spatial variation. Overall, the majority of the coefficients for Central, Southern Italy and Island are positive and significant, indicating that migrants arriving in those regions are more likely to remain than those arriving in the North, especially among longer-stayers. The only exceptions are the coefficients for arrival in the South at higher quantiles, which become negative. This may point to the limited capacity of the South to retain migrants in the long run with respect to the North, probably due to limited long-term opportunities. Having established the main determinants of migration duration in the full sample, we now turn to potential heterogeneities across migrant groups by focusing on the five largest nationalities.

## 5.3 Return migration for the five largest nationalities

To explore heterogeneity in migration duration across the ethnic groups, we estimate the parametric model for individuals born in Romania, Albania, Morocco, China, and Ukraine. We estimate two complementary models: a parametric survival model using time-varying covariates and a competing risks model focusing on return to the country of origin.

While individual characteristics such as education or legal status are not observed in registry data, country of origin might help capturing systematic variation in migrants' skill composition, migration motives, institutional pathways, and access to labor market segments. Disaggregating the analysis by nationality therefore provides a meaningful and tractable way to explore heterogeneity in migration duration and return behavior. Limiting the number of countries of origin to the top five, allows us to retain sufficient sample size to estimate complex duration models.

### 5.3.1. Parametric survival model with time-varying covariates

Table 13 presents the results of the parametric estimates. As mentioned, given the smaller sample size we are able to use time varying covariates in these estimates. These estimates can also be interpreted as additional evidence of the robustness of the results obtained using time invariant covariates.

The estimates are largely consistent with those obtained for the whole sample using the value at arrival of the different covariates.

The results also show some heterogeneity in migration duration and its characteristics across the five largest nationalities present in Italy.

**Table 13. Estimation results – Parametric survival analysis for the main 5 nationalities**

| | (1)<br>Romania | (2)<br>Albania | (3)<br>Morocco | (4)<br>China | (5)<br>Ukraine |
|---|---|---|---|---|---|
| Female | 0.354*** | 0.869*** | 1.009*** | 0.434*** | 0.835*** |
| | (0.0120) | (0.0431) | (0.0380) | (0.0417) | (0.0623) |
| Age of entry | -0.0128*** | -0.0279*** | -0.0198*** | -0.0282*** | -0.0140*** |
| | (0.000504) | (0.00148) | (0.00178) | (0.00205) | (0.00193) |
| Political instability | 4.115*** | -10.25*** | 1.246*** | 0.671* | -0.514*** |
| | (0.446) | (0.660) | (0.338) | (0.348) | (0.120) |
| Relative cost of living | -30.40*** | -176.3** | -102.7*** | -109.6*** | -39.06*** |
| | (0.878) | (73.13) | (2.045) | (4.002) | (1.047) |
| Log differential between Italian and origin country's GDP | -1.032*** | -12.14*** | 5.377*** | -1.367*** | -2.100*** |
| | (0.302) | (0.503) | (0.600) | (0.430) | (0.506) |
| Income per worker in Italian LLMs | 0.338*** | 1.544*** | 0.900*** | 2.384*** | 2.578*** |
| | (0.0438) | (0.169) | (0.145) | (0.161) | (0.188) |
| Average rent in Italian LLMs | -0.00494 | 0.0239* | -0.0777*** | 0.0210 | -0.108*** |
| | (0.00418) | (0.0139) | (0.0137) | (0.0132) | (0.0182) |
| Urban | 0.0451*** | -0.196*** | -0.385*** | -0.265*** | 0.0756 |
| | (0.0144) | (0.0612) | (0.0499) | (0.0693) | (0.0872) |
| Major socio-economic region (NUTS1) of arrival: Center | 0.0135 | -0.191*** | 0.333*** | -0.820*** | 0.0231 |
| | (0.0140) | (0.0493) | (0.0554) | (0.0508) | (0.0681) |
| Major socio-economic region (NUTS1) of arrival: Islands | 0.746*** | 0.465*** | 1.645*** | 2.085*** | 1.917*** |
| | (0.0217) | (0.0831) | (0.0730) | (0.0911) | (0.0871) |
| Major socio-economic region (NUTS1) of arrival: South | 0.798*** | 1.333*** | 1.356*** | 3.127*** | 1.863*** |
| | (0.0281) | (0.176) | -0.385*** | (0.152) | (0.250) |
| *Observations:* | *693,003* | *147,826* | *167,459* | *136,301* | *164,777* |

*Notes: Standard errors in parentheses, * p<0.10 ** p<0.05 *** p<0.01*

As in the overall sample, age at entry consistently reduces the duration of stay across all groups, confirming the robustness of this determinant. The strongest effect is found for Chinese migrants. Gender gaps are more heterogeneous: the largest differences between men and women are observed among Albanians, Moroccans, and Ukrainians. This suggests that gendered migration strategies differ across groups, even though the direction of the effect (men leaving earlier) remains broadly consistent.

Political instability in the country of origin shows mixed effects across groups. It increases duration for Romanian, Moroccan, and Chinese migrants—suggesting a protective effect in contexts where return is constrained—but decreases it for Albanians and Ukrainians, where instability may interact with circular migration patterns. These apparently divergent effects thus reflect group-specific migration strategies rather than contradictory findings.

Economic factors also exhibit substantial influence, albeit with some heterogeneity across groups. The relative cost of living exerts a strong negative effect across all groups, confirming that, overall, a lower cost of living in the country of origin increases the likelihood of out-migrating.

The (log) differential between Italian GDP and that of the origin country has heterogeneous effects. While it is negative in most cases—suggesting that a larger income gap accelerates exit—it turns positive for Moroccans, pointing to stronger long-term attachment to Italy when economic returns are high. This heterogeneity highlights that economic incentives are filtered through group-specific expectations and strategies.

Concerning local factors in Italy, higher income per worker in the LLM is associated with longer stays for all five groups. Housing costs, instead, have divergent effects: higher rents shorten stays for Moroccans and Ukrainians, but lengthen them for Albanians. This apparent contradiction suggests that housing markets may operate differently depending on networks and settlement patterns—being a barrier for some groups but an indicator of opportunity for others.

Regional and urban–rural patterns also vary. For Romanians, urban residence is associated with longer stays, probably due to greater job opportunities. For Albanians, Moroccans, and the Chinese, instead, cities are linked to shorter durations, reflecting that urban areas also facilitate international mobility and thus out-migration. Across all groups, settlement in the South or Islands generally corresponds to longer stays, while Central Italy shows mixed patterns (e.g., Albanians and Chinese leaving sooner, Moroccans staying longer). Taken together, these results confirm that some determinants (such as age and local income) operate consistently across groups, while others (political instability, rents, urban residence) show significative differences across nationalities.

### 5.3.2. Competing risks model with time-varying covariates

Turning to return migration dynamics, Table 14 reports the results for the five largest nationalities from competing risks models estimating the sub-hazard of return to the country of origin, relative to remaining in Italy or moving elsewhere. This framework captures not only the destination of exit but also its timing and competing pathways. In general, positive coefficients suggest that the risk of return migration – the event of interest in this model – is greater than the risk of competing events (migration to a third country or departure without a specified destination). Negative coefficients imply the opposite, indicating a higher relative importance of competing events.

Female migrants face significantly lower risks of return compared to male migrants across all groups, with the strongest effects for Ukrainians and Albanians. Age at entry positively predicts return migration: older migrants are more likely to return, consistent with shorter planning horizons and weaker integration incentives. Political instability at origin has heterogeneous effects: it increases the hazard of return for Albanians and Ukrainians, but reduces it for Moroccans and Chinese. Rather than a contradiction, this divergence mirrors the survival model findings and suggests that instability may either constrain return (protective effect) or foster circular mobility, depending on group-specific contexts. Relative cost of living is a strong and positive predictor of return for all groups except Romanians. GDP differentials affect return mostly for Romanians: a higher income gap with Italy increases their likelihood of return. This could reflect target-saving strategies among Romanian migrants. Higher local incomes in Italy reduce return probabilities across all groups, suggesting that better host-country opportunities discourage return. Higher rental costs reduce return risk for Romanians and Albanians. Urban residence reduces the risk of return for most groups, consistent with better access to services or onward mobility opportunities. Regional location within Italy also shapes return probabilities. Residence in the Center, South or Islands is associated with lower hazards of return compared to residence in Northern Italy, especially for Romanians and Ukrainians. This likely reflects both lower mobility and higher community embeddedness.

**Table 14. Competing risks**

| | (1)<br>Romania | (2)<br>Albania | (3)<br>Morocco | (4)<br>China | (6)<br>Ukraine |
|---|---|---|---|---|---|
| Female | -0.123*** | -0.315*** | -0.303*** | -0.0659 | -0.499*** |
| | (0.0269) | (0.0933) | (0.100) | (0.0851) | (0.0984) |
| Age of entry | 0.0148*** | 0.0324*** | 0.0375*** | 0.0168*** | 0.0111*** |
| | (0.00118) | (0.00319) | (0.00469) | (0.00432) | (0.00372) |
| Political Instability | -1.729** | 10.36*** | -4.861*** | -5.394*** | 0.750*** |
| | (0.855) | (1.432) | (0.877) | (0.568) | (0.210) |
| Relative cost of living | 2.012 | 262.9* | 46.58*** | 29.98*** | 13.02*** |
| | (1.627) | (150.5) | (4.827) | (8.551) | (1.334) |
| Log differential between Italian and origin country’s GDP | 2.719*** | -0.897 | -8.804*** | 0.778 | -0.589 |
| | (0.600) | (1.034) | (1.627) | (0.900) | (0.567) |
| Income per worker in Italian LLMs | -1.070*** | -1.231*** | -1.722*** | -1.149*** | -1.914*** |
| | (0.0876) | (0.331) | (0.345) | (0.263) | (0.256) |
| Average rent in Italian LLMs | -0.0372*** | -0.133*** | -0.0519 | 0.0392 | -0.0162 |
| | (0.0103) | (0.0304) | (0.0421) | (0.0247) | (0.0368) |
| Urban | -0.393*** | -0.493*** | -0.227* | -0.657*** | -0.545*** |
| | (0.0307) | (0.112) | (0.119) | (0.118) | (0.121) |
| Major socio-economic region (NUTS1) of arrival: Center | -0.322*** | 0.0676 | -0.231 | 0.00655 | -0.509*** |
| | (0.0319) | (0.104) | (0.152) | (0.0979) | (0.121) |
| Major socio-economic region (NUTS1) of arrival: Islands | -1.070*** | -1.480*** | -0.923*** | -1.431*** | -0.950** |
| | (0.0702) | (0.526) | (0.269) | (0.415) | (0.399) |
| Major socio-economic region (NUTS1) of arrival: South | -0.942*** | -0.628*** | -1.105*** | -0.467** | -1.587*** |
| | (0.0525) | (0.192) | (0.203) | (0.185) | (0.161) |
| *Observations* | *693,003* | *147,826* | *167,459* | *136,301* | *164,777* |

*Notes: Standard errors in parentheses, * p<0.10 ** p<0.05 *** p<0.01*

The analysis of the five largest nationalities shows that while some determinants of migration duration—such as age at entry, gender, and local income levels—operate consistently across groups, others (political instability, housing costs, and urban residence) display marked heterogeneity. Rather than representing contradictions, these differences reflect the diversity of migration strategies, return constraints, and integration trajectories across communities. This underscores the importance of combining aggregate and group-specific analyses to capture the full complexity of migrant retention and return dynamics in Italy.

