DISCUSSION PAPER SERIES

DP17668

HOME, SWEET HOME: RETURNS TO RETURNING IN THE AGE OF MASS MIGRATION

Olof Ejermo, Kerstin Enflo, Björn Eriksson and Erik Prawitz

ECONOMIC HISTORY

ISSN 0265-8003

HOME, SWEET HOME: RETURNS TO RETURNING IN THE AGE OF MASS MIGRATION

Olof Ejermo, Kerstin Enflo, Björn Eriksson and Erik Prawitz

Discussion Paper DP17668
Published 14 November 2022
Submitted 14 November 2022

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

HOME, SWEET HOME: RETURNS TO RETURNING IN THE AGE OF MASS MIGRATION

Abstract

Studying migrants from Sweden to the United States, we provide new evidence on return migration during the Age of Mass Migration. Focusing on a sample of migrants and stayers observed in childhood, we document limited effects on income and occupational upgrading, but large effects on wealth. Male returnees held about twice as much wealth as stayers and about 40 percent more than staying brothers. These effects were likely driven by accumulated savings overseas, rather than inheritance or an income premium back home. For female returnees, wealth effects are of similar magnitude, but appear to be realized primarily through marriage.

JEL Classification: N13, F22, O15, J24, J62

Keywords: N/A

Olof Ejermo - olof.ejermo@ekh.lu.se
Department of Economic History, Lund University

Kerstin Enflo - kerstin.enflo@ekh.lu.se
Department of Economic History, Lund University and CEPR

Björn Eriksson - bjorn.eriksson@ekh.lu.se
Department of Economic History, Lund University

Erik Prawitz - erik.prawitz@ifn.se
Research Institute of Industrial Economics (IFN), Sweden

Acknowledgements
We acknowledge generous funding by the Ragnar Söderberg foundation (grant no. E18/16). Prawitz also acknowledges generous support from the Johan and Jakob Söderbergs Stiftelse.

Home, Sweet Home: Returns to Returning in the Age of Mass Migration*

Olof Ejermo[†1], Kerstin Enflo[‡2], Björn Eriksson[§2], and Erik Prawitz[¶3]

[1]Lund University, RATIO & Stockholm School of Economics
[2]Lund University
[3]Linnaeus University & Research Institute of Industrial Economics

October 7, 2022

Abstract

Studying migrants from Sweden to the United States, we provide new evidence on return migration during the Age of Mass Migration. Focusing on a sample of migrants and stayers observed in childhood, we document limited effects on income and occupational upgrading, but large effects on wealth. Male returnees held about twice as much wealth as stayers and about 40 percent more than staying brothers. These effects were likely driven by accumulated savings overseas, rather than inheritance or an income premium back home. For female returnees, wealth effects are of similar magnitude, but appear to be realized primarily through marriage.

Keywords: emigration, returnees, selection, return location, occupational and social upgrading, income and wealth

JEL Classification: N13, F22, O15, J24, J62

*We acknowledge generous funding by the Ragnar Söderberg foundation (grant no. E18/16). Prawitz also acknowledges generous support from the Johan and Jakob Söderbergs Stiftelse.

[†]olof.ejermo@ekh.lu.se
[‡]kerstin.enflo@ekh.lu.se
[§]bjorn.eriksson@ekh.lu.se
[¶]erik.prawitz@ifn.se

1 Introduction

During the Age of Mass Migration (c. 1850-1920) an estimated 30 million Europeans left for the New World. The ebb of migration also generated a significant return flow. Studies indicate that as many as 40–60 percent of the cross-Atlantic emigrants eventually returned to Europe (Kuznets and Rubin, 1954; Wyman, 1996; Hatton and Ward, 2019), and some suggest even higher rates (Bandiera, Rasul, and Viarengo, 2013). The historical figure is by no means exceptional. Current estimates suggest that half of the world's migrants part from their destination country within five years, and that many choose to go back home (Dumont et al., 2008). Despite overwhelming evidence of persistently large circular migration flows, there is surprisingly scarce evidence on the returns to return migration.[1]

This study provides empirical evidence for the historical case of Sweden. Similarly to many other European countries, Sweden saw migrants leave en masse. More than 1 million of its inhabitants left the country 1860–1914, many with the hope of a richer life in America.[2] It is believed that about a fifth of Swedish migrants eventually returned, but beyond rough aggregate figures and anecdotal evidence, little is known of them. By linking individuals between Swedish full-count censuses, emigration and immigration registers, and death records, we trace the life trajectories of both male and female migrants. This allows us to compare return migrants with individuals that stayed in Sweden, controlling for a rich number of confounding individual pre-migration characteristics. In particular, by comparing siblings with different migration histories, we study the returns to returning in terms of outcomes measured in adulthood, such as occupation, income, wealth, marriage, and location. To our best knowledge this material provides the most detailed exposition of the lives and characteristics of the return migrants from the Age of Mass Migration that has been performed to date.

A core topic in the migration literature is 'self-selection'. In other words, the choice to both emigrate and eventually return home is viewed as the outcome of individual preferences and characteristics rather than as an exogenous event. We therefore start by giving an overview of the potential selection patterns among

[1] For the Age of Mass Migration, studies by Abramitzky, Boustan, and Eriksson (2019) on Norwegian return migrants and Fernihough and O'Grada (2019) on the Irish are recent exceptions.

[2] In fact, Swedish children dubbed the new land as *'mer rika*, literally meaning "more rich" in Swedish (Wyman, 1996).

emigrants and returnees before investigating the returns to returning. To measure selection, we concentrate on the parent's occupational class (observed in childhood), which is less likely to be affected by the migration decision of the individual as compared to the occupational class of the migrants themselves. Consistent with empirical evidence from this historical period of relatively open migration, we find negative emigrant selection (Abramitzky, Boustan, and Eriksson, 2012; Mokyr and Ó Gráda, 1982; Connor, 2019; Spitzer and Zimran, 2018).[3] This is to a large extent driven by an underrepresentation of individuals born in white-collar families. However, in contrast to earlier evidence (Ward, 2017; Abramitzky, Boustan, and Eriksson, 2019), we find that return migrants were positively selected among emigrants in terms of family backgrounds.[4] Our results suggest the return of 'the best of the worst', which squares well with a Roy model of migration, given that Sweden was more unequal in terms of income relative the United States during the period (see Borjas, 1987; Borjas and Bratsberg, 1996).

During the remainder of our analysis, we study the returnees' endeavors once settled back in Sweden. In light of the observed selection patterns, we focus on a subset of individuals observed in their childhood homes in the full-count census of 1880. This not only allows us to compare individuals from similar childhood backgrounds, we may also rule out between-family selection by comparing siblings that migrated to the United States after 1880 (and later returned) to those that stayed in Sweden. We consider a set of outcomes in the census of 1910, when the return migrants were about 30–45 years old, and later in the 1930 census, when they had reached the age of 50–65.

First, studying the economic returns to migration, we find that male returnees did not earn significantly more than their staying brothers.[5] This holds both when measuring the income in terms of occupational income scores in the 1910 census as well as when studying their actual income in the 1930 census.[6] Similarly, when we analyze inter-generational mobility by ranking both fathers and

[3]Spitzer and Zimran (2018) find negative emigrant selection at the national level, but positive selection at the local level.

[4]An exception is Fernihough and O'Grada (2019), who finds positively selected Irish return migrants when comparing them to the Irish 1911 census population.

[5]In fact, if not adjusting agricultural income for in-kind compensation (following Collins and Wanamaker, 2022), male returnees earn less than their staying brothers in 1910.

[6]While not robust to sibling fixed effects, we find a positive association between returnees and income in 1930 when including fixed effects for the childhood municipality and social class of the household head in 1880.

children in terms of their income status in the population, we see relatively little occupational upgrading among returnees, although patterns differ by gender. Male returnees are about as likely to experience occupational upgrading as their staying brothers, while female returnees were more upwardly mobile than their staying sisters.[7] Female returnees saw both higher absolute and relative intergenerational mobility as measured in rank-rank correlations similar to Chetty et al. (2014). In particular, female returnees born to low or unskilled fathers saw the highest social returns, to a large extent driven by marrying a white-collar husband. While male returnees from the lower echelons of society were able to advance to being landowning farmers, we do not see any general upward mobility in terms of occupational status as they are slightly less likely to advance in non-agricultural occupations.

Second, while we find little evidence of any substantial returns in terms of occupation or income, at least for men, we document large effects on wealth. Returnee men recorded in the 1930 census held roughly twice as much wealth as stayers from the same place and social origin, and almost 50 percent more than their staying brother. Similarly, spouses of female returnees were wealthier than the spouses of their staying sisters.

Exploring underlying mechanisms, we provide suggestive evidence that the differential effect for men regarding wealth is not due to occupational differences or inheritance related to birth order. Instead, the positive effect on wealth is driven entirely by return migrants with longer stays abroad. Similarly, male returnees with longer stays were also more likely to be landowning farmers and to be located in rural areas in their greater childhood region.[8] While longer stays in the United States may have enabled migrants to accumulate both more savings and human capital, the fact that we do not find any positive effects for occupational income is suggestive of a key role for wealth accumulation. Together, our results are consistent with a story of status-upgrading among male returnees into wealthy land-owners that took place within the agricultural sector, not easily captured in income scores or occupational titles.[9]

Our paper is closely related to the literature on migration in the Age of Mass

[7] We calculate the income rank of daughters using the income score of their spouses.

[8] However, consistent with a limited role for the inheritance of their childhood farm, returnees were not more likely than stayers to be found in their childhood municipality.

[9] Female returnees display a contrasting picture, as the positive effect on their spouse's wealth is entirely driven by short-term stays abroad. Their wealth increase is more likely driven by differential marriage patterns. In line with this, we find that female returnees (as well as male) were more attractive on the marriage market, more broadly capturing their higher social status.

Migration (see, e.g., Hatton, Williamson, et al., 1998; Bohlin and Eurenius, 2010; Abramitzky, Boustan, and Eriksson, 2012, 2013, 2014; Sequeira, Nunn, and Qian, 2017).[10] This literature has long focused on the receiving country, but a growing number of studies have documented the effects also on the sending countries (see, e.g., Hatton, Williamson, et al., 1998; Karadja and Prawitz, 2019; Andersson, Karadja, and Prawitz, 2021). We contribute to this literature by taking a close look at the individuals that eventually returned home, their life trajectories and their economic and social returns to returning. While the emigration literature generally has documented large returns to emigrating, our study provides evidence that some of these returns were accumulated and brought back to the Old World.[11]

More narrowly, our paper relates to the literature on return migration (see, e.g., Dustmann and Kirchkamp, 2002; Saxenian, 2007; Rooth and Saarela, 2007; Dustmann, Fadlon, and Weiss, 2011; Wahba and Zenou, 2012) and, in particular, to studies in the context of the Age of Mass Migration (Ward, 2016; Abramitzky, Boustan, and Eriksson, 2019; Fernihough and O'Grada, 2019). Most closely related to our paper is Abramitzky et al. (2019) who study the return migration of Norwegian emigrants to the United States. They find that return migrants were negatively selected among emigrants both in the US labor market and, in contrast to our study, in terms of pre-emigration characteristics. However, upon returning they fared better than permanent stayers despite being negatively selected (see also Abramitzky et al., 2012). Similarly to our findings they document that return migrants were more likely than stayers to own a farm.

We contribute to the literature in several ways. First, unlike many historical migration samples, our data is drawn from records that were updated on a running basis by the church, implying that names, birth dates and birth parishes were recorded at high precision. The detailed information in our sources allow for a linking rate of 60-70% between sources, which is much higher than the typical 20-30% reported for other countries (Long, 2005; Long and Ferrie, 2013; Abramitzky et al., 2012). We are able to identify a large number of return migrants by linking emigration records to the complete Swedish death index. Our implied (lower bound) return rate is 14 % of all emigrants, defined as the share of emigrants which may be identified as having returned by virtue of being recorded

[10] Abramitzky and Boustan (2017); Hatton and Ward (2019) provide overviews.

[11] For example, Abramitzky et al. (2012) finds a return of emigration of about 70 percent as measured in income scores, and in a recent working paper using linked Swedish and US census data, Castillo (2022) finds similar returns to Swedish emigrants in the US.

as deceased in Sweden.[12] Our implied return rate comes close to the commonly cited 19% for the case of Sweden (Wyman, 1996, p. 12), suggesting that we have been able to trace back the vast majority of all return migrants that appear in the Swedish records.[13]

Second, to the best of our knowledge, this is the first paper to present estimates of the intergenerational mobility of return migrants in the Age of Mass Migration.[14] By linking individuals to their family background, and especially by adding information about the occupational status of the household head, we are able to compare migrating to staying siblings to obtain a cleaner estimate on how they fared in terms of intergenerational mobility. With our detailed records we are also able to include women in the analysis – a large but frequently neglected group in historical migration studies. Women are often difficult to link across longitudinal samples due to name changes associated with marriage. In the Swedish sources women appear with their maiden name, even after marriage, which allows for women to be linked between different sources to nearly the same extent as men.

Third, we provide estimates on the monetary returns to returning, not only based on occupational income scores, but in terms of both registered income and wealth. We believe that these outcome measures provide a major advantage since individuals earn different wages within occupations, and this is not captured in occupational income scores (something that is especially true for the large group of returnees found in agriculture). In addition, registered wealth allows for a richer picture of the monetary returns to returning, since it holds information about lasting wealth effects from income accumulated abroad.

2 Background

Mid-19$^{\text{th}}$ century Sweden was a rural and poor country in the periphery of Europe. With about 70% of its population in agriculture and a GDP per capita

[12]To provide robustness to the sensitivity of this result, we count all duplicates links to the death index (which we discard to err on the side of caution in our main sample) as one possible returnee. It is comforting for our sample, that this has virtually no effect on the calculated returnee rate.

[13]Bandiera et al. (2013) find that Russian, Irish and Scandinavian migrants were less likely to return compared to migrants from Southern Europe, and Ward (2017) reports that only 14% of those arriving at Ellis island in the early 20$^{\text{th}}$ century eventually returned despite at least 20% stated an intention to do so. Thus, many Scandinavians unexpectedly stayed in the United States, perhaps because outcomes were better than expected.

[14]Abramitzky et al. (2021) studies the intergenerational mobility of immigrants to the United States during our era.

far below the European average it had just started to make the journey towards modern economic growth (Schön and Krantz, 2012). Necessary economic adjustments, such as trade and business liberalizations as well as the abolishment of internal passport laws had just been implemented. The combination of liberal migration laws and a backward economy resulted in a massive flow towards the New World. Emigrants left in waves, with a first peak that was sparked by several years of bad harvests in the late 1860s. Between 1850 and 1920 more than 1 million Swedes left, a significant loss for a country of about 5 million inhabitants.

The emigrants, as well as the stayers, are well documented in historical sources. As a result of the superior taxing and military organisation of the 17$^{\text{th}}$ century, the Swedish population was carefully registered by its authorities. In each parish, local priests was in charge of recording all inhabitants with dates of births, deaths, marriages in church books that have formed the basis for the world's oldest running population records. Migrations in and out of the parish was also noted in the registers, often identifying individuals by their name, place and date of birth.

Out of the many emigrants, aggregate statistics suggest that about one fifth returned (see Wyman, 1996). Return rates increased towards the late 19$^{\text{th}}$ century as means of transport improved and emigrants increasingly could work for short periods in blue-collar industries in US cities. By that time Sweden had transformed into a fast growing industrial nation, with high social and geographical mobility patterns (Berger et al., 2021; Enflo et al., 2014). Upon return, many travellers encountered a population that had achieved substantial success at home, and the question was whether the returnees had fared better or worse than the stayers. Although folklore and popular works (Henricson and Lindblad, 1995) emphasize success cases, this could result from more noticable individuals, leading to biased perceptions.

3 Data

We combine several historical data sources: full-count censuses 1880–1910, a 33% sample of the 1930 census, the US census of 1900, death registers, and migration records.[15] These are linked together at the individual level to create high-quality longitudinal samples of stayers, emigrants and return migrants.

[15]The sample of the 1930 census were obtained from the Swedish National Archives. It covers all individuals resident in parishes which have so far been completely digitized and coded.

3.1 Linking data sources

At the heart of our study lies the linking of return migrants between sources. To identify the returnees, we link emigration records to the Swedish Death Index (Federation of Swedish Genealogical Societies, 2018) of all people deceased in Sweden, available from 1860 onwards, and to the censuses. Since both the emigration registers and the death index report birth year, month and date, unlike the censuses which only report birth year, this allows for more return migrants to be identified compared to only using the censuses.[16] Once individuals are linked to the death index, the detailed recording of birth parish allows for linking individuals back to the censuses with high precision.

In general, to identify the same individual in different sources we rely on probabilistic linking methods. In order for an individual to be linked between two sources the following criteria must be met: (a) match exactly on a number of index variables (b) meet a threshold for the similarity between names (c) constitute an unambiguous link between two individuals.

The linking procedure starts with identifying variables suitable for matching individuals. In order to minimize bias, only variables that are time-invariant over the life course should be considered (see Ruggles 2006). We use sex, birth date and birth place as index variables, meaning that two records have to match exactly on these variables in order to be considered a candidate for further evaluation. Importantly, birth place is recorded at the parish level, of which there were about 2,500 in 19th century Sweden. Moreover birth years are very accurate, which mitigates potential problems of false positives which may substantially overestimate migrant social mobility (Massey, 2017; Bailey et al., 2020), especially since many migrant sources are plagued by self-reported ages from individuals lacking numeracy skills (i.e. problems with age-heaping) and extreme name-spelling variations among migration officers unfamiliar with foreign naming practices (Anbinder et al., 2021). The detail and accuracy of the indexing variables means that each emigrant is only ever compared to a small number of possible matches.

The second step in the linking process involves separating true links from false

[16] However, when studying emigrant selection patterns in Section 4, we link individuals between the emigration records (Emibas) to the 1880, 1890, 1900 and 1910 full-count nominal censuses (The Swedish National Archives et al., 2011a,b, 2014, 2016). Although it would be possible to add the links between the emigration register and the death index to add more observations, this would lead to an unrepresentative sample, because identification of emigrants in the census would in these cases be conditional on return migration. See Appendix 8 for a more detailed discussion.

among the candidates generated by the initial matching on the index variables. To adjudicate between the potential matches, we rely on comparisons of first and last names. To reduce the influence of minor differences in spelling or transcription errors, we first standardize names by removing nobility prefixes, patronymic suffixes and all non-alphabetic characters.

Still, names are recorded with a certain degree of imprecision in the sources. Allowance must therefore be made for the fact that the name of the same individual may not be written identically in the emigration register, censuses and the death index. We therefore employ the Jaro-Winkler (JW) algorithm to estimate the similarity of first and last names. The JW algorithm assigns a similarity score between 0 (no similarity) and 1 (identical) for two text strings by comparing common characters, common character pairs and transpositions. Moreover, initial characters weigh more heavily in the algorithm and accounts for the fact that irregularities are more common at the end of longer strings than in shorter.

We consider individuals linked if the JW-score for both the first and the last names exceeds a given threshold, and that the link is non-ambiguous, i.e., we require that no other plausible competing link exists. Given the problematic nature of false positive links, we choose to err on the conservative and discard all ambiguous links. This means that links are only retained if an emigrant is linked to one individual in the censuses or death index. It should however be noted that we allow for multiple emigration records to be linked to one census or death index observation since individuals could have migrated multiple times.

We complement the linked migration records with links between the censuses and the death index. Each census has been linked to the subsequent census, and the death index, following the procedure described above. The links are combined to create a panel that follows individuals across the censuses, ending with the death event in case where a link has been made to the death index. For a more detailed description of the census and death index linking see Eriksson (2015); Dribe and Eriksson (2018).

In terms of linkage rates our algorithm performs well. Table 1 shows that the linkage rate between the sources ranges between 60% and 70%. These rates are comparable to what is typically achieved when linking between Swedish historical censuses (Eriksson, 2015; Wisselgren et al., 2014) and exceeds rates achieved for US and UK censuses (23% in both Long, 2005; Long and Ferrie, 2013). It is also notably higher than the rate achieved on Norwegian data by Abramitzky et al. (2012, 29%).

We also use links from the list of Swedish emigrants to the US census in 1900 (Ruggles et al., 2021) as created in Castillo (2022). Different from the linking of the Swedish censuses and the Death index, this linking had to rely only on one name (in Swedish censuses often three names) and the US census only states "Sweden" as birthplace for Swedish migrants (Castillo, 2022), which lowers the linking rate compared to those that rely on Swedish data.[17]

3.2 Migrant definition and migration rates

As noted above, we define a return migrant as an individual that has both been observed in the emigration register and in the death index.[18]

Our return migrant definition allows us to calculate return rates by departure year. Figure 1 graphs the number of emigrants to the US together with an estimation of the share that eventually returned by emigration year.[19] Following our definition that returnees need to be found in the death index, roughly 14 percent returned during the entire period. This should be viewed as a lower bound as not all returnees have been found. However, this is not far from the often quoted 19 percent (Wyman, 1996, p. 12) and in line with evidence based on those arriving at Ellis island in the early 20$^{\text{th}}$ century (Ward, 2017).[20]

As shown in Figure 1, the tendency to return varied considerably throughout the period. Emigrants which departed during the peak periods were less likely to return, but over time a trend shows that the share identified as having returned increases from around 10 to more than 20 percent. This may suggest that the nature of emigration changed from being primarily permanent to increasingly becoming a temporary move for many emigrants. Noteworthy, the trend is also in line with the falling transportation costs and increased wage convergence between Sweden and the US over the same period.

