Migrants' self-selection in the early stages of modern economic growth, Spain (1880–1930)[†]

By FRANCISCO J. BELTRÁN TAPIA and SANTIAGO DE MIGUEL SALANOVA[*]

Drawing on a large database from the register of inhabitants of Madrid, this article confirms that the literacy levels of internal migrants moving to the Spanish capital city in the late nineteenth and early twentieth century were higher than that of those who remained in their provinces of origin. This article also explores the different factors influencing the nature and intensity of the selection process. The empirical exercise stresses that the stock of previous migration was a fundamental factor in allowing less literate individuals to join the migration process as well. Interestingly, distance to Madrid hardly affected the profile of male migrants, but it was a strong influence on female migration, although its importance diminished over time. Lastly, the results presented here show that other internal destinations were attracting different types of migrants, often resulting in negative self-selection.

M ost of the academic attention on historical migration has been devoted to the forces driving this process and its subsequent consequences both in sending and receiving regions.[1] Recent research, however, has started to closely examine the characteristics of migrants and address the potential self-selection within sending societies. This research usually finds that rural-to-urban migrants were typically positively self-selected: on average, those who did move abroad were healthier, carried more human capital, and were possibly more enterprising than those who stayed behind.[2] British urban migrants, for instance, came from the best of the rural labour pool between 1851 and 1881.[3] Similarly, among mid-nineteenth-century seamen born outside London, the taller, the literate, and the numerate were more

[1] For overseas migration see, among others, Taylor and Williamson, 'Convergence'; Hatton and Williamson, *Age of mass migration*; O'Rourke and Williamson, *Globalization*. For examples on internal migration, see Baines, *Migration*; idem, 'Population'; Boyer, 'Labour migration'; Boyer and Hatton, 'Migration'; Grant, *Migration and inequality*.

[2] The evidence on the historical selectivity of international migrants, however, is mixed. On this issue, see Hatton, 'Cliometrics'; Abramitzky, Platt Boustan, and Eriksson, 'Huddled masses'; Stolz and Baten, 'Brain drain'; Spitzer and Zimran, 'Migrant self-selection'; Kosack and Ward, 'Who crossed the border?'.

[3] Long, 'Rural–urban migration'.

likely to migrate to London.[4] Likewise, after analysing individual-level data of male workers in London around 1930, it appears that the slightly higher wages and lower levels of unemployment enjoyed by migrants living in London relative to native Londoners were due to differences in average skill levels and personal characteristics.[5] These studies, focusing on Britain, tend to confirm the idea that the propensity to migrate is higher among those who enjoyed higher levels of human capital.[6]

However, the lack of historical research on less developed countries prevents the reaching of more general conclusions about the elements shaping the decision to migrate. This article seeks to contribute to the existing literature by looking at the human capital of internal migrants in late nineteenth- and early twentieth-century Spain. Apart from testing whether internal migrants were positively or negatively selected, the article also examines the different factors influencing the nature and intensity of the selection process. In order to address these issues, data at both the individual and the provincial level are used. On the one hand, a very large database from the register of inhabitants of Madrid, the *Padrón municipal de habitantes*, makes it possible to compute the literacy levels of migrants to Madrid classified by their province of origin for 1880, 1905, and 1930. This information, which distinguishes between male and female migrants, is then compared to the average literacy levels of those provinces derived from population censuses.

The literacy gap, the difference between both rates, confirms the presence of positive selection. Given that the degree of migrants' selectivity varies by region, the article next explores the determinants of migrants' selectivity, for both men and women. The analysis carried out stresses that the stock of previous migration, the so-called family and friends effect, was a fundamental factor allowing less literate individuals to join the migration process as well. Remarkably, distance to Madrid hardly affected the profile of male migrants but it was a strong influence on female migration. However, its importance diminished over time, and eventually even reversed, in response to improvements in the transportation infrastructure. Lastly, the article also shows that other internal destinations, different from Madrid, were attracting other types of migrants, often resulting in negative self-selection. In this regard, urban centres in the same province or in neighbouring regions usually attracted, on average, less literate migrants.

Unveiling the nature of the migrants is not only important in order to increase our understanding of the characteristics of migration, but is also of crucial importance to correct a potential bias affecting the literature dealing with human capital. Studies attempting to assess the ability of different regions to promote educational attainments may be misleading because of the failure to account adequately for both differences in migratory flows and/or their patterns. Given that the human capital of migrants is not observed in the statistics, average literacy rates in the sending region would actually be higher or lower if migratory flows had not taken place, a problem that also applies to biological living standards. Therefore, without properly considering the effect of migratory patterns, inferences about each region's success or failure in fostering human capital may be flawed.

---

[4] Humphries and Leunig, 'Dick Whittington'.
[5] Hatton and Bailey, 'Natives and migrants'.
[6] Sandberg, 'Ignorance'. For Germany, see Grant, *Migration and inequality*.

The rest of the article is organized as follows. Section I describes migratory patterns in Spain during the period of study. Section II presents the data employed here and confirms that migrants going to Madrid were positively selected. Section III then discusses the potential determinants of migrants' selectivity, and section IV empirically tests the importance of the different factors driving the observed differences in migrants' literacy gap, for both men and women. Finally, the conclusion summarizes the main findings.

I

Although internal migration flows were surely a feature of pre-industrial Spanish society, the range and intensity of these movements increased considerably during the second half of the nineteenth century and especially during the early twentieth century.[7] The demographic, social, and economic changes taking place during this period triggered internal mobility, especially from rural areas to the developing urban and industrial areas. Immigration predominantly concentrated in a few destinations due to the spatial concentration of industrial activities.[8] In 1930, the provinces of Madrid and Barcelona accounted for 45.8 per cent of the total internal migration.[9] Only the areas around Bilbao and Seville can also be considered important destinations, but their importance lagged far behind and was mostly restricted to neighbouring provinces. The dynamism of these areas contrasted with the low growth or immobility of other regions. The thriving urban agglomerations generated an extensive range of employment opportunities and widened the wage gap between the capital and those areas, which encouraged migration. Although migration abroad also boomed during this period, especially between 1890 and 1914,[10] it is true nonetheless that, from an international perspective, Spain experienced delayed and less intense migratory flows.

The inability of the countryside to release a more substantial part of its labour force to the non-agricultural sectors has been related to the backwardness of the rural sector and the weak pull from urban and industrial areas.[11] In addition, migration rates, both internal and abroad, were regionally diverse.[12] The low out-migration rates experienced by southern Spain, for instance, have been repeatedly stressed by the literature. Although demographic pressures contributed to higher internal migration rates, only the existence of dynamic urban centres providing job opportunities permitted absorption of the population surplus.[13] In this sense, in a period where a modern system of internal transportation was still in its infancy, the distance to attractive destinations appears to have constituted a critical

---

[7] Reher, *Town and country*; Mikelarena Peña, 'Los movimientos migratorios'; Erdozáin and Mikelarena Peña, 'Las cifras'; Silvestre, 'Internal migrations in Spain, 1877–1930'. Temporary migration was also frequent, especially to fulfil seasonal rural employment. On this issue, see Silvestre, 'Temporary internal migrations'.

[8] Silvestre, 'Viajes de corta distancia'; Paluzie, Pons, Silvestre, and Tirado, 'Migrants and market potential'.

[9] Silvestre, 'Internal migrations in Spain', p. 240.

[10] Sánchez-Alonso, *Las causas*; idem, 'Those who left'.

[11] Silvestre, 'Internal migrations in Spain'.

[12] Mikelarena Peña, 'Los movimientos migratorios'; Sánchez-Alonso, *Las causas*; Silvestre, 'Internal migrations in Spain'.

[13] Silvestre, 'Internal migrations in Spain', p. 244.

factor in shaping migratory patterns.[14] Low income levels, on the other hand, constrained overseas emigration.[15] Likewise, higher literacy rates were positively related to emigration levels, a relationship that remains unclear in the case of internal migration.[16] Lastly, prior migration, the so-called friends and relatives effect, was crucial in sustaining migratory chains, due to the role of these social networks as providers of information and assistance prior to and after the move.

According to the literature, Spanish internal migrants were predominantly young male adults of rural origin, but not necessarily skilled.[17] Regarding educational levels, literacy rates have been shown to be unrelated to migration levels. Núñez, however, disagrees.[18] According to this author, not only did education facilitate labour mobility, but also the decision to migrate was not taken at the individual level but within the household and, therefore, the educational level of those who did not migrate (elder males and females of all ages) influenced migration as well.[19] In any case, the fact that provincial differences in literacy rates may (or may not) explain migration rates does not preclude that migrants were not positively self-selected. It is plausible that, even within less literate provinces, those with higher human capital were more likely to migrate. The theoretical reasons for this are clear: higher literacy rates not only allow acquisition of the necessary information about potential destinations, but also increase the potential returns of migration.[20] To our knowledge, the stronger evidence of positive selection has been provided by Quiroga, who, drawing on military records, compares those recruits who remained resident in their province of origin with those who did not.[21] This author finds that average literacy rates for internal migrants between 1893 and 1899 were 24 percentage points higher than for those who stayed in their province of origin. Apart from a high proportion of students, these recruits were also more likely to work in modern sectors as liberal professionals or white collar workers. The rest of this article analyses whether or not positive selection is what actually characterized Spanish migratory patterns in the late nineteenth and early twentieth century.