## 6. Conclusions

This paper has provided new evidence on the duration of stay and return migration of immigrants in Italy between 2011 and 2022, using a novel dataset that reconstructs individual migration histories from administrative population registers. By combining parametric survival models, split-cure models, quantile regressions for interval-censored data, and competing risks models, we provide new evidence on the determinants of return and onward migration, and on the role of local conditions within the host country.

Several key findings emerge. First, our results confirm that migration to Italy is relatively permanent: the majority of immigrants remain in the country after more than a decade, with exit rates concentrated in the early years. About 80 percent of the 2011 arrival cohort was still present after twelve years and later cohorts behave in a similar way. Second, individual characteristics matter. Male and older migrants are more likely to leave earlier, while women and younger migrants tend to remain longer, although quantile regressions reveal more complex patterns that suggest differentiated strategies and constraints across the distribution of stays. Third, economic incentives play a central role. Higher relative costs of living in Italy reduce the length of stay, while larger income gaps in favor of Italy prolong it. These findings are robust across estimation methods and consistent with target-saving and consumption-based models of migration. Political instability in origin countries, by contrast, has a heterogeneous impact: it matters more for long-term stays, and its effects vary substantially across nationality groups.

Fourth, our analysis underscores the significance of subnational conditions within Italy. Local labor markets, housing costs, and broader regional disparities emerge as strong and independent determinants of migration duration. Higher local incomes encourage longer stays, while high rents accelerate departure. Migrants in Southern Italy and the Islands tend to remain longer than those in the North, even after controlling for economic variables. This highlights the role of community integration, and institutional differences across regions: a dimension largely neglected in the return migration literature, which usually treats host countries as homogeneous spaces.

Finally, our exploration of the five largest nationality groups shows both commonalities and important heterogeneity. Age at entry and relative cost of living consistently affect return migration across all groups, but gender gaps, the role of political instability, and the influence of local conditions vary substantially. Competing risks estimates further confirm that return and onward migration are shaped by distinct logics: while return to origin is strongly linked to demographic factors and relative costs of living, onward migration is more sensitive to urban residence and local labor market opportunities.

Methodologically, we show the value of applying a set of duration models to administrative data characterized by censoring and heterogeneity. The split cure model captures the fraction of permanent stayers, quantile regressions identify differential effects across the distribution of stays, and competing risks analysis distinguishes between return and onward migration. Together, these approaches provide a richer and more nuanced understanding of migrant behavior than any single method alone.

Taken together, these findings contribute to a more nuanced understanding of temporary and return migration. They suggest that migrants' decisions are not only shaped by macroeconomic and political conditions across borders, but also by where within the host country they reside and integrate. This underscores the need to move beyond national averages and to take seriously the role of local contexts in shaping migration trajectories.

These findings have important policy implications. They suggest that integration policies cannot be designed solely with permanent settlement in mind but must also accommodate temporary and onward trajectories. They further underline the importance of subnational variation: policies aimed at migrant retention or integration will need to consider regional labor markets, housing costs, and institutional differences within Italy.

Several limitations should be acknowledged. Our administrative data, while comprehensive, lack information on education, household structure, and employment, which are likely to play an important role in migration decisions. The presence of administrative cancellations also complicates the measurement of exact departure dates, introducing interval censoring. While our estimation strategy mitigates some of these issues, future research would benefit from richer data that combine registry information with individual-level socio-economic variables.

Future work should also further investigate the interaction between local conditions and migrants' life-cycle trajectories, including family reunification, naturalization, and labor market participation. Linking administrative data to other sources could shed light on the role of remittances, return intentions, and onward migration strategies. Comparative analyses with other Southern European countries would also help assess to what extent our findings reflect Italy’s specific institutional and economic setting or broader regional patterns.

## 7. References


Abramitzky, R., Boustan, L., Jácome, E., & Pérez, S. (2021). Intergenerational mobility of immigrants in the United States over two centuries. *American Economic Review, 111*(2), 580-608.

Accetturo, A., Manaresi, F., Mocetti, S., & Olivieri, E. (2014). Don't Stand so close to me: the urban impact of immigration. *Regional Science and Urban Economics, 45*, 45-56.

Adda et al. (2022). The Dynamics of Return Migration, Human Capital Accumulation, and Wage Assimilation. *Review of Economic Studies*, 89, 2841–2871.

Albert, Christoph, and Joan Monras. (2022). "Immigration and Spatial Equilibrium: The Role of Expenditures in the Country of Origin." *American Economic Review* 112 (11): 3763–3802.

Anghel et al. (2016). International Migration, Return Migration, and Their Effects: A Comprehensive Review on the Romanian Case. *IZA DP* No. 10445.

Bauloz, C. et al. (2024). Gender and migration: Trends, gaps and urgent action. In: World Migration Report 2024 (M. McAuliffe and L.A. Oucho, eds.). International Organization for Migration (IOM), Geneva.

Bijwaard, G. E. (2010). Immigrant Migration Dynamics Model for The Netherlands. *Journal of Population Economics*, 23(4), 1213–1247.

Bijwaard, G. E., & Wahba, J. (2014). The Impact of Labor Market Dynamics on the Return Migration of Immigrants. *The Review of Economics and Statistics*, 96(3), 483–494.

Bijwaard, G. E., & Wahba, J. (2023). Return versus onward migration: Go back or move on?. *Review of Income and Wealth*, *69*(3), 640-667.

Chernozhukov, Victor, and Han Hong**.** (2003) An MCMC approach to classical estimation. *Journal of Econometrics*, 115(2), 2003: 293–346.

Chiurco, L., et al. (2023). Indicatori di integrazione dei cittadini con background migratorio residenti in Italia. INAPPREPORT

Cobb-Clark, D. A., & Stillman, S. (2013). Return Migration and the Age Profile of Retirement Among Immigrants. *IZA Journal of Migration*, 2(20).

De Vreyer, P., Gubert, F., & Robilliard, A. S. (2010). Are There Returns to Migration Experience? An Empirical Analysis Using Data on Return Migrants and Non-Migrants in West Africa. *Annals of Economics and Statistics*, 307–328.

Dustmann, C. (1996). Return Migration: The European Experience. *Economic Policy*, April 1996, pp. 214–242.

Dustmann, C. (2003). Return Migration, Wage Differentials, and the Optimal Migration Duration. *European Economic Review*, 47, 353–369.

Dustmann, C., & Görlach, J. (2016). The Economics of Temporary Migrations. *Journal of Economic Literature*, 54(1), 98–136.

Dustmann C. & Preston I. P. (2019). Free Movement, Open Borders, and the Global Gains from Labor Mobility. *Annual Review of Economics*, vol. 11(1).

Dustmann, C., & Weiss, Y. (2007). Return Migration: Theory and Empirical Evidence from the UK. *British Journal of Industrial Relations*, 45(2), 236–256.

Farcomeni, A. and M. Geraci (2020). Multi-state quantile regression models. *Statistics in Medicine 39*, 45–56.

Fine, J.P. and Gray, R.J. (1999). A proportional hazards model for the subdistribution of a competing risk. *JASA* 94:496-509.

Friedman, M. (1982). Piecewise exponential models for survival data with covariates. *The Annals of Statistics 10*, 101–113.

Frumento, P. (2022). A quantile regression estimator for interval-censored data. *The International Journal of Biostatistics 19*, 81–96.

Galor, O., & Stark, O. (1990). Migrants' Savings, the Probability of Return Migration, and Migrants' Performance. *International Economic Review*, 31(2), 463–467.

Guo G. (1993). Event-History Analysis for Left-Truncated Data. *Sociological Methodology*, Vol. 23 (1993), pp. 217-243

INAPP (2023). Indicatori di integrazione dei cittadini con background migratorio residenti in Italia

ISTAT (2024). Benessere e disuguaglianze in Italia. https://www.istat.it/wp-content/uploads/2024/11/Benessere-disuguaglianze.pdf

Kalbfleisch, J. D., & Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data (2nd ed.).* Wiley.

Kindar, M. (2009). Labor Market Outcomes, Savings Accumulation, and Return Migration. *Labor Economics*, 16, 418–428.

Lambert, P. C. (2007). Modeling of the cure fraction in survival studies. *The Stata Journal*, *7*(3), 351-375.

Latouche, A., et al. (2007). Misspecified regression model for the subdistribution hazard of a competing risk. *Statistics in medicine*, *26*(5), 965-974.

Lawless, J. F. (2003). *Statistical Models and Methods for Lifetime Data (2nd ed.).* Wiley.

Maller, R. A., & Zhou, X. (1997). *Survival Analysis with Long-Term Survivors*. Wiley.

Martin, R., & Radu, D. (2012). Return Migration: The Experience of Eastern Europe. *International Migration*, 50(6), 109–128.

OECD (2017). *Immigrant Integration and the Role of Return Migration*. Chapter in *International Migration Outlook 2017*. OECD Publishing, Paris. https://doi.org/10.1787/migr_outlook-2017-en.

OECD (2020). *International Migration Outlook 2020*. OECD Publishing, Paris. https://doi.org/10.1787/ec98f531-en.

Ottaviano, G. I., & Peri, G. (2008). *Immigration and national wages: Clarifying the theory and the empirics* (No. w14188). National Bureau of Economic Research.

Peng, Y, Taylor, JMG. 2014. *Cure models. In Handbook of Survival Analysis, Handbooks of Modern Statistical Methods series*, ed. J Klein, H van Houwelingen, JG Ibrahim, TH Scheike, chapter 6, pages 113-134. Boca Raton, FL : Chapman & Hall

Portnoy, S. (2003). Censored Regression Quantiles. *Journal of the American Statistical Association*, 98(464), 1001–1012. https://doi.org/10.1198/016214503000000954

Ramos, F. (1992). *Out-Migration and Return Migration of Puerto Ricans.* In: *Immigration and the Workforce: Economic Consequences for the United States and Source Areas*, pp. 49–66.

Sá, F. (2015). Immigration and house prices in the UK. *The Economic Journal*, *125*(587), 1393-1424.

Saiz, A. (2007). Immigration and housing rents in American cities. *Journal of urban Economics*, *61*(2), 345-371.

Schmidt, C. M. (1994). The Country of Origin, Family Structure, and Return Migration of Germany's Guest-Workers. *Vierteljahrshefte zur Wirtschaftsforschung*, 63(1/2), 119–125.

Tezcan, S. (2019). Return Home? Determinants of Return Migration Intention Amongst Turkish Immigrants in Germany. *Geoforum*, 98, 189–201.

Vanderkamp, J. (1971). Migration Flows, Their Determinants and the Effects of Return Migration. *Journal of Political Economy*, 79(5), 1012–1030.

Ward, Z. (2017). Birds of Passage: Return Migration, Self-Selection, and Immigration Quotas. *Explorations in Economic History*, 64, 37–52.