We next explore how migration rates varied spatially within Sweden. Figure 2 draws maps of the geographical distribution of cumulative emigration and return migration flows 1880-1910 expressed in relation to population size by municipality in 1880. Map (A) shows emigration from the origin, (B) returnees by origin, and (C) by return destination. The emigration patterns observed in (A) are to a

[17] Roughly 16% of the Swedish-born men in the US Census of 1900 are linked to the emigrant register.

[18] We define an emigrant in the census as an individual linked to the emigration register.

[19] The number of emigrants come from the official statistics (BISOS).

[20] While Bandiera et al. (2013) suggest a much higher return rate for Europeans in general, they find that Russian, Irish and Scandinavian migrants were less likely to return compared to migrants from Southern Europe.

large extent mirrored by the share of return migrants in the population (C). As a consequence, locations which initially lost large numbers because of emigration, eventually ended up with a sizable share of return migrants in the population.[21]

3.3 Childhood sample

To compare individuals with similar childhood backgrounds, we focus on a subset of individuals below 15 years of age and observed with a household head in the 1880 full-count census. In order to estimate the returns to return migration in adulthood, in which we compare returnees from North America to permanent stayers in Sweden (i.e. non-migrants), we focus on a sample of individuals linked to the censuses of 1910 and 1930 when these individuals are 30–45 and 50–65 years old, respectively.[22] To account for any potential sample selection bias, we show that our main results are robust to weighting the sample by the inverse probability that an individual is observed in the 1910 census using demographic observables on the household head in 1880 in Appendix Table C.4.

3.4 Income, wealth and other outcomes in adulthood

The census data 1880-1910 provides us with information about occupation, occupational class, location and marital status. Moreover, using our links to the death index, we obtain the year of death for all linked individuals that die in Sweden. Lastly, for 46 percent of identified return migrants, we obtain information on the length of time abroad, using the time of departure and arrival.[23]

In addition to the above-mentioned information, the census of 1930 also includes information on individual income and wealth. This provides us with a rare opportunity to measure the monetary returns to returning in terms of both observed income (flow) and wealth (stock). Figures B.1 and B.2 provides the distribution of income and wealth in the 1930 census for both returnees and permanent stayers in Sweden, and for both men and women, for our main sample of analysis consisting of children observed in the 1880 census.

[21]The geographical pattern of emigration is well-known in Swedish historiography with large emigration rates observed primarily in south–central Sweden, along the Swedish–Norwegian border, and in local well known hot-spots including the island of Öland and the Bjäre peninsula (Persson, 2007). By contrast urbanized regions, including Stockholm, Malmoe and Gothenburg, had very little emigration relative to the population.

[22]In general, we include birth-year fixed effects to compare individuals from the same cohort.

[23]For cases of repeat migration we use the earliest date of emigration and the last date of return migration that we observe.

With the observed data on income, occupation and location in the 1930 census, we may construct income scores to use in the earlier censuses. This enables us to both rank fathers of prospective emigrants in the 1880 census and men in adulthood found in the 1910 census.[24] To construct income scores, we follow Ward (2020) and create county-occupation cells using the three digit code from the Historical International Standard Classification of Occupations (HISCO). When cells include at least 30 men, we assign the mean value of income to the county-occupation combination. When cells have less than 30 men, we continue by assigning the mean national income to this occupation as long as there is at least 30 men with this three digit occupation. In cases where there are fewer than 30 men, we use the mean income at the one-digit level of HISCO.

For both incomes in the 1930 census and as a last step in the creation of income scores in earlier censuses, we adjust agricultural earnings upwards to address that some compensation was in-kind.[25]

4 Migrant selection

4.1 Estimating equations

We estimate migrant selection by using the occupation of the household head in childhood as recorded in the census of 1880. This allows us to assess selection in terms of observables that are less prone to be endogenous to the migration decision itself. For instance, individuals planning to migrate may be drawn to other occupations than individuals planning to stay.[26]

The baseline estimation model is a linear probability model of the form:

$$Emigrate_i = \sum_{j=1}^{12} \beta_j HC_{j,i} + \chi_i + \upsilon_i, \tag{1}$$

where $Emigrate_i$ is a dummy variable that takes the value 1 if individual i ever emigrates. Here, $HC_{j,i}$ are a set of indicator variables capturing the social class of the household head in 1880, measured by $j = 1, ..., 12$ main categories of the

[24]Since female wage earners are relatively few, we cannot construct reliable income scores for women.

[25]In particular, we follow Collins and Wanamaker (2022) and inflate the earnings of both farmers and farm hands by 35 and 19 percent, respectively. In Appendix B.1, we show the robustness of our results when altering this adjustment.

[26]To provide a fuller picture of occupational selection, we nevertheless document migrant selection in terms of the occupation of the individual herself for a sample of all working-age individuals in Appendix 8, Section B.1. An advantage of this is that we can assess selection over time during for different emigration decades.

HISCLASS-scheme.[27] Moreover, we include a set of fully saturated individual controls in χ_i: indicators for being the eldest brother or sister, respectively, as well as age and birth-order fixed effects.[28]

Similarly, we study return migrant selection by estimating the corresponding model for the sample of emigrants:

$$Return_i = \sum_{j=1}^{12} \gamma_j HC_{j,i} + \chi_i + \nu_i, \tag{2}$$

where $Return_i$ is a dummy variable that takes the value 1 if an emigrant returned to Sweden. Thus, we compare return migrants with never-returning emigrants who stayed in America. In both models we calculate standard errors clustered at the level of the household head.

4.2 Results

4.2.1 Emigrant selection

The upper part of Figure 3 documents the selection of emigrants based on the occupational class of the household head observed in 1880, prior to emigrating. Results are presented separately for men (panel A) and women (panel B). The omitted category is medium-skilled industrial households heads.

Starting with the estimates shown in colored circles, which depict the results when using regression equation 1 without any control variables, there is a general pattern of negative selection for both men and women. At large, the negative selection is driven by an under-representation of individuals from backgrounds with non-manual household heads. However, it is interesting to note that *within* both manual agricultural classes and industrial classes there is evidence of positive selection. A potential explanation is that individuals from the poorest backgrounds were financially constrained. Adding birth-year and birth-order fixed effects, indicators for being the eldest brother or sister, respectively, as well as fixed effects for the childhood municipality does little to affect our estimates.[29]

[27]The HISCLASS-scheme (Van Leeuwen and Maas, 2011) is based on the HISCO coding scheme (Van Leeuwen et al., 2002) and consists of 12 occupation-based classes which are grouped according to economic sector, whether the occupation is manual or non-manual, its skill level and level of supervision.

[28]Birth order is calculated by ranking children living in the same household in 1880 by birth year. This is admittedly imperfect as some children may have already left the household, but provides us with the best reasonable proxy of birth order.

[29]Appendix Tables B.1 and B.2 display the results in table format.

Estimating selection in terms of the occupational income scores of the household heads, providing us with a more general coefficient of the selection pattern, suggests that emigrants stemmed from family backgrounds with less labor income (see Appendix Tables B.1–B.2). Thus, the overall pattern conforms to typical selection models (see, e.g., Borjas, 1987) with more skilled individuals being less likely to emigrate. This result is also in line with earlier work indicating negative selection of emigrants from the European periphery (Mokyr and Ó Gráda, 1982; Abramitzky et al., 2012; Spitzer and Zimran, 2018; Connor, 2019).

4.2.2 Return migrant selection

The lower parts of Figure 3 displays the selection pattern of return migrants separated for men (panel C) and women (panel D). The pattern suggests positive return selection based on the pre-emigration characteristics of the family household head. The pattern is most pronounced among returnees from manual agricultural backgrounds, where those with unskilled household heads were less likely to return than farmers, and non-manual industrial backgrounds, in which high skilled non-manual workers are some of the most likely to return. In general, adding our set of controls has only minor effects on our estimates.

The positive selection pattern of return migrants is confirmed using occupational income scores, as we find that returnees came from family backgrounds with 3–4 percent higher income scores, as compared to emigrants that stayed overseas in the United States. The results are found in Appendix Tables B.1–B.2.

How were return migrants selected in terms of their characteristics in the US labor market? To explore this, we study the selection patterns in a sub-sample of Swedish emigrants linked to the US census of 1900 (see Appendix Table B.4).[30] We find suggestive evidence of negative return migrant selection in the US labor market, such that migrants that were less successful were more likely to return.[31]

To sum up, while we find signs of negative selection in the US labor market, the selection pattern in terms of *pre-emigration* childhood characteristics in Sweden may be best described as the return of "the best of the worst". This pattern partly off-set the negative outbound selection, such that the return migrants

[30] We thank Castillo (2022) for graciously sharing these data.

[31] The limited sample size calls for caution in interpreting these results, however. Indeed, estimating selection patterns in terms of their father's income score, suggests that the group that is linked to the US census is a selected sample of emigrants.

only compared mildly worse than the staying Swedish population in terms of pre-migrant characteristics.[32]

5 Returns to returning

To study how the return migrants fared after resettling in Sweden, we now turn to our sub-sample of individuals observed as children (below 15 years) with a parent in the full-count census of 1880.

5.1 Estimating equations

We estimate the effects of being a return migrant on different outcomes, measured in adulthood, Y_i, by employing the following equation:

$$Y_{ih} = \beta Returnee_{ih} + \chi_{ih} + \delta_h + \varepsilon_{ih}, \tag{3}$$

for an individual i from childhood household h. In most specifications, we measure Y_{ih} in the census year of 1910, when the individuals are between 30 and 45 years old. Our main variable of interest is $Returnee_{ih}$, which is an indicator variable capturing if the individual has ever migrated and returned home prior to 1910 with stayers in the control group.

To compare individuals from similar upbringings, we either include fixed effects for the childhood municipality and the social class of the household head, or more restrictively, as in equation (3), we include "sibling fixed effects", δ_h, based on the household head in the childhood census of 1880. The latter allows us to estimate within family estimates, absorbing both similar nature and nurture characteristics. In all models, we calculate cluster-robust standard errors at the level of the household head in 1880.[33]

To control for selection within the family, we include a set of fully saturated individual controls in χ_{ih}: indicators for sex, indicators for being the eldest brother or sister, respectively, as well as age and birth-order fixed effects. There are some differences between returnees and stayers in terms of pre-migrant characteristics that we document in Appendix Table B.3. Although return migrants are comparable to stayers in terms of father's income score in 1880, they are somewhat

[32]The selection pattern of return migrants compared to stayers in terms of characteristics observed in the household 1880 is reported in Appendix Figure B.10.

[33]In Appendix Table C.5, we show that our results are robust to alternative standard error calculations.

older than stayers (roughly 1–2 years), and less likely to be the eldest sibling.[34]

Our main sample includes both men and women. To allow for differential effects, we either include an interaction between our sex identifier and $Returnee_i$ or use a split sample approach.

5.2 Results

5.2.1 Income and occupational status

What were the economic returns to returning? We start by studying returnees' income and occupational trajectories at home compared to stayers. Table 2 shows the results from estimating equation 3 with occupational income scores for 1910 as the outcome variable. The results pertain to the returns observed for individuals in their prime working age (30-45 years old). Panel A provides results for men and panel B provides results for women's spouses.[35] Starting with men, the results do not not reveal any income premium from returning.[36] In fact, the occupational income score is slightly lower among returnees than stayers. Column 1 estimates the raw relationship using no controls, suggesting that returnees had about 2.6 percent less income as compared to stayers. However, as documented in the previous section, returnees were differently selected than stayers, such that differences in income score could stem from inherent characteristics rather than from the migration experience itself. Therefore, as a first step, we control for our set of individual controls (age and birth-order fixed effects, and indicators for being eldest brother and eldest sister, respectively). This only strengthens the negative association.

The benefit of being able to observe individuals in their childhood, however, is that we are able to control also for parental background. In column 3, we add fixed effects for the municipality and the social class of the household head observed in the 1880 census and in column 4, we introduce fixed effects for the household head (sibling fixed effects), enabling us to control for any unobserved

[34] Absorbing across family variation by including sibling fixed effects, however, it is the other way around. On average, male and female returnees are 10 and 5 percent more likely to be eldest brother and sister, respectively. Studying the full distribution in Figure B.3, broken down by sex and return migrant status, no stark differences are visible.

[35] Recall that we have too few wage earner among women to construct reliable income scores.

[36] In Appendix Figure C.1, we display how robust results are to altering the adjustment to agricultural income. It shows that the negative effect is robust to a vast range of adjustments. In fact, the less farmer's income is adjusted upwards the more negative is the effect on men's income scores. This is in line with farmers being over-represented among male returnees.

migrant selection across families.[37] Both specifications reduce the negative magnitudes, but do not alter the qualitative relationship. In other words, this is consistent with negative selection across families, such that migrants stemmed from families with lower income.

Turning to women, where we focus on the spouses of female returnees, we find a positive effect for the specifications with no controls, individual controls and childhood household head controls. Similarly to men, the coefficient increases when we introduce childhood characteristics fixed effects in column 3 (municipality and the social class of the household head). The coefficient in column 3 suggests that among married women, the spouses of returnees had occupations with 2.6 percent higher wages compared to spouses of stayers. Adding sibling fixed effects in column 4 drastically increases the standard errors, however, such that results are no longer statistically significant at conventional levels.

A natural question concerns how our results regarding occupational income scores relates to any differences in the underlying occupational sector between returnees and stayers. To investigate this, we follow the scheme in Section 4 and categorize households in three broad occupational classes: agriculture, manual industrial occupations, or non-manual occupations.[38] Figure 5A displays results from regressing indicators for each of these types on return migrant status in separate regressions for our most demanding sibling fixed effects specification. The returnee coefficients for women are displayed in blue diamonds and men in red. While women returnees do not seem to be too different from stayers, male returnees are more likely than stayers to work in agriculture and less likely to work in the other two categories (although the relationship with non-manual household occupations is non-significant at the 5 percent level). Thus, a potential explanation to the decrease in male returnee wages is the fact that they were less likely to be found in the rapidly expanding industrial sector.

In panel B we further document that differences in occupational sectors corresponds well to the location of returnees vis-à-vis stayers. Panel B shows that while female returnees did not differ significantly from stayers, male returnees were more likely to reside in agricultural locations and less likely to reside in

[37] In the latter specification, we drop singleton observations giving us a smaller sample. In Appendix Table C.1, we document that coefficients without sibling fixed effects are similar when studying a subsample of individuals with at least one brother or sister for men and women, respectively.

[38] We define household occupation such that the individual or the spouse has an occupation within the class. This implies that roughly two thirds of women in the sample are given a household occupation through their husband.

towns as compared to stayers.

Occupational income scores may not tell the full story of individuals' income trajectory, since they cannot capture potential wage differentials within occupations in a specific location. Since we have access to actual taxable income in the 1930 census, we are however fortunate enough to be able to replicate the 1910 results in columns 1–4 for actual observed incomes.[39] We display these results in columns 5–6. Comparing individuals with fathers of the same social class in column 5, both male and female returnees appear to be much better off than stayers. Some of this effect appear to be driven by the fact that they are more prone to report an income.[40] However, the positive relation between return migrant status and income is not robust to introducing sibling fixed effects.

5.2.2 Intergenerational mobility

Next, we turn to investigating the social mobility of returnees when we compare their occupations in 1910 with their father's occupation in 1880. While we found little evidence of substantial economic returns to returning in terms of income above, it is an open question to what extent returnees climbed higher than their parents on the social ladder. Ranking individuals by their income score, the results are displayed as binned scatter plots of, so-called, rank-rank correlations of either father-son pairs or father-daughter's spouse pairs, both for returnees and stayers in Figure 4. While stayer estimates are more precise, male returnees appear to roughly line up with stayers.

Married returnee daughters, for which we compare their spouses' income to their fathers, appear to be somewhat more mobile in absolute terms as compared to married stayer daughters. This is seen more formally in Table 3, which documents a positive and significant coefficient on our returnee indicator throughout columns 5 to 7. Comparing women with similar upbringings in column 7, suggests that returnee women increased their position in the income distribution with 2.5 percentiles. However, adding sibling fixed effects, lowers estimates and increases standard errors. Overall, while the levels of our estimates for the male sample are somewhat sensitive to how we calculate farmer's income, as is seen in Appendix Figure C.3, it is clear that the returnee premium in terms of income appear to be higher for women than for men.

[39] Recall that our income scores only measure the mean income for county-3 digit Hisco cells.

[40] Appendix Table B.6, columns 1–4, shows the relationship between returnees and employment. Employment is defined as an indicator equal to one if the individual has any registered occupation (in 1910) or any income (in 1930), and zero otherwise.

Figure 4 documents little evidence of differences in both absolute and relative mobility between stayers and returnees for men. More formally, the latter can be seen from the mostly statistically insignificant slope coefficient interactions with the returnee indicator in columns 2–4 in Table 3. For women, however, the slope coefficient for returnees is somewhat flatter than for stayers. This is also seen in panel B of the same table, which documents a negative slope coefficient interaction ranging between -0.05 and -0.03, the latter insignificant when including fixed effects for the social class of the household head in 1880. Comparing sisters in column 8, we find a negative coefficient of -0.12, which is statistically significant at the ten percent level.

These results provide suggestive evidence that return migration may have been a path to upward mobility for at least women. In Table 4, we take a closer look at a subsample of individuals born to lower- or unskilled fathers.[41] We define indicators for occupational mobility, both downward and upward as well as absolute, where we compare father-son pairs in panel A and father-daughter pairs (again using the occupation of daughter's spouses) in panel B. While there is suggestive evidence of upward mobility for men as long as we omit our sibling fixed effects, estimates are in general insignificant when we compare brothers. While male returnees from poor backgrounds were more likely to advance upwards into farming, they appear less likely than stayers to advance to skilled manual and non-manual occupations (columns 7–9).

As expected, results for women are stronger, both with and without sibling fixed effects. Columns 1–2 in panel B documents that female returnees were 4.1–6.6 percent more mobile as compared to stayers. This is driven entirely by upward mobility as seen from columns 3–6. Moreover, in contrast to male returnees, these returning women from poorer backgrounds were not more inclined than stayers to turn to farming. Instead, upward mobility is driven by female return migrants marrying husbands in non-manual occupations as seen in columns 7–9.

Taken together, we do not find any stark effects on upward mobility for male returnees. For female returnees, there is some evidence of success of women from poorer background and in terms of marrying husbands in non-manual occupations.

[41] Appendix Table B.7 documents these relationships for the entire sample of individuals observed in their childhood home 1880.

5.2.3 Social status

So far, our results on the returns to returning have centered around incomes and occupational incomes scores. Another way to look at social status is to investigate the returnees potential success on the marriage market. Historically, social status was closely related to a person's marriage status and lack of resources (economic or social) could result in difficulties finding a partner (Van Bavel et al., 1998; Dribe and Lundh, 2009). Table 5 investigates if returnees were more likely to marry than stayers. Columns 1–2 estimates the relationship between an indicator for being married in 1910 and our returnee indicator with childhood household ans sibling fixed effects, respectively. They document a weak positive relationship between returnees and marriage. To rule out returnees that were too old, or at least older than the average age of marriage once returning, column 3–6 focuses on the subset of individuals that were observed in the 1900 Swedish census and, in the case of returnees, had returned before 1900. In the last two columns (5–6), we restrict the sample to only individuals that were single in 1900, allowing us to focus on a group of individuals who married later in life.[42] Within these sub-samples, there is a stronger and positive relationship with returnees being roughly 6–10 percent more likely to being married in 1910. The results indicate that returnees enjoyed some kind of social advantage compared to stayers that we may not pick up in incomes or occupational incomes scores.

5.2.4 Wealth

Similarly to other developing nations, 19$^{\text{th}}$ century Sweden was still predominantly rural. Indeed, more than half of the workforce was occupied in agriculture, where income is notably hard to measure, making monetary returns to return migration possibly hard to detect. Luckily, the census of 1930 reported the taxable wealth of individuals allowing us to compare any accumulated capital differences between returnees and stayers.[43] The results are reported in Table 6. Column 1 suggests that returnees owned almost twice the wealth of stayers of the same social background in childhood. The magnitudes are similar for men, women spouses and for women's households. Adding sibling fixed effects in column 2 effectively diminishes this coefficient to about half and doubles standard errors,

[42] Since our sample includes individuals born in 1865–80, they were single at the corresponding ages of 20–35.

[43] Appendix Figure B.5 documents that return migrants do not display any differential death rates. This is comforting, since it could affect attrition rates differentially between returnees and stayers.

for men and women's households, but still suggest that returnee households possessed some 40–60 percent more wealth than stayers. The drop in coefficient size is expected since wealth may in part be shared among brothers.

For women's spouses, we find similarly large effects of return migration on wealth. These estimates are also more stable when comparing only sisters, potentially due to that the wealth of spouses is less prone to spillovers. Estimates when studying the combined household wealth of women are similar, but to a greater extent affected by the inclusion of sibling fixed effects.

To what extent are our wealth results driven by the occupational category they ended up in after returning to their homeland? This is difficult to answer since occupational choice is endogenous, but we provide suggestive evidence in columns 3 and 4 of Table 6. When controlling for occupation in 1910 the coefficient diminishes some 10–30 percent among both men and women. The coefficients are less robust to including sibling fixed effects in column 4. Taken together, evidence however suggests that the large wealth returns we find among returnees are not mainly driven by different occupational choices after return. Instead, a non-negligible share of the association is attributed to the fact that returnees accumulated more wealth compared to stayers in the same occupational class.