II

Together with Barcelona, Madrid was by far the largest Spanish city in the late nineteenth and early twentieth century. From a population of 279,379 inhabitants in 1860, Madrid rapidly grew to 518,656 in 1900 and reached 948,401 in 1930.[22] Only Barcelona showed comparable figures at the time.[23] A significant part of

---

[14] Pons, Paluzie, Silvestre, and Tirado, 'Testing'; Silvestre, 'Internal migrations in Spain'. For a recent analysis of the Spanish transport infrastructure, see Herranz-Loncán, 'Spanish infrastructure stock'; idem, 'Spatial distribution'.

[15] Sánchez-Alonso, 'Those who left'; idem, 'European emigration'.

[16] While Silvestre, 'Internal migrations in Spain', p. 435, does not find literacy rates to be correlated with migration flows during the 1920s, Núñez, 'Within the European periphery', p. 639, argues that education was crucial in promoting labour mobility.

[17] Silvestre, 'Internal migrations in Spain', p. 245.

[18] Núñez, 'Within the European periphery', p. 639.

[19] A study of mountainous areas also finds a positive correlation between educational attainments and out-migration. See Collantes Gutiérrez, 'Las disparidades educativas'.

[20] Gould, 'European inter-continental emigration'; Sandberg, 'Ignorance'.

[21] Quiroga, 'Literacy', p. 600.

[22] Reher, 'Ciudades', pp. 27–9.

[23] Barcelona's population grew from 237,994 inhabitants in 1860 to 529,486 in 1900 and 998,382 in 1930.

this growth took place by absorbing population from the rest of Spain.[24] The dynamism of Madrid relied on being the country's capital and, therefore, the centre of commerce and distribution. Although Madrid had no direct access to the sea, the design of the transportation and communication infrastructure, which placed Madrid at the centre of the hub from which a series of radial links emanated, consolidated its importance.[25] In this sense, while migrants to Madrid came from almost every corner of Spain, migrants to Barcelona mostly originated in its neighbouring Catalonian provinces and other eastern regions.[26] Migration to other potential attractive destinations within Spain, such as Bilbao, Valencia, or Seville, was mostly restricted to nearby provinces.

This article draws on a large sample of individuals taken from the *Padrón municipal de habitantes* which provides individual-level information about the inhabitants of Madrid.[27] Some of these records have been recently computerized, comprising around 250,000 observations distributed between the *Padrones* of 1880, 1905, and 1930.[28] According to these records, only a relatively small fraction of those living in Madrid were actually born in that city (ranging between 33.1 and 42.3 per cent, depending on the date analysed), thus making this source of information an invaluable tool for the study of migration patterns. Importantly for this article, the *Padrón* includes information on the birthplace and the ability to read and write for each observation, which permits the calculation of the literacy levels of those individuals living in Madrid classified by their province of origin.

The data nonetheless present some limitations. First, the sample is not random but includes all individuals residing in the city centre. Given that living downtown was more expensive than on the outskirts, our sample may overestimate the human capital of Madrid's inhabitants. This is actually the case if literacy rates are compared to those obtained from the Population Census: average literacy levels from Madrid's Population Census are around 0.5 to 5.1 percentage points lower than those found here.[29] As will become clear later, these numbers are relatively low in comparison with the literacy gap we find between migrants moving to Madrid and those remaining in the province of origin. In any case, an important number of low-skilled workers also lived there, which ensured a high degree of socio-economic diversity.[30] Second, the number of individuals coming from certain provinces is relatively small, which may somewhat affect the results for those areas. In order to address this issue, the empirical exercise carried out here will be also replicated focusing only on those provinces which supply a sufficiently large number of individuals to our sample. Similarly, it should be noted that there is no information on stage migration, so we do not know whether the move between the birthplace and Madrid took place in one step or comprised several steps. There is evidence,

[24] Pallol Trigueros, Carballo Barral, and Vicente Albarrán, 'Inmigración'; Vicente, 'Los motores'; Otero Carvajal and Pallol Trigueros, 'El Madrid moderno'.

[25] Herranz-Loncán, 'Spatial distribution'.

[26] Oyon, Maldonado, and Griful, *Barcelona*; Silvestre, 'Viajes de corta distancia'.

[27] This source can be found in the Archivo de Villa de Madrid, Sección Estadística, *Padrón Municipal de Habitantes de Madrid* (1880, 1905, and 1930).

[28] Otero Carvajal and Pallol Triqueros, 'El Madrid moderno'.

[29] While our records yield literacy rates of 76.6, 86.1, and 95.0% in 1880, 1905, and 1930 respectively, the Madrid population censuses of 1887, 1900, and 1930 yield literacy rates of 75.4, 81.0, and 94.5% respectively; Dirección General del Instituto Geográfico y Estadístico, *Censo de la población* (1891, 1902, 1932).

[30] Miguel Salanova, 'Las raíces'; idem; 'Bajo los tejados'.

however, suggesting that stage migration was unimportant.[31] Lastly, we do not have data on return migration, so the importance of that process is unknown. It is plausible that, on average, the least skilled were more likely to return if the move was not successful, which would strengthen the nature of the migrants' positive selection. Moreover, migrants could have become literate once they were already living in Madrid, thus also overestimating their original literacy levels. Given that our sample contains information on the year these migrants arrived in Madrid, we are able to control for these issues.

Migrants tended to be younger and, in a period of increasing literacy, exhibited higher literacy rates than the general population. For comparability reasons, we therefore only focus on those individuals aged 16 to 30.[32] Tables S1 and S2 in the online appendix present the number of observations and the average literacy rates of male and female migrants within that age group living in Madrid, classified by their province of origin.[33] The regional picture and its evolution over time are illustrated in figure 1. While the literacy rates of male migrants are remarkably high, female migrants show significantly lower levels of literacy, as well as wider regional variation.

The gender literacy gap constituted a structural feature of Spanish educational attainments throughout the nineteenth and early twentieth century.[34] The period under study coincides with growing literacy levels, which also contributed to narrowing the educational differences between men and women. Nevertheless, compared to other European countries, Spain suffered from low levels of education due to the delay in the transition to universal literacy. Moreover, although literacy rates were increasing everywhere during this period, regional differences persisted well into the twentieth century.[35] The regional pattern arising from the literacy levels of migrants living in Madrid, however, does not conform to the image portrayed by Núñez on the basis of population censuses.[36] On the one hand, and contrary to the wide provincial variation apparent in national statistics, the literacy levels of male migrants were homogeneously high. On the other hand, female migrants from southern Spain and Catalonia, as well as some other northern regions, consistently performed better than individuals from other provinces, although the difference diminished over time. In this regard, the underperformance of women born in the neighbouring provinces of Madrid is also noteworthy.

Migrants are obviously not likely to be a representative sample of the original population, so it is interesting to compare the literacy levels of those individuals migrating to Madrid with that of those staying in their provinces of origin by using

---

[31] Sánchez-Alonso, *Las causas*, p. 31; Silvestre, 'Internal migrations in Spain', pp. 245–6. Studying overseas migration using micro-level data, Moya, *Cousins and strangers*, shows, on the contrary, that rural–urban staging prior to crossing the Atlantic was common. However, these movements were most likely to have taken place within the province of origin and therefore it does not affect the analysis carried out by Silvestre or the one followed here.

[32] The image depicted here does not change if we instead rely only on those individuals aged 21 to 30.

[33] Foreign inhabitants have also been excluded from the sample, together with those observations whose place of origin was unclear. Although the number of foreign-born migrants was relatively significant (between 2.0 and 2.8% of the sample depending on the year), those with unknown birthplaces only represent 0.08% of the sample in 1880 (and indeed a much lower figure at the other two dates).

[34] Núñez, *La fuente*; Tortella, 'Patterns'.

[35] The reasons behind these dissimilar paths have been analysed elsewhere. See Núñez, *La fuente*; idem, 'Within the European periphery'; Beltrán Tapia, 'Enclosing literacy'.

[36] Núñez, *La fuente*.