Yahirun, J. J. (2014). Take Me “Home”: Return Migration among Germany’s Older Immigrants. *International Migration*, 52(4).

Yu, Binbing, & Peng, Yingwei. (2008). Mixture cure models for multivariate survival data. *Computational Statistics & Data Analysis*, 52(3), 1524-1532. https://doi.org/10.1016/j.csda.2007.04.018

## Appendices

### Appendix A. Dataset description

Return migration is one of the most difficult aspects of international migration to observe and measure accurately. This is largely due to the limited obligations for emigrants to report their departure to national authorities, which leads to significant undercounting in official statistics. As a result, most international organizations rely on indirect measures, such as Labor Force Surveys (LFS) and household surveys, to estimate return migration flows. These sources typically collect self-reported migration histories, often retrospectively, asking individuals whether they have previously lived abroad.

Recent OECD reports, such as *Immigrant Integration and the Role of Return Migration* (2017), have highlighted the importance of leveraging longitudinal, individual-level data from population registers, civil records, or social insurance systems to better understand return dynamics, particularly for temporary or circular migrants (OECD, 2020). While OECD and other international bodies call for registry-based approaches rather than survey-based sources, only a few countries have implemented comprehensive systems capable of tracking international emigration over time.

We use administrative microdata from Italy's municipal population register, accessed through ISTAT's ADELE platform. This dataset enables us to construct a rich, temporally precise picture of international migration episodes for foreign-born residents in Italy, disaggregated by date of registration, citizenship, age, and duration of stay. It complements international survey-based indicators by offering a more comprehensive and granular perspective on emigration behavior and return propensities.

In details, the dataset we employ contains individual-level records of all entries and exits formally registered by Italian municipalities between 2011 and 2022. These records include both voluntary notifications (i.e., self-declared moves) and administrative removals (e.g., cancellations due to presumed emigration), allowing us to reconstruct detailed longitudinal migration histories for millions of individuals.

This administrative strategy offers three main advantages over traditional survey data:

1. Full Coverage: The dataset includes all registered residents, enabling analysis of the entire foreign-born population.

2. High Temporal Precision: Each migration episode includes both the event date and the date of registry update, allowing us to track migration durations and flows with high resolution.

3. Rich Demographic Detail: Records contain information on sex, age, country of birth, citizenship, and municipality of residence, supporting fine-grained disaggregated analyses.

However, working with administrative data also entails specific challenges, especially regarding privacy constraints and data linkage. Full personal identifiers are not available, so we construct synthetic IDs using demographic characteristics (e.g., date of birth, sex, and citizenship). We apply strict disambiguation filters to avoid duplicate or implausible matches, dropping cases with multiple potential linkages or inconsistent movement histories. Furthermore, we distinguish between voluntary cancellations, likely indicative of return or onward migration, and administrative cancellations, which offer only upper bounds on duration and lack destination information.

After extensive cleaning and filtering, our final dataset includes approximately 3.7 million unique foreign-born individuals whose migration episodes into and out of Italy can be tracked consistently across time.

To validate the quality and representativeness of our reconstruction, we compare our aggregated migration flows with ISTAT's official statistics published through the ISTAT Data Browser. These include the registrations from abroad (immigrations) and the cancellations for abroad (emigrations). We perform this comparison using two distinct subsamples:

1. Foreign-born individuals (classified by country of birth)
2. Foreign citizens (classified by citizenship), which is the standard used in ISTAT's published migration statistics

Tables A1 and A2 below summarize these selected yearly comparisons.

**Table A1. Data sets comparison - Foreign-born individuals (*classified by country of birth*)**

| Year | Emigration *(microdata)* | Emigration *(reconstruction)* | Emigration *(ISTAT data)* | Immigration *(microdata)* | Immigration *(reconstruction)* | Immigration *(ISTAT data)* |
|---|---|---|---|---|---|---|
| 2015 | 62390 | 56556 | 44696 | 254121 | 237170 | 262929 |
| 2016 | 66189 | 57802 | 42553 | 270126 | 252763 | 262929 |
| 2017 | 69612 | 61930 | 40551 | 310793 | 288178 | 301071 |
| 2018 | 71479 | 63461 | 40228 | 298180 | 279464 | 285500 |
| 2019 | 91242 | 75621 | 57485 | 280595 | 270129 | 264571 |
| 2020 | 64533 | 55777 | 38934 | 187619 | 180153 | 191766 |
| 2021 | 71170 | 61177 | 64093 | 247612 | 238071 | 243607 |
| 2022 | 67043 | 57867 | 50679 | 316108 | 304833 | 336495 |

*Notes: The microdata from the ADELE database include all the foreigners cancelled/registered in Italy (8,610,884 obs). Following our hypothesis, we obtain a reconstructed dataset of 8,432,845 observations. Administrative cancellations are not included. Aggregated data on cancellation/registration of foreigners are publicly available through the ISTAT Data Browser*

**Table A2. Data sets comparison - Foreign citizens (*classified by citizenship*)**

| Year | Emigration *(microdata)* | Emigration *(reconstruction)* | Emigration *(ISTAT data)* | Immigration *(microdata)* | Immigration *(reconstruction)* | Immigration *(ISTAT data)* |
|---|---|---|---|---|---|---|
| 2015 | 44695 | 38700 | 44696 | 250014 | 233109 | 250006 |
| 2016 | 42553 | 36555 | 42553 | 262926 | 247345 | 262926 |
| 2017 | 40550 | 34259 | 40551 | 301054 | 278708 | 301071 |
| 2018 | 40227 | 33936 | 40228 | 285485 | 267415 | 285500 |
| 2019 | 57484 | 44907 | 57485 | 264567 | 255348 | 264571 |
| 2020 | 38934 | 32216 | 38934 | 191762 | 184348 | 191766 |
| 2021 | 64093 | 52756 | 64093 | 243606 | 234570 | 243607 |
| 2022 | 50679 | 42075 | 50679 | 336490 | 325129 | 336495 |

*Notes: The microdata from the ADELE database include all the foreigners cancelled/registered in Italy (8,225,952 obs). Following our hypothesis, we obtain a reconstructed dataset of 4,244,201 observations. Administrative cancellations are not included. Aggregated data on cancellation/registration of foreigners are publicly available through the ISTAT Data Browser*

In the cleaned and reconstructed dataset, both emigration and immigration figures are slightly lower than ISTAT data. This is consistent with our decision to exclude ambiguous records. Indeed, these differences reflect a conservative reconstruction strategy that prioritizes accuracy and traceability over exhaustive coverage. Notably, when we compute migration flows directly from the original ADELE microdata, without applying filters or synthetic ID construction, the alignment with ISTAT figures is nearly exact, confirming the reliability of the source data.

## Appendix B. Balance test with the original sample

Table B1 and B2 report balance tests related to our sample and the original one, for arrivals and departures respectively. The first three columns refer to the 2011 cohort, while the last three to the total sample. Tests are reported for the total sample of individuals and for the five main nationalities. While some differences are statistically significant, they are rather small in size.

**Table B1. Arrivals**

| | **2011 cohort** | | | **Total sample** | | |
|---|---|---|---|---|---|---|
| | **Our sample** | **Original sample** | **Difference** | **Our sample** | **Original sample** | **Difference** |
| **Total** | | | | | | |
| Males | .455 | .470 | -0.0151*** | .515 | .524 | -0.0409*** |
| Age | 33.647 | 33.598 | 0.0494 | 33.395 | 33.436 | -0.00854*** |
| **Romanian** | | | | | | |
| Males | .394 | .392 | 0.00171 | .381 | .375 | 0.00693*** |
| Age | 33.683 | 33.633 | 0.0499 | 36.308 | 36.310 | -0.00264 |
| **Albanian** | | | | | | |
| Males | .450 | .457 | -0.00720 | .462 | .465 | -0.00215 |
| Age | 31.100 | 30.989 | 0.111 | 33.531 | 33.452 | 0.0795 |
| **Moroccan** | | | | | | |
| Males | .540 | .584 | -0.0440*** | .491 | .505 | -0.0140*** |
| Age | 31.691 | 33.059 | -1.368*** | 32.391 | 34.698 | -2.307*** |
| **Chinese** | | | | | | |
| Males | .484 | .495 | -0.0109 | .479 | .482 | -0.00318 |
| Age | 33.081 | 32.842 | 0.240* | 34.161 | 34.032 | 0.129** |
| **Ukrainian** | | | | | | |
| Males | .227 | .223 | 0.00426 | .249 | .247 | 0.00251 |
| Age | 40.417 | 40.535 | -0.117 | 37.613 | 37.730 | -0.117 |

**Table B2. Departures**

| | 2011 cohort | | | Total sample | | |
|---|---|---|---|---|---|---|
| | **Our sample** | **Original sample** | **Difference** | **Our sample** | **Original sample** | **Difference** |
| **Total sample** | | | | | | |
| Males | .474 | .474 | 0.000535 | .563 | .570 | 4.440*** |
| Age | 35.659 | 37.912 | -2.254*** | 43.668 | 39.228 | -0.00730*** |
| **Romanian** | | | | | | |
| Males | .453 | .434 | 0.0196* | .451 | .452 | -0.000988 |
| Age | 36.297 | 37.783 | -1.486*** | 44.713 | 40.527 | 4.186*** |
| **Albanian** | | | | | | |
| Males | .583 | .594 | -0.0114 | .600 | .628 | -0.0284*** |
| Age | 42.704 | 38.636 | 4.068*** | 45.968 | 38.524 | 7.443*** |
| **Moroccan** | | | | | | |
| Males | .548 | .572 | -0.0233 | .634 | .662 | -0.0280*** |
| Age | 38.364 | 38.520 | -0.156 | 45.734 | 40.519 | 5.216*** |
| **Chinese** | | | | | | |
| Males | .472 | .477 | -0.00425 | .538 | .539 | -0.000823 |
| Age | 34.857 | 35.727 | -0.870 | 45.154 | 39.114 | 6.040*** |
| **Ukrainian** | | | | | | |
| Males | .272 | .261 | 0.0109 | .237 | .241 | -0.00398 |
| Age | 44.297 | 45.089 | -0.792 | 50.683 | 46.328 | 4.355*** |

## Appendix C. Additional descriptive statistics

### Arrivals and departure by province

Figure C1 and C3 show for each Italian province the average number (between 2011 and 2022) of yearly arrivals and departures, considering the sample of all individuals in the 15-69 age range when the movement to and from Italy happened. On average, the province that received more immigrants was Rome, followed by Milan in the North of Italy and Naples in the South. Despite general lower numbers, the situation is similar when we look at the provinces from which immigrants returned home or left the country. Figure C2 and C4 show the number of arrivals and departures by municipality, for the cohort of arrivals in 2011 and for the whole sample (from 2011 to 2022).