The relationship between migrant status and wealth holds also at the extensive margin, thus mitigating any concern that our results are driven by a few very wealthy individuals. This is seen from the last columns of Appendix Table B.6, which display results from estimating the difference between returnees and stayers in the propensity of having any recorded wealth at all.[44]

5.2.5 Wealth accumulation among returnees

We did not find evidence of productivity differences between returnees and stayers measured by income (at least not for men), suggesting it is unlikely that the established differences in wealth were accumulated after returning. Since we also do not find returnees crowing in locations or occupations with high monetary returns compared to stayers, we propose that the large wealth differences between returnees and stayers are most likely to stem from two main sources: (i) inheritance of valuable assets at home; or (ii) wealth accumulation abroad in the

[44]Appendix Table C.3 also documents that our intensive margin results are robust to winsorizing wealth at the 5th or 10th percentile. Returnee men are about 10 percent more likely to possess any wealth as compared to stayers from the same class and 5 percent more likely as compared to their staying brothers. For women, where we display results for their combined household wealth, the corresponding estimates are 8.5 and 4, albeit the latter is statistically insignificant.

United States.

Although we are not able to observe inheritance directly, our evidence points us in the direction of excluding inheritance as the main driver of our results. Firstly, inheritance practices of the time stipulated that the eldest son was most likely to inherit. Therefore, we should observe a differential effect on wealth for eldest sons if wealth returns of returnees are driven by inheritance. Moreover, since the most valuable asset to inherit in rural areas was the family farm, it is also informative to study any differential effects on being a landowning farmer. In Table 7, we test the relationship between the type of agricultural occupation of the household, return migrant status, and an indicator for being the eldest son. We run regressions with an indicator for being a general landowning farmer in columns 1–3. The evidence suggests that returnees were indeed more likely to be landowning farmers than working in any other agricultural category. Thus, the positive relationship between returnees and agricultural households in Figure 5 is entirely driven by landowning farmers. However, this effect is not driven by first-born sons, as column 3 documents a statistically insignificant relationship for our interaction variable. For completeness, we show also the interaction for the other two agricultural categories: subsistence farmer in column 4, and agricultural landless worker in column 5. Both are statistically insignificant.

The second indication for the limited role of inheritance, is the locations pattern of the returnees. In column 6, we test to what extent return migrants were more often found in their childhood municipality. If the wealth effect is driven by inheritance of the family farm, we should here expect a positive relationship between our outcome and our interaction variable. However, while we do see that eldest brothers are more likely to be found in their home municipality (third row), we do not find that any differential effect for eldest brothers that are returnees. In fact, returnees are not even more likely to be found in their home municipality in general.[45]

Although we do not find any differential effects for eldest sons in terms of farm holdings or childhood location, we cannot exclude that returning eldest sons have both inherited and sold their family farm. Lastly, we therefore show results for

[45]In contrast, Appendix Table B.5 show that returnees are somewhat more likely to return to their home region as measured by the wider county. There is no differential effect for eldest brothers, however. We can also document that the difference in location patterns between male return migrants and stayers is not related to age. Appendix Figure B.4 (panel A) documents in a binned scatter plot that male return migrants are more prone to reside in their home county compared to stayers for more or less all ages. Moreover, female return migrants appear similar to stayers across all age groups as seen in panel B of the same figure.

wealth in the 1930 census in column 7. While we document a positive relationship between returnees and wealth, we do not find any positive interaction between returnees and our indicator for being the eldest son. If anything the relationship is negative, albeit insignificant.

Finding little evidence of any prominent role for inheritance, we turn to studying the importance of wealth accumulation overseas. Leveraging the idea that individuals staying abroad for a longer time also had more time to accumulate savings, we test if individuals staying longer in the US before returning are more likely to hold larger a wealth stock. Table 8 displays results from regressing outcomes in adulthood on two indicators for short and long US stays, respectively, with stayers as the omitted category.[46] In our main specifications, we define long stays as being abroad for at least five years.[47] The picture that emerges from Table 8 suggests that the documented relationships for men from above is driven entirely by returnees with longer US stays. In contrast, male returnees with short stays are not much more likely than stayers to be landowning farmers, to be located in rural locations, or having more wealth. Moreover, this is also robust to controlling for the emigration period, such that it is not driven by the timing of emigration. Our results are thus consistent with the notion that returnees accumulated capital abroad during longer stays in the US and returned to their home regions to buy a farm.[48]

6 Conclusion

Millions of European emigrants returned to the Old World during the Age of Mass Migration, but for long relatively little has been known about the return migrants and how they fared socially and economically once returned. This paper contributes towards filling this knowledge gap by making use of detailed historical census data for the case of Sweden, together with migration and death records, linked across time and space.

[46] We have information on the length of time abroad for about half of our returnees (46%). Returnees without this information are dropped from the analysis in Table 8.

[47] In our childhood sample, five years represent the median stay among returnees. In Appendix Table C.2, we document that results are robust to using the mean stay of 7 years as the cutoff (instead of the median).

[48] In contrast to men, the length of stay for female returnees did not play a substantial role, except for wealth. However, for the latter, female returnees display the opposite pattern. It is only those that stayed in the US for a shorter period of time that experienced an increase in wealth as compared to stayers. A potential explanation could be that female returnees with shorter stays are more likely to be married in 1910. Moreover, female emigrants in the US worked in other sectors than men, such as services, which may have enabled women to accumulate savings at a faster pace.

Our results suggest that the returns to making the journey across the Atlantic and back again were substantial. But they did not mainly manifest as income premiums or social mobility back home. Instead, we suggest that the accumulation of savings in the United States resulted in significantly higher wealth stocks among returnees compared to staying brothers. Upon return, male returnees were more likely to relocate to rural areas and become landowning farmers. While not earning higher labor income than their staying brothers, their social status was demonstrated in higher wealth as well as a larger success on the marriage market.

Female returnees were comparatively more footloose. They ended up in locations and occupational sectors not significantly different from their staying sisters. However, studying the occupational income score of their spouse, they rank higher in the income distribution than stayers and shows signs of greater both absolute and relative mobility. In particular, women from poorer origins advanced upwards from their father's place in the distribution by marrying a white-collar husband.

Our study contributes to the growing literature on the returns to migration. While most previous studies have documented large returns on the receiving job market, our study shows that some of these returns were brought back to sending regions in the form of accumulated wealth. We conclude by calling for more work to assess the footprint of these returnees, and their capital savings, to the locations they returned to.

7 Figures

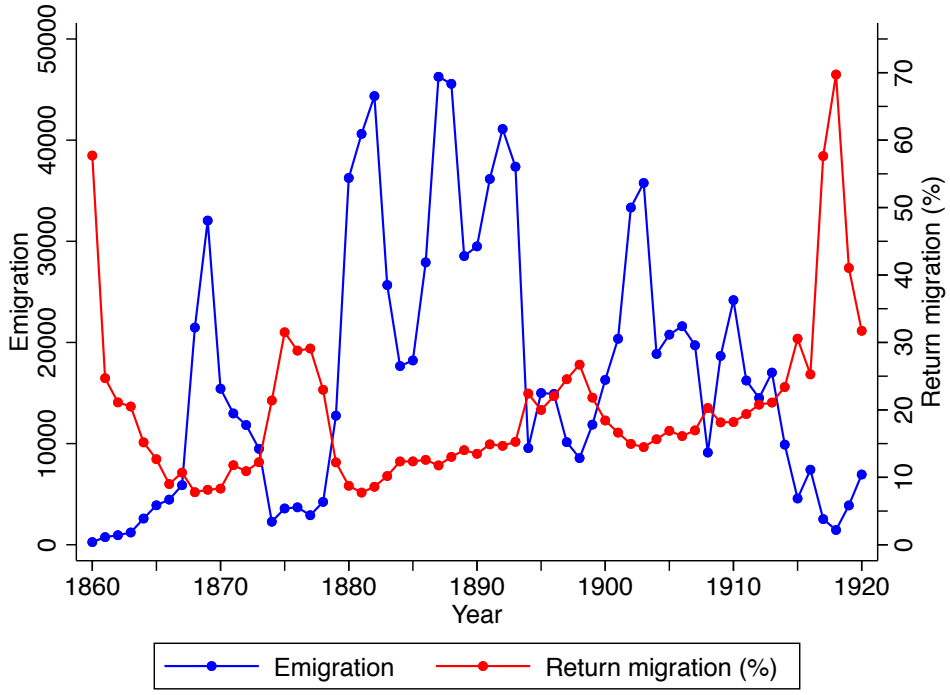
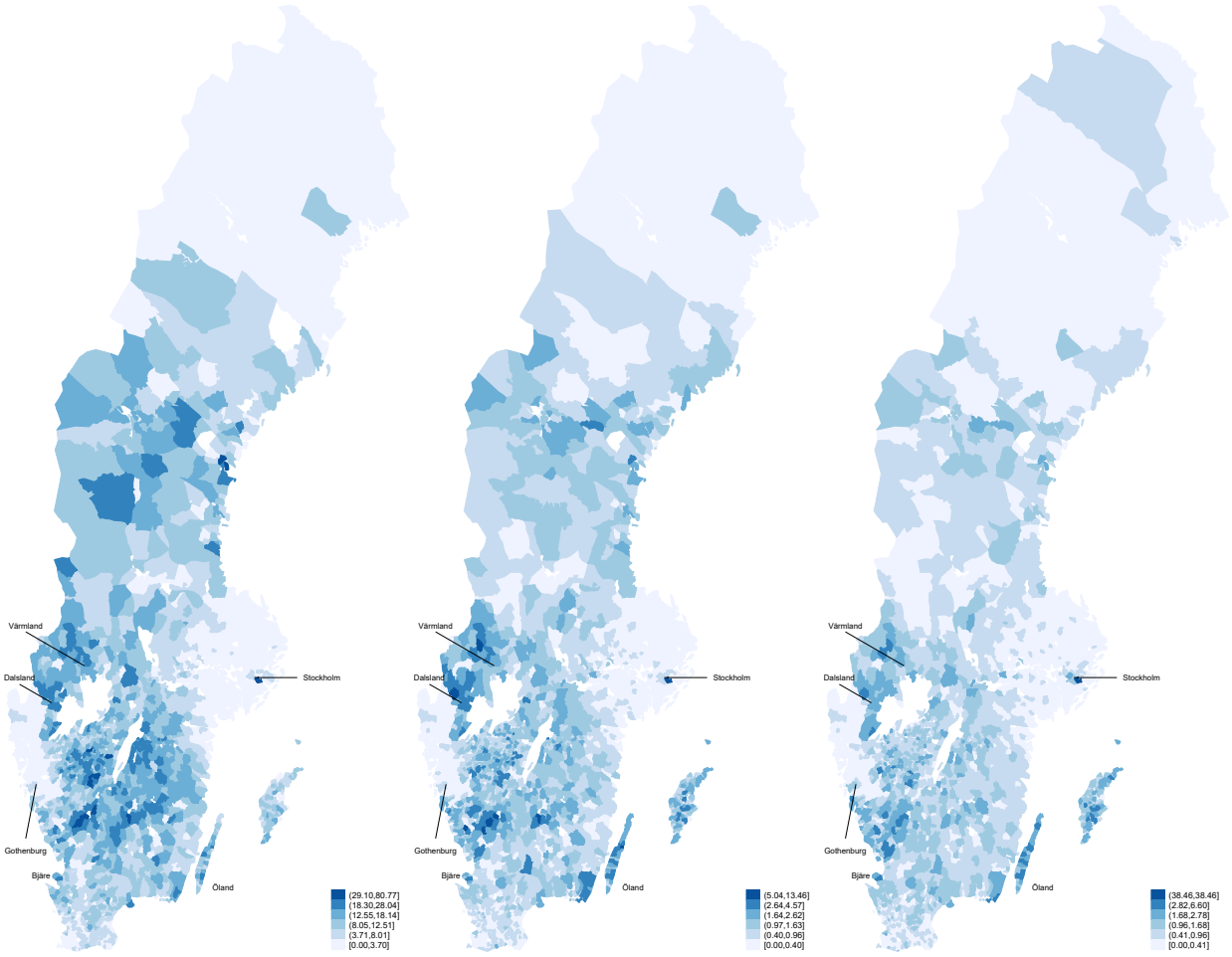


FIGURE 1:
EMIGRATION VOLUME AND RETURN MIGRATION SHARES 1860–1920

Notes: The figure shows the number of US–Sweden emigrants by departure year and the share of emigrants linked to the death index.

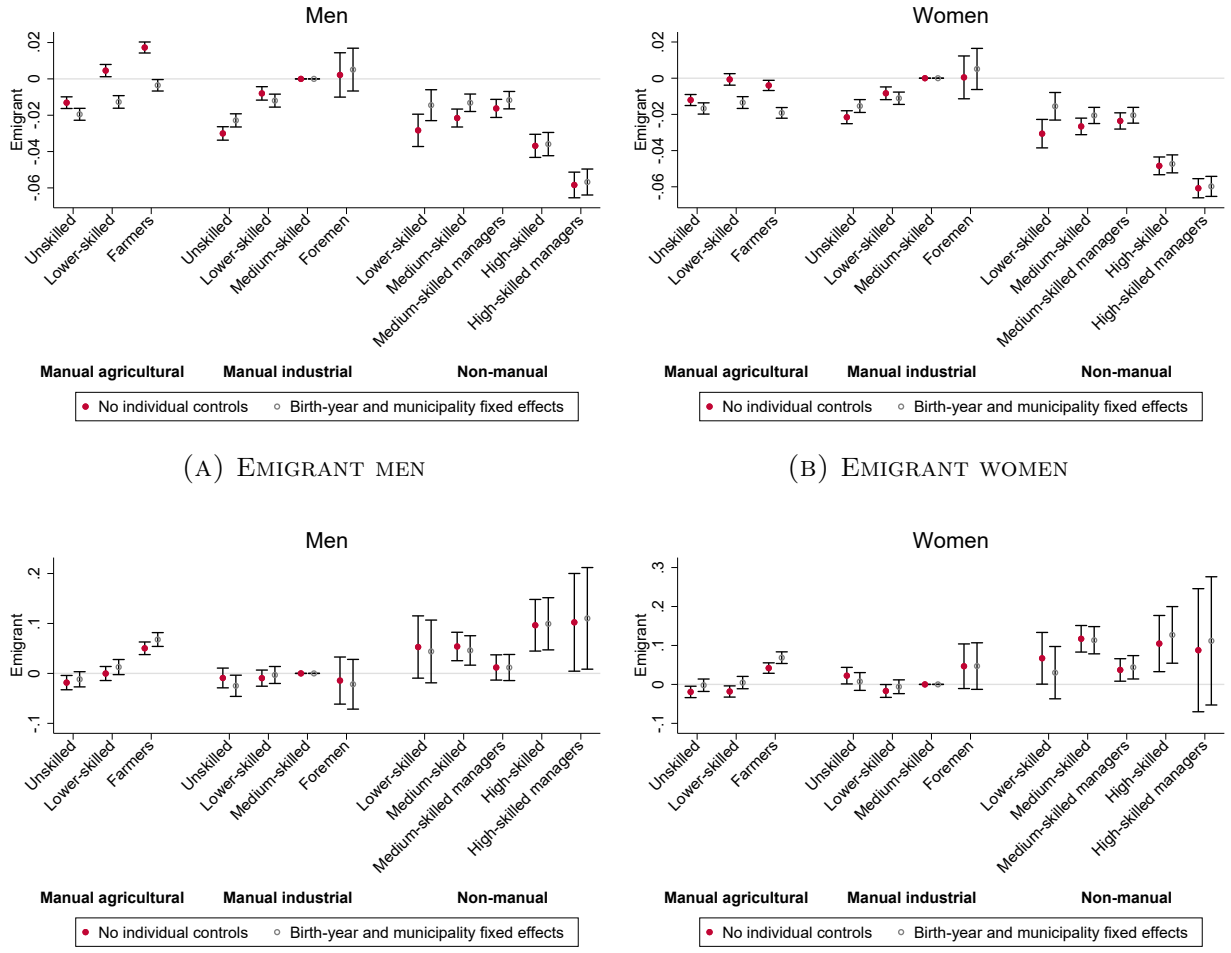
§



(A) Emigration/population (B) Returnees-origin/pop. (C) Returnees-destin./pop.

Figure 2:
Cumulative emigration and returnee flows by municipality 1880-1910

Notes: (A) and (C) show the cumulative flow of emigrants and returnees 1880-1910, related to municipality population in 1880. (B) shows returnees 1880-1910 by municipality of *origin* divided by population 1880 and (C) shows returnees 1880-1910 by municipality of *destination* divided by population 1880. Legend categories are based on clustered k-means.



(A) Emigrant men

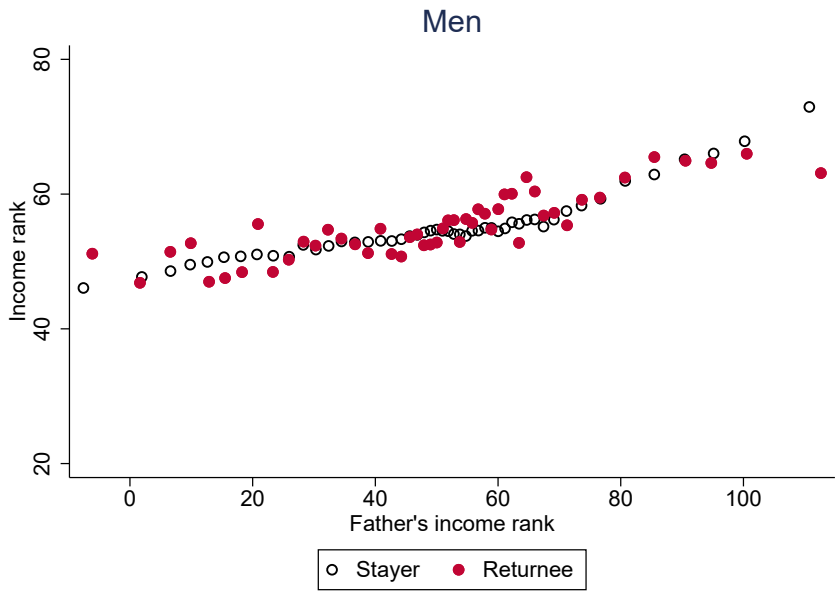
(B) Emigrant women

(C) Return migrant men

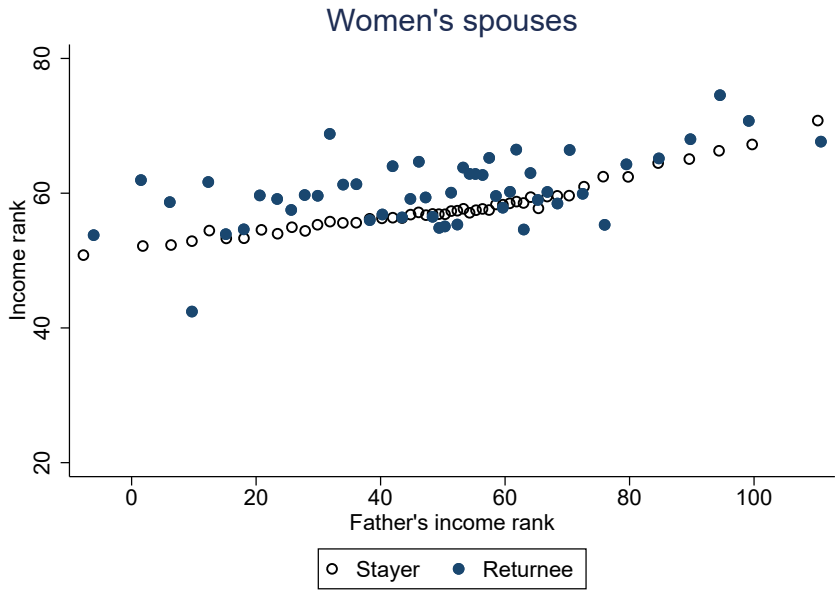
(D) Return migrant women

Figure 3:
Migrant selection by occupational class of household head in 1880

Notes: OLS regressions for linked emigrant childhood sample by sex. These figures display the relationships between migrant status and a set of indicator variables capturing occupational class for the household head in childhood. Figure A and B display the probability of emigrating, while figure C and D display the probability of returning, conditional on emigrating. Birth-year and municipality fixed effects include fixed effects for the birth-year of both the household head and the son/daughter, and the childhood municipality measured in 1880. Medium-skilled manual workers are set as the reference group. Bars represent 95% confidence levels. Standard errors clustered at the level of the household head.