MEN   WOMEN
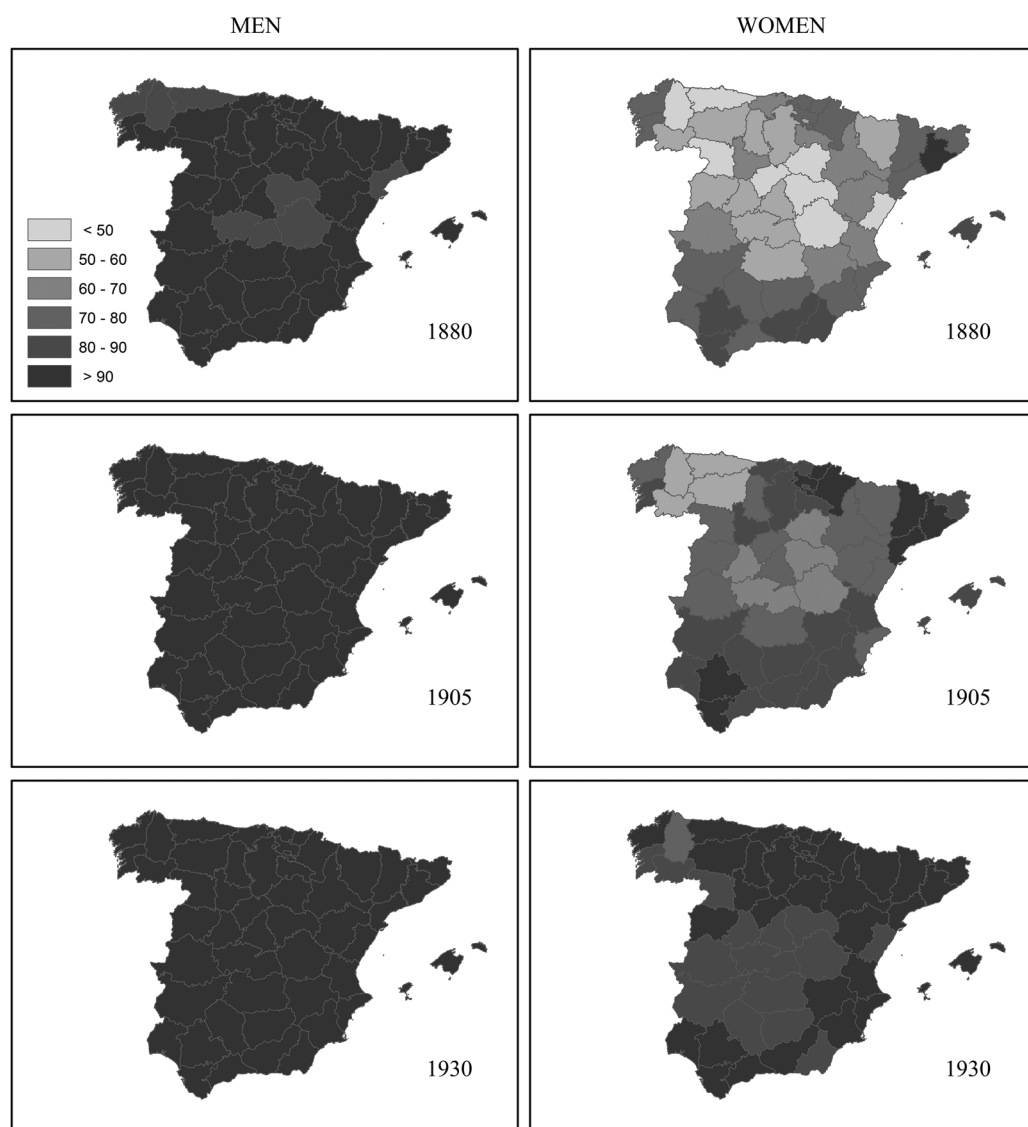
Figure 1.   *Migrants' literacy, individuals aged 16–30 living in Madrid, by province of origin (%)*

Source: See online app. tab. S5. Actual figures are reported in online app. tabs. S1 and S2.

information from the population censuses. The literacy gap, the relative difference between those groups, provides a first approximation to the issue under study:[37]

$$LIT\_GAP_{it} = MAD\_LIT_{it} - PROV\_LIT_{it} \qquad (1)$$

[37] It should be noted that the timing of the *Padrones* and the *Censos de Población* does not match perfectly for the first two periods. Due to data availability, we are only able to establish the following comparisons: 1880–7 and 1900–5.
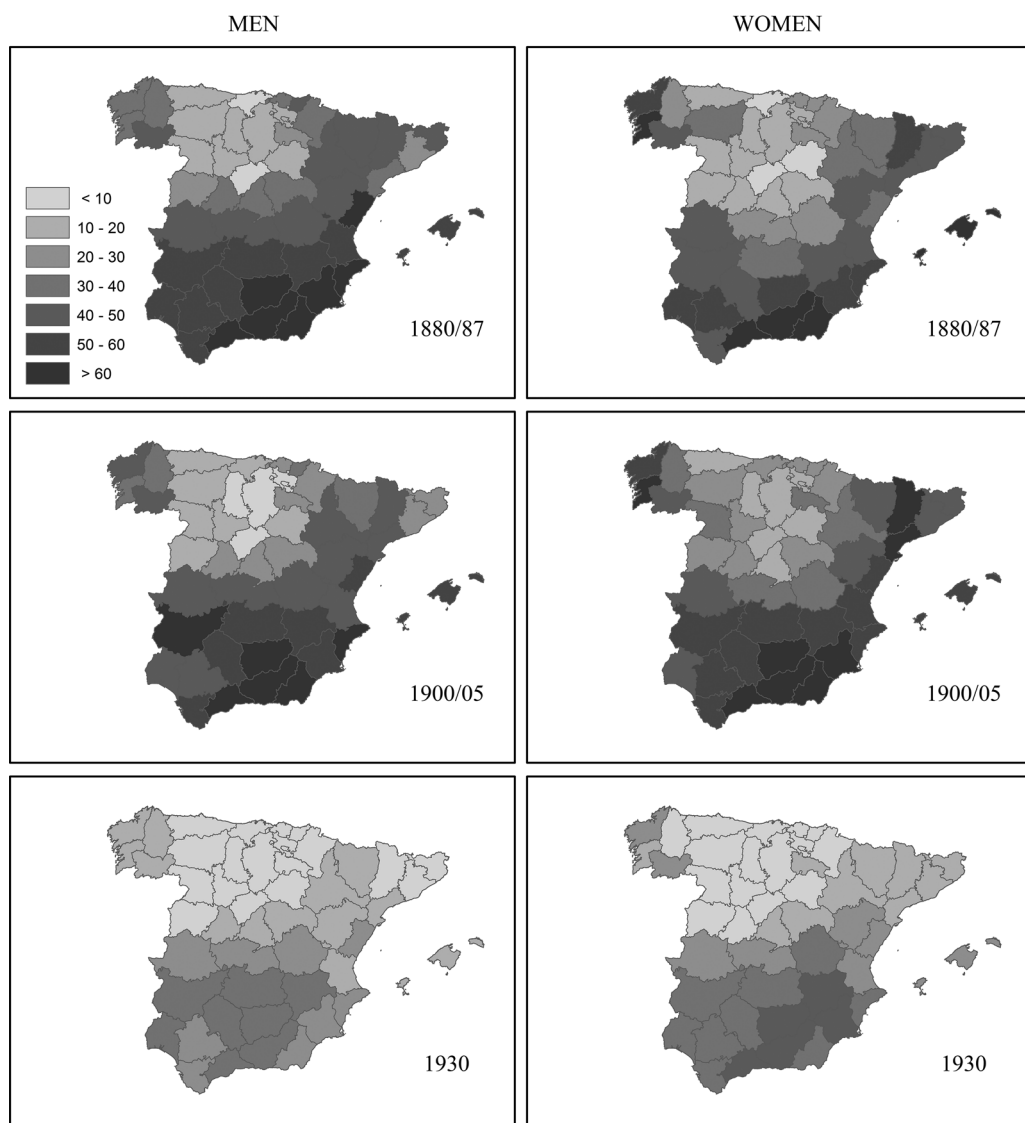
Figure 2. *Migrants' literacy gap, individuals aged 16–30, by province of origin (percentage points)*

The statistics are reported in table S3 in the online appendix. As shown in figure 2, the average literacy rates of migrants living in Madrid, classified by birthplace (MAD_LIT_{it}), are generally much higher than those who were living in their province of origin ($PROV\_LIT_{it}$), thus confirming that migrants, at least those moving to Madrid, were positive selected. Nevertheless, the literacy gap not only significantly decreased over time, but also showed wide geographical

variation.[38] With the exception of the Galician provinces, the literacy gap, both
male and female, is clearly smaller in north-western Spain. Broadly speaking, this
regional configuration inverts the picture arising from the population censuses:
where literacy rates in the place of origin are higher, the migrants' literacy gap
is lower, and vice versa. This result confirms the importance of education in
the decision to migrate. Literate individuals not only tended to be better able
to collect information about the potential destinations, but also enjoyed higher
expected returns from migrating. In order to shed more light on the processes that
shaped migrants' self-selection, the following sections exploit the regional variation
presented here to explore the factors influencing the selectivity of internal migrants.