**Figure C1. Average number of arrivals in Italian provinces (whole sample)**

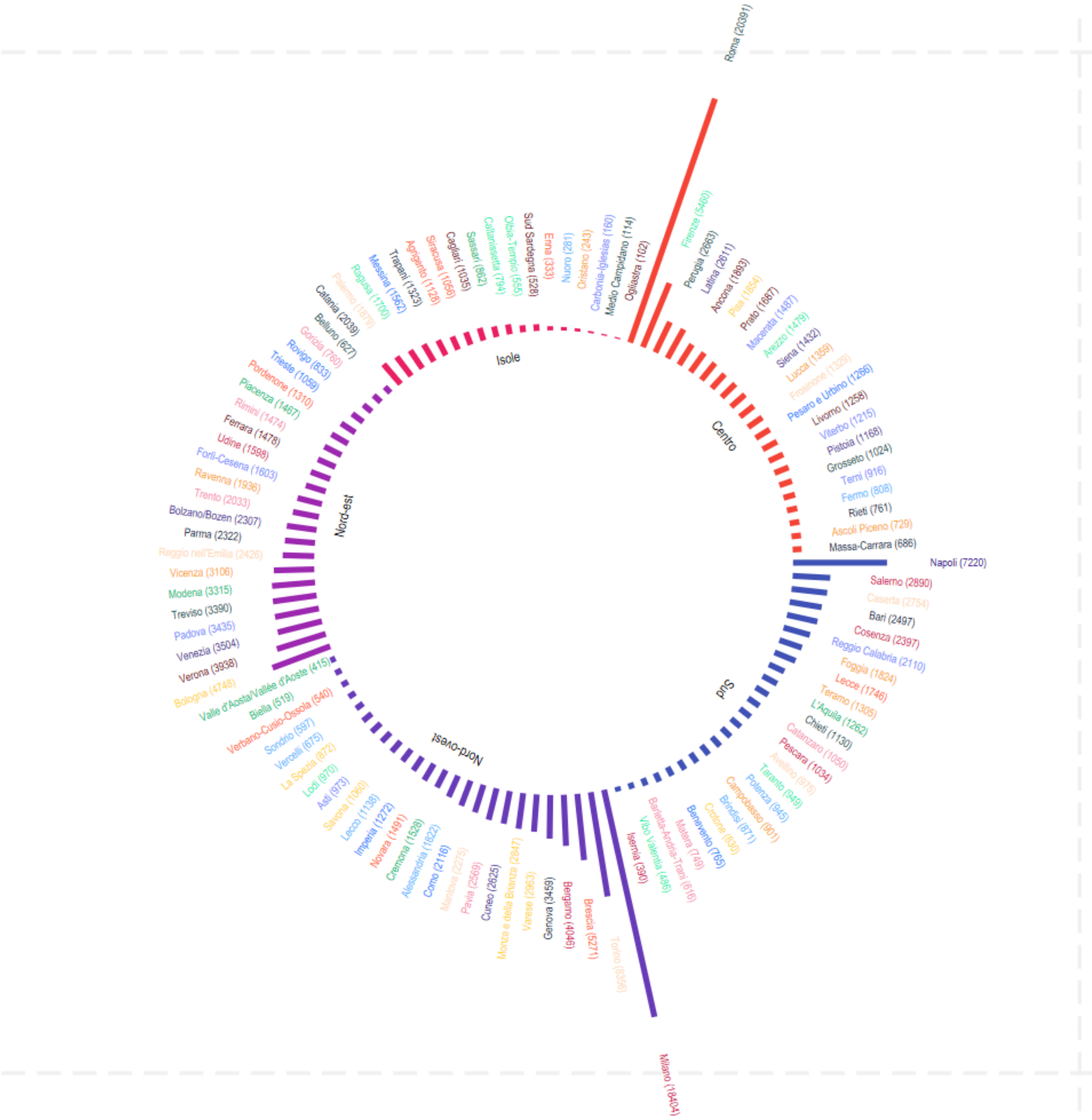


**Figure C2. Number of arrivals by municipality**

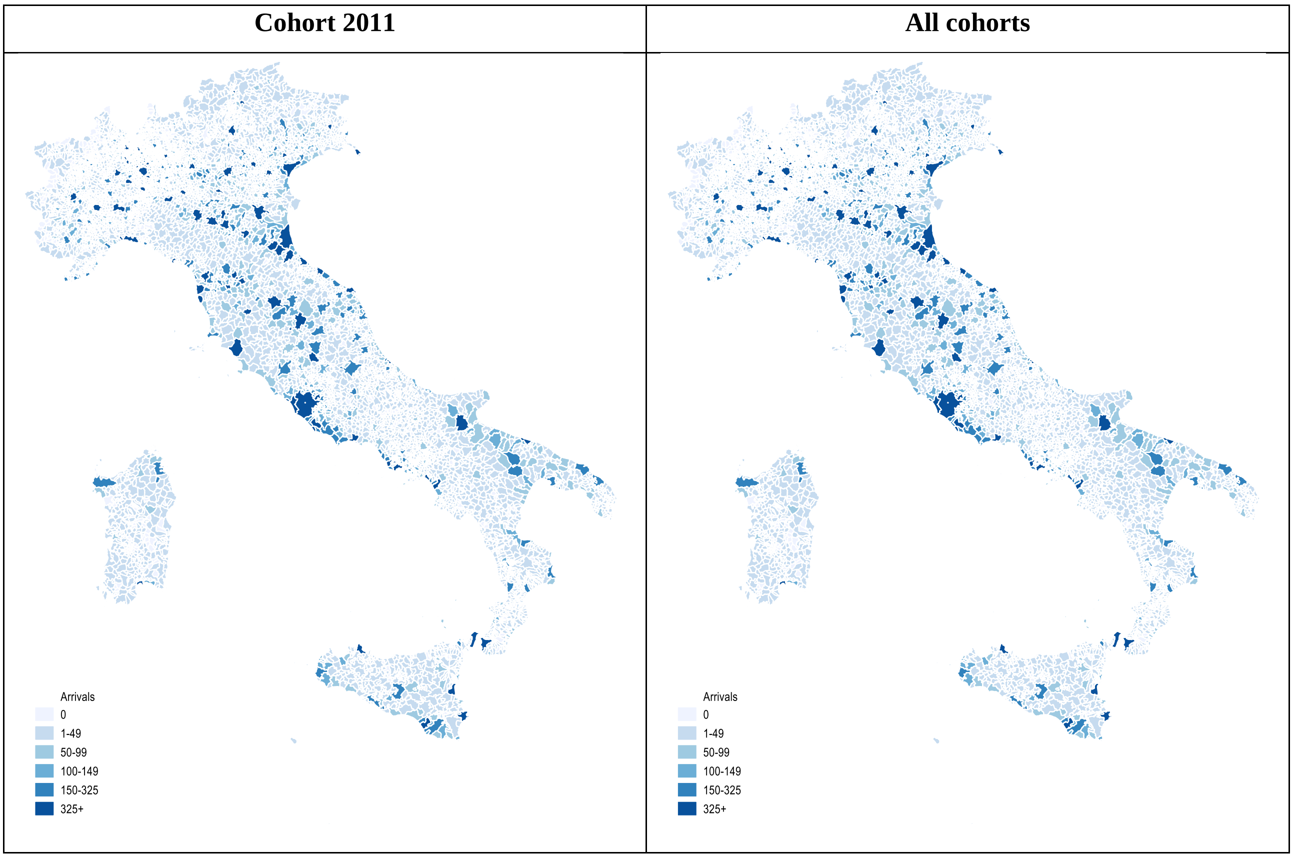


**Figure C3. Average number of departures from Italian provinces (whole sample)**

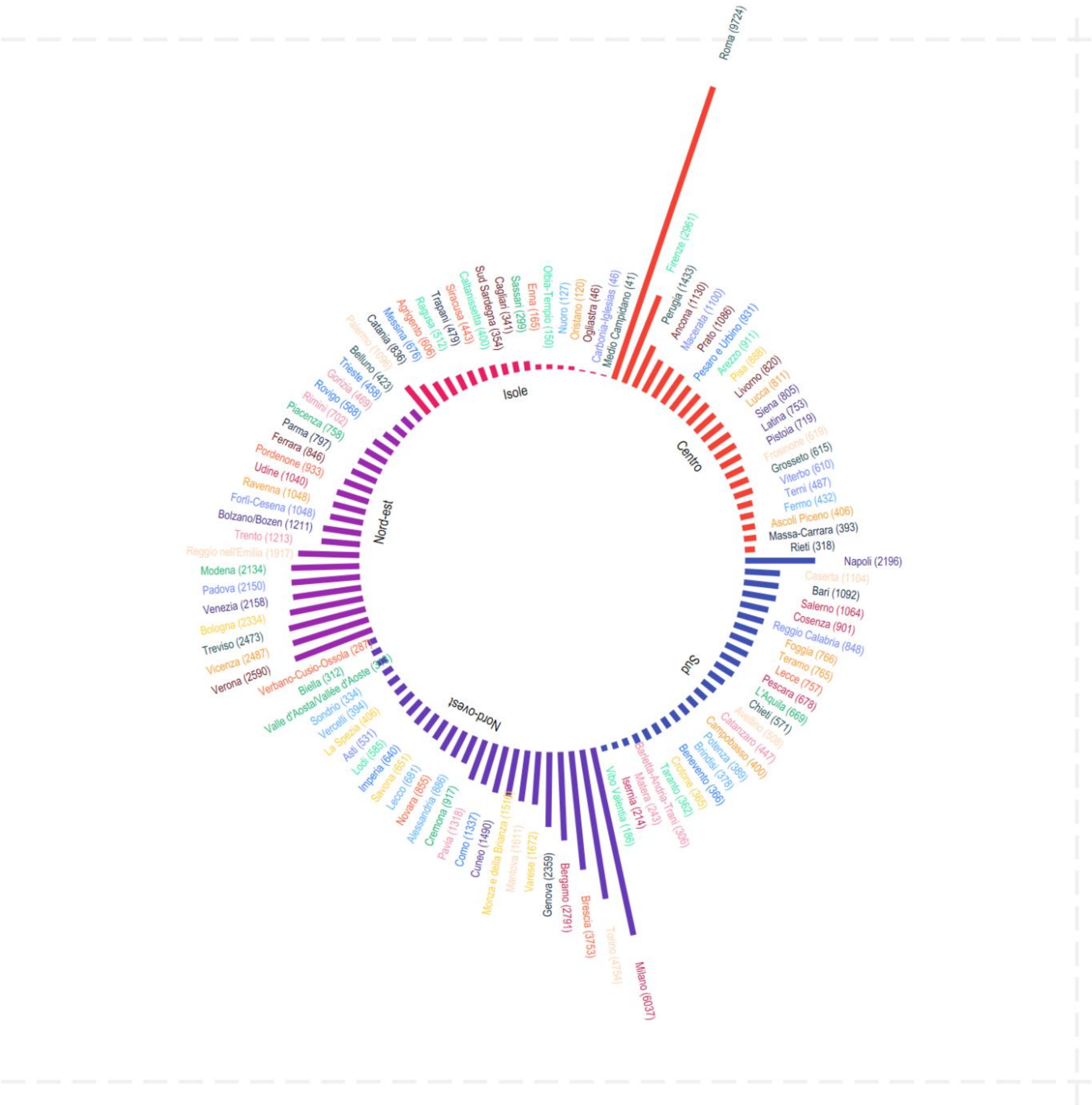
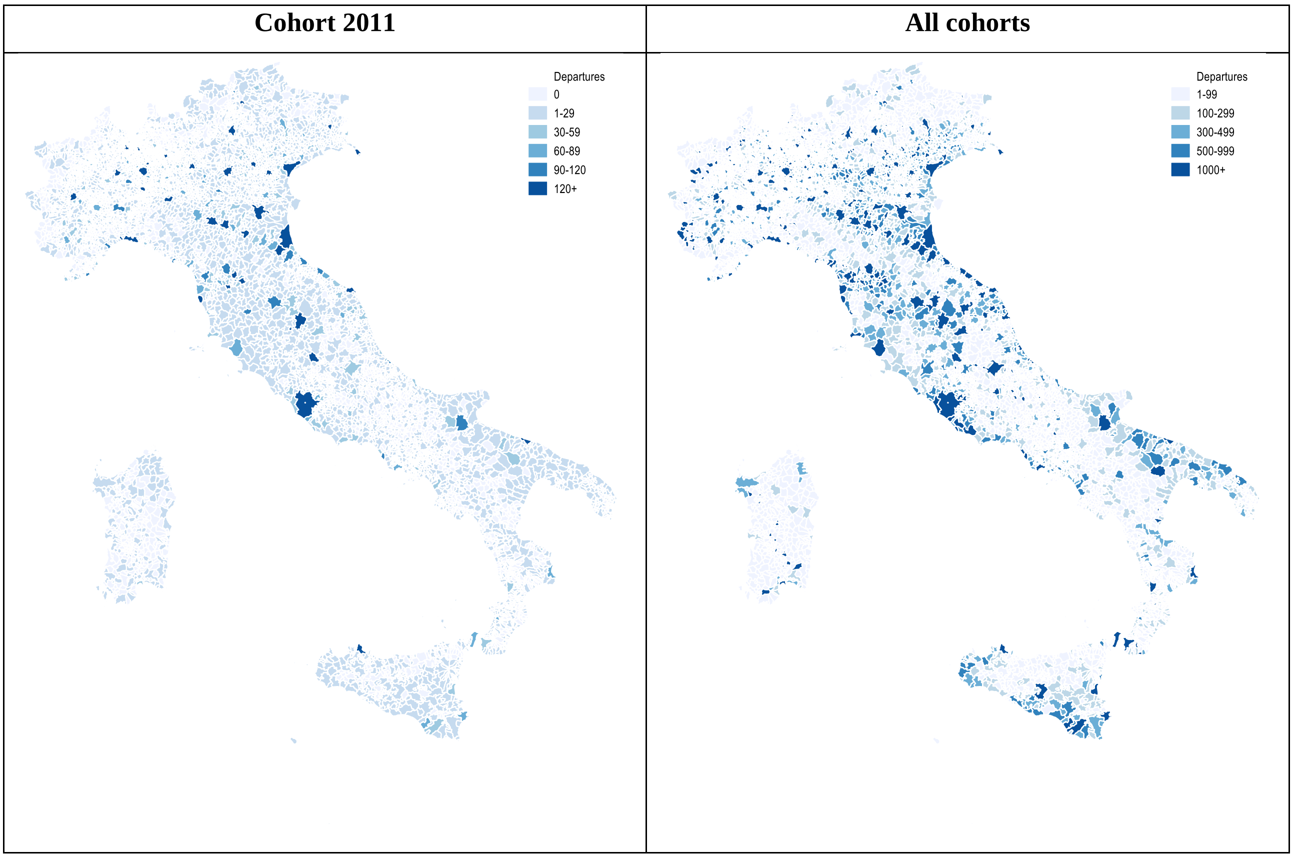