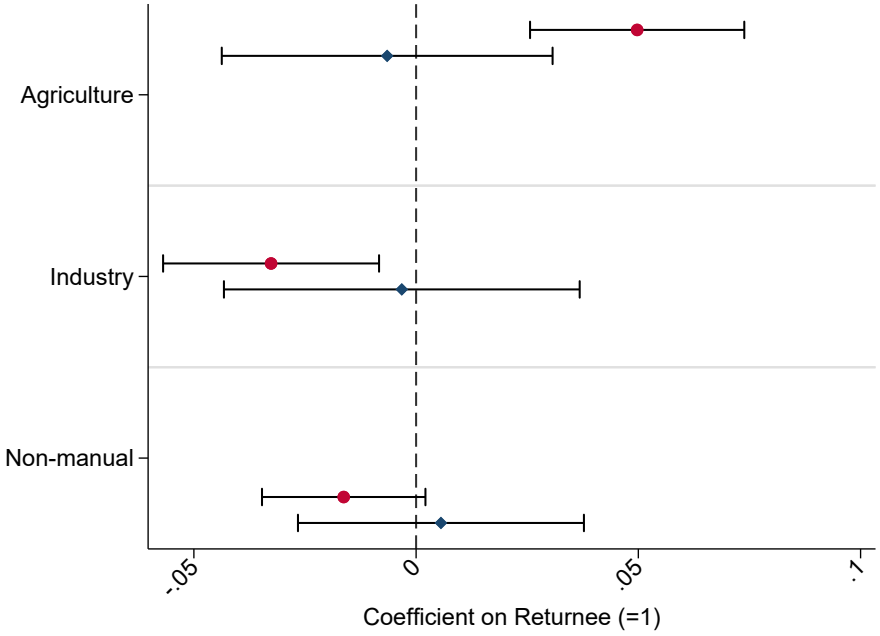
(A) Men



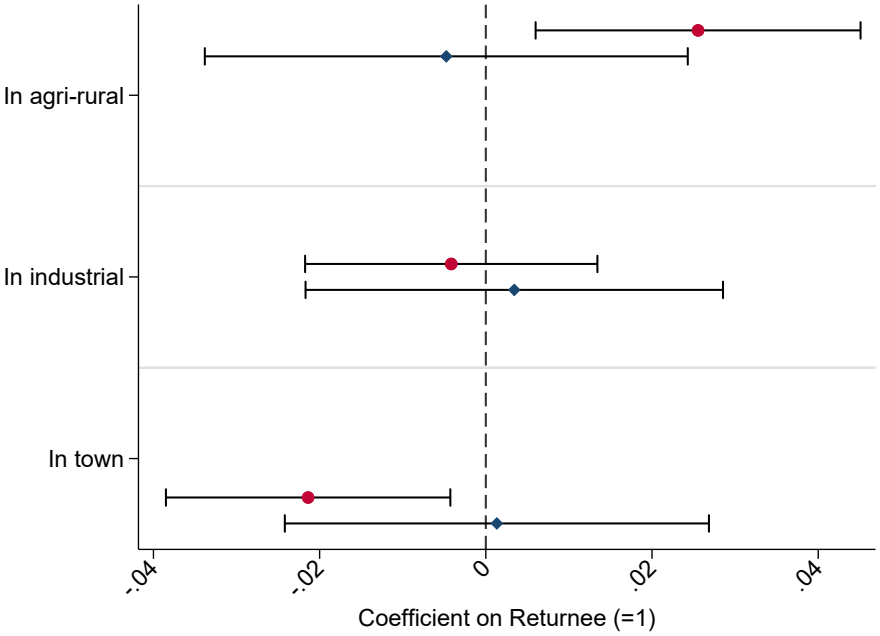
(B) Women

Figure 4:
Intergenerational mobility: rank-rank correlations

Notes: Binned scatter plots by sex for linked childhood sample. These figures display the correlation between the income rank of: the father and son (in A), and the father and the daughter's spouse (in B). Income is measured in 1880 for fathers and in 1910 for their adult children. Controls include childhood municipality, age and birth-order fixed effects, and indicators for being eldest brother and eldest sister, respectively. Returnees in solid circles.



(A) Household occupation



(B) Location

Figure 5:
Household's occupation and location

Notes: OLS regressions by sex for linked childhood sample. All outcomes are defined as indicator variables. Estimates are displayed with red circle for men and with blue diamond for women. All regressions include individual controls and sibling fixed effects. Bars represent 95% confidence levels. Standard errors clustered at the childhood household head-level.

8 Tables

Table 1: Emigrant and census linkage rates

(a) Census links

		Linked to preceding census		Linked to succeeding census		Linked to death index	
Census	A	B	C = B/A	D	E	F	G = F/A
1880	4,624,807	-	-	2,910,969	62.9	2,765,240	59.8
1890	4,843,074	2,910,969	60.1	3,285,579	67.8	3,259,449	67.3
1900	5,199,887	3,285,579	63.2	3,718,504	71.5	3,828,560	73.6
1910	5,586,353	3,718,504	66.6	-	-	4,210,626	75.4
Total	20,254,121	9,915,052	63.4	9,915,052	67.60	14,063,875	69.4

(b) Emigrant links

		Linked to preceding census		Linked to death index		Linked to immigration register	
Period	A	B	C = B/A	D	E	F	G = F/D
-1880	187,835	-	-	-	-	-	-
1881-1890	278,045	172,931	62.2	18,098	6.5	6,507	36.0
1891-1900	141,845	94,409	66.6	14,395	10.1	6,316	43.9
1901-1910	113,776	76,661	67.4	12,037	10.6	6,309	52.4
1911-1920	36,996	25,820	69.8	4,940	13.4	2,663	53.9
Total	758,497	369,821	64.8	49,470	8.7	21,795	44.1

Table 2: Returnees and income

Dependent variable:	Income score 1910 (ln)				Income 1930 (ln)	
Panel A: Men	(1)	(2)	(3)	(4)	(5)	(6)
Returnee	-0.026***	-0.034***	-0.012**	-0.019*	0.142**	0.006
	(0.005)	(0.005)	(0.005)	(0.010)	(0.067)	(0.162)
Observations	314,094	314,093	314,092	147,290	78,511	22,334
Mean dep. var.	7.44	7.44	7.44	7.44	6.08	6.13
Panel B: Women's spouses	(1)	(2)	(3)	(4)	(5)	(6)
Returnee	0.020**	0.019**	0.026***	0.018	0.277**	-0.097
	(0.009)	(0.009)	(0.009)	(0.019)	(0.116)	(0.379)
Individual controls	No	Yes	Yes	Yes	Yes	Yes
Childhood hh controls	No	No	Yes	No	Yes	No
Sibling FE	No	No	No	Yes	No	Yes
Observations	218,023	218,023	218,022	79,446	49,267	9,196
Mean dep. var.	7.48	7.48	7.48	7.48	6.10	6.09

Notes: OLS regressions for linked childhood sample. Returnee is an indicator variable equal to one if the individual has emigrated and returned by 1910, and zero otherwise. Individual controls include age and birth-order fixed effects, and indicators for being eldest brother and eldest sister, respectively. Childhood household controls include fixed effects for the childhood municipality and social class in 1880. Sibling FE are fixed effects for the household head in 1880. Standard errors clustered at the level of the household head. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

Table 3: Intergenerational occupational mobility

Dependent variable:	Income rank			
Panel A: Men	(1)	(2)	(3)	(4)
Returnee	-1.409***	-1.409***	0.177	-0.519
	(0.358)	(0.358)	(0.349)	(0.723)
Father's income rank	0.247***	0.247***	0.116***	
	(0.002)	(0.002)	(0.003)	
Returnee x Father's income rank		0.002	0.004	-0.016
		(0.016)	(0.015)	(0.029)
Observations	300,414	300,414	300,413	144,480
Mean dep. var.	54.99	54.99	54.99	55.14
Panel B: Women's spouses	(1)	(2)	(3)	(4)
Returnee	1.991***	1.889***	2.483***	1.488
	(0.606)	(0.602)	(0.587)	(1.400)
Father's income rank	0.227***	0.228***	0.077***	
	(0.003)	(0.003)	(0.003)	
Returnee x Father's income rank		-0.053**	-0.033	-0.121*
		(0.025)	(0.024)	(0.062)
Individual controls	Yes	Yes	Yes	Yes
Childhood hh controls	No	No	Yes	No
Sibling FE	No	No	No	Yes
Observations	208,645	208,645	208,644	78,110
Mean dep. var.	57.77	57.77	57.77	57.62

Notes: OLS regressions for the linked childhood sample by sex. Returnee is an indicator variable equal to one if the individual is linked to the emigrant lists 1880-1910 and observed in the 1910 census, and zero otherwise. Father's income rank is the demeaned percentile of the father's income in 1880. Individual controls include age and birth-order fixed effects, and indicators for being eldest brother and eldest sister, respectively. Childhood household controls include fixed effects for the childhood municipality and the social class in 1880. Sibling FE are fixed effects for the household head in 1880. Standard errors clustered at the level of the household head. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

TABLE 4: INTERGENERATIONAL OCCUPATIONAL MOBILITY AMONG CHILDREN OF LOW-SKILLED FATHERS

Dependent variable:	Mobile		Downward		Upward				
					Any		Farmer	Skilled	Non-manual
Panel A: Men	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Returnee	0.022**	-0.020	-0.033***	-0.021	0.055***	0.001	0.066***	-0.027	-0.006
	(0.010)	(0.020)	(0.009)	(0.020)	(0.012)	(0.025)	(0.019)	(0.020)	(0.016)
Observations	118,683	54,084	118,683	54,084	118,683	54,084	54,084	54,084	54,084
Mean dep. var.	0.79	0.79	0.15	0.16	0.63	0.63	0.10	0.18	0.09
Panel B: Women's spouses	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Returnee	0.041***	0.066**	-0.066***	-0.020	0.108***	0.086**	0.048	0.046	0.070**
	(0.013)	(0.031)	(0.010)	(0.029)	(0.015)	(0.038)	(0.037)	(0.042)	(0.034)
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Childhood hh controls	Yes	No	Yes	No	Yes	No	No	No	No
Sibling FE	No	Yes	No	Yes	No	Yes	Yes	Yes	Yes
Observations	88,426	32,582	88,426	32,582	88,426	32,582	32,582	32,582	32,582
Mean dep. var.	0.82	0.82	0.13	0.14	0.69	0.68	0.14	0.19	0.09

Notes: OLS regressions for linked childhood sample with low or unskilled fathers. Occupational mobility is captured by an indicator variable constructed by comparing the rank of the social class of the father in 1880 and his adult children in 1910. Returnee is an indicator variable equal to one if the individual has emigrated and returned by 1910, and zero otherwise. Individual controls include age and birth-order fixed effects, and indicators for being eldest brother and eldest sister, respectively. Childhood household controls include fixed effects for the childhood municipality and the social class in 1880. Sibling FE are fixed effects for the household head in 1880. Standard errors clustered at the level of the household head. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

Table 5: Returnees and marriage

Dependent variable:	Married (=1)					
	All		Returnees before 1900		Single 1900	
Panel A: Men	(1)	(2)	(3)	(4)	(5)	(6)
Returnee	0.026***	0.013	0.072***	0.062***	0.102***	0.095***
	(0.006)	(0.013)	(0.008)	(0.017)	(0.013)	(0.031)
Observations	314,111	147,308	311,502	145,384	200,103	71,563
Mean dep. var.	0.68	0.68	0.68	0.68	0.55	0.52
Panel A: Women	(1)	(2)	(3)	(4)	(5)	(6)
Returnee	0.044***	0.001	0.098***	0.048**	0.128***	0.099**
	(0.008)	(0.017)	(0.010)	(0.022)	(0.018)	(0.049)
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes
Childhood hh controls	Yes	No	Yes	No	Yes	No
Sibling FE	No	Yes	No	Yes	No	Yes
Observations	328,701	156,195	327,237	155,167	180,776	60,179
Mean dep. var.	0.67	0.66	0.67	0.66	0.47	0.43

Notes: OLS regressions for linked childhood sample. Columns 1–2 include the full sample, while columns 3–6 drops returnees returning after 1900 with columns 5–6 additionally dropping individuals observed as married in the 1900 census. Returnee is an indicator variable equal to one if the individual has emigrated and returned by 1910, and zero otherwise. Individual controls include age and birth-order fixed effects, and indicators for being eldest brother and eldest sister, respectively. Childhood household controls include fixed effects for the childhood municipality and social class in 1880. Sibling FE are fixed effects for the household head in 1880. Standard errors clustered at the level of the household head. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

Table 6: Returnees and wealth 1930

Dependent variable:	Wealth 1930 (ln)			
Panel A: Men	(1)	(2)	(3)	(4)
Returnee	0.973***	0.404*	0.668***	0.296
	(0.094)	(0.230)	(0.089)	(0.218)
Observations	78,511	22,334	78,511	22,334
Mean dep. var.	5.24	5.51	5.24	5.51
Panel B: Women's spouses	(1)	(2)	(3)	(4)
Returnee	0.950***	0.984**	0.676***	0.872*
	(0.174)	(0.486)	(0.159)	(0.491)
Observations	49,267	9,196	49,267	9,196
Mean dep. var.	5.39	5.57	5.39	5.57
Panel C: Women's household	(1)	(2)	(3)	(4)
Returnee	0.823***	0.547	0.256	1.378
	(0.145)	(0.368)	(0.406)	(1.282)
Individual controls	Yes	Yes	Yes	Yes
Childhood hh controls	Yes	No	Yes	No
Sibling FE	No	Yes	No	Yes
Social class 1910 FE	No	No	Yes	Yes
Observations	83,379	23,538	13,183	1,448
Mean dep. var.	4.09	4.21	3.11	3.33

Notes: OLS regressions for linked childhood sample. Returnee is an indicator variable equal to one if the individual has emigrated and returned by 1910, and zero otherwise. Individual controls include age and birth-order fixed effects, and indicators for being eldest brother and eldest sister, respectively. Childhood household controls include fixed effects for the childhood municipality and social class in 1880. Sibling FE are fixed effects for the household head in 1880. Social class 1910 FE are fixed effects for social class in 1910. Standard errors clustered at the level of the household head. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

Table 7: Male returnees and inheritance

Dependent variable:	Landowning farmer			Subsistence	Worker	In home muni	Wealth (ln)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Returnee	0.114***	0.047***	0.052***	-0.003	0.001	0.005	0.482*
	(0.006)	(0.012)	(0.014)	(0.004)	(0.013)	(0.014)	(0.261)
Returnee x Eldest brother			-0.014	0.008	0.011	-0.021	-0.256
			(0.023)	(0.009)	(0.021)	(0.024)	(0.449)
Eldest brother	0.024***	0.016***	0.016***	0.004**	-0.001	0.014***	0.053
	(0.002)	(0.004)	(0.004)	(0.002)	(0.005)	(0.005)	(0.123)
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Childhood hh controls	Yes	No	No	No	No	No	No
Sibling FE	No	Yes	Yes	Yes	Yes	Yes	Yes
Observations	289,256	130,228	130,228	130,228	130,228	147,308	22,334
Mean dep. var.	0.20	0.19	0.19	0.02	0.21	0.44	5.51

Notes: OLS regressions for male linked childhood sample. Returnee is an indicator variable equal to one if the individual has emigrated and returned by 1910, and zero otherwise. Individual controls include age and birth-order fixed effects, and indicators for being eldest brother and eldest sister, respectively. Childhood household controls include fixed effects for the childhood municipality and the social class in 1880. Sibling FE are fixed effects for the household head in 1880. Standard errors clustered at the level of the household head. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

Table 8: Returnees by time abroad

Dependent variable:	Landowning farmer		In rural location		Wealth (ln)	
Panel A: Men	(1)	(2)	(3)	(4)	(5)	(6)
Returnee <5 Years Abroad	0.017	0.019	0.030	0.028	0.400	0.471
	(0.022)	(0.023)	(0.021)	(0.021)	(0.423)	(0.435)
Returnee ≥ 5 Years Abroad	0.088***	0.094***	0.072***	0.065***	1.032**	1.176**
	(0.027)	(0.029)	(0.019)	(0.020)	(0.495)	(0.519)
Observations	128,394	128,394	145,201	145,201	21,859	21,859
Mean dep. var.	0.19	0.19	0.55	0.55	5.49	5.49
Panel B: Women	(1)	(2)	(3)	(4)	(5)	(6)
Returnee <5 Years Abroad	0.022	0.031	0.014	0.027	1.766**	1.873**
	(0.035)	(0.037)	(0.032)	(0.032)	(0.848)	(0.836)
Returnee ≥ 5 Years Abroad	0.047	0.069*	0.023	0.051*	0.812	1.130
	(0.034)	(0.037)	(0.027)	(0.029)	(0.683)	(0.785)
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes
Sibling FE	Yes	Yes	Yes	Yes	Yes	Yes
Emigration period FE	No	Yes	No	Yes	No	Yes
Observations	115,086	115,086	155,061	155,061	23,264	23,264
Mean dep. var.	0.18	0.18	0.53	0.53	4.20	4.20

Notes: OLS regressions for linked childhood sample. Returnee < 5 Years Abroad is an indicator variable equal to one if the individual has emigrated and returned by 1910 and has spent less than 5 years abroad, and zero otherwise. Returnee ≥ 5 Years Abroad is an indicator variable equal to one if the individual has emigrated and returned by 1910 and has more than 5 years abroad, and zero otherwise. The omitted category is permanent stayers. Individual controls include age and birth-order fixed effects, and indicators for being eldest brother and eldest sister, respectively. Childhood household controls include fixed effects for the childhood municipality and the social class in 1880. Sibling FE are fixed effects for the household head in 1880. Emigration period FE are fixed effects for the decade of emigration. Standard errors clustered at the level of the household head. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

References

Abramitzky, R. and L. Boustan (2017). Immigration in american economic history. *Journal of economic literature 55*(4), 1311–45.

Abramitzky, R., L. Boustan, and K. Eriksson (2019). To the new world and back again: Return migrants in the age of mass migration. *ILR Review 72*(2), 300–322.

Abramitzky, R., L. Boustan, E. Jacome, and S. Perez (2021, February). Intergenerational mobility of immigrants in the United States over two centuries. *American Economic Review 111*(2), 580–608.

Abramitzky, R., L. P. Boustan, and K. Eriksson (2012). Europe's tired, poor, huddled masses: Self-selection and economic outcomes in the age of mass migration. *American Economic Review 102*(5), 1832–56.

Abramitzky, R., L. P. Boustan, and K. Eriksson (2013). Have the poor always been less likely to migrate? evidence from inheritance practices during the age of mass migration. *Journal of Development Economics 102*, 2–14.

Abramitzky, R., L. P. Boustan, and K. Eriksson (2014). A nation of immigrants: Assimilation and economic outcomes in the age of mass migration. *Journal of Political Economy 122*(3), 467–506.

Anbinder, T., D. Connor, C. Ó Gráda, and S. Wegge (2021). The problem of false positives in automated census linking: Evidence from nineteenth-century New York's Irish immigrants. *CAGE working paper no. 568*.

Andersson, D., M. Karadja, and E. Prawitz (2021). Mass migration and technological change. *Journal of the European Economic Association*.

Bailey, M., C. Cole, M. Henderson, and C. Massey (2017). How well do automated linking methods perform? lessons from us historical data. Technical report, National Bureau of Economic Research.

Bailey, M. J., C. Cole, M. Henderson, and C. Massey (2020, December). How well do automated linking methods perform? lessons from us historical data. *Journal of Economic Literature 58*(4), 997–1044.

Bandiera, O., I. Rasul, and M. Viarengo (2013). The making of modern America: Migratory flows in the age of mass migration. *Journal of Development Economics 102*, 23–47.

Berger, T., P. Engzell, B. Eriksson, and J. Molinder (2021). Social mobility in Sweden before the welfare state. Technical report.

Bohlin, J. and A.-M. Eurenius (2010). Why they moved—emigration from the Swedish countryside to the United States, 1881–1910. *Explorations in Economic History 47*(4), 533–551.

Borjas, G. J. (1987). Self-selection and the earnings of immigrants. *American Economic Review 77*(4), 531–55.

Borjas, G. J. and B. Bratsberg (1996). Who leaves? the outmigration of the foreign-born. *The Review of Economics and Statistics 78*(1), 165–176.

Castillo, M. (2022). The land of opportunity? returns to migration for Swedish migrants during the age of mass migration. Technical Report 258, Department of Economic History, Lund University.

Chetty, R., N. Hendren, P. Kline, and E. Saez (2014). Where is the land of opportunity? the geography of intergenerational mobility in the United States. *The Quarterly Journal of Economics 129*(4), 1553–1623.

Collins, W. and M. Wanamaker (2022). African American intergenerational economic mobility since 1880. *American Economic Journal: Applied Economics*.

Connor, D. S. (2019). The cream of the crop? geography, networks, and Irish migrant selection in the age of mass migration. *The Journal of Economic History 79*(1), 139–175.

Dribe, M. and B. Eriksson (2018). Socioeconomic status and adult life expectancy in early twentieth-century Sweden: Evidence from full-count micro census data.

Dribe, M. and C. Lundh (2009). Partner choice and intergenerational occupational mobility: the case of nineteenth-century rural Sweden. *Continuity and Change 24*(3), 487–512.

Dumont, J.-C., G. Spielvogel, et al. (2008). Return migration: A new perspective. *International migration outlook 2008*, 161–222.

Dustmann, C., I. Fadlon, and Y. Weiss (2011). Return migration, human capital accumulation and the brain drain. *Journal of Development Economics 95*(1), 58–67.

Dustmann, C. and O. Kirchkamp (2002). The optimal migration duration and activity choice after re-migration. *Journal of Development Economics 67*(2), 351–372.

Enflo, K., C. Lundh, and S. Prado (2014). The role of migration in regional wage convergence: Evidence from Sweden 1860–1940. *Explorations in Economic History 52*, 93–110.

Eriksson, B. (2015). *Dynamic Decades. A Micro Perspective on Late Nineteenth Century Sweden*. Ph. D. thesis, Lund University: Lund Studies in Economic History 72.

Federation of Swedish Genealogical Societies (2018). Swedish Death Index 7. Solna: Federation of Swedish Genealogical Societies.

Fernihough, A. and C. O'Grada (2019). Across the sea to ireland: Return atlantic migration before the first world war. *Available at SSRN 3491676*.

Hatton, T. J. and Z. Ward (2019). International migration in the Atlantic economy 1850–1940. In C. Diebolt and M. Haupert (Eds.), *Handbook of Cliometrics*, pp. 301–329. Springer.

Hatton, T. J., J. G. Williamson, et al. (1998). *The age of mass migration: Causes and economic impact*. Oxford University Press on Demand.

Helgertz, J., J. Price, J. Wellington, K. J. Thompson, S. Ruggles, and C. A. Fitch (2021). A new strategy for linking u.s. historical censuses: A case study for the ipums multigenerational longitudinal panel. *Historical Methods: A Journal of Quantitative and Interdisciplinary History 0*(0), 1–18.

Henricson, I. and H. Lindblad (1995). *Tur och retur Amerika: utvandrare som förändrade Sverige*. Fischer.

Karadja, M. and E. Prawitz (2019). Exit, voice, and political change: Evidence from Swedish mass migration to the United States. *Journal of Political Economy 127*(4), 1864–1925.

Kuznets, S. and E. Rubin (1954). *Immigration and the Foreign Born*. NBER books.