III

Relying on a panel data set at the provincial level in three different periods (1880–
7, 1900–5, and 1930) and following the literature on this topic,[39] we specify the
following equation in order to explain the differences in the literacy gap and assess
the relative importance of the potential determinants involved in the decision to
migrate:

$$LIT\_GAP_{it} = \alpha + \beta' X_{it} + \gamma' Z_{it} + \alpha_t + u_{it} \qquad (2)$$

The first set of variables, X_{it}, refers to the different benefits and costs which may
influence migratory behaviour. In this regard, the most obvious economic driver
of migration is wage differentials.[40] Given that wages at destination, Madrid, are
fixed, we include agricultural wages at origin in the model to account for the
relative attraction of the capital. Higher wages in the sending region are expected
to keep prospective migrants at home, which would reduce the observed literacy
gap. Borjas, however, argues that relative inequality also influences the selectivity
of migrants.[41] If inequality in the sending region is low, high-profile potential
migrants are more likely to move to more unequal regions where the expected pay-
off is going to be larger. On the contrary, if inequality is lower at the destination,
negative selection would ensue. In order to capture this effect, we add the level of
inequality in the province of origin to the model. Inequality is measured using the
Williamson index, which computes the distance between income per worker and
the unskilled wage in order to compare the bottom part of the distribution to the
average income. Given that, according to the Borjas model, higher inequality at
origin would retain high-skilled migrants at their provinces of origin and therefore
decrease the literacy gap, we expect both variables to be negatively related. Other
circumstances may have also helped to determine migrants' selectivity, so we have
also considered population growth in earlier decades in order to capture other

---

[38] Note from online app. tab. S2 that, although quantitatively small, some provinces in 1930 even present a
negative difference, suggesting the presence of negative selection at that date.

[39] Different studies have addressed this issue focusing on historical migration abroad. See, for instance, among
others, Stolz and Baten, 'Brain drain'; Abramitzky et al., 'Huddled masses'; Kosack and Ward, 'Who crossed the
border?'; Spitzer and Zimran, 'Migrant self-selection'.

[40] Chiswick, 'Are immigrants favorably self-selected?'.

[41] Borjas, 'Self-selection'. For a recent application of Borjas's model to international migration, see Stolz and
Baten, 'Brain drain'.

aspects defining the underlying context. Where demographic pressures were high, it is likely that wider segments of the population would consider the possibility of migration.

On the other hand, the costs of moving, involving financial but also information and psychological costs, have been shown to inhibit migration.[42] Information about distant destinations was costly and sometimes unreliable. More skilled individuals were not only more likely to be better informed about potential opportunities, but their expected returns were also potentially higher. By disincentivizing low-skilled migrants, we expect that the difference between literacy rates increased as we move further from Madrid. The period under study, however, coincided with vast improvements in the transportation infrastructure.[43] In order to capture the reduction in transportation costs, distance is also interacted with time dummies. Likewise, migration costs are reduced when relatives and friends are already living in the destination.[44] By sending information and remittances, as well as providing temporary accommodation and help, these social networks encourage chain migrations and allow less able individuals to make the move. The so-called family and friends effect is measured as the stock of previous migrants coming from the same province of origin.

It is, however, crucial to stress that, while we only have information on individuals migrating to Madrid, migrants did not exclusively move there. It is likely that the type of migrant going elsewhere was different which, in turn, not only influenced the potential selectivity of those migrants going to Madrid, but also, by abandoning the province of origin, those migrants going elsewhere also affected average literacy levels in the sending region. The model therefore includes Z_{it}, a set of variables that, as explained below, capture the pull of other potential destinations.

On the one hand, the literature has considered that the characteristics of Spanish migrants going abroad, principally to Latin America, differed from those of internal migrants, due to the higher costs and risks involved in this type of migration.[45] Therefore positive selection is likely to have been stronger for international movers. If that was the case, average literacy levels in the province of origin would be pushed downwards by those leaving, because those individuals would no longer be reflected in the official statistics and therefore the literacy gap with those going to Madrid would increase. The provincial rates of migration abroad are therefore included in the model. On the other hand, apart from Madrid, there were other important internal destinations available. Most internal migration nonetheless concentrated in a few destinations, due to the spatial concentration of industrial activities. Barcelona was the other main pole of attraction within Spain.[46] Although Silvestre argues that Madrid and Barcelona were somewhat interchangeable destinations, it may be the

---

[42] Silvestre, 'Internal migrations in Spain'; Sánchez-Alonso, 'European emigration'.
[43] Herranz-Loncán, 'Spanish infrastructure stock'.
[44] Sánchez-Alonso, 'Those who left', p. 738.
[45] Ibid.; Silvestre, 'Internal migrations in Spain'. In this sense, while Sánchez-Alonso finds a positive correlation between regional literacy rates and overseas migration rates, Silvestre shows no significant relationship between that variable and internal migration rates. Furthermore, given that migration abroad is negatively correlated with internal migration at the provincial level during the 1920s, Silvestre argues that both types of migration were substitutes. The fact that income constraints prevented overseas migration to a greater extent than internal migration reinforces that conclusion.
[46] In 1930, 35.9% of the inhabitants of Barcelona had been born in another province, a figure that had grown from 19.6% in 1877; Silvestre, 'Internal migrations in Spain', p. 240.

case that they attracted different migrant profiles.[47] Fortunately, the provincial origin of the migrants going to Barcelona has been studied: apart from other Catalonian provinces, they mostly came from the other eastern regions of Baleares, Aragón, Valencia, and Murcia, together with a small but growing fraction from eastern Andalusia.[48] In order to control for this issue, the relative importance of the population in the sending province going to Barcelona is taken into account in the analysis.

Although their relative importance as major destinations was much smaller, other regions, such as Bilbao and Seville, also experienced a significant inflow of migrants.[49] Importantly, the regional scope of these migratory movements was restricted. Apart from the province itself, most migrants in the industrializing area of Bilbao at the end of the nineteenth century came from its neighbouring provinces.[50] A similar situation, although in a different economic context and involving smaller numbers, was taking place in Seville, and even in other cities such as Valencia, Zaragoza, or Valladolid.[51] It is difficult to assume that the type of migrants going to these areas was the same as those individuals going to Madrid. In order to capture this potential bias, migration to these places is captured by constructing dummy variables for each of these alternative migratory basins. For each migratory basin, the dummy variable takes a value of 1 in each of the neighbouring provinces around that destination.[52] Although we are aware that these cases do not exhaust the potential destinations for internal migrants, the relative importance of the population flows to other areas was minimal.[53]

Alternatively, migrants could move but not leave their province of origin if enough opportunities were available. In this sense, the pull of the capital or other important cities within the province could attract rural dwellers and retain potential out-migrants. Similarly, a strong urban pull would also attract migrants from other neighbouring provinces. The urbanization rate, measured as the fraction of the provincial population living in cities bigger than 5,000 inhabitants, is therefore included in the specification in order to account for this factor. If migration to the cities is positively selected, more urbanized provinces would draw more literate migrants, who would otherwise be likely to have abandoned that province. This process would keep average literacy rates relatively high in that province and

---

[47] Ibid., p. 240. For the occupational attainment of migrants going to Barcelona, see Silvestre, Ayuda, and Pinilla, 'Occupational attainment'.

[48] Oyon et al., *Barcelona*.

[49] Not only was their size much lower than that of Madrid or Barcelona but their population was less diverse. In 1930, 24.9 and 15.3% of their population had been born in another province respectively; Silvestre, 'Viajes de corta distancia', p. 257; idem, 'Las migraciones interiores', p. 167.

[50] González Portilla and García Abad, 'Migraciones interiores'. In 1930, migrants to Vizcaya mostly came from Álava, Burgos, Guipuzcoa, and Logroño; Silvestre, 'Las migraciones interiores', pp. 167–8.

[51] Silvestre, 'Las migraciones interiores', pp. 167–8. Migrants going to Seville mostly originated in Badajoz, Cádiz, Córdoba, and Huelva. Other potential destinations, such as Zaragoza and Valladolid, also mostly attracted people from neighbouring provinces: Teruel, Huesca, Soria, Navarra, and Logroño, in the case of Zaragoza, and Palencia, Zamora, Segovia, Salamanca, and León, in the case of Valladolid.

[52] It appears that the booming area around Bilbao attracted a significant number of migrants from farther regions; González Portilla and García Abad, 'Migraciones interiores'. Therefore, its migratory basin has been extended to the provinces that González Portilla and García Abad enumerate as more important: Burgos, Álava, Cantabria, La Rioja, Guipúzcoa, Asturias, and Navarra. The results nonetheless hardly change if a more restrictive set of provinces is employed.