**Figure C4. Number of departures by municipality**



## Arrivals by regions and additional statistics

**Table C1. Share of arrivals by region**

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Panel A. Arrivals in 2011** | | | | | | | |
| Abruzzo | 0.0198 | 0.0258 | 0.0319 | 0.0222 | 0.0190 | 0.0121 | 0.0143 |
| Basilicata | 0.0058 | 0.0046 | 0.0134 | 0.0042 | 0.0013 | 0.0022 | 0.0034 |
| Calabria | 0.0244 | 0.0060 | 0.0390 | 0.0354 | 0.0109 | 0.0229 | 0.0197 |
| Campania | 0.0659 | 0.0230 | 0.0498 | 0.0767 | 0.0525 | 0.2562 | 0.0580 |
| Emilia-Romagna | 0.0978 | 0.1333 | 0.0782 | 0.1389 | 0.1233 | 0.1031 | 0.0961 |
| Friuli-Venezia Giulia | 0.0180 | 0.0192 | 0.0200 | 0.0079 | 0.0132 | 0.0128 | 0.0191 |
| Lazio | 0.1303 | 0.0756 | 0.1447 | 0.0361 | 0.0674 | 0.1139 | 0.1466 |
| Liguria | 0.0293 | 0.0580 | 0.0236 | 0.0432 | 0.0191 | 0.0179 | 0.0299 |
| Lombardia | 0.2238 | 0.2074 | 0.1267 | 0.2042 | 0.2179 | 0.2690 | 0.2678 |
| Marche | 0.0283 | 0.0349 | 0.0279 | 0.0246 | 0.0696 | 0.0197 | 0.0252 |
| Molise | 0.0028 | 0.0011 | 0.0054 | 0.0027 | 0.0008 | 0.0011 | 0.0021 |
| Piemonte | 0.0719 | 0.0893 | 0.0983 | 0.1328 | 0.0731 | 0.0305 | 0.0558 |
| Puglia | 0.0294 | 0.0456 | 0.0454 | 0.0248 | 0.0163 | 0.0088 | 0.0244 |
| Sardegna | 0.0110 | 0.0020 | 0.0199 | 0.0088 | 0.0093 | 0.0059 | 0.0086 |
| Sicilia | 0.0437 | 0.0193 | 0.0633 | 0.0453 | 0.0250 | 0.0102 | 0.0420 |
| Toscana | 0.0734 | 0.1312 | 0.0816 | 0.0596 | 0.1324 | 0.0285 | 0.0646 |
| Trentino-Alto Adige/ Südtirol | 0.0182 | 0.0263 | 0.0136 | 0.0138 | 0.0078 | 0.0114 | 0.0217 |
| Umbria | 0.0168 | 0.0277 | 0.0240 | 0.0188 | 0.0148 | 0.0124 | 0.0130 |
| Valle d'Aosta/ Vallée d'Aoste | 0.0018 | 0.0025 | 0.0032 | 0.0042 | 0.0006 | 0.0007 | 0.0010 |
| Veneto | 0.0878 | 0.0673 | 0.0900 | 0.0956 | 0.1256 | 0.0607 | 0.0869 |
| **Panel B. from 2011 to 2022** | | | | | | | |
| Abruzzo | 0.0211 | 0.0264 | 0.0253 | 0.0240 | 0.0154 | 0.0151 | 0.0201 |
| Basilicata | 0.0076 | 0.0101 | 0.0117 | 0.0050 | 0.0028 | 0.0033 | 0.0070 |
| Calabria | 0.0307 | 0.0089 | 0.0382 | 0.0378 | 0.0079 | 0.0253 | 0.0322 |
| Campania | 0.0653 | 0.0275 | 0.0476 | 0.0729 | 0.0671 | 0.1913 | 0.0664 |
| Emilia-Romagna | 0.0929 | 0.1301 | 0.0889 | 0.1305 | 0.1025 | 0.1248 | 0.0851 |
| Friuli-Venezia Giulia | 0.0211 | 0.0199 | 0.0247 | 0.0080 | 0.0105 | 0.0200 | 0.0221 |
| Lazio | 0.1177 | 0.0553 | 0.1399 | 0.0365 | 0.0751 | 0.1094 | 0.1278 |
| Liguria | 0.0298 | 0.0550 | 0.0224 | 0.0396 | 0.0172 | 0.0236 | 0.0297 |
| Lombardia | 0.2000 | 0.1823 | 0.1356 | 0.1941 | 0.2257 | 0.2342 | 0.2152 |
| Marche | 0.0276 | 0.0368 | 0.0274 | 0.0236 | 0.0469 | 0.0202 | 0.0262 |
| Molise | 0.0058 | 0.0019 | 0.0041 | 0.0035 | 0.0007 | 0.0023 | 0.0073 |
| Piemonte | 0.0760 | 0.0926 | 0.0897 | 0.1319 | 0.0609 | 0.0458 | 0.0691 |
| Puglia | 0.0389 | 0.0502 | 0.0486 | 0.0324 | 0.0168 | 0.0138 | 0.0373 |
| Sardegna | 0.0141 | 0.0023 | 0.0180 | 0.0097 | 0.0071 | 0.0130 | 0.0151 |
| Sicilia | 0.0529 | 0.0335 | 0.0620 | 0.0428 | 0.0149 | 0.01376 | 0.0578 |
| Toscana | 0.0779 | 0.1377 | 0.0731 | 0.0624 | 0.2039 | 0.0429 | 0.0676 |
| Trentino-Alto Adige/ Südtirol | 0.0194 | 0.0333 | 0.0147 | 0.0154 | 0.0061 | 0.0173 | 0.0207 |
| Umbria | 0.0160 | 0.0201 | 0.0214 | 0.0172 | 0.0123 | 0.0154 | 0.0145 |
| Valle d'Aosta/ Vallée d'Aoste | 0.0019 | 0.0026 | 0.0027 | 0.0052 | 0.0005 | 0.0011 | 0.0014 |
| Veneto | 0.0842 | 0.0736 | 0.1041 | 0.1076 | 0.1058 | 0.0673 | 0.0842 |

**Table C2. Share of arrivals (in 2011) - Sex**

| Panel A. By major socio-economic region (NUTS1) | | | | | |
|---|---|---|---|---|---|
| | **Males** | **Females** | **Males** | **Females** | ***Tot.*** |
| North-east | 0.2118 | 0.2300 | 0.4342 | 0.5658 | ***1*** |
| North-west | 0.3318 | 0.3225 | 0.4616 | 0.5384 | ***1*** |
| Center | 0.2496 | 0.2481 | 0.4560 | 0.5440 | ***1*** |
| South | 0.1484 | 0.1478 | 0.4556 | 0.5444 | ***1*** |
| Islands | 0.0585 | 0.0515 | 0.4860 | 0.5140 | ***1*** |
| ***Tot*** | ***1*** | ***1*** | | | |
| **Panel B. By region** | | | | | |
| Abruzzo | 0.0191 | 0.0203 | | | |
| Basilicata | 0.0053 | 0.0062 | | | |
| Calabria | 0.0263 | 0.0228 | | | |
| Campania | 0.0669 | 0.0651 | | | |
| Emilia-Romagna | 0.0963 | 0.0990 | | | |
| Friuli-Venezia Giulia | 0.0164 | 0.0194 | | | |
| Lazio | 0.1407 | 0.1216 | | | |
| Liguria | 0.0289 | 0.0296 | | | |
| Lombardia | 0.2314 | 0.2174 | | | |
| Marche | 0.0255 | 0.0307 | | | |
| Molise | 0.0025 | 0.0030 | | | |
| Piemonte | 0.0699 | 0.0736 | | | |
| Puglia | 0.0283 | 0.0303 | | | |
| Sardegna | 0.0091 | 0.0125 | | | |
| Sicilia | 0.0493 | 0.0391 | | | |
| Toscana | 0.0687 | 0.0772 | | | |
| Trentino-Alto Adige/ Südtirol | 0.0173 | 0.0189 | | | |
| Umbria | 0.0147 | 0.0186 | | | |
| Valle d'Aosta/ Vallée d'Aoste | 0.0016 | 0.0019 | | | |
| Veneto | 0.0818 | 0.0928 | | | |
| ***Tot*** | ***1*** | ***1*** | | | |