Long, J. (2005). Rural-urban migration and socioeconomic mobility in victorian britain. *The Journal of Economic History 65*(1), 1–35.

Long, J. and J. Ferrie (2013). Intergenerational occupational mobility in great britain and the United States since 1850. *American Economic Review 103*(4), 1109–37.

Massey, C. G. (2017). Playing with matches: An assessment of accuracy in linked historical data. *Historical Methods: A Journal of Quantitative and Interdisciplinary History 50*(3), 129–143.

Mokyr, J. and C. Ó Gráda (1982). Emigration and poverty in prefamine Ireland. *Explorations in Economic History 19*(4), 360 – 384.

Persson, M. (2007). *Coming Full Circle? Return Migration and the Social Mobility on the Bjäre Peninsula 1860-1930*. Ph. D. thesis, Department of History, Lund University.

Rooth, D.-O. and J. Saarela (2007). Selection in migration and return migration: Evidence from micro data. *Economics Letters 94*(1), 90–95.

Ruggles, S., C. A. Fitch, R. Goeken, D. Hacker, M. A. Nelson, E. Roberts, M. Schouweiler, and M. Sobek (2021). Ipums ancestry full count data: Version 3.0 [dataset]. *Minneapolis, MN: IPUMS 10*, D010.

Saxenian, A. (2007). *The new argonauts: Regional advantage in a global economy*. Harvard University Press.

Schön, L. and O. Krantz (2012). Swedish historical national accounts 1560–2010. Technical report.

Sequeira, S., N. Nunn, and N. Qian (2017). Migrants and the making of america: The short-and long-run effects of immigration during the age of mass migration. Technical report, National Bureau of Economic Research.

Spitzer, Y. and A. Zimran (2018). Migrant self-selection: Anthropometric evidence from the mass migration of italians to the United States, 1907–1925. *Journal of Development Economics 134*, 226 – 247.

Swedish Emigrant Institute and Federation of Swedish Genealogical Societies (2005). EMIBAS. Växjö: Swedish Emigrant Institute.

The Swedish National Archives, Umeå University, and the Minnesota Population Center (2011a). National sample of the 1890 census of Sweden, version 2.0. Minneapolis: Minnesota Population Center [distributor]. https://www.nappdata.org/napp/.

The Swedish National Archives, Umeå University, and the Minnesota Population Center (2011b). National sample of the 1900 census of Sweden, version 3.0. Minneapolis: Minnesota Population Center [distributor]. https://www.nappdata.org/napp/.

The Swedish National Archives, Umeå University, and the Minnesota Population Center (2014). National sample of the 1880 census of Sweden, version 2.0. Minneapolis: Minnesota Population Center [distributor]. https://www.nappdata.org/napp/.

The Swedish National Archives, Umeå University, and the Minnesota Population Center (2016). National sample of the 1910 census of Sweden, version 1.0. Minneapolis: Minnesota Population Center [distributor]. https://www.nappdata.org/napp/.

Van Bavel, J., H. Peeters, and K. Matthijs (1998). Connections between intergenerational and marital mobility: A case study: Leuven, 1830–1910. *Historical Methods: A Journal of Quantitative and Interdisciplinary History 31*(3), 122–134.

Van Leeuwen, M. H. and I. Maas (2011). *HISCLASS: A historical international social class scheme*. Leuven: Leuven University Press.

Van Leeuwen, M. H., I. Maas, and A. Miles (2002). *HISCO: Historical international standard classification of occupations*. Leuven: Leuven University Press.

Wahba, J. and Y. Zenou (2012). Out of sight, out of mind: Migration, entrepreneurship and social capital. *Regional Science and Urban Economics 42*(5), 890–903.

Ward, Z. (2016). There and back (and back) again: repeat migration to the United States, 1897–1936. Unpublished paper, Australian National University.

Ward, Z. (2017). Birds of passage: Return migration, self-selection and immigration quotas. *Explorations in Economic History 64*, 37–52.

Ward, Z. (2020). Internal migration, education, and intergenerational mobility: Evidence from american history. *Journal of Human Resources*.

Wisselgren, M. J., S. Edvinsson, M. Berggren, and M. Larsson (2014). Testing methods of record linkage on Swedish censuses. *Historical Methods: A Journal of Quantitative and Interdisciplinary History 47*(3), 138–151.

Wyman, M. (1996). *Round-trip to America: The immigrants return to Europe, 1880–1930*. Cornell University Press.

Appendix A: Definitions and data appendix

A.1 Historical sources

Our primary source of interest is the emigration register EMIBAS (Swedish Emigrant Institute and Federation of Swedish Genealogical Societies, 2005) which includes the majority of emigrants that left Sweden during the age of mass migration. We complement these data with the Swedish nominal full count 1880, 1890, 1900 and 1910 censuses distributed through IPUMS International (The Swedish National Archives et al., 2014, 2011a,b, 2016), a sample of the 1930 census provided by The Swedish National Archives, and the Swedish death index (Federation of Swedish Genealogical Societies, 2018).

The Swedish historical censuses are of very high quality, both in terms of detail and accuracy. The main reason for this is that enumeration was not based on self-reports to census takers. Instead, the demographic information is drawn from parish records that were kept updated on a running basis by the church. Every ten years the parish registers were used to create extracts which in turn were compiled into the decennial censuses by the Central Bureau of Statistics. Similarly, the emigration records and the death index are based on migration and death registers which were also administered at the parish level. As a consequence of this system, the Swedish censuses, emigration registers and death index provides a detailed and accurate representation of the population, migrants and decedents. In particular, is information about date and place of birth unusually detailed and accurate. Birthplace is recorded at the parish level, a much finer geographic unit than for example US states (there were approximately 2,500 parishes in Sweden in 1880). In addition, since individual information was continuously updated in the parish registers, birth years do not suffer from recall error, something which is evident from the lack of age-heaping in the sources. These aspects means that the prospects of identifying the same individual between sources are unusually good.

In terms of coverage, the 1880-1910 censuses and the death register virtually includes the complete population at the time of the census and all decedents between 1863-2016. The digitization and coding of 1930 census has only been partially completed. To date, 1,964,303 individuals out of a population of 6,142,191 has been fully digitized and coded by the Swedish National Archives. It is important to note that the sample only includes individuals residing in parishes for which the complete population has been digitized, coded and checked. The

Table A.1: Sample size by county, 1930 census

County	Population	Sample	Share
Stockholms län	767,292	50,034	6.5
Uppsala län	138,201	138,275	100.1
Södermanlands län	189,192	17,065	9.0
Östergötlands län	309,911	309,906	100.0
Jönköpings län	231,536	33,994	14.7
Kronobergs län	155,551	86,260	55.5
Kalmar län	231,410	185,507	80.2
Gotlands län	57,458	57,462	100.0
Blekinge län	144,879	70,152	48.4
Kristianstads län	245,928	63,302	25.7
Malmöhus län	510,607	48,494	9.5
Hallands län	150,285	25,483	17.00
Göteborgs- och Bohus län	457,010	87,101	19.1
Älsvborgs län	313,311	163,865	52.3
Skaraborgs län	242,325	16,932	7.0
Värmlands län	269,998	239,609	88.7
Örebro län	219,208	23,799	10.9
Västmanlands län	161,723	47,972	29.7
Kopparbergs län	249,717	0	0.0
Gävleborgs län	279,716	0	0.0
Västernorrlands län	278,562	31,333	11.3
Jämtlands län	134,514	49,600	36.9
Västerbottens län	204,031	203,999	100.0
Norrbottens län	199,826	14,159	7.1
Total	6142191	1,964,303	32.0

sample thus only contain complete households, allowing for the identification of spouses, children and other individuals residing in the same household as the individual of interest. In terms of geographic coverage the sample is unbalanced across regions. Table X provides at summary of the distribution of the sample across regions.

The emigration register contain the majority of all emigrants, but is not as complete as the censuses and the death index. The official statistics (ref BISOS) reported that 1,021,306 emigrants left for North America between 1860-1920, a number which is generally considered an underestimation. The equivalent number of emigrants recorded in EMIBAS is 741,559. Figure A.1 plots the number of emigrants by departure year published by the official statistics together with the numbers contained in EMIBAS. The numbers registered in both sources follows a common general trend and are very similar in terms of volume during

the height of the emigration, after which the total numbers of emigrants and the share thereof registered in EMIBAS gradually diminishes.

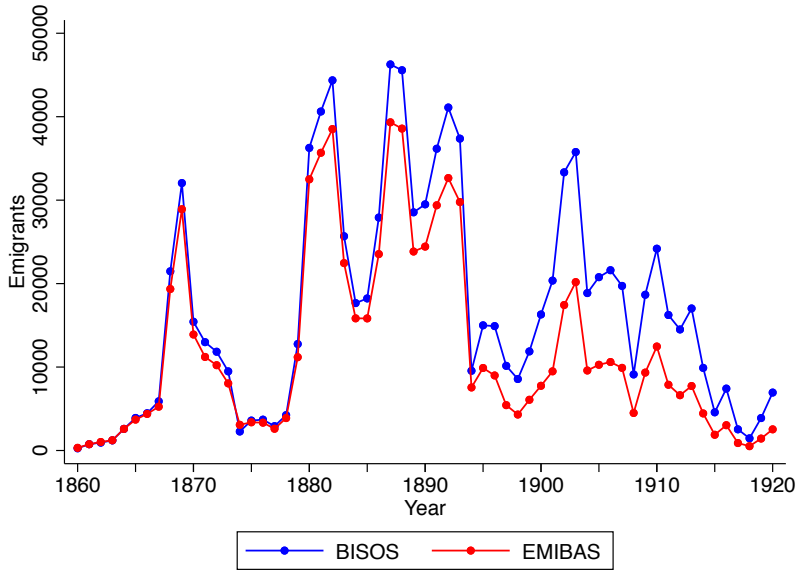


FIGURE A.1: EMIGRATION 1860-1920

Notes: This figure shows the number of emigrants by departure year recorded in the official statistics (BISOS) and the emigration registers (EMIBAS).

A.2 Linking procedure

In this appendix, we provide a more detailed description of the procedure used to link individuals between the migration registers, censuses and the death index. The aim of the linking procedure is to create representative longitudinal samples of the Swedish population which identifies emigrants and return migrants. Our method is fully automated and unlike some recent approaches (Helgertz et al., 2021; Bailey et al., 2020) does not rely on manually created training data, thus ensuring that samples and results are fully replicable.

The linking procedure starts with identifying variables suitable for matching individuals. In order to not introduce bias, only variables that are time-invariant over the life course should be considered (see Ruggles 2006).[49] Disqualified variables therefore include information such as current location of residence and civil

[49]Recent approaches to record linking using sources of inferior quality have made use of time-invariant variables, including information about family members, in order to improve linkage rate and quality (Helgertz et al., 2021). However, since the underlying Swedish data is of such high quality, we are able to achieve both high linkage rates and quality by relying on basic time-invariant variables.

Table A.2: Linking variables by source

	1880 Census	1890 Census	1900 Census	1910 Census	1930 Census	Death index	Emigration register	Immigration register
Sex	YES	YES	YES	YES	YES	YES	YES	YES
Birth year	YES	YES	YES	YES	YES	YES	YES	YES
Birth month	-	-	-	-	YES	YES	YES	-
Birth day	-	-	-	-	YES	YES	YES	-
Birth parish	YES	YES	YES	YES	YES	YES	YES	-
Names	YES	YES	YES	YES	YES	YES	YES	YES
Period	1880	1890	1900	1910	1930	1881–2016	-1920	1881-
Observations	4,624,825	4,843,782	5,200,111	5,586,360	x	x	570,622	x

or occupational status. We use sex, birth date and birth place as index variables, meaning that two records have to match exactly in order to be considered a candidate. The emigration register, censuses and death index all includes detailed and high quality information about birth place, date of birth, first and last names and sex which enables identification of individuals between the sources. Importantly, birth place is recorded at the parish level, of which there were about 2,500 in 19th century Sweden, making it a very small geographic unit. Moreover birth years do not suffer from age heaping which is otherwise common in historical records, which allows for the exact year (or date) of birth to be used. As a result, each emigrant is only ever compared to a relatively small number of permissible matches. Additionally, for obvious reasons, we exclude any matches in which the date of departure predate a census year or postdate the date of death recorded in the death index.

The second, and the most critical step, in the linking process involves separating true links from false among all candidates generated by the initial matching on the index variables. To adjudicate which of these potential matches in fact is the same individual, we rely on comparisons of first and last names. However, because names are recorded with a certain degree of imprecision in the sources, due to transcription errors (both in the original sources and as a result of the digitization process) or spelling variations. Allowance must therefore be made for the fact that the name of the same individual may not be identical in the emigration register, censuses and the death index. To reduce the influence of minor differences in spelling or transcription errors, we first standardize names by removing nobility prefixes, patronymic suffixes and all non-alphabetic characters. To allow for the fact that even standardized names may differ between censuses

for the same individual, we employ the Jaro-Winkler (JW) algorithm to estimate the similarity of first and last names recorded for potential matches. The JW algorithm assigns a similarity score between 0 (no similarity) and 1 (identical) for two text strings by comparing common characters, common character pairs and transpositions. Moreover adjustments are made for when strings have the same initial characters and accounts for the fact that irregularities are more common in longer strings than in shorter.

We consider individuals linked if the JW for both the first and the last names exceeds a given threshold, and that the link is non-ambiguous. That is, we require that no other plausible competing link exists. Given the problematic nature of false positive links, we choose to err on the conservative and make no attempt to establish which of the ambiguous links are true or false, but instead treat all ambiguous links as suspicious and discard them. In practice this means that links are only retained if an emigrant is linked to only one individual in the censuses or death index. It should however be noted that we allow for multiple emigration records to be linked to one census or death index observation since individuals could have migrated multiple times and may therefore appear more than once in the emigration register.

When choosing the thresholds which the JW similarity scores must exceed, a trade off exists between the resulting sample size and the quality of matches. Prioritizing a high number of matches by lowering the threshold and thereby the required similarity between names, increases the risk of introducing false positives. This, in turn, might lead to measurement error that will bias estimates of migrant selection, and also create a false impression of high occupational mobility (Bailey et al., 2017). An overly restrictive similarity threshold, on the other hand, reduces the number of false positives but results in a smaller sample that might be an unrepresentative subset of the full population. We thus need to find an optimal threshold for the JW similarity score that maximises the number of linked individuals, while maintaining a low rate of false positives.

To identify an optimal JW threshold, we use secondary characteristics to evaluate the quality of links at different threshold levels for the JW similarity score. The quality of links is evaluated by considering the share of matches that we can confirm using information on additional first ("middle") names that are not used to generate the original link. We define a link as confirmed if middle name initials match. Moreover we consider various thresholds for what we classify to be competing links and subsequently choose thresholds on the basis of the share

confirmed and the number of links made.

We use the above method to link the emigration register to the censuses, and the emigration register to the death index. When linking the emigration register to the censuses we link every censuses to the emigrants which departed in the following ten year period. I.e. the 1880 census is linked to emigrants which departed 1881-1890, the 1890 census is linked to emigrants which departed 1891-1900 etc. Sex, birth year, and birth parish are used as index variables. When linking the emigration register to the death index we only link emigrants whose year of departure precedes the year of death. Sex, date of birth, and birth parish are used to index the data. We begin by linking individuals with two or more first names, after which we remove these links from the pool of candidates and proceed with linking any remaining candidates using only one first name.

A.3 Complementary linked data

We complement the linked emigration data above with links between the censuses and the death index. Each census has been linked to the subsequent census, and the death index, following the procedure described above. The links are combined to create a panel which follows individuals across the censuses, ending with the death event in case where a link has been made to the death index. For a more detailed decription of the census and death index linking see Eriksson (2015); Dribe and Eriksson (2018)

A.4 Assessing the Linked Sample

In total we are able to link 369,821 US emigrants to the censuses (64.8 per cent of all US emigrants recorded in EMIBAS 1880-1920). These numbers compare well to what is typically achieved when linking historical sources. However, a more important issue than the number of links made is how well the resulting linked samples represents the emigrant population, since differential matching rates based on demographic or economic characteristics may introduce bias in the subsequent analysis. Naturally, because emigration and return migration are in itself selective processes, our linked sample of emigrants will never be a representative sample of the Swedish population recorded in the censuses, nor will emigrants that returned and were recorded as deceased in the death index be representative of emigrants as a whole. The sample of emigrants linked to the census should however be representative of all emigrants. Therefore, our main

focus is to compare linked emigrants to all emigrants.

In terms of departure years, the linked sample represents the emigrants recorded in the registers well. The procedure links a somewhat higher share of emigrants over time, and emigrants which departed nearer in time to each preceding census. In the years immediately following a census the linkage rates exceed 70 per cent, and never falls below 55 per cent in a given year.

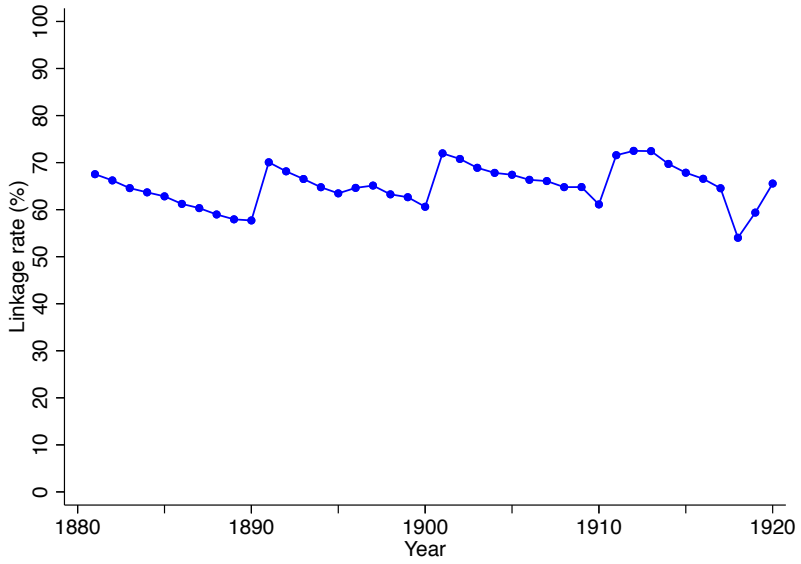


FIGURE A.2: EMIGRATION REGISTER LINKAGE RATES, 1880-1920

Notes: This figure shows the number of emigrants by departure year recorded in the official statistics (BISOS) and the emigration registers (EMIBAS).

The linked sample is also representative in terms of the geographic origins of emigrants in the emigration register. The geographic distribution across the Swedish provinces is similar, but with a lower share of links made for Stockholm. That we link fewer emigrants from Stockholm is to be expected since the city's population size means that it is more difficult to identify links among all possible candidates. The underrepresentation of Stockholm has been noted in previous work linking Swedish censuses (Eriksson, 2015).

A.5 Variable definitions

Occupational status

We employ the HISCLASS scheme (Van Leeuwen and Maas 2011) in order to code occupations into distinct classes. HISCLASS is based on the HISCO coding

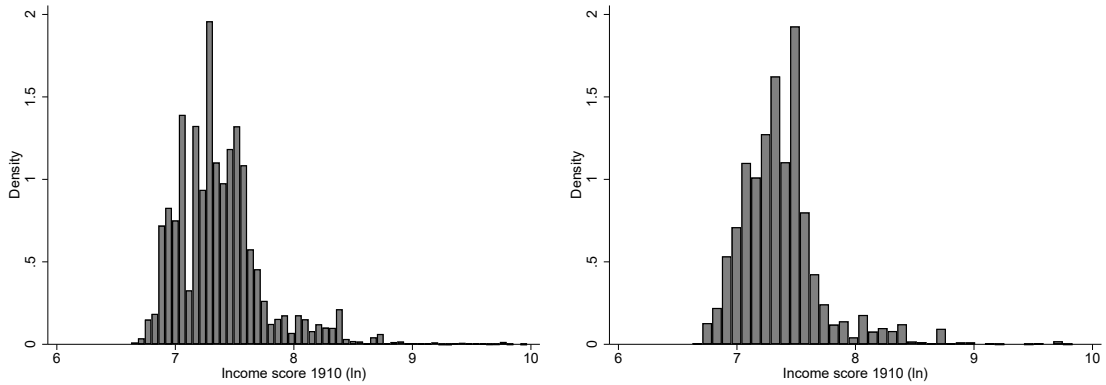
scheme (Van Leeuwen, Maas and Miles 2002) and classifies occupational information into 12 classes which are grouped according to economic sector, whether the occupation is manual or non-manual, and the level of skill and supervision.