[53] For a detailed analysis of the relative importance of the different origins and destinations, see Silvestre, 'Viajes de corta distancia'.

subsequently reduce the literacy gap with those countrymen moving to Madrid. However, short-distance rural–urban migration could also have attracted less skilled individuals, so the coefficient on urbanization may turn out to be positive. Given the relative importance of agro-towns, especially in southern Spain,[54] the importance of industrialization is also added to the model using the gross value-added by non-agricultural activities per capita.

Temporary internal migrations, referring to any kind of non-permanent movement, were also commonplace during this period.[55] The process of economic modernization, or rather its moderate pace, was not able appreciably to alter the importance of traditional temporary flows inherited from the preindustrial period. These movements mainly affected unskilled workers in the agricultural and the service sectors, so these flows, if they lasted long enough, might have had an effect on the literacy gap.[56] In order to account for that, temporary out-migration rates at the provincial level are therefore included in the model.

It is also important to note that the intensity of migratory flows may also be affecting our specification. In areas with large migration flows, the decision to move spreads down the skill distribution, so the literacy gap should be lower there. The net migration flow in each province, measured using the inter-census balance method, is thus employed as a control. Likewise, the inclusion of time dummies, α_t, allows us to capture whether the literacy gap was subject to country-wide changes over the period of study.

Lastly, given that being literate appeared to be an important factor in the migration process, the average literacy level of the population is considered. However, as it is employed to compute the literacy gap, this variable also appears on the left-hand side of the equation and is therefore endogenous. To address this issue, we will also re-specify our model by using migrants' literacy level as the dependent variable and including literacy in province of origin as an additional explanatory variable. This robustness test provides further insights into migrants' selectivity because, by controlling for the literacy levels of the sending regions, it effectively addresses the sources of the selectivity of migrants going to Madrid.

Summary statistics of all the variables employed, together with a description of how these variables have been constructed, are reported in tables S4 and S5 in the online appendix.

IV

Tables 1 and 2 present the results of estimating equation (2) for men and women respectively.[57] We are interested in variation between provinces, so a random effects model is employed.[58] Column 1 reports the baseline specification, which includes the variables capturing the benefits and costs of migration, as well as those reflecting

---

[54] Reher, 'Ciudades', p. 11.

[55] Silvestre, 'Temporary internal migrations'.

[56] Although temporary migration in the industrial sector was probably less important, this type of migration also supplied rural and urban industries and, especially, mining areas; ibid., pp. 544–5.

[57] Given that, for some of the variables of interest, it is not possible to have information on the province of Madrid excluding the capital city, that province is left out from the analysis. Similarly, certain data are lacking for the Canary Islands. The model thus relies on 47 provinces.

[58] Estimating a fixed-effects model would only exploit variation within-provinces across time.

Table 1. *Determinants of migrants' selectivity, men, 1880–1930*

			Dependent variable				
			Male migrants' literacy gap			*Male migrants' literacy*	
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*
Benefits and costs							
Agricultural wages	−0.06**	−0.06**	−0.07**	−0.07**	−0.02	0.01	0.02*
	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.01)	(0.01)
Inequality	−1.38	−1.31	−0.79	−0.74	−1.08	0.32	0.12
	(1.05)	(1.01)	(0.98)	(0.99)	(0.88)	(0.30)	(0.34)
Population growth	−0.13*	−0.13*	−0.13*	−0.14*	−0.18**	−0.02	−0.02
	(0.08)	(0.07)	(0.07)	(0.08)	(0.08)	(0.02)	(0.02)
Distance to Madrid	−0.93	0.50	−2.72	−3.30	−0.84	−1.04	−1.85
	(4.12)	(4.12)	(4.13)	(4.28)	(4.21)	(1.46)	(1.57)
*d_1905	0.67	0.65	0.69	0.75	0.15	1.26	1.05
	(1.77)	(1.79)	(1.85)	(1.87)	(1.96)	(1.19)	(1.05)
*d_1930	−4.28	−4.60*	−4.00	−3.95	−5.40*	−1.06	−1.78*
	(2.80)	(2.65)	(2.84)	(2.91)	(3.14)	(1.15)	(1.07)
Family and friends	−16.50***	−14.11**	−15.85***	−16.70***	−12.94**	−2.59*	−3.40**
	(5.76)	(5.65)	(5.44)	(5.58)	(5.09)	(1.56)	(1.55)
Alternative destinations							
Migration abroad	0.27	0.26	0.24	0.24	0.40**	−0.01	0.10
	(0.24)	(0.24)	(0.24)	(0.24)	(0.20)	(0.10)	(0.08)
Urbanization	0.35***	0.32***	0.22**	0.22**	0.19*	0.05**	0.04*
	(0.10)	(0.10)	(0.09)	(0.09)	(0.11)	(0.02)	(0.02)
Industrialization	−0.01	−0.01	−0.01	−0.01*	−0.00	−0.00	−0.00*
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Temporary migration	−0.31	−0.28	−0.42	−0.44	−0.37	0.06	0.07
	(0.43)	(0.43)	(0.41)	(0.43)	(0.42)	(0.16)	(0.17)
Migration to Barcelona	−1.03**	−1.10**	−1.09**	−0.95*	−1.21*	0.13	0.45*
	(0.48)	(0.52)	(0.49)	(0.49)	(0.62)	(0.15)	(0.23)
d_Bilbao basin	−13.59***	−12.31***	−14.23***	−13.64***	−16.71***	0.99	1.60
	(3.15)	(3.18)	(3.89)	(3.97)	(4.55)	(0.86)	(1.16)
d_Valencia basin		12.46***	8.93***	8.65***	7.84**	0.30	0.11
		(3.08)	(3.10)	(3.07)	(3.21)	(0.70)	(0.75)
d_Seville basin		5.79	5.51	5.56	5.77	-0.59	−0.23
		(3.90)	(3.52)	(3.54)	(4.88)	(0.76)	(1.09)
d_Zaragoza basin			−3.09	−2.97	−3.31	1.31**	1.40**
			(4.58)	(4.47)	(5.51)	(0.56)	(0.70)
d_Valladolid basin			−19.67***	−19.53***	−20.19***	0.60	0.90
			(3.46)	(3.54)	(3.84)	(0.75)	(0.94)
Controls							
Net migration				−0.10	−0.06	−0.09	−0.17*
				(0.17)	(0.20)	(0.09)	(0.10)
Male literacy in origin						0.03	0.03
						(0.03)	(0.03)
d_1905	−5.80	−5.60	−5.59	−5.74	−2.98	−5.23	−4.41
	(9.66)	(9.76)	(10.14)	(10.24)	(10.50)	(6.35)	(5.59)
d_1930	3.03	5.10	3.14	3.55	9.87	9.72	13.12**
	(15.04)	(14.16)	(15.35)	(15.47)	(16.32)	(6.21)	(5.57)
Observations	141	141	141	141	117	141	117
R²	0.680	0.711	0.790	0.791	0.757	0.511	0.548

Notes: Random effects estimation (47 provinces). Robust standard errors are reported in parentheses; *, **, and *** denote significance at the 10, 5, and 1% level. For simplicity, the intercept is not reported. Distance to Madrid is expressed in natural logs.
Source: See online app. tab. S5.