**Table C3. Share of arrivals (in 2011) – Age (entry) categories**

| **Panel A. By major socio-economic region (NUTS1)** | | | |
|---|---|---|---|
| | **15-29** | **30-49** | **50-69** |
| North-east | 0.2347 | 0.2052 | 0.2377 |
| North-west | 0.3331 | 0.3281 | 0.2973 |
| Center | 0.2422 | 0.2515 | 0.2627 |
| South | 0.1354 | 0.1583 | 0.1554 |
| Islands | 0.0546 | 0.0568 | 0.0469 |
| **Panel B. By region** | | | |
| Abruzzo | 0.0202 | 0.0191 | 0.0209 |
| Basilicata | 0.0059 | 0.0058 | 0.0053 |
| Calabria | 0.0248 | 0.0252 | 0.0198 |
| Campania | 0.0523 | 0.0762 | 0.0770 |
| Emilia-Romagna | 0.1024 | 0.0903 | 0.1096 |
| Friuli-Venezia Giulia | 0.0187 | 0.0165 | 0.0214 |
| Lazio | 0.1241 | 0.1390 | 0.1195 |
| Liguria | 0.0310 | 0.0272 | 0.0307 |
| Lombardia | 0.2247 | 0.2315 | 0.1901 |
| Marche | 0.0281 | 0.0269 | 0.0347 |
| Molise | 0.0030 | 0.0026 | 0.0029 |
| Piemonte | 0.0755 | 0.0678 | 0.0746 |
| Puglia | 0.0293 | 0.0295 | 0.0295 |
| Sardegna | 0.0098 | 0.0117 | 0.0125 |
| Sicilia | 0.0448 | 0.0451 | 0.0344 |
| Toscana | 0.0726 | 0.0707 | 0.0865 |
| Trentino-Alto Adige/ Südtirol | 0.0192 | 0.0169 | 0.0194 |
| Umbria | 0.0175 | 0.0149 | 0.0220 |
| Valle d'Aosta/ Vallée d'Aoste | 0.0020 | 0.0016 | 0.0019 |
| Veneto | 0.0943 | 0.0816 | 0.0873 |

**Table C4. Italian wages (in €) at provincial level**

| | Total sample | Albanian | Romanian | Moroccan | Chinese | Ukrainian |
|---|---|---|---|---|---|---|
| North-east | 36402 | 35831 | 36197 | 36473 | 36302 | 36096 |
| North-west | 40791 | 39284 | 39058 | 38972 | 40440 | 41982 |
| Center | 37579 | 36173 | 36905 | 35505 | 35936 | 38427 |
| South | 30254 | 30617 | 29243 | 29880 | 31143 | 31463 |
| Islands | 30178 | 29880 | 29511 | 30408 | 31272 | 30885 |

**Table C5. Italian wages (in €) at LLM level**

| | Total sample | Albanian | Romanian | Moroccan | Chinese | Ukrainian |
|---|---|---|---|---|---|---|
| North-east | 36402 | 35831 | 36197 | 36473 | 36302 | 36096 |
| North-west | 40791 | 39284 | 39058 | 38972 | 40440 | 41982 |
| Center | 37579 | 36173 | 36905 | 35505 | 35936 | 38427 |
| South | 30254 | 30617 | 29243 | 29880 | 31143 | 31463 |
| Islands | 30178 | 29880 | 29511 | 30408 | 31272 | 30885 |

**Table C6. Consumer price space index (2021) – Base 100=Italy**

| **Region** | **Value** |
|---|---|
| Abruzzo | 93.75 |
| Basilicata | 94.72 |
| Calabria | 98.25 |
| Campania | 90.47 |
| Emilia-Romagna | 103.80 |
| Friuli-Venezia Giulia | 102.67 |
| Lazio | 99.59 |
| Liguria | 104.77 |
| Lombardia | 105.1 |
| Marche | 98.95 |
| Molise | 95.80 |
| Piemonte | 101.49 |
| Puglia | 97.58 |
| Sardegna | 102.15 |
| Sicilia | 99.16 |
| Toscana | 102.46 |
| Trentino | 101.57 |
| Alto Adige | 105.33 |
| Umbria | 98.85 |
| Valle d'Aosta/ Vallée d'Aoste | 102.29 |
| Veneto | 103.00 |

**Source: ISTAT**

### Multinomial logit: initial macro-area choice and co-national networks

Here we report the baseline-category multinomial logit of the macro-area of initial settlement (reference: North-west). The baseline nationality is Albania. Continuous covariates are standardized (z-scores), so coefficients refer to a one-standard-deviation (+1 SD) change. The model includes: (i) nationality main effects (how each community's odds of choosing a macro-area differ from Albanians, ceteris paribus), (ii) a common network slope (for Albanians), and (iii) nationality×network interactions. Summed together, the baseline slope and each interaction yield the total network effect for that nationality in that macro-area (reported via Wald tests)[13].

[13] Coefficients in Table C7 are log-odds (β). Exponentiating β gives the odds ratio (OR). For interactions, interpretation is additive in β and multiplicative in OR. The network effect for a given nationality×area equals: (βnetwork for Albanians) + (β{nat×network}); the corresponding OR is exp of that sum. This is what Panel B in Table C8 tests via Wald restrictions (reporting OR and 95% CIs).

**Table C7 — Multinomial logit: log-odds coefficients (all macro-areas vs North-west).**

| | North-east | Center | South | Islands |
|---|---|---|---|---|
| Constant | -0.013 | 0.429*** | 0.022 | 0.262*** |
| | (0.043) | (0.035) | (0.033) | (0.028) |
| China (vs Albania) | 0.101*** | 0.464*** | 0.056*** | 0.018** |
| | (0.037) | (0.031) | (0.018) | (0.009) |
| Morocco (vs Albania) | 0.341*** | -1.201*** | -0.110*** | -0.574*** |
| | (0.012) | (0.014) | (0.026) | (0.036) |
| Romania (vs Albania) | -0.456*** | -0.051 | -0.154*** | 0.298*** |
| | (0.039) | (0.033) | (0.040) | (0.045) |
| Ukraine (vs Albania) | 0.284*** | -0.547*** | 0.618*** | -0.407*** |
| | (0.039) | (0.039) | (0.029) | (0.046) |
| Co-national network (z) | 0.606*** | -0.009 | 0.003 | 1.075*** |
| | (0.032) | (0.018) | (0.009) | (0.012) |
| China × network (z) | 0.138** | -5.804*** | 0.674*** | 0.818*** |
| | (0.061) | (0.116) | (0.103) | (0.093) |
| Morocco × network (z) | 1.059*** | 0.156 | -0.696*** | 0.005 |
| | (0.086) | (0.124) | (0.099) | (0.036) |
| Romania × network (z) | -0.006 | -0.487*** | -3.020*** | 2.457*** |
| | (0.019) | (0.028) | (0.025) | (0.060) |
| Ukraine × network (z) | 0.830*** | 0.928*** | -0.039 | 0.719*** |
| | (0.115) | (0.119) | (0.103) | (0.142) |
| Age (z) | -0.043 | 0.498*** | -2.428*** | 0.251*** |
| | (0.034) | (0.044) | (0.063) | (0.055) |
| Average rent in Italian | 0.066 | 0.528*** | 1.744*** | -0.167*** |
| LLMs (z) | (0.052) | (0.043) | (0.050) | (0.053) |
| Female | -1.341*** | -0.617*** | -0.433*** | -0.608*** |
| | (0.015) | (0.027) | (0.038) | (0.044) |
| Income per worker (log) in | -0.081*** | -0.005 | -0.065*** | -3.014*** |
| Italian LLMs (z) | (0.024) | (0.012) | (0.016) | (0.021) |
| Urban | 1.183*** | 0.152** | -0.465*** | -0.223*** |
| | (0.034) | (0.064) | (0.070) | (0.057) |

*Notes: the Table reports the log-odds coefficients (β) with standard errors beneath; asterisks denote significance (*** p<0.01, ** p<0.05, * p<0.10).*

Signs and significance of the coefficients are consistent with the descriptive patterns described in the main text: wages and urbanization generally favor northern areas, while higher housing costs dampen relative odds, especially where rents are steep. “Nationality × network” coefficients foreshadow some heterogeneity: for Eastern-European groups, network responsiveness tends to be stronger in the Center and North-east, whereas effects are muted or negative in the South and Islands for some communities. However, the magnitude and significance of the total network effect for each nationality × area—i.e., the sum of the baseline slope and the nationality-specific deviation—are reported in Panel B of Table C8 via Wald tests.

**Table C8 - Multinomial logit: nationality main effects (with all the controls held fixed).**

| Area | China | Morocco | Romania | Ukraine |
|---|---|---|---|---|
| Panel A: Relative odds of choosing area (vs North-west, reference: Albania) | | | | |
| North-east | 1.11*** | 1.41*** | 0.63*** | 1.33*** |
| | [1.03, 1.19] | [1.37, 1.44] | [0.59, 0.68] | [1.23, 1.43] |
| Center | 1.59*** | 0.30*** | 0.95 | 0.58*** |
| | [1.50, 1.69] | [0.29, 0.31] | [0.89, 1.01] | [0.54, 0.62] |
| South | 1.06*** | 0.90*** | 0.86*** | 1.86*** |
| | [1.02, 1.09] | [0.85, 0.94] | [0.79, 0.93] | [1.75, 1.97] |
| Islands | 1.02** | 0.56*** | 1.35*** | 0.67*** |
| | [1.00, 1.04] | [0.52, 0.60] | [1.23, 1.47] | [0.61, 0.73] |
| Panel B: Total network effect (Wald tests) | | | | |
| North-east | 1.59*** | 0.90* | 1.01 | 1.51*** |
| | [1.49, 1.69] | [0.79, 1.01] | [0.84, 1.21] | [1.40, 1.63] |
| Center | 1.83*** | 1.19 | 1.00 | 1.76*** |
| | [1.77, 1.90] | [0.95, 1.49] | [0.78, 1.28] | [1.64, 1.87] |
| South | 0.85*** | 0.99 | 0.53*** | 0.68*** |
| | [0.83, 0.86] | [0.80, 1.21] | [0.44, 0.65] | [0.64, 0.71] |
| Islands | 0.50*** | 1.14 | 1.26*** | 0.48*** |
| | [0.49, 0.51] | [0.95, 1.37] | [1.17, 1.36] | [0.43, 0.54] |

*Notes: the Table reports the odds ratios, with 95% confidence intervals beneath; asterisks denote significance (*** p<0.01, ** p<0.05, * p<0.10).*

Panel A summarizes nationality main effects as odds ratios: values above one indicate a stronger inclination toward that macro-area versus North-west (holding the network at its mean), values below one indicate a relative avoidance. Patterns are consistent with the descriptive shares in the main text: several groups are more likely to settle in the North-east and the Center, while Ukrainians display a comparatively stronger tilt toward the South.