Income and wealth

The 1930 census includes individual level information about income and wealth recorded in intervals of 100 SEK (income) and 1000 SEK (wealth) rounded down. The primary source used by Statistics Sweden for compiling information about income and wealth for the census was individually declared taxes. In those cases were no tax was declared, information about incomes were supplemented by consulting wage lists provided by employers (74.1 per cent of the reported income and wealth were based on individual tax declarations, while the remaining 25.9 per cent were based on information provided by employers). As a result,

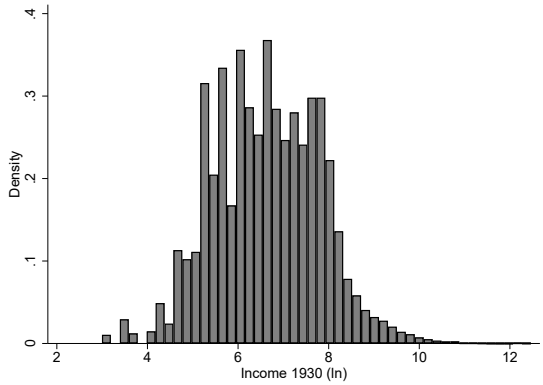
As a general rule, wives' and children's incomes were added to their husbands or fathers, unless the wife or child themselves had a recorded occupation. Hence,

Appendix B: Additional material

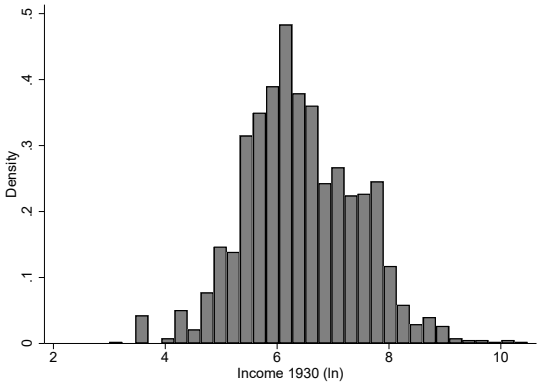


(A) Income score 1910 (stayer men)

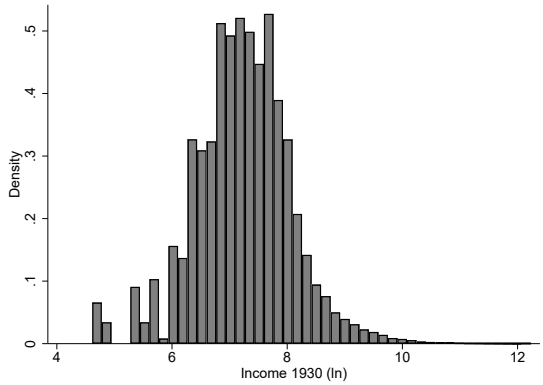
(B) Income score 1910 (returnee men)



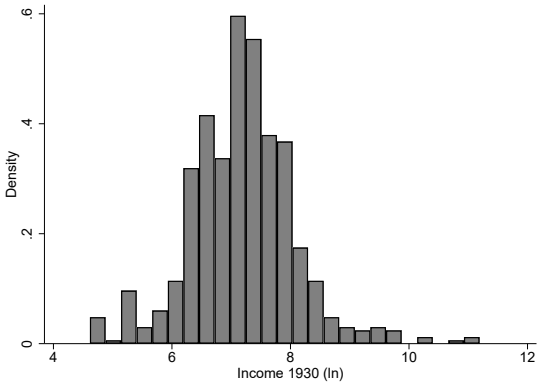
(C) Income 1930 (stayer men)



(D) Income 1930 (returnee men)



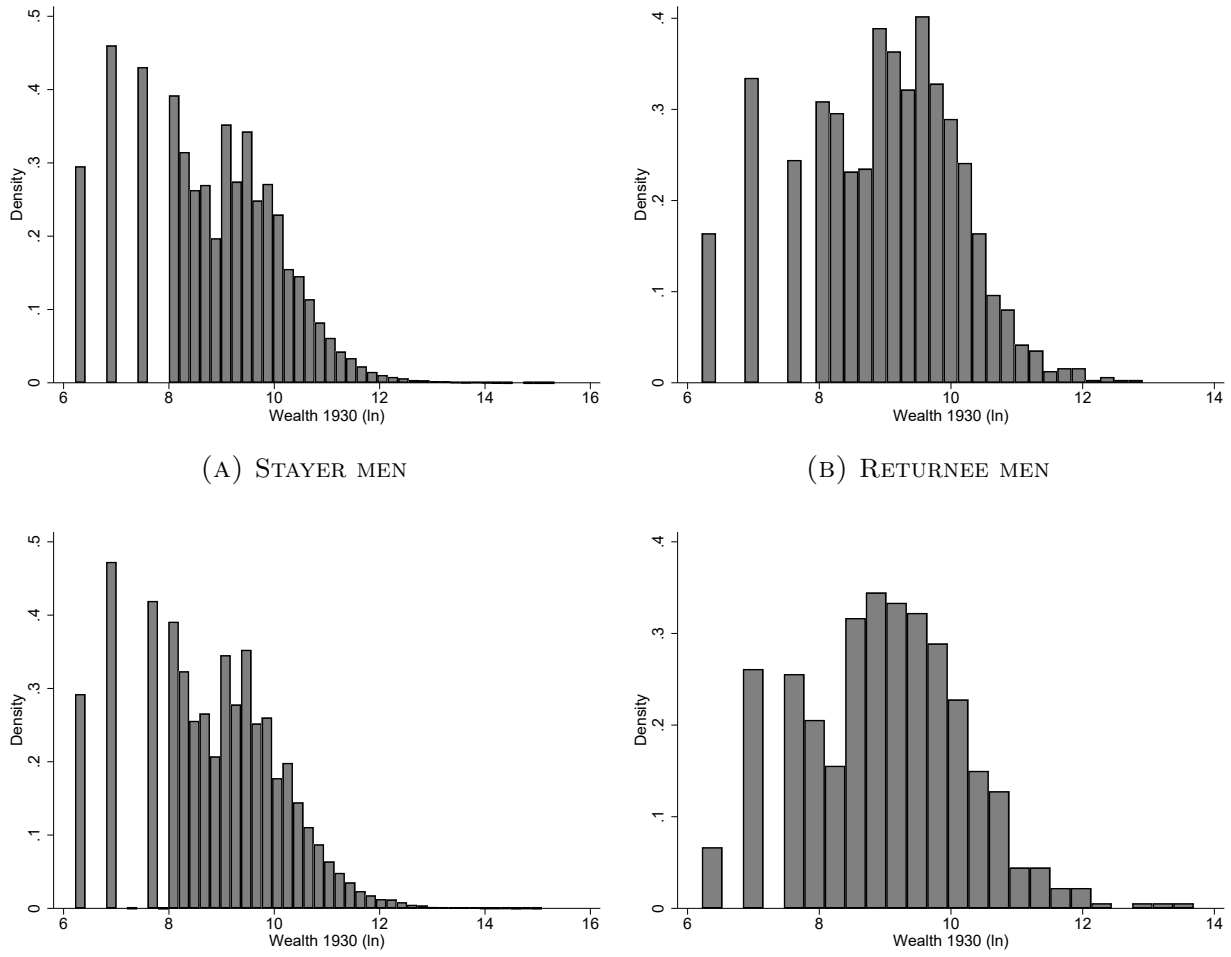
(E) Income 1930 (stayer women)



(F) Income 1930 (returnee women)

Figure B.1: Distribution of income score 1910 and income 1930

Notes: This figure displays the distribution of income scores in 1910 and income in 1930, both in natural logarithms, for men and women by returnee status in the linked childhood sample. Income scores 1910 are only calculated for men. Individuals with zero income in 1930 are omitted. They represent 26.58% and 40.77% for male returnees and stayers, respectively, and 84.77% and 86.57% for female returnees and stayers, respectively.



(A) Stayer men

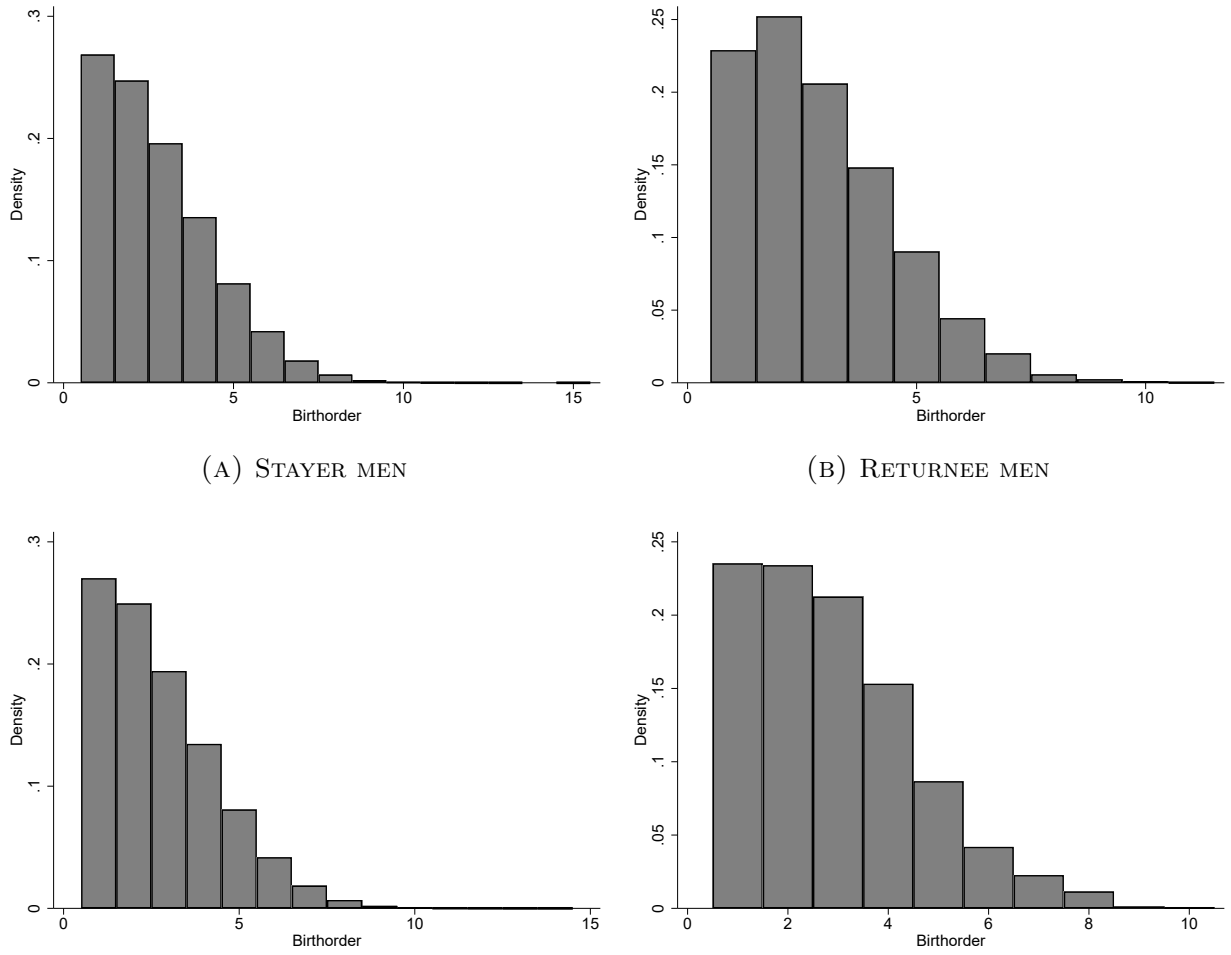
(B) Returnee men

(C) Stayer women

(D) Returnee women

Figure B.2: Distribution of wealth 1930

Notes: This figure displays the distribution of wealth in 1930, in natural logarithms, for men and women by returnee status in the linked childhood sample. Individuals with zero income are omitted. They represent 18.32% and 17.14% for male returnees and stayers, respectively, and 83.30% and 82.29% for female returnees and stayers, respectively.



(A) Stayer men

(B) Returnee men

(C) Stayer women

(D) Returnee women

Figure B.3: Distribution of birth order

Notes: This figure displays the distribution of birth order, as observed in the 1880 census, among returnees and stayers by sex.

Table B.1: Migrant selection (men)

Dependent variable:	Emigrant						Returnee					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Father income score (ln)	-0.0219***	-0.0044***	-0.0055***				0.0314***	0.0345***	0.0364***			
	(0.0010)	(0.0011)	(0.0011)				(0.0059)	(0.0063)	(0.0063)			
Unskilled (agri.)				-0.0131***	-0.0195***	-0.0177***				-0.0184**	-0.0117	-0.0138*
				(0.0016)	(0.0017)	(0.0017)				(0.0073)	(0.0078)	(0.0078)
Lower-skilled (agri.)				0.0046***	-0.0117***	-0.0124***				-0.0002	0.0126	0.0123
				(0.0017)	(0.0018)	(0.0018)				(0.0072)	(0.0077)	(0.0077)
Farmers				0.0173***	-0.0036**	-0.0045***				0.0503***	0.0678***	0.0701***
				(0.0015)	(0.0016)	(0.0016)				(0.0064)	(0.0071)	(0.0071)
Unskilled (non-agri.)				-0.0300***	-0.0217***	-0.0214***				-0.0091	-0.0248**	-0.0261**
				(0.0019)	(0.0019)	(0.0018)				(0.0101)	(0.0108)	(0.0108)
Lower-skilled (non-agri.)				-0.0080***	-0.0096***	-0.0119***				-0.0095	-0.0037	-0.0021
				(0.0019)	(0.0018)	(0.0018)				(0.0082)	(0.0087)	(0.0087)
Foremen				0.0022	0.0061	0.0045				-0.0143	-0.0227	-0.0182
				(0.0062)	(0.0060)	(0.0060)				(0.0241)	(0.0253)	(0.0253)
Lower-skilled (non-manual)				-0.0283***	-0.0147***	-0.0135***				0.0528*	0.0434	0.0435
				(0.0045)	(0.0044)	(0.0043)				(0.0318)	(0.0320)	(0.0320)
Medium-skilled				-0.0215***	-0.0146***	-0.0127***				0.0539***	0.0462***	0.0469***
				(0.0025)	(0.0025)	(0.0025)				(0.0145)	(0.0150)	(0.0150)
Medium-skilled managers				-0.0163***	-0.0129***	-0.0113***				0.0119	0.0114	0.0122
				(0.0025)	(0.0024)	(0.0024)				(0.0129)	(0.0133)	(0.0133)
High-skilled				-0.0369***	-0.0409***	-0.0347***				0.0964***	0.1001***	0.1014***
				(0.0032)	(0.0033)	(0.0033)				(0.0264)	(0.0266)	(0.0266)
High-skilled managers				-0.0584***	-0.0598***	-0.0561***				0.1023**	0.1118**	0.1117**
				(0.0036)	(0.0036)	(0.0037)				(0.0499)	(0.0523)	(0.0517)
Individual controls	No	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes
Childhood controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Observations	649,124	649,119	649,117	621,895	621,888	621,886	57,358	57,201	57,201	53,658	53,498	53,498
Mean dep. var.	0.09	0.09	0.09	0.09	0.09	0.09	0.21	0.21	0.21	0.21	0.21	0.21

Notes: OLS regressions for male linked childhood sample. In columns 1–6, the dependent variable is equal to one if an individual ever emigrates. In columns 7–12, the dependent variable is equal to one if an individual ever returns to Sweden and zero if the individual emigrates, but never returns. Occupations are measured for their household head in childhood. Indicators for medium-skilled manual workers are omitted. Individual controls include age and birth-order fixed effects, and indicators for being eldest brother and eldest sister, respectively. Childhood household controls include fixed effects for the childhood municipality and the social class in 1880. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

TABLE B.2: MIGRANT SELECTION (WOMEN)

Dependent variable:	Emigrant						Returnee					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Father income score (ln)	-0.0225***	-0.0105***	-0.0114***				0.0348***	0.0376***	0.0385***			
	(0.0009)	(0.0010)	(0.0010)				(0.0064)	(0.0070)	(0.0070)			
Unskilled (agri.)				-0.0121***	-0.0166***	-0.0152***				-0.0193***	-0.0020	-0.0031
				(0.0016)	(0.0016)	(0.0016)				(0.0075)	(0.0081)	(0.0082)
Lower-skilled (agri.)				-0.0007	-0.0131***	-0.0133***				-0.0182**	0.0047	0.0047
				(0.0016)	(0.0017)	(0.0017)				(0.0073)	(0.0081)	(0.0081)
Farmers				-0.0040***	-0.0192***	-0.0201***				0.0419***	0.0684***	0.0703***
				(0.0014)	(0.0015)	(0.0015)				(0.0068)	(0.0076)	(0.0077)
Unskilled (non-agri.)				-0.0216***	-0.0148***	-0.0142***				0.0224**	0.0078	0.0068
				(0.0018)	(0.0018)	(0.0018)				(0.0108)	(0.0116)	(0.0117)
Lower-skilled (non-agri.)				-0.0083***	-0.0096***	-0.0110***				-0.0168**	-0.0069	-0.0055
				(0.0018)	(0.0017)	(0.0017)				(0.0084)	(0.0090)	(0.0090)
Foremen				0.0005	0.0054	0.0045				0.0466	0.0468	0.0479
				(0.0060)	(0.0058)	(0.0058)				(0.0292)	(0.0306)	(0.0305)
Lower-skilled (non-manual)				-0.0306***	-0.0160***	-0.0145***				0.0670**	0.0289	0.0300
				(0.0040)	(0.0039)	(0.0039)				(0.0338)	(0.0342)	(0.0342)
Medium-skilled				-0.0266***	-0.0217***	-0.0203***				0.1171***	0.1132***	0.1138***
				(0.0023)	(0.0023)	(0.0023)				(0.0174)	(0.0179)	(0.0178)
Medium-skilled managers				-0.0236***	-0.0210***	-0.0204***				0.0371**	0.0443***	0.0445***
				(0.0023)	(0.0022)	(0.0022)				(0.0147)	(0.0153)	(0.0153)
High-skilled				-0.0484***	-0.0501***	-0.0464***				0.1048***	0.1290***	0.1275***
				(0.0025)	(0.0025)	(0.0025)				(0.0368)	(0.0371)	(0.0371)
High-skilled managers				-0.0608***	-0.0613***	-0.0594***				0.0877	0.1133	0.1135
				(0.0027)	(0.0028)	(0.0029)				(0.0806)	(0.0842)	(0.0839)
Individual controls	No	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes
Childhood controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Observations	632,716	632,713	632,713	607,133	607,127	607,127	42,456	42,269	42,269	39,715	39,528	39,528
Mean dep. var.	0.07	0.07	0.07	0.07	0.07	0.07	0.17	0.17	0.17	0.18	0.18	0.18

Notes: OLS regressions for female linked childhood sample. In columns 1–6, the dependent variable is equal to one if an individual ever emigrates. In columns 7–12, the dependent variable is equal to one if an individual ever returns to Sweden and zero if the individual emigrates, but never returns. Occupations are measured for their household head in childhood. Indicators for medium-skilled manual workers are omitted. Individual controls include age and birth-order fixed effects, and indicators for being eldest brother and eldest sister, respectively. Childhood household controls include fixed effects for the childhood municipality and the social class in 1880. Standard errors clustered at the level of the household head. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

Table B.3: Pre-migration characteristics in the linked childhood sample

	Men			Women		
	(1)	(2)	(3)	(4)	(5)	(6)
Age	1.893***	1.888***	1.515***	1.075***	1.052***	0.996***
	(0.056)	(0.056)	(0.116)	(0.074)	(0.075)	(0.152)
Birth order	0.117***	0.026	-0.468***	0.147***	0.102***	-0.327***
	(0.022)	(0.022)	(0.041)	(0.030)	(0.029)	(0.056)
Eldest brother/sister	-0.046***	-0.032***	0.100***	-0.038***	-0.029***	0.053***
	(0.007)	(0.007)	(0.015)	(0.009)	(0.009)	(0.020)
Dependent variable:	Log dist. to birthpl.	Birth county	High GDP	High growth	Urban	Married
	(1)	(2)	(3)	(4)	(5)	(6)
Returnee	0.618***	-0.099***	0.214***	0.100***	0.047***	0.282***
	(0.017)	(0.003)	(0.004)	(0.005)	(0.003)	(0.004)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	26,282	30,708	30,708	30,708	30,708	30,708
Mean dep. var.	1.51	0.83	0.43	0.42	0.20	0.36

Notes: OLS regressions for the linked childhood sample by sex. Each entry displays the coefficient on Returnee from a separate regression. The dependent variable is displayed Returnee is an indicator variable equal to one if the individual is linked to the emigrant lists 1880-1910 and observed in the 1910 census, and zero otherwise. Individual controls include age and birth-order fixed effects, and indicators for being eldest brother and eldest sister, respectively. Childhood household controls include fixed effects for the childhood municipality and the social class in 1880. Sibling FE are fixed effects for the household head in 1880. Standard errors clustered at the level of the household head. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

Table B.4: Returnee selection in the US and Swedish labor markets

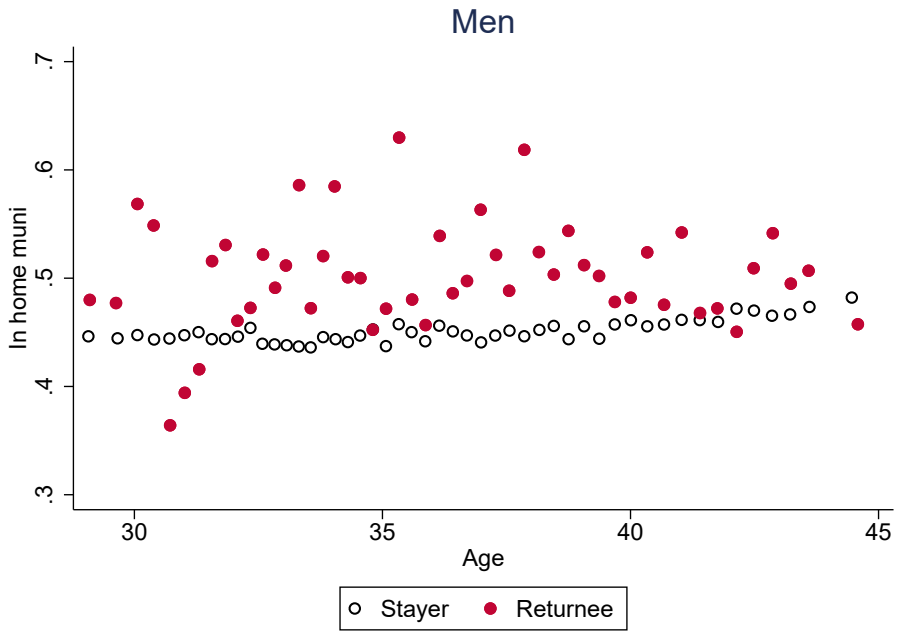
Dependent variable:	Returnee					
	US labor market				Swedish labor market	
	(1)	(2)	(3)	(4)	(5)	(6)
Own US income score (ln)	-0.006*	-0.007**	-0.007**	-0.007**		
	(0.003)	(0.003)	(0.003)	(0.003)		
Father income score (ln)					0.036***	0.005
					(0.006)	(0.010)
US state FE	No	Yes	Yes	Yes	No	No
Age controls	No	No	Yes	Yes	Yes	Yes
Swedish class FE	No	No	No	Yes	No	No
Swedish county FE	No	No	No	No	Yes	Yes
Observations	7,314	7,311	7,311	7,311	57,358	7,042
Mean dep. var.	0.07	0.07	0.07	0.07	0.21	0.07

Notes: OLS regressions. Columns 1–4 display results for male emigrants from the linked childhood sample observed in the US census 1900 with an occupation. Column 5–6 display results using the income scores of their fathers in Sweden. Columns 6 restricts the analysis to individuals observed in the US census. The dependent variable is equal to one if an individual returns to Sweden and zero if the individual has emigrated, but not returned. US state fixed effects are measured in the US census 1900. Age controls is a second order polynomial in age measured either in the US census 1900 (columns 1–4) or the Swedish census 1880 (columns 5–6). Swedish class fixed effects and county fixed effects are measured for the household head in childhood in 1880. Standard errors clustered at the level of the household head. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

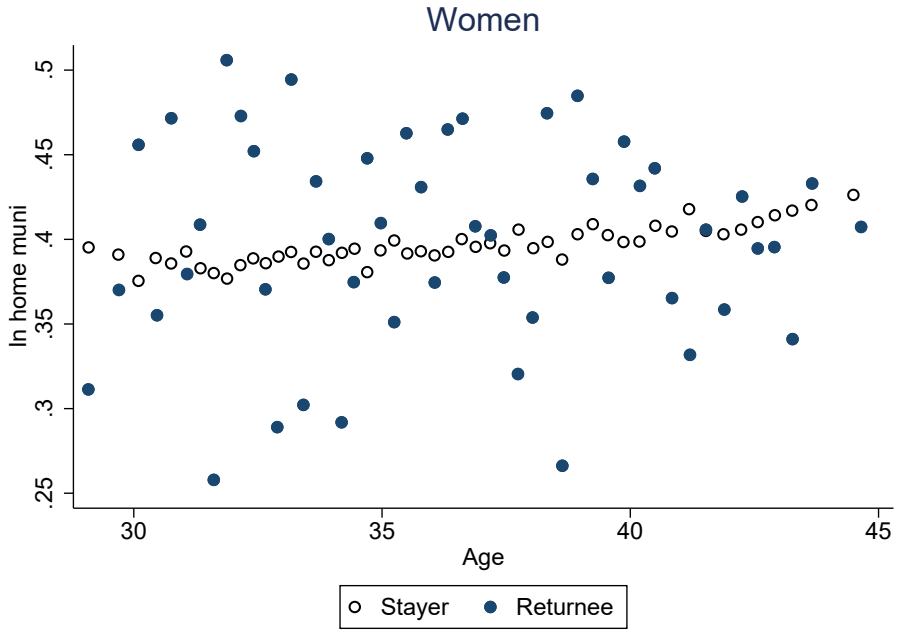
Table B.5: Returnees returning home?