Table 2. *Determinants of migrants' selectivity, women, 1880–1930*

	Dependent variable						
	Female migrants' literacy gap					Female migrants' literacy	
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*
Benefits and costs							
Agricultural wages	−0.12**	−0.11**	−0.12***	−0.12***	−0.03	−0.05*	−0.01
	(0.05)	(0.05)	(0.04)	(0.05)	(0.04)	(0.03)	(0.04)
Inequality	−1.51	−1.43	−1.03	−1.07	−0.96	0.30	0.05
	(1.16)	(1.14)	(1.12)	(1.15)	(1.09)	(0.87)	(0.97)
Population growth	0.02	0.02	0.01	0.02	−0.05	−0.05	−0.10
	(0.11)	(0.11)	(0.11)	(0.11)	(0.09)	(0.08)	(0.07)
Distance to Madrid	11.03***	11.74***	9.54**	9.61**	12.10***	6.16	7.33*
	(4.19)	(4.33)	(4.49)	(4.52)	(4.18)	(3.92)	(3.74)
*d_1905	−5.19***	−5.20***	−5.16***	−5.27***	−4.54***	−6.64***	−6.28***
	(1.54)	(1.54)	(1.59)	(1.58)	(1.57)	(1.66)	(1.81)
*d_1930	−10.12***	−10.30***	−9.87***	−9.86***	−8.77***	−10.85***	−9.51***
	(3.09)	(3.06)	(3.16)	(3.17)	(3.13)	(2.71)	(2.88)
Family and friends	−15.22***	−14.04**	−14.79***	−14.64***	−10.64**	−15.35***	−13.24**
	(5.82)	(5.95)	(5.97)	(6.06)	(5.03)	(5.89)	(5.19)
Alternative destinations							
Migration abroad	0.20	0.19	0.17	0.18	-0.07	0.21	0.05
	(0.16)	(0.16)	(0.16)	(0.16)	(0.17)	(0.14)	(0.16)
Urbanization	0.23***	0.20**	0.12*	0.13*	0.13	0.16**	0.22***
	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.07)	(0.07)
Industrialization	−0.01***	−0.01***	−0.01***	−0.01***	−0.01**	−0.00	−0.01
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Temporary migration	−0.88*	−0.86*	−0.94**	−0.97**	−0.95**	−0.18	−0.28
	(0.47)	(0.46)	(0.45)	(0.45)	(0.46)	(0.36)	(0.37)
Migration to Barcelona	0.58	0.54	0.47	0.42	−0.09	1.01**	0.05
	(0.48)	(0.50)	(0.50)	(0.55)	(0.50)	(0.43)	(0.40)
d_Bilbao basin	−14.66***	−14.13***	−15.93***	−16.14***	−17.97***	−2.93	−5.27
	(3.00)	(3.06)	(3.50)	(3.66)	(3.92)	(3.89)	(4.51)
d_Valencia basin		5.17	2.58	2.62	5.43	−2.69	1.50
		(3.30)	(3.35)	(3.38)	(3.61)	(3.13)	(2.64)
d_Seville basin		3.73	3.45	3.25	4.03	2.70	1.39
		(2.57)	(2.50)	(2.46)	(3.08)	(1.80)	(2.20)
d_Zaragoza basin			−2.48	−2.44	−3.55	−0.89	−1.54
			(4.30)	(4.37)	(4.98)	(3.18)	(3.57)
d_Valladolid basin			−15.62***	−15.62***	−15.28***	−7.16**	−7.36**
			(3.75)	(3.73)	(3.87)	(3.53)	(3.63)
Controls							
Net migration				0.09	0.25	-0.12	0.06
				(0.17)	(0.17)	(0.14)	(0.18)
Female literacy in origin						0.45***	0.53***
						(0.09)	(0.10)
d_1905	34.08***	34.18***	34.20***	34.59***	30.69***	46.14***	43.63***
	(8.19)	(8.19)	(8.48)	(8.50)	(8.32)	(9.29)	(10.00)
d_1930	44.41***	45.56***	44.37***	43.72**	37.56**	66.90***	57.13***
	(16.61)	(16.41)	(17.01)	(17.05)	(16.65)	(16.47)	(17.37)
Observations	141	141	141	141	120	141	120
R²	0.767	0.770	0.813	0.816	0.807	0.801	0.810

Notes: As for tab. 1.
Source: See online app. tab. S5.

the importance of the most important alternative destinations. While columns 2 and 3 extend the model to incorporate other migratory basins, column 4 controls for the importance of migratory flows. As explained in section II, the number of migrants in Madrid coming from certain provinces is relatively small, so column 5 replicates the exercise, but focusing only on those provinces for which we have more than 30 individuals in our sample. Lastly, we re-specify the model by using migrants' literacy as the dependent variable and including literacy rates in the province of origin as an additional explanatory variable: while column 6 employs all the observations, column 7 focuses again only on those provinces which have at least 30 migrants in our sample. All specifications include time-fixed effects which capture general factors that may be affecting the migrants' literacy gap as time goes by. The high R-squared values obtained confirm that the model is able to explain a large part of the variation in the dependent variable.

Benefits and costs of migration

High agricultural wages, as expected, help to reduce the literacy gap for both men and women, suggesting that wealthier rural regions were better able to retain part of their skilled labour force. Alternatively, it may be the case that poverty constraints, which prevented the poor from out-migrating and subsequently depressed average literacy levels in the place of origin, were a lesser burden in these areas. Although the coefficient on inequality shows the expected negative sign, it is however statistically insignificant, so the Borjas model hardly contributes to an explanation of the selectivity of internal migrants in nineteenth- and early twentieth-century Spain.[59] Demographic pressures appear to have had an effect, albeit weak, on male migrants' profiles (but not on women's): where a growing population was straining the available resources, a wider section of the lower ranks of the male population was more likely to move away. Interestingly, the family and friends effect clearly allowed less literate individuals to migrate. Its relative influence, as also evidenced by the standardized coefficients reported in tables S6 and S7 in the online appendix, is quite sizeable: a 1 percentage point increase in the stock of previous migrants reduces the literacy gap by more than 10 percentage points. These results confirm the importance that the literature has usually attached to this variable.[60]

Interestingly, distance to Madrid only shows the expected positive effect on the literacy gap in the case of women. Although the more literate were more likely to have better information and were also more likely to be able to compensate for the high economic and non-monetary costs of moving long distances, this constraint does not seem to have affected male migrants. This finding is likely to be related to the role attached to women in such a highly patriarchal society,[61] as well as the higher risks that long distances would have involved for female migrants during this period. As its interaction with the time dummies testifies, the impact of distance on women's self-selection was greatly diminished by 1905

[59] The standardized coefficients reported in online app. tabs. S6 and S7 confirm the low relative importance of this variable. Reviewing other studies, Abramitzky et al., 'Huddled masses', p. 1836, explain that the Borjas model has found mixed support in contemporary immigrant flows.
[60] See, for instance, Beine, Docquier, and Özden, 'Diasporas'.
[61] On the role of women in Spanish society during this period, see Borderías, Pérez-Fuentes, and Sarasúa, 'Gender inequalities'; Beltrán Tapia and Martínez-Galarraga, 'Land access inequality'.

and it was actually reversed by 1930, a pattern that also affected male migrants. The extension of the road and railway infrastructure, by allowing the poor in more distant regions to get to Madrid easily, appears to have facilitated the relocation of labour significantly during this period.

Alternative destinations

The information conveyed by the coefficients on the variables that capture the possibility of going to alternative destinations is worth analysing, since it provides hints about the selectivity of other types of migration, different from those moving to Madrid. On the one hand, although it has been argued that overseas migrants differed from internal migrants due to the higher costs and risks involved in this type of migration,[62] we find only weak support for this thesis, at least relative to those going to Madrid. Focusing on male migrants, the positive sign of the coefficient on this variable means that, in those areas where migration abroad was higher, the difference between the literacy of those who went to Madrid and those who did not move was even larger. This implies that, by carrying superior levels of human capital than those staying, international migrants depressed average literacy levels at home when they left. This coefficient, however, is only statistically significant in column 5, which casts doubt on this interpretation.[63] The estimated coefficient is even smaller in the case of women, which probably reflects their lesser involvement in overseas migration.

On the other hand, the coefficient on urbanization is positive and highly significant, for both men and women. This result suggests that, although urban agglomerations also generated a demand for high-skilled jobs, they predominantly created low-skilled occupations. By absorbing a relatively larger pool of unskilled workers from neighbouring areas, expanding cities actually kept average literacy rates in that province low. It should be noted that, in addition to hosting manufacturing and commercial jobs, towns and cities had an intense demand for construction workers and domestic service.[64] The fact that the positive and significant coefficient on urbanization is also visible in columns 6 and 7 implies that cities were also able to supply Madrid with skilled individuals. As explained above, the existence of large agro-towns, especially in southern Spain, indicates that the importance of manufacturing should be added to the model. The negative sign of this variable implies that regions where manufacturing was important managed to some extent to retain or attract more literate individuals, an effect which is especially visible in the case of women.

The fact that migrants' profiles depended on the labour market opportunities of the receiving area is confirmed by looking at the effect of the possibility of migrating to other potential destinations. The literacy gap with their countrymen living in Madrid is greatly diminished in those provinces which supplied labour to Barcelona, Bilbao, or Valladolid. By moving there, these male migrants were also indirectly raising average literacy rates in their provinces of origin, so those

[62] Sánchez-Alonso, 'Those who left'; Silvestre, 'Internal migrations in Spain'.
[63] In cols. 2 to 5, the *p*-value nonetheless revolves around 0.14, which is close to being statistically significant at the 10% level. Measurement errors may be inflating standard errors and preventing a more accurate estimation.
[64] Reher, *Town and country*; Sarasúa, *Criados, nodrizas y amos*.

provinces were thus releasing a large proportion of unskilled migrants. This finding is consistent with micro-level analysis of the characteristics of migrants feeding the industries thriving around Barcelona and Bilbao.[65] The case of Valencia, where the sign of the coefficient is the opposite, offers an interesting counterpoint. The different economic structure of Valencia seems to have attracted a relatively skilled population given that the literacy gap increased in the sending province.[66] Mostly a commercial city, Valencia lacked the industrial fabric that characterized Barcelona or Bilbao, so this Mediterranean city was attracting individuals who were more positively selected, on average, than the other two cities. Analysing the migrants' literacy gap for women confirms these patterns, with the exception of those regions where Barcelona was the main destination. While disproportionately attracting low-skilled male workers, the positive sign on the coefficient of that variable in table 2 indicates that female migrants moving to Barcelona were, on average, above the middle part of the skilled distribution of their provinces of origin. Although not statistically significant, this result contrasts with that for men, thus pointing again to the peculiar economic structure of this city. Lastly, temporary migration flows seem to have reduced the literacy gap, a result which is more evident in the case of women. By leaving in search of employment, unskilled or low-skilled workers pushed up average literacy levels in the sending region and consequently reduced the literacy gap with those who had migrated to Madrid.