Panel B reports the total co-national network effect per +1 SD by nationality×area. Where the OR exceeds one (and is significant), denser local communities pull new arrivals of the same nationality toward that macro-area; where it is below one, networks are associated with a relative shift away (or a reinforced preference for North-west). Heterogeneity is substantial. Eastern-European groups (Ukrainians, Romanians, Albanians) typically exhibit positive and significant network pull in the Center and North-east. For the South and Islands, network associations are often neutral or negative for some groups (e.g., Albanians, Romanians). For Chinese, effects are generally muted, suggesting other covariates (wages, urban structure) dominate area selection. For Moroccans the effect is geographically specific (e.g., pull in the Islands; push in the South).

## Appendix D. Robustness: results for other cohorts

### D.1. Cohort of 2013

We now replicate the analysis for the cohort of individuals arrived in Italy in 2013. Table D1 reports the results of the parametric survival model. The findings are largely consistent with those reported in Table 10, with two notable exceptions: the effects of the average rent and living in an urban area, are now positive and no longer statistically significant.

**Table D1. Estimation results – Parametric survival analysis (2013)**

| | Total sample |
|---|---|
| Female | 0.840*** |
| | (0.0272) |
| Age of entry | -0.00252** |
| | (0.00109) |
| Political Instability | -0.0278 |
| | (0.0409) |
| Relative costs of living | -0.344*** |
| | (0.0376) |
| Log differential between Italian and origin country's GDP | 0.260*** |
| | (0.0332) |
| Income per worker in Italian LLMs (log) | 1.582*** |
| | (0.0938) |
| Average rent in Italian LLMs | 0.0104 |
| | (0.00938) |
| Urban | 0.0372 |
| | (0.0385) |
| Major socio-economic region (NUTS1) of arrival: Center | 0.263*** |
| | (0.0332) |
| Major socio-economic region (NUTS1) of arrival: Islands | 1.257*** |
| | (0.0679) |
| Major socio-economic region (NUTS1) of arrival: South | 1.151*** |
| | (0.0496) |
| *Observations* | *208,410* |

Turning to the quantile regression model, the results from reported in Table D2 reveal some differences compared to the estimates in Table 12 in the main text.

**Table D2. Estimation results – Quantile regression analysis (2013 cohort)**

| | (1)<br>10% | (2)<br>25% | (3)<br>50% | (4)<br>75% | (5)<br>95% |
|---|---|---|---|---|---|
| Female | 2.174*** | -0.012 | -0.493*** | -1.308*** | -3.579*** |
| | (0.082) | (0.031) | (0.036) | (0.023) | (0.012) |
| Age of entry | -0.0002 | -0.0002 | 0.0003 | 0.0009 | 0.0003 |
| | (0.003) | (0.001) | (0.001) | (0.0010) | (0.0006) |
| Political instability | -0.020 | -0.002 | -0.030 | -0.078** | -0.173*** |
| | (0.135) | (0.044) | (0.052) | (0.033) | (0.016) |
| Average rent in Italian LLMs | 0.002 | 0.030** | 0.132*** | 0.311*** | 0.779*** |
| | (0.026) | (0.011) | (0.013) | (0.008) | (0.004) |
| Relative cost of living | -0.061 | -0.127** | -0.405*** | -0.853*** | -1.697*** |
| | (0.110) | (0.041) | (0.054) | (0.044) | (0.032) |
| Income per worker in Italian | 0.454 | 0.497*** | 1.441*** | 2.989*** | 5.416*** |
| LLMs (log) | (0.309) | (0.099) | (0.134) | (0.114) | (0.092) |
| Log differential between Italian | -0.040 | 0.088*** | 0.166*** | 0.322*** | 0.787*** |
| and origin country's GDP | (0.078) | (0.025) | (0.030) | (0.019) | (0.010) |
| Urban | -0.126 | 0.095** | 0.373*** | 0.793*** | 1.245*** |
| | (0.107) | (0.044) | (0.057) | (0.046) | (0.030) |
| Major socio-economic region | 0.002 | 0.040 | -0.025 | -0.138*** | -0.521*** |
| (NUTS1) of arrival: Center | (0.114) | (0.038) | (0.043) | (0.026) | (0.011) |
| Major socio-economic region | 1.396*** | 0.166*** | 0.200*** | 0.289*** | 0.435*** |
| (NUTS1) of arrival: South | (0.158) | (0.047) | (0.059) | (0.041) | (0.025) |
| Major socio-economic region | 1.691*** | 0.038 | -0.268** | -0.649*** | -0.781*** |
| (NUTS1) of arrival: Islands | (0.223) | (0.062) | (0.095) | (0.092) | (0.059) |
| Constant | -6.638** | 4.778*** | -3.974** | -18.230*** | -38.150*** |
| | (3.212) | (1.023) | (1.402) | (1.194) | (0.969) |

*Observations: 208,410*

*Notes: Standard errors in parentheses, * p<0.10 ** p<0.05 *** p<0.01. The number of observations remains the same across the five columns as this type of regressions takes into consideration censored data exploiting information from all observations (censored and uncensored) to obtain a more robust estimate.*

Starting with gender, the 2013 results confirm a strong heterogeneity across quantiles, though with even more pronounced effects than in 2011. This indicates that women were substantially more likely to exit quickly among the cohort of arrivals in 2013. At higher quantiles, the effect remains strongly negative, suggesting that long-term stayers were more likely to be male, and this gender gap widened slightly over time.

In terms of age at entry, the 2011 model showed a clear non-linear pattern, with negative coefficients at low quantiles and positive at higher ones. In contrast, the 2013 estimates are very close to zero and insignificant across all quantiles, suggesting a diminished role of age as a predictor of migration duration.

Also for the political instability the effects are insignificant and substantially small, especially across the lower quantiles, while the coefficients at the higher quantiles (75th and 95th) show a growing negative impact, in contrast with the (positive) effects from the previous cohort, but with magnitudes substantially smaller than those in 2011. This suggests that geopolitical push factors may have become somewhat less important in explaining especially short stays, possibly due to tighter migration policy enforcement.

Relative cost of living maintains a consistent negative effect in 2013, confirming and amplifying the pattern seen in 2011. This reinforces the idea that high local costs discourage long-term permanence.

As for the (log) differential between Italian and origin country's GDP, its positive effect becomes weaker in 2013, especially at higher quantiles. This indicates that origin-country wealth still plays a role in shaping return or onward migration, but its magnitude may have declined over time.

Local income per worker continues to be a key determinant of migrant retention. In 2013, the effects are even larger than in 2011, underlining a strong and increasing economic pull effect.

Similarly, the effect of average rent becomes even stronger and more consistent in 2013, particularly across the higher quantiles. This confirms that high-rent areas, often associated with stronger labor markets, are increasingly associated with longer migrant stays, particularly for the upper part of the distribution.

Urban residence shows a striking shift. In 2011, the urban dummy had a consistent negative effect on stay duration across all quantiles. In 2013, however, the sign becomes positive at the 25th and remains positive and significant through the 95th. This reversal suggests that urban areas, perhaps due to improved services or labor access, became more conducive to longer-term migrant retention between 2011 and 2013.

Looking at macro-area effects, the Centre areas' coefficients shift from positive (or insignificant) in 2011 to slightly negative in 2013 at higher quantiles, indicating a reduced retention capacity on long stays with respect to the North. The South, on the other hand, shows modestly positive and consistent effects in 2013, contrasting with the non-significant or negative coefficients at upper quantiles in 2011. This suggests improved retention dynamics in the South. However, the Islands show the opposite trend: while they had large positive effects in 2011, their coefficients turn negative at higher quantiles in 2013, suggesting a decline in long-term settlement in insular areas.

In summary, the comparative analysis between 2011 and 2013 shows important evolutions in the factors influencing migrant retention in Italy. Economic variables remain central, with stronger positive effects over time. However, the influence of geopolitical origin factors and macro-regional dummies shows more variation, with signs of regional rebalancing and changes in urban versus rural dynamics. Urban areas, in particular, appear to have transitioned from spaces of circulation in 2011 to more stable destinations in 2013.

Lastly, the results of the parametric model for the five nationalities, presented in Table D3, largely mirror those in Table 13. However, there are a few exceptions. Arriving in Central Italy is now positive and significant. For Albanians, living in an urban area and relative cost of living are now negative and significant. For Chinese nationals, living in urban areas is now positive and significant, indicating that staying in urban areas increases the median time to the event (leaving Italy); at the same time, the effect of political instability is now negative and significant. For Ukrainians, age now shows a positive and significant effect on migration duration. Some coefficients are not significant anymore for this cohort.

**Table D3. Estimation results – Parametric survival analysis (2013)**

| | (1)<br>Romania | (2)<br>Albania | (3)<br>Morocco | (4)<br>China | (5)<br>Ukraine |
|---|---|---|---|---|---|
| Female | 0.210*** | 0.496*** | 0.858*** | 0.298*** | 0.803*** |
| | (0.0148) | (0.0452) | (0.0461) | (0.0498) | (0.0838) |
| Age of entry | -0.00502*** | -0.0104*** | -0.00884*** | -0.0101*** | 0.00599** |
| | (0.000564) | (0.00150) | (0.00205) | (0.00219) | (0.00243) |
| Political instability | 3.876*** | -4.076*** | -6.730*** | -14.98*** | -3.092*** |
| | (0.490) | (0.867) | (0.882) | (0.926) | (0.227) |
| Relative cost of living | -30.52*** | -413.9*** | -86.88*** | -73.14*** | -39.73*** |
| | (0.936) | (69.10) | (2.469) | (5.439) | (1.448) |
| Log differential between Italian and origin country’s GDP | -0.0732 | -8.153*** | 15.34*** | -2.591*** | -0.441 |
| | (0.337) | (0.959) | (1.427) | (0.751) | (0.657) |
| Income per worker in Italian LLMs | 0.321*** | 0.0150 | 0.613*** | 1.761*** | 1.524*** |
| | (0.0524) | (0.183) | (0.176) | (0.196) | (0.260) |
| Average rent (LLM) | -0.00762 | -0.0250* | -0.0262 | 0.0763*** | -0.0362 |
| | (0.00556) | (0.0148) | (0.0177) | (0.0165) | (0.0237) |
| Urban | 0.212*** | -0.071 | -0.459*** | 0.225** | 0.585*** |
| | (0.0177) | (0.0655) | (0.0617) | (0.0953) | (0.0959) |
| Major socio-economic region (NUTS1) of arrival: Center | 0.0911*** | -0.152*** | 0.141** | -0.794*** | 0.00840 |
| | (0.0171) | (0.0530) | (0.0673) | (0.0627) | (0.0870) |
| Major socio-economic region (NUTS1) of arrival: Islands | 0.321*** | 0.568*** | 1.359*** | 3.217*** | 2.019*** |
| | (0.0329) | (0.187) | (0.131) | (0.247) | (0.309) |
| Major socio-economic region (NUTS1) of arrival: South | 0.449*** | -0.00861 | 1.080*** | 1.206*** | 0.925*** |
| | (0.0268) | (0.0885) | (0.0849) | (0.106) | (0.112) |
| *Observations:* | *413,376* | *95,289* | *110,535* | *123,358* | *105,643* |

*Notes: Standard errors in parentheses, * p<0.10 ** p<0.05 *** p<0.01*

### D.2. Cohort of 2016

Lastly, this section of the appendix presents additional results based on the cohort of individuals who entered Italy in 2016. It is worth mentioning that, obviously, this is the cohort for which the observation period is shortest, so the results may be less representative than previous cohorts.