Dependent variable:	In home muni			In home region		
Panel A: Men	(1)	(2)	(3)	(4)	(5)	(6)
Returnee	0.034***	-0.001	0.005	0.050***	0.016	0.019*
	(0.007)	(0.012)	(0.014)	(0.005)	(0.010)	(0.012)
Returnee x Eldest brother			-0.021			-0.011
			(0.024)			(0.019)
Observations	314,111	147,308	147,308	314,111	147,308	147,308
Mean dep. var.	0.45	0.44	0.44	0.77	0.77	0.77
Panel B: Women	(1)	(2)	(3)	(4)	(5)	(6)
Returnee	-0.010	-0.005	-0.004	-0.010	-0.017	-0.014
	(0.009)	(0.016)	(0.019)	(0.008)	(0.014)	(0.016)
Returnee x Eldest sister			-0.001			-0.011
			(0.034)			(0.027)
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes
Childhood hh controls	Yes	No	No	Yes	No	No
Sibling FE	No	Yes	Yes	No	Yes	Yes
Observations	328,701	156,195	156,195	328,701	156,195	156,195
Mean dep. var.	0.40	0.38	0.38	0.76	0.75	0.75

Notes: OLS regressions for linked childhood sample. Returnee is an indicator variable equal to one if the individual has emigrated and returned by 1910, and zero otherwise. Individual controls include age and birth-order fixed effects, and indicators for being eldest brother and eldest sister, respectively. Childhood household controls include fixed effects for the childhood municipality and social class in 1880. Sibling FE are fixed effects for the household head in 1880. Standard errors clustered at the level of the household head. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$. Sample of individuals born 1866-80.



(A) Men



(B) Women

Figure B.4:
Residing in home municipality by age

Notes: Binned scatter plots by sex for linked childhood sample. These figures display the correlation between an indicator for if the individual resides in his/her home municipality (1880) and age. Controls include childhood municipality and birth-order fixed effects, and indicators for being eldest brother and eldest sister, respectively. Returnees in solid circles.

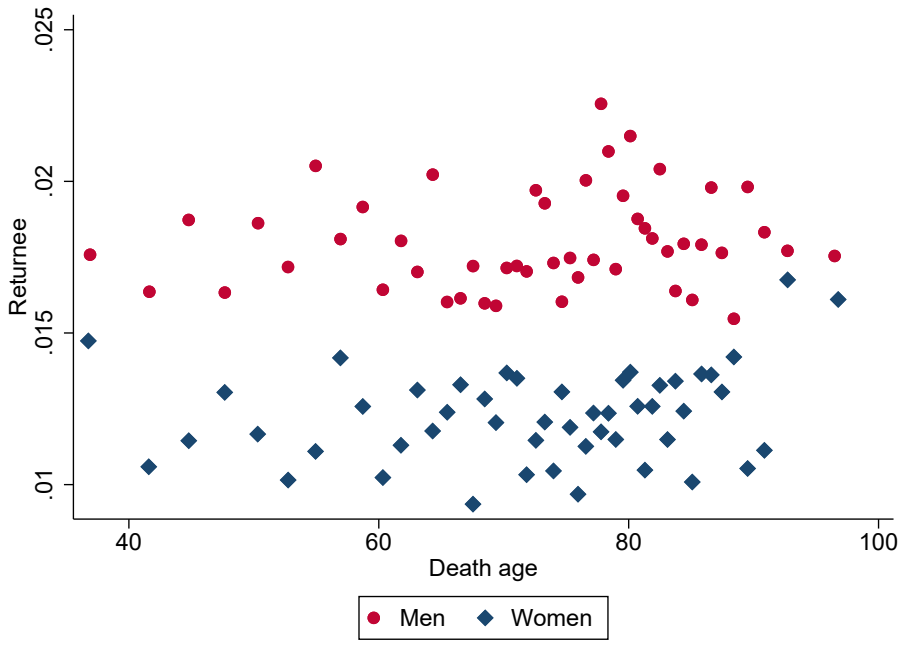


FIGURE B.5:
RETURNEES AND DEATH AGE BY SEX

Notes: Binscatter plots for the linked childhood sample. Controls include childhood municipality, age and birth-order fixed effects, and indicators for being eldest brother and eldest sister, respectively.

Table B.6: Extensive margin for labor income and wealth in 1910 and 1930

Dependent variable:	Employed 1910 (=1)		Employed 1930 (=1)		Any wealth 1930 (=1)	
Panel A: Men	(1)	(2)	(3)	(4)	(5)	(6)
Returnee	0.005	0.008	0.019**	0.000	0.103***	0.048*
	(0.004)	(0.008)	(0.009)	(0.022)	(0.010)	(0.026)
Observations	314,111	147,308	78,511	22,334	78,511	22,334
Mean dep. var.	0.92	0.92	0.83	0.84	0.60	0.62
Panel B: Women	(1)	(2)	(3)	(4)	(5)	(6)
Returnee	-0.030***	-0.001	-0.013	0.000	0.085***	0.042
	(0.007)	(0.014)	(0.012)	(0.031)	(0.016)	(0.041)
Observations	328,701	156,195	83,379	23,538	83,379	23,538
Mean dep. var.	0.19	0.20	0.18	0.19	0.46	0.48
Panel C: Women's spouses	(1)	(2)	(3)	(4)	(5)	(6)
Returnee	0.021**	-0.023	0.030**	-0.026	0.090***	0.088
	(0.008)	(0.017)	(0.015)	(0.054)	(0.019)	(0.055)
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes
Childhood hh controls	Yes	No	Yes	No	No	No
Sibling FE	No	Yes	No	Yes	No	No
Observations	328,701	156,195	49,267	9,196	49,267	9,196
Mean dep. var.	0.64	0.64	0.83	0.83	0.61	0.63

Notes: OLS regressions for the linked childhood sample by sex. Returnee is an indicator variable equal to one if the individual is linked to the emigrant lists 1880-1910 and observed in the 1910 census, and zero otherwise. Individual controls include age and birth-order fixed effects, and indicators for being eldest brother and eldest sister, respectively. Childhood household controls include fixed effects for the childhood municipality and the social class in 1880. Sibling FE are fixed effects for the household head in 1880. Standard errors clustered at the level of the household head. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

Table B.7: Intergenerational occupational mobility in full childhood sample

Dependent variable:	Mobile		Downward		Upward	
Panel A: Men	(1)	(2)	(3)	(4)	(5)	(6)
Returnee	-0.045***	-0.015	-0.040***	-0.004	-0.005	-0.011
	(0.007)	(0.013)	(0.006)	(0.013)	(0.006)	(0.013)
Observations	255,112	119,086	255,112	119,086	255,112	119,086
Mean dep. var.	0.68	0.69	0.28	0.29	0.40	0.40
Panel B: Women's spouses	(1)	(2)	(3)	(4)	(5)	(6)
Returnee	0.016*	0.024	-0.049***	-0.012	0.066***	0.036
	(0.010)	(0.023)	(0.009)	(0.020)	(0.010)	(0.023)
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes
Childhood hh controls	Yes	No	Yes	No	Yes	No
Sibling FE	No	Yes	No	Yes	No	Yes
Observations	187,385	69,925	187,385	69,925	187,385	69,925
Mean dep. var.	0.69	0.69	0.25	0.26	0.44	0.43

Notes: OLS regressions for the linked childhood sample by sex. Returnee is an indicator variable equal to one if the individual is linked to the emigrant lists 1880-1910 and observed in the 1910 census, and zero otherwise. Individual controls include age and birth-order fixed effects, and indicators for being eldest brother and eldest sister, respectively. Childhood household controls include fixed effects for the childhood municipality and the social class in 1880. Sibling FE are fixed effects for the household head in 1880. Standard errors clustered at the level of the household head. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

B.1 Occupational selection for all working-age individuals

To complement our selection estimates, measured in terms of the occupational status of the household head in 1880, we document occupational selection in the full working-age population in terms of their own occupation prior to emigrating.

In particular, we estimate migrant selection by using observables in the full-count census prior to emigrating in each decade following a linear probability model of the form:

$$Emigrate_{i\tau} = \sum_{j=1}^{12} \beta_j HC_{j,it} + \phi_t + \upsilon_{it}, \tag{4}$$

where $Emigrate_{i\tau}$ is a dummy variable that takes the value 1 if individual i emigrates in τ, with $t < \tau < t + 10$, i.e., between two censuses $t, ..., t + 10$, and zero otherwise.[50] $HC_{j,it}$ are a set of indicator variables capturing how emigrants were selected in terms of their occupational status in t, measured by $j = 1, ..., 12$ main categories of the HISCLASS-scheme. In our main specifications, we omit the category with medium-skilled workers. To account for unobserved temporal variation, we include census-year fixed effects, ϕ_t.

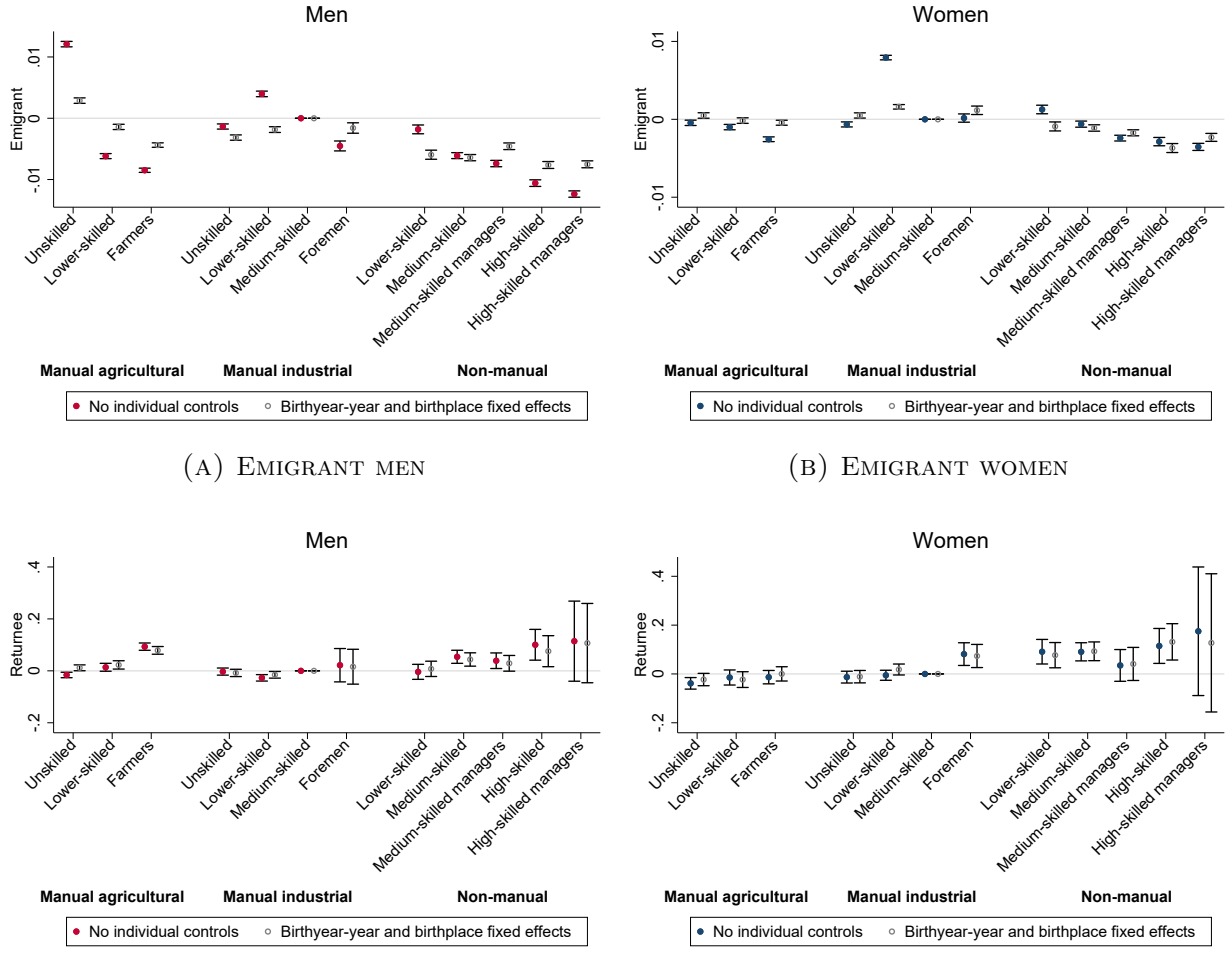
Similarly, we study return migrant selection by estimating the corresponding model:

$$Return_{i\tau} = \sum_{j=1}^{12} \gamma_j HC_{j,it} + \phi_t + \nu_{it}, \tag{5}$$

where $Return_{i\tau}$ is a dummy variable that takes the value 1 if a return migrant i emigrated in τ, with $t < \tau < t + 10$, i.e., between two censuses $t, ..., t + 10$, and 0 otherwise. Thus, we compare return migrants with never-returning emigrants in the census year prior to emigration.

In additional specifications to 4 and 5, we include different demographic characteristics: birth-year and birth-parish fixed effects as well as indicator variables for if an individual was married, resided in an urban location, and resided in her birth county. We use heteroskedasticity-robust standard errors in both models.

[50] Repeat migrants are only treated as emigrants at his or her first-time emigration event.



(A) Emigrant men

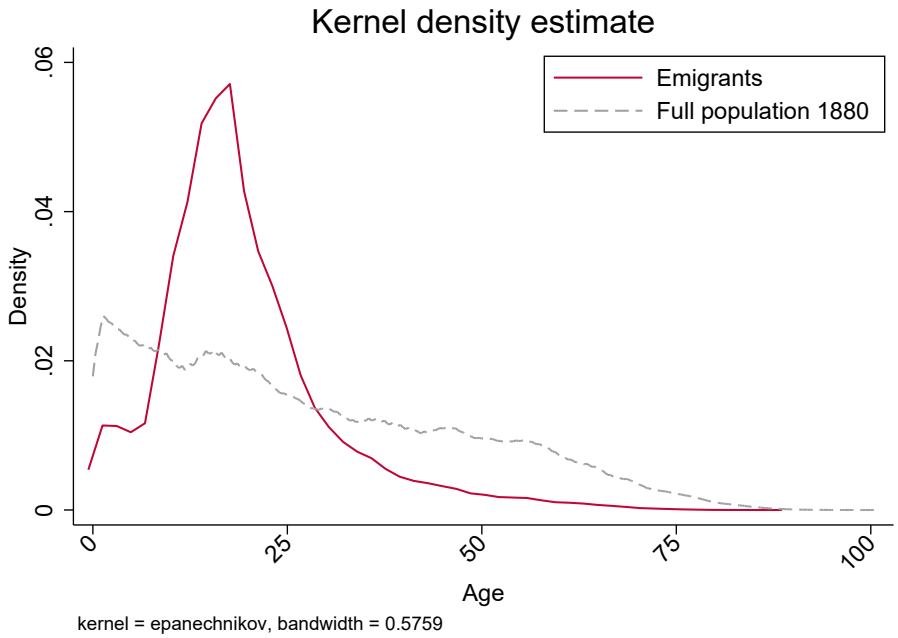
(B) Emigrant women

(C) Return migrant men

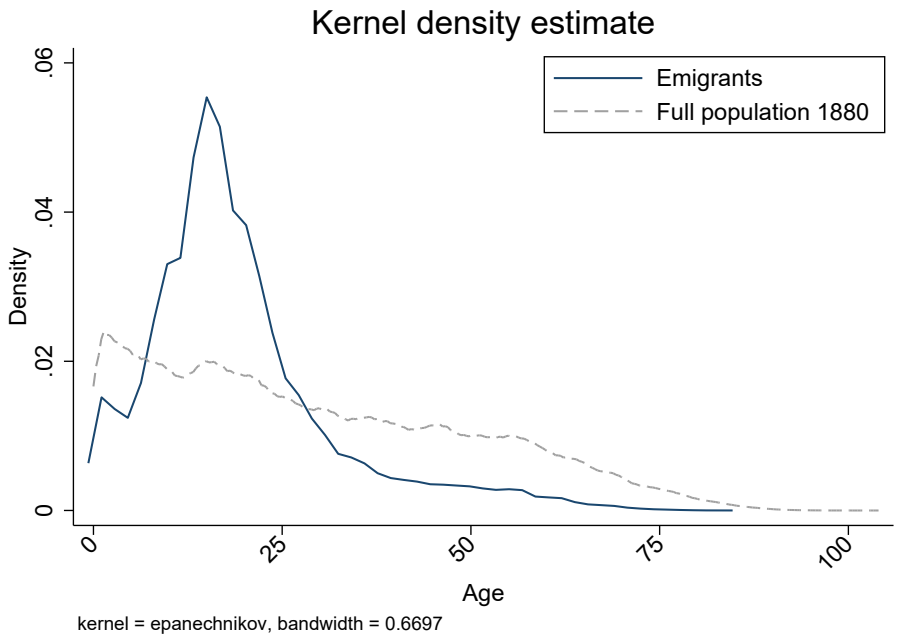
(D) Return migrant women

Figure B.6:
Migrant selection by occupational group 1880–1910

Notes: OLS regressions for full working-age sample by sex. These figures display the relationships between migrant status and a set of indicator variables capturing occupational class for men and women separately. Figure A and B display the probability of emigrating, while figure C and D display the probability of returning, conditional on emigrating. All occupations are measured in the census prior to emigrating. All estimates include (census) year fixed effects. The occupations of women are imputed using the occupation of the spouse, whenever it is missing for women. Medium-skilled manual workers are set as the reference group. Bars represent 95% confidence levels. Robust standard errors.



(A) Men



(B) Women

FIGURE B.7: AGE DENSITY OF EMIGRANTS BY SEX

Notes: This figure displays kernel density plots using the Epanechnikov kernel.