Other factors

As hypothesized in the previous section, if migration is positively selected, the pool of qualified migrants decreases as migration increases. This negative relationship between migration flows and the male literacy gap turns out to be very weak when confronted with our data. We should bear in mind, however, that this variable captures all types of migration, not just migrants going to Madrid, so it conflates different types of movers. What it is important to stress here is that the results discussed above are barely altered when including this variable as a control.

The results reported in columns (6) and (7) show that, as expected, literacy rates in origin are positively related with the literacy levels of migrants living in Madrid. Remarkably, this relationship is much stronger in the case of women.[67] Moreover, R-squared values are also significantly smaller for men in these specifications. Our model is thus able to explain a larger part of the variance of the literacy of female migrants going to Madrid. This finding suggests that men's decision to migrate contained a higher degree of randomness or, in other words, that a wider range of choices was opened to them.

Lastly, the time dummies also evidenced that migrants' selectivity evolved differently for men and women. While the male literacy gap in columns 1 to 5

---

[65] Oyon et al., *Barcelona*; González Portilla, ed., *Los orígenes*; García Abad, *Historias de emigración*. For a comparative analysis, see also Silvestre, 'Las migraciones interiores', pp. 177–8.

[66] Valencia's economic growth during this period was based on the growth of existing craft industries and of sectors supporting agricultural activities (agricultural machinery, chemicals, and shipbuilding). On this issue, see Nadal, 'El desenvolupament'; Sorribes, *Crecimiento económico*; idem, 'La transición urbana'.

[67] Although tab. 1 shows that the coefficient on male literacy in origin is not statistically significant, the robustness test performed in the next subsection confirms that that relationship is statistically significant (see online app. tab. S8).

and male migrants' literacy in columns 6 and 7 hardly changed between 1880 and 1930, the coefficients on the time dummies in the case of women show that the selectivity of female migrants increased during this period. While female literacy rates were very low at the end of the nineteenth century, the expansion of female literacy from then onwards allowed that literate women could increasingly join the migratory process.

Robustness test

Before concluding, it is worth stressing that the literacy rates of migrants moving to Madrid might be overestimated, thus potentially biasing our results. On the one hand, we do not have information on return migration and it is probable that the less literate were more likely to leave Madrid if the move was not successful. On the other hand, instead of being able to read and write before moving, migrants could have become literate once they were already living in Madrid. Given that our sample contains information on the year these individuals arrived in Madrid, we are able to compute the time they had been residing there, and to test the robustness of our results, controlling for these issues.

In order to do so, we rely on the individual-level information about migrants living in Madrid and estimate the likelihood of being literate, taking into account the time since they arrived there, measured in years. We also consider the potential non-linearities associated with this type of variable by including its square. Table S8 in the online appendix presents the results of estimating logit regressions for men and women for each of our periods. As evidenced there, years in Madrid increased the probability of being literate, although at a decreasing rate. All these specifications include provincial dummies, so we then predict the migrants' literacy levels net of the influence of time residing in Madrid. Lastly, we use these estimations to obtain average literacy levels by province of origin and re-compute the literacy gap. Tables S9 and S10 in the online appendix replicate the empirical exercise reported in tables 1 and 2 but using the new estimates. It should be noted that the number of observations is now reduced because, first, if all migrants that come from a particular province are literate, it is impossible to compute the influence of the time spent in Madrid on their likelihood of being literate, so those observations are dropped; and second, because the information on migrants' length of residence in Madrid is not available for all individuals in our sample. Despite the loss of degrees of freedom (especially in the case of the table for men), the results largely confirm our previous discussion.

V

Internal migrants moving to Madrid, the Spanish capital city, during the late nineteenth and early twentieth century were positively self-selected. Analysing the variation in the literacy gap between these migrants and those who stayed behind shows that the stock of previous migration, the so-called family and friends effect, was a strong factor in allowing less literate individuals to join the migratory process. Interestingly, while the distance to Madrid barely influenced the profile of male migrants moving there, it clearly accentuated female positive selection, especially

up to the late nineteenth century. That effect nonetheless was largely mitigated during the early twentieth century and eventually even reversed as the transport infrastructure improved and allowed less skilled migrants in more distant regions to get to Madrid more easily.

The empirical analysis carried out here also evidences that internal migratory patterns were highly complex, and that migrants' selectivity depended on the context of both the sending and the receiving regions. In this regard, the existence of other urban centres in the same province or in neighbouring regions usually attracted, on average, less literate migrants. Taken together, the distance to Madrid and the existence of different migratory basins, including the possibility of rural–urban migration within the same province, generally confirms the presence of two types of internal migration: short-distance migration, which was dominated by unskilled workers who supplied the labour force of growing neighbouring cities, and medium- and long-distance migration, which was more prone to positive selection in terms of literacy. The effect of distance, however, appears to diminish, and even reverse, over time as new and cheaper means of transport became available.

Footnote references

Abramitzky, R., Platt Boustan, L., and Eriksson, K., 'Europe's tired, poor, huddled masses: self-selection and economic outcomes in the age of mass migration', *American Economic Review*, 102 (2012), pp. 1832–56.

Baines, D., *Migration in a mature economy: emigration and internal migration in England and Wales*, 1861–1900 (Cambridge, 1985).

Baines, D., 'Population, migration and regional development, 1870–1939', in R. Floud and D. McCloskey, eds., *The economic history of Britain since 1700, 2: 1860–1939* (Cambridge, 1994), pp. 29–61.

Beine, M., Docquier, F., and Özden, Ç., 'Diasporas', *Journal of Development Economics*, 95 (2011), pp. 30–41.

Beltrán Tapia, F. J., 'Enclosing literacy? Common lands and human capital in Spain, 1860–1930', *Journal of Institutional Economics*, 9 (2013), pp. 491–515.

Beltrán Tapia, F. J. and Martínez-Galarraga, J., 'Land access inequality and education in pre-industrial Spain', *University of Oxford discussion papers in economic and social history*, 137 (2015).

Borderías, C., Pérez-Fuentes, P., and Sarasúa, C., 'Gender inequalities in family consumption: Spain 1850–1930', in T. Addabbo, M.-P. Arrizabalaga, C. Borderías, and A. Owens, eds., *Gender inequalities, households and the production of well-being in modern Europe* (Farnham, 2010), pp. 179–96.

Borjas, G. J., 'Self-selection and the earnings of immigrants', *American Economic Review*, 77 (1987), pp. 531–53.

Boyer, G. R., 'Labour migration in southern and eastern England, 1861–1901', *European Review of Economic History*, 1 (1997), pp. 191–215.

Boyer, G. R. and Hatton, T. J., 'Migration and labour market integration in late nineteenth-century England and Wales', *Economic History Review*, L (1997), pp. 697–734.

Chiswick, B. R., 'Are immigrants favorably self-selected? An economic analysis', in C. B. Brettell and J. F. Hollifield, eds., *Migration theory: talking across disciplines* (New York, 2000), pp. 61–74.

Collantes Gutiérrez, F., 'Las disparidades educativas en la España rural contemporánea, 1860–2000: un análisis comparado de las comarcas montañosas', *Revista de Demografía Histórica*, XXII (2004), pp. 15–52.

Erdozáin Azpilicueta, P. and Mikelarena Peña, F., 'Las cifras de activos agrarios de los censos de población españoles del período 1877–1991. Un análisis crítico', *Boletín de la Asociación de Demografía Histórica*, XVII (1999), pp. 89–113.

García Abad, R., *Historias de emigración: factores de expulsión y selección de capital humano en la emigración a la ría de Bilbao (1877–1935)* (Bilbao, 2005).

González Portilla, M., ed., *Los orígenes de una metrópolis industrial: la ría de Bilbao* (Bilbao, 2001).

González Portilla, M. and García Abad, R., 'Migraciones interiores y migraciones en familia durante el ciclo industrial moderno el área metropolitana de la ría de Bilbao', *Scripta Nova*, X (2006), http://www.ub.edu/geocrit/sn/sn-218-67.htm (accessed on 4 Feb. 2014).

Gould, J. D., 'European inter-continental emigration: the role of "diffusion" and "feedback"', *Journal of European Economic History*, 9 (1980), pp. 267–317.

Grant, O., *Migration and inequality in Germany, 1870–1913* (Oxford, 2005).