Table D4 displays the outcomes of the basic parametric survival analysis. Compared to the previous cohorts, the only two covariates that change effect on the duration are the political instability and living in an urban area.

**Table D4. Estimation results – Parametric survival analysis (2016)**

| | Total sample |
|---|---|
| Female | 0.667*** |
| | (0.0236) |
| Age of entry | -0.00899*** |
| | (0.000915) |
| Political Instability | -0.181*** |
| | (0.0262) |
| Relative costs of living | -0.165*** |
| | (0.0369) |
| Log differential between Italian and origin country GDP | 0.192*** |
| | (0.0179) |
| Income per worker in Italian LLMs (log) | 0.633*** |
| | (0.0852) |
| Average rent in Italian LLMs | 0.0336*** |
| | (0.00846) |
| Urban | 0.0913*** |
| | (0.0295) |
| Major socio-economic region (NUTS1) of arrival: Center | 0.257*** |
| | (0.0278) |
| Major socio-economic region (NUTS1) of arrival: Islands | 0.243*** |
| | (0.0524) |
| Major socio-economic region (NUTS1) of arrival: South | 0.223*** |
| | (0.0401) |
| *Observations* | *195,051* |

Table D5 presents the results of the quantile regressions for the last cohort under analysis.

**Table D5. Estimation results – Quantile regression analysis (2016 cohort)**

| | (1)<br>10% | (2)<br>25% | (3)<br>50% | (4)<br>75% | (5)<br>95% |
|---|---|---|---|---|---|
| Female | 2.821*** | 0.178*** | 0.354*** | 0.668*** | 1.356*** |
| | (0.059) | (0.027) | (0.023) | (0.021) | (0.022) |
| Age of entry | -0.014*** | 0.001 | 0.009*** | 0.021*** | 0.052*** |
| | (0.002) | (0.001) | (0.001) | (0.001) | (0.001) |
| Political instability | -0.564*** | 0.058* | 0.279*** | 0.725*** | 2.679*** |
| | (0.067) | (0.028) | (0.023) | (0.019) | (0.015) |
| Average rent in Italian LLMs | 0.007 | 0.045*** | 0.144*** | 0.338*** | 0.918*** |
| | (0.017) | (0.010) | (0.009) | (0.007) | (0.006) |
| Relative cost of living | -0.400*** | -0.271*** | -0.864*** | -1.704*** | -2.480*** |
| | (0.048) | (0.040) | (0.042) | (0.046) | (0.052) |
| Income per worker in Italian LLMs (log) | 1.316*** | 0.200* | 0.419*** | 0.669*** | 0.765*** |
| | (0.151) | (0.094) | (0.084) | (0.082) | (0.092) |
| Log differential between Italian and origin country's GDP | -0.163*** | 0.120*** | 0.322*** | 0.672*** | 1.823*** |
| | (0.042) | (0.021) | (0.018) | (0.016) | (0.017) |
| Urban | 0.016 | 0.288*** | 0.936*** | 1.803*** | 1.843*** |
| | (0.044) | (0.035) | (0.033) | (0.035) | (0.037) |
| Major socio-economic region (NUTS1) of arrival: Center | 0.367*** | 0.200*** | 0.598*** | 1.476*** | 4.913*** |
| | (0.056) | (0.032) | (0.026) | (0.020) | (0.014) |
| Major socio-economic region (NUTS1) of arrival: South | 0.315*** | -0.103* | -0.449*** | -0.905*** | -0.828*** |
| | (0.067) | (0.043) | (0.038) | (0.038) | (0.040) |
| Major socio-economic region (NUTS1) of arrival: Islands | 0.314*** | 0.132* | 0.309*** | 0.624*** | 1.458*** |
| | (0.100) | (0.062) | (0.051) | (0.045) | (0.041) |
| Constant | -13.880*** | 4.286*** | 0.927 | -3.572*** | -9.922*** |
| | (1.557) | (0.986) | (0.882) | (0.869) | (0.981) |
| *Observations: 195,051* | | | | | |

*Notes: Standard errors in parentheses, * p<0.10 ** p<0.05 *** p<0.01. the number of observations remains the same across the five columns as this type of regressions takes into consideration censored data exploiting information from all observations (censored and uncensored) to obtain a more robust estimate.*

The gender effect continues to be significant across all years. In 2016, the coefficient at the 10th percentile is even larger than in 2013, and much stronger than in 2011, suggesting an increasingly disproportionate share of early exits among female migrants. Interestingly, the effect becomes more positive even at higher quantiles in 2016, in contrast with the negative coefficients observed in 2011 and partially in 2013, pointing to a significant transformation in long-term female migration dynamics. The role of age at entry consolidates its inverted-U pattern. In 2011 and 2013, the coefficients at low quantiles were close to zero or negative and turned positive with increasing quantiles. This pattern is still evident in 2016, but the effects are stronger, especially at higher quantiles, confirming that younger migrants are more likely to stay longer over time. The political instability effect shows an amplification in 2016. This suggests that geopolitical stability at origin plays an increasingly important role in shaping long-term migrant retention. The shift from a weak or negative effect at lower quantiles to a large positive one at upper quantiles is consistent across all years, but most pronounced in 2016. The relative cost of living remains a robust predictor of early exits. Its negative impact is consistent across years and grows more pronounced over time, suggesting affordability constraints are increasingly shaping long-term retention. The (log) differential between Italian and origin country's GDP continues to positively affect retention, especially at higher quantiles. This supports the view that migrants from wealthier origins are less inclined to stay long-term, possibly due to greater opportunities for return or onward migration. Local income per worker continues to be a strong pull factor. While the effect size peaked in 2013, the 2016 coefficient is slightly lower at the 95th quantile. This drop in magnitude might reflect a broader saturation or diminishing marginal effect, but the direction and significance remain robust across all quantiles and years. Average rent maintains its positive association with stay duration. The 2016 coefficients continue the trend seen in 2013 and amplify it. This reflects a reinforcing association between staying longer and residing in economically vibrant (and expensive) areas. Urban residence shows a sharp break in trend. From being negatively associated with stay duration in 2011, the urban dummy becomes strongly positive in 2013 and even more so in 2016, especially at higher quantiles. This transformation suggests that cities have evolved from transitory migration hubs into more stable destinations, likely due to improved infrastructure, services, and economic opportunities. The macro-regional effects also evolve notably. The Centre's influence strengthens substantially, while the South continues to present a mixed profile, with positive effects at low quantiles, but negative effects at the upper quantiles. As for the Islands, their effect rebounds in 2016 to be positive at the 95th quantile, after being quite negative in 2013 and strongly positive in 2011.

Overall, the comparative picture from 2011 to 2016 cohorts reveals a gradual intensification of economic and geopolitical drivers of retention, an increased feminization of long-term stayers, and a growing centrality of urban areas and certain regions (notably the Centre). These findings underscore the need to consider both the temporal and spatial dynamics of migration decisions in policy and planning.

Finally, Table D6 shows that, for this last cohort, the results by nationalities change. For Romanians, the GDP differential has a positive effect; political instability has a negative effect on migration duration; and arriving in the Centre has a positive and significant coefficient. Among Albanians, residing in urban areas and GDP differentials are now linked to a longer time to event. For Moroccans, arriving in Central Italy is associated with a reduction in survival time, meaning they leave Italy sooner compared to immigrants arriving in the North. The same happens now for political instability in the origin country. The coefficients related to the relative cost of living and living in urban areas for the Chinese group are now positively associated with the time to event, indicating a delay in the time to leave Italy. On the opposite, political instability is now linked to negative effect on migration duration. For Ukrainians, the GDP differential is now positively associated with migration duration. Lastly, also in this table, some coefficients that were significant in Table 13 now turn insignificant.

**Table D6. Estimation results – Parametric survival analysis (2016)**

| | (1)<br>Romania | (2)<br>Albania | (3)<br>Morocco | (4)<br>China | (5)<br>Ukraine |
|---|---|---|---|---|---|
| Female | 0.247*** | 0.391*** | 0.663*** | 0.503*** | 0.535*** |
| | (0.0209) | (0.0440) | (0.0507) | (0.0529) | (0.0847) |
| Age of entry | -7.23e-05 | -0.0133*** | -0.0193*** | -0.0239*** | 0.0400*** |
| | (0.000771) | (0.00141) | (0.00200) | (0.00213) | (0.00334) |
| Political instability | -1.635** | -0.0696 | -3.051*** | -17.81*** | -5.202*** |
| | (0.642) | (0.789) | (0.717) | (0.877) | (0.657) |
| Relative cost of living | -35.16*** | -1,543*** | -57.70*** | 41.27*** | -8.503* |
| | (1.230) | (203.9) | (3.104) | (8.577) | (4.441) |
| Log differential between Italian and origin country's GDP | 5.457*** | 2.543 | 27.89*** | -8.742*** | 5.749*** |
| | (0.562) | (1.642) | (1.762) | (0.899) | (0.814) |
| Income per worker in Italian LLMs | 0.396*** | 0.0767 | 0.160 | 2.664*** | 0.348 |
| | (0.0798) | (0.186) | (0.196) | (0.221) | (0.306) |
| Average rent in Italian LLMs | -0.0126 | 0.0207 | -0.0399** | 0.159*** | -0.0659** |
| | (0.00794) | (0.0159) | (0.0199) | (0.0191) | (0.0280) |
| Urban | 0.382*** | 0.106* | -0.321*** | 0.424*** | 0.824*** |
| | (0.0248) | (0.0593) | (0.0660) | (0.0926) | (0.116) |
| Major socio-economic region (NUTS1) of arrival: Center | 0.0839*** | 0.0104 | -0.242*** | -0.00769 | -0.523*** |
| | (0.0250) | (0.0523) | (0.0708) | (0.0600) | (0.100) |
| Major socio-economic region (NUTS1) of arrival: Islands | 0.114** | 1.393*** | 0.573*** | 2.288*** | 0.161 |
| | (0.0480) | (0.203) | (0.133) | (0.267) | (0.257) |
| Major socio-economic region (NUTS1) of arrival: South | 0.245*** | 0.332*** | 0.594*** | 1.706*** | -0.00842 |
| | (0.0396) | (0.0864) | (0.0949) | (0.127) | (0.138) |
| *Observations:* | 230,777 | *71,289* | *61,754* | *63,597* | *48,938* |

*Notes: Standard errors in parentheses, * p<0.10 ** p<0.05 *** p<0.01*