Hatton, T. J., 'The cliometrics of international migration: a survey', *Journal of Economic Surveys*, 24 (2010), pp. 941–69.

Hatton, T. J. and Bailey, R. E., 'Natives and migrants in the London labour market, 1929–1931', *Journal of Population Economics*, 15 (2002), pp. 59–81.

Hatton, T. J. and Williamson, J. G., *The age of mass migration: causes and economic impact* (New York, 1998).

Herranz-Loncán, A., 'The Spanish infrastructure stock, 1844–1935', *Research in Economic History*, 23 (2005), pp. 85–129.

Herranz-Loncán, A., 'The spatial distribution of Spanish transport infrastructure between 1860 and 1930', *Annals of Regional Science*, 41 (2007), pp. 189–208.

Humphries, J. and Leunig, T., 'Was Dick Whittington taller than those he left behind? Anthropometric measures, migration and the quality of life in early nineteenth century London?', *Explorations in Economic History*, 46 (2009), pp. 120–31.

Kosack, E. and Ward, Z., 'Who crossed the border? Self-selection of Mexican migrants in the early twentieth century', *Journal of Economic History*, 74 (2014), pp. 1015–44.

Long, J., 'Rural–urban migration and socioeconomic mobility in Victorian Britain', *Journal of Economic History*, 65 (2005), pp. 1–35.

Miguel Salanova, S. de, 'Las raíces de una metrópoli: el centro financiero de Madrid a principios del siglo XX', *Hispania Nova*, 10 (2012), http://hispanianova.rediris.es/10/articulos/10a008.pdf (accessed on 12 March 2014).

Miguel Salanova, S. de, 'Bajo los tejados de Madrid: segregación residencial en el primer tercio del siglo XX', in VV. AA., *I Congreso Histórico Internacional. As cidades na historia: populaçao, 4: Cidade industrial* (Guimaraes, 2013), pp. 237–63.

Mikelarena Peña, F., 'Los movimientos migratorios interprovinciales en España entre 1877 y 1930: áreas de atracción, áreas de expulsión, periodización cronológica y cuencas migratorias', *Cuadernos Aragoneses de Economía*, 3 (1993), pp. 213–40.

Moya, J. C., *Cousins and strangers: Spanish immigrants in Buenos Aires*, 1850–1930 (Berkeley, Calif., 1998).

Nadal, J., 'El desenvolupament de l'economia valenciana a la segona meïtat del segle XIX: una via exclusivament agrària?', *Recerques*, 19 (1987), pp. 115–32.

Núñez, C. E., *La fuente de la riqueza: educación y desarrollo económico en la España contemporánea* (Madrid, 1992).

Núñez, C. E., 'Within the European periphery: education and labor mobility in twentieth-century Spain', *Paedagogica Historica*, 39 (2003), pp. 621–49.

O'Rourke, K. H. and Williamson, J. G., *Globalization and history: the evolution of a nineteenth-century Atlantic economy* (Cambridge, Mass., 1999).

Otero Carvajal, L. E. and Pallol Trigueros, R., 'El Madrid moderno, capital de una España urbana en transformación, 1860–1931', *Historia Contemporánea*, 39 (2009), pp. 541–88.

Oyon, J. L., Maldonado, J., and Griful, E., *Barcelona 1930: Un atlas social* (Barcelona, 2001).

Pallol Trigueros, R., Carballo Barral, B., and Vicente Albarrán, F., 'Inmigración y mercado en el Madrid de la segunda mitad del siglo XIX', *Revista de Demografía Histórica*, XXVIII (2010), pp. 131–66.

Paluzie, E., Pons, J., Silvestre, J., and Tirado, D. A., 'Migrants and market potential in Spain over the twentieth century: a test of the new economic geography', *Spanish Economic Review*, 11 (2009), pp. 243–65.

Pons, J., Paluzie, E., Silvestre, J., and Tirado, D. A., 'Testing the new economic geography: migrations and industrial agglomerations in Spain', *Journal of Regional Science*, 47 (2007), pp. 289–313.

Quiroga, G., 'Literacy, education and welfare in Spain (1893-1954)', *Paedagogica Historica*, 39 (2003), pp. 559–619.

Reher, D. S., *Town and country in pre-industrial Spain: Cuenca*, 1550–1870 (Cambridge, 1990).

Reher, D. S., 'Ciudades, procesos de urbanización y sistemas urbanos en la Península Ibérica, 1550–1991', in M. Guàrdia, F. J. Monclús, and J. L. Oyón, eds., *Atlas histórico de ciudades europeas: Península Ibérica* (Barcelona, 1994), pp. 1–29.

Sánchez-Alonso, B., *Las causas de la emigración española*, 1880–1930 (Madrid, 1995).

Sánchez-Alonso, B., 'European emigration in the late nineteenth century: the paradoxical case of Spain', *Economic History Review*, LIII (2000), pp. 309–30.

Sánchez-Alonso, B., 'Those who left and those who stayed behind: explaining emigration from the regions of Spain, 1880–1914', *Journal of Economic History*, 60 (2000), pp. 730–55.

Sandberg, L. G., 'Ignorance, poverty and economic backwardness in the early stages of European industrialization: variations on Alexander Gerschenkron's grand theme', *Journal of European Economic History*, 11 (1982), pp. 675–97.

Sarasúa, C., *Criados, nodrizas y amos. El servicio doméstico en la formación del mercado de trabajo madrileño*, 1758–1868 (Madrid, 1994).

Silvestre, J., 'Viajes de corta distancia: una visión espacial de las migraciones interiores en España, 1877–1930', *Revista de Historia Económica*, XIX (2001), pp. 247–86.

Silvestre, J., 'Internal migrations in Spain, 1877–1930', *European Review of Economic History*, 9 (2005), pp. 233–65.

Silvestre, J., 'Las migraciones interiores durante la modernización económica de España, 1860–1930', *Cuadernos Económicos de ICE*, 69 (2006), pp. 157–82.

Silvestre, J., 'Temporary internal migrations in Spain, 1860–1930', *Social Science History*, 31 (2007), pp. 539–74.

Silvestre, J., Ayuda, M. I., and Pinilla, V., 'The occupational attainment of migrants and natives in Barcelona, 1930', *Economic History Review*, 68 (2015), pp. 985–1015.

Sorribes, J., *Crecimiento económico, burguesía y crecimiento urbano en la Valencia de la Restauración*, 1874–1931 (Madrid, 1983).

Sorribes, J., 'La transición urbana: métodos y resultados. Valencia 1874–1931', in J. García Delgado, ed., *Las ciudades en la modernización de España: los decenios interseculares* (Madrid, 1992), pp. 197–222.

Spitzer, Y. and Zimran, A., 'Migrant self-selection: anthropometric evidence from the mass migration of Italians to the United States, 1907–1925', *research paper* (2015), http://aez.econ.northwestern.edu/spitzer_zimran_italian_stature.pdf (accessed on 15 October 2015).

Stolz, Y. and Baten, J., 'Brain drain in the age of mass migration: does relative inequality explain migrant selectivity?', *Explorations in Economic History*, 49 (2012), pp. 205–20.

Taylor, A. M. and Williamson, J. G., 'Convergence in the age of mass migration', *European Review of Economic History*, 1 (1997), pp. 27–63.

Tortella, G., 'Patterns of economic retardation and recovery in south-western Europe in the nineteenth and twentieth centuries', *Economic History Review*, XLVII (1994), pp. 1–21.

Vicente, F. 'Los motores del crecimiento demográfico de Madrid (1860–1930). Flujos migratorios y procesos de segregación en los nuevos espacios urbanos', in A. Pareja, ed., *El capital humano en el mundo urbano: experiencias desde los padrones municipales (1850–1930)* (Bilbao, 2011), pp. 259–82.

Official publications

Dirección General del Instituto Geográfico y Estadístico, *Censo de la población de España, 1887* (Madrid, 1891).

Dirección General del Instituto Geográfico y Estadístico, *Censo de la población de España, 1900* (Madrid, 1902).

Dirección General del Instituto Geográfico, Catastral y de Estadística, *Censo de la población de España, 1930* (Madrid, 1932).

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Table S1. Migrants' literacy rates in Madrid, men aged 16–30 (by province of origin)

Table S2. Migrants' literacy rates in Madrid, women aged 16–30 (by province of origin)

Table S3. Literacy gap between migrants and non-movers, by province of origin (percentage points)

Table S4. Summary statistics

Table S5. Description of the variables employed

Table S6. Determinants of migrants' selectivity, men (standardized coefficients)

Table S7. Determinants of migrants' selectivity, women (standardized coefficients)

Table S8. Migrants' literacy and time living in Madrid, aged 16–30, logit regression

Table S9. Determinants of migrants' selectivity controlling for time residing in Madrid, men

Table S10. Determinants of migrants' selectivity controlling for time residing in Madrid, women