TABLE B.8: MIGRANT SELECTION (MEN)

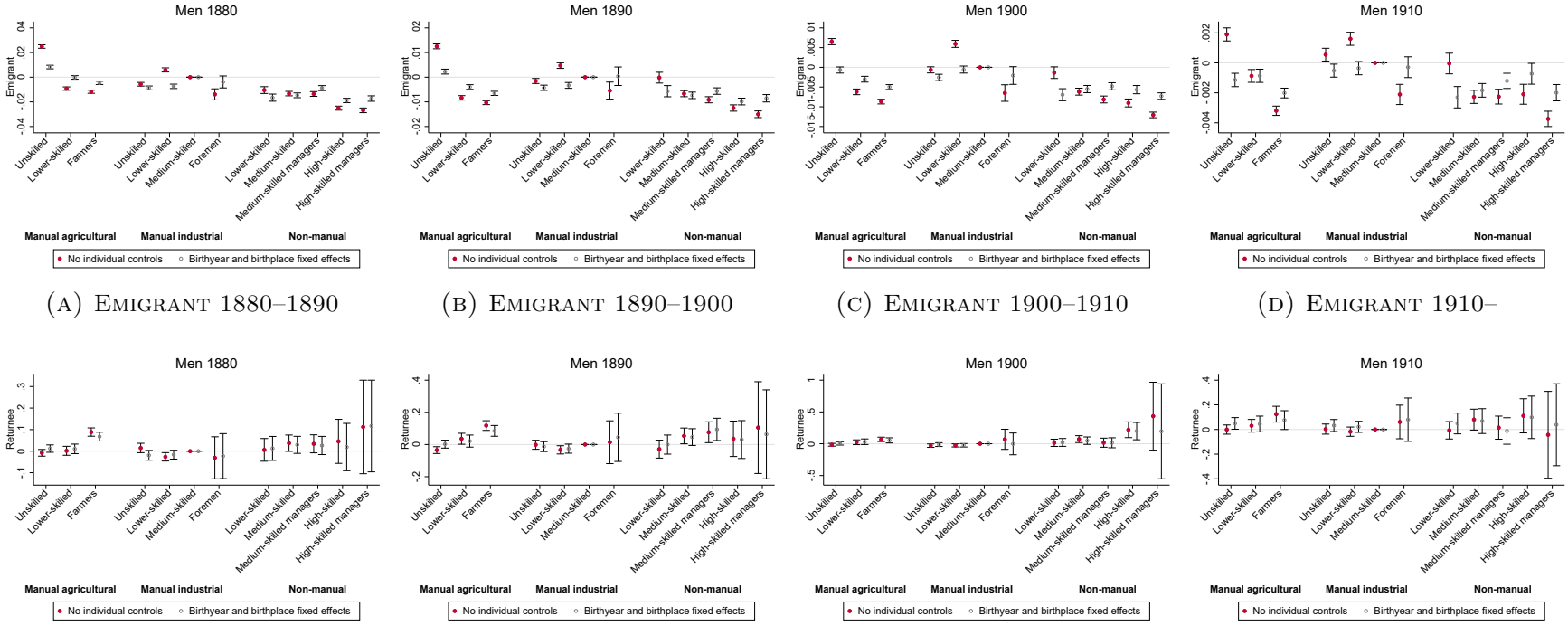
Dependent variable:	Emigrant				Returnee			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Unskilled	0.0121***	0.0114***	0.0065***	0.0050***	-0.0159***	-0.0125**	-0.0050	-0.0176***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0053)	(0.0056)	(0.0057)	(0.0060)
Lower-skilled	-0.0062***	-0.0077***	-0.0025***	-0.0029***	0.0137*	0.0191**	0.0158*	0.0094
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0078)	(0.0080)	(0.0081)	(0.0083)
Farmers	-0.0085***	-0.0105***	-0.0045***	-0.0050***	0.0932***	0.0808***	0.0741***	0.0668***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0072)	(0.0074)	(0.0075)	(0.0077)
Unskilled	-0.0013***	-0.0008***	-0.0027***	-0.0031***	-0.0029	-0.0164**	-0.0143**	-0.0167**
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0070)	(0.0072)	(0.0072)	(0.0072)
Lower-skilled	0.0040***	0.0035***	-0.0005*	-0.0006**	-0.0270***	-0.0254***	-0.0224***	-0.0241***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0063)	(0.0065)	(0.0065)	(0.0065)
Foremen	-0.0045***	-0.0049***	0.0006	0.0012***	0.0217	0.0346	0.0265	0.0314
	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0328)	(0.0339)	(0.0339)	(0.0338)
Lower-skilled	-0.0018***	-0.0013***	-0.0058***	-0.0059***	-0.0039	-0.0152	-0.0088	-0.0101
	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0147)	(0.0150)	(0.0150)	(0.0150)
Medium-skilled	-0.0061***	-0.0062***	-0.0060***	-0.0060***	0.0541***	0.0499***	0.0484***	0.0480***
	(0.0002)	(0.0003)	(0.0003)	(0.0003)	(0.0128)	(0.0130)	(0.0131)	(0.0131)
Medium-skilled managers	-0.0074***	-0.0072***	-0.0044***	-0.0043***	0.0390**	0.0378**	0.0317**	0.0294*
	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0153)	(0.0154)	(0.0154)	(0.0154)
High-skilled	-0.0106***	-0.0111***	-0.0080***	-0.0081***	0.1004***	0.0939***	0.0879***	0.0895***
	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0302)	(0.0303)	(0.0304)	(0.0303)
High-skilled managers	-0.0124***	-0.0120***	-0.0077***	-0.0074***	0.1143	0.1152	0.1049	0.1062
	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0787)	(0.0774)	(0.0773)	(0.0777)
Age 15–29			0.0228***	0.0210***			-0.0460***	-0.0583***
			(0.0002)	(0.0002)			(0.0042)	(0.0049)
Age 45–			-0.0081***	-0.0081***			-0.0515***	-0.0503***
			(0.0001)	(0.0001)			(0.0061)	(0.0061)
Married				-0.0041***				-0.0219***
				(0.0001)				(0.0046)
Urban location				-0.0019***				-0.0110**
				(0.0002)				(0.0056)
In birth county (=1)				0.0010***				0.0256***
				(0.0001)				(0.0045)
Census-year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Birth-parish FE	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Observations	4,637,489	4,548,664	4,548,664	4,548,664	66,839	66,737	66,737	66,737
Mean dep. var.	0.01	0.01	0.01	0.01	0.22	0.22	0.22	0.22

Notes: OLS regressions for full male working-age sample. In columns 1–4, the dependent variable is equal to one if an individual emigrates in the following ten year period and zero if not (return migrants are excluded from the analysis after their return). In columns 5–8, the dependent variable is equal to one if an individual returns to Sweden in the following ten year period and zero if the individual has emigrated, but not returned, in the same period. All determinants are measured in the census prior to emigrating. Indicators for medium-skilled manual workers and the age category 30-44 years are omitted. Robust standard errors are given in parentheses. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

Table B.9: Migrant selection (women)

Dependent variable:	Emigrant				Returnee			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Unskilled	-0.0005***	-0.0008***	-0.0005**	-0.0010***	-0.0385***	-0.0242*	-0.0238*	-0.0251*
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0121)	(0.0128)	(0.0128)	(0.0130)
Lower-skilled	-0.0010***	-0.0008***	-0.0006***	-0.0009***	-0.0145	-0.0173	-0.0194	-0.0203
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0157)	(0.0164)	(0.0164)	(0.0164)
Farmers	-0.0026***	-0.0031***	-0.0007***	-0.0015***	-0.0132	0.0033	-0.0019	-0.0032
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0140)	(0.0146)	(0.0148)	(0.0150)
Unskilled	-0.0007***	-0.0009***	-0.0006***	-0.0022***	-0.0130	-0.0153	-0.0116	-0.0164
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0123)	(0.0129)	(0.0129)	(0.0131)
Lower-skilled	0.0079***	0.0079***	0.0044***	0.0023***	-0.0054	-0.0043	0.0101	-0.0006
	(0.0001)	(0.0002)	(0.0001)	(0.0002)	(0.0105)	(0.0110)	(0.0112)	(0.0122)
Foremen	0.0002	0.0001	0.0013***	-0.0019***	0.0812***	0.0756***	0.0767***	0.0664***
	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0237)	(0.0243)	(0.0242)	(0.0247)
Lower-skilled	0.0013***	0.0015***	-0.0019***	-0.0042***	0.0911***	0.0650**	0.0746***	0.0627**
	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0257)	(0.0263)	(0.0264)	(0.0267)
Medium-skilled	-0.0006***	-0.0005**	-0.0011***	-0.0027***	0.0906***	0.0924***	0.0969***	0.0899***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0190)	(0.0195)	(0.0196)	(0.0199)
Medium-skilled managers	-0.0024***	-0.0024***	-0.0014***	-0.0014***	0.0349	0.0369	0.0335	0.0342
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0332)	(0.0340)	(0.0342)	(0.0342)
High-skilled	-0.0029***	-0.0028***	-0.0034***	-0.0046***	0.1147***	0.1247***	0.1320***	0.1245***
	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0365)	(0.0377)	(0.0377)	(0.0379)
High-skilled managers	-0.0035***	-0.0035***	-0.0020***	-0.0021***	0.1748	0.1427	0.1323	0.1315
	(0.0002)	(0.0003)	(0.0003)	(0.0003)	(0.1344)	(0.1400)	(0.1407)	(0.1409)
Age 15–29			0.0170***	0.0154***			-0.0341***	-0.0377***
			(0.0002)	(0.0002)			(0.0059)	(0.0062)
Age 45–			-0.0033***	-0.0036***			-0.0013	-0.0009
			(0.0001)	(0.0001)			(0.0085)	(0.0085)
Married				-0.0056***				-0.0172**
				(0.0001)				(0.0073)
Urban location				0.0009***				0.0098*
				(0.0002)				(0.0054)
In birth county (=1)				-0.0000				0.0024
				(0.0001)				(0.0053)
Census-year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Birth-parish FE	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Observations	3,684,664	3,620,907	3,620,907	3,620,907	35,896	35,743	35,743	35,743
Mean dep. var.	0.01	0.01	0.01	0.01	0.14	0.14	0.14	0.14

Notes: OLS regressions for full female working-age sample. In columns 1–4, the dependent variable is equal to one if an individual emigrates in the following ten year period and zero if not (return migrants are excluded from the analysis after their return). In columns 5–8, the dependent variable is equal to one if an individual returns to Sweden in the following ten year period and zero if the individual has emigrated, but not returned, in the same period. All determinants are measured in the census prior to emigrating. Indicators for medium-skilled manual workers and the age category 30-44 years are omitted. Robust standard errors are given in parentheses. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

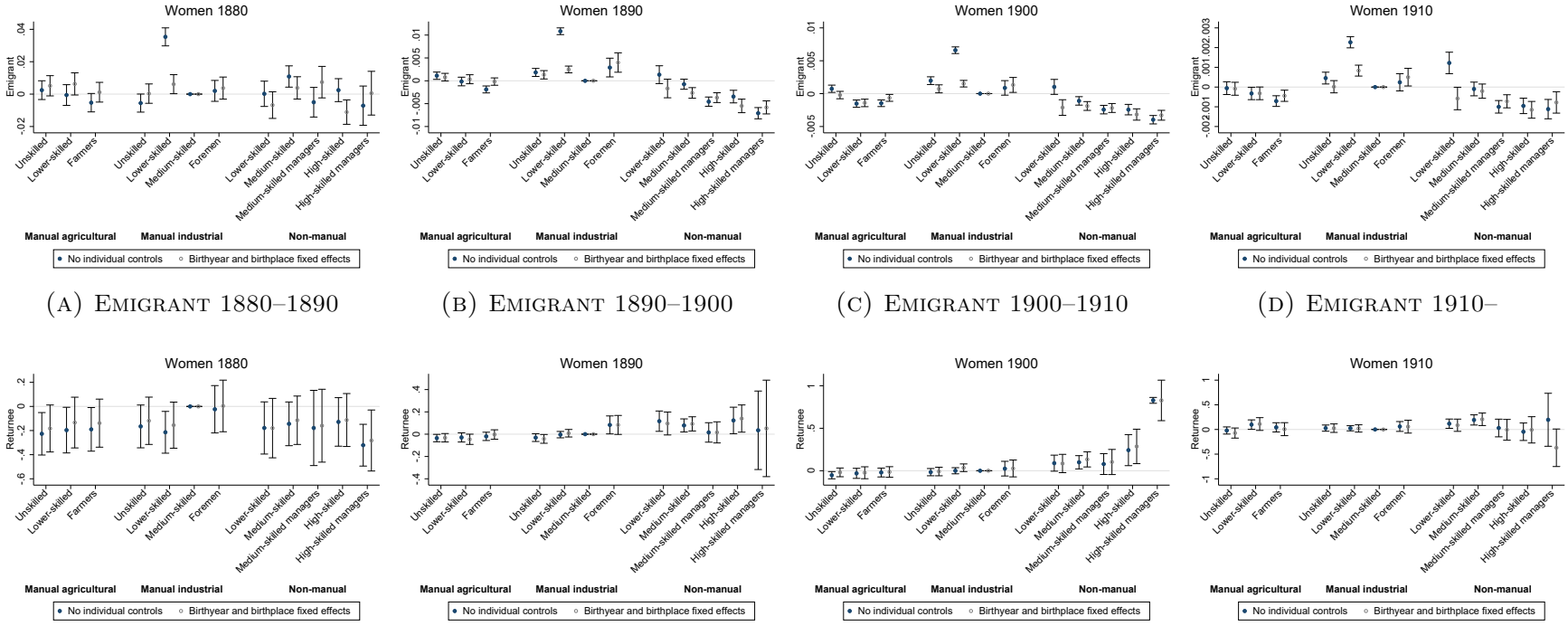


(A) Emigrant 1880–1890 (B) Emigrant 1890–1900 (C) Emigrant 1900–1910 (D) Emigrant 1910–

(E) Returnee 1880–1890 (F) Returnee 1890–1900 (G) Returnee 1900–1910 (H) Returnee 1910–

Figure B.8:
Decadel male migrant selection by occupational group

Notes: OLS regressions for full male working-age sample by census year. These figures display the relationships between migrant status and a set of indicator variables capturing occupational class for men and women separately. Figure A–D display the probability of emigrating, while figure E–H display the probability of returning, conditional on emigrating. All occupations are measured in the census prior to emigrating. The occupations of women are imputed using the occupation of the spouse, whenever it is missing for women. Medium-skilled manual workers are set as the reference group. Bars represent 95% confidence levels. Robust standard errors.

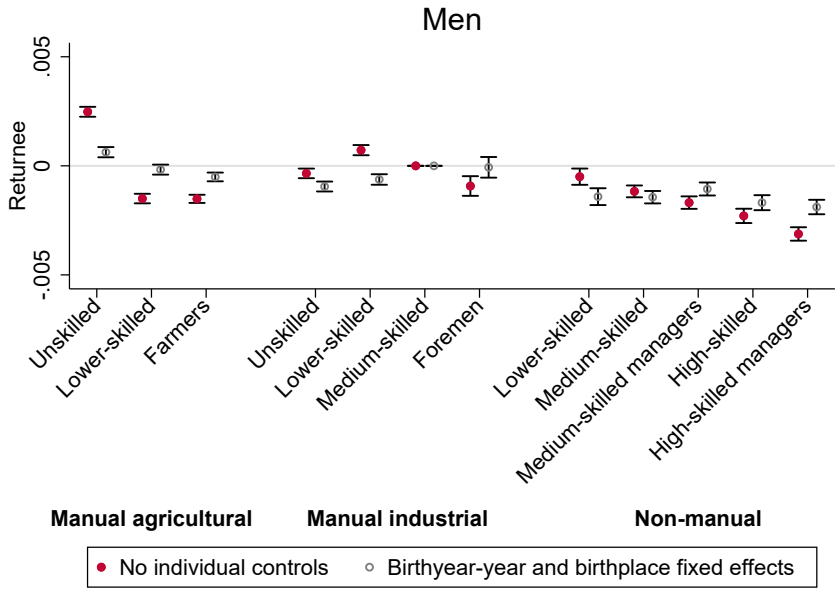


(A) Emigrant 1880–1890 (B) Emigrant 1890–1900 (C) Emigrant 1900–1910 (D) Emigrant 1910–

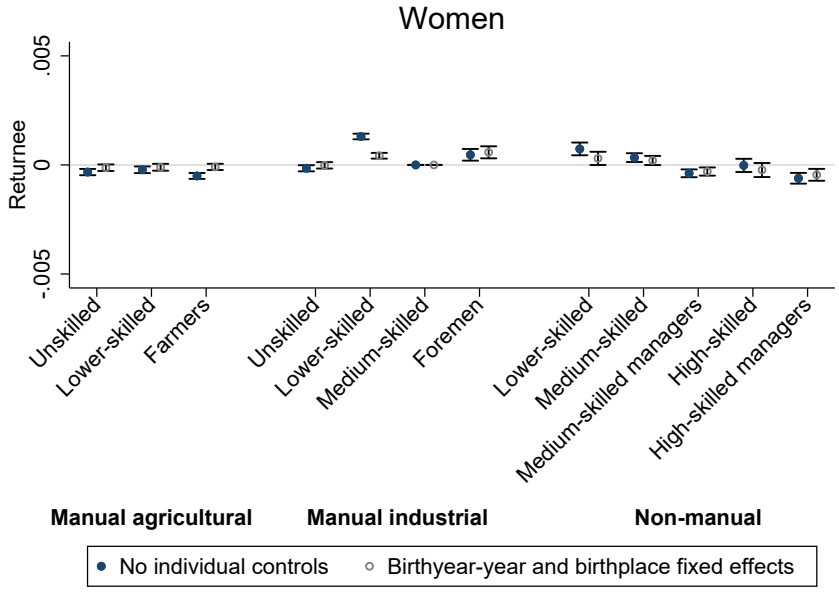
(E) Returnee 1880–1890 (F) Returnee 1890–1900 (G) Returnee 1900–1910 (H) Returnee 1910–

Figure B.9:
Decadel female migrant selection by occupational group

Notes: OLS regressions for full female working-age sample by census year. These figures display the relationships between migrant status and a set of indicator variables capturing occupational class for men and women separately. Figure A–D display the probability of emigrating, while figure E–H display the probability of returning, conditional on emigrating. All occupations are measured in the census prior to emigrating. The occupations of women are imputed using the occupation of the spouse, whenever it is missing for women. Medium-skilled manual workers are set as the reference group. Bars represent 95% confidence levels. Robust standard errors.



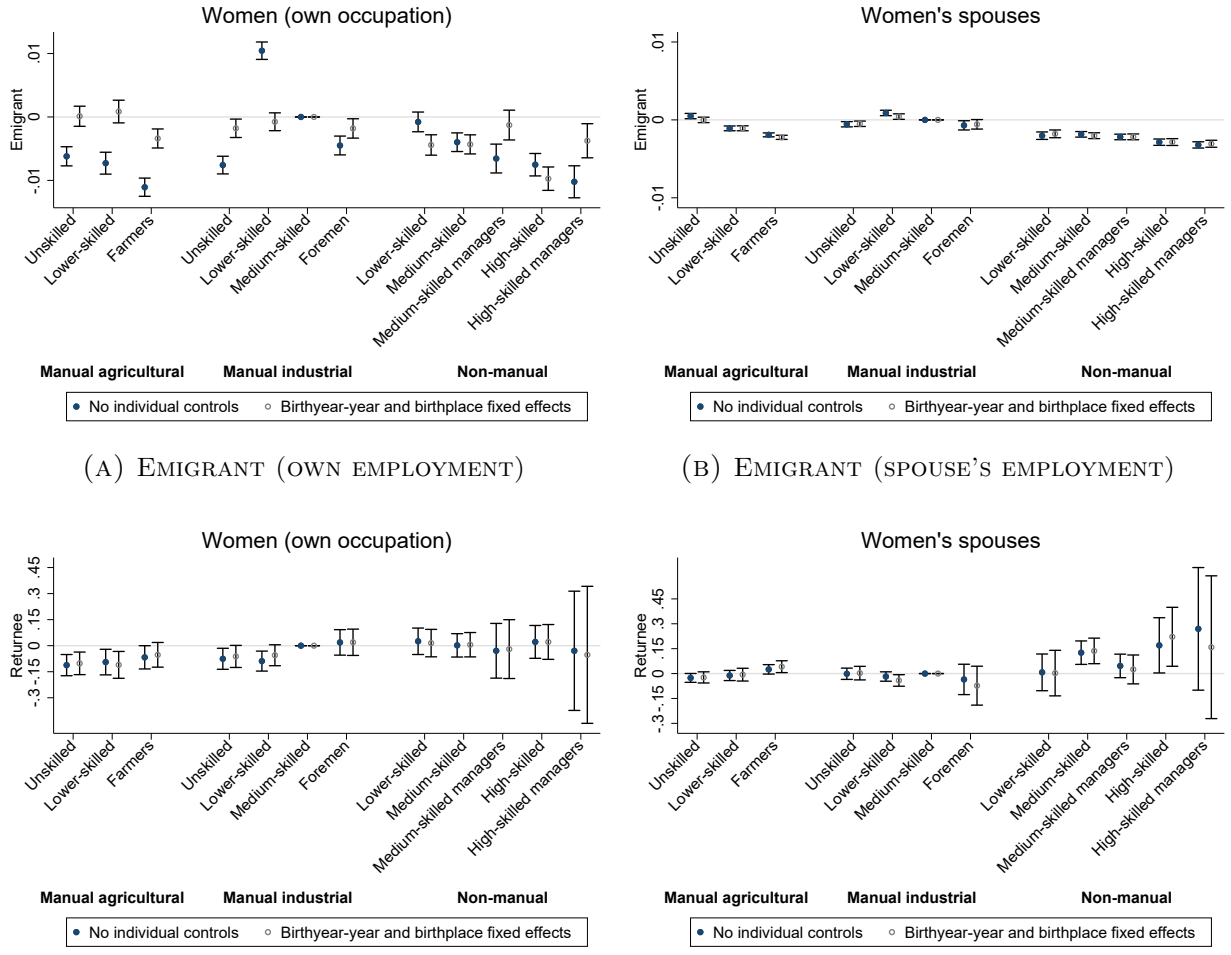
(A) MEN



(B) WOMEN

FIGURE B.10:
RETURNEE VS PERMANENT STAYER SELECTION

Notes: OLS regressions for full working-age sample by sex. These figures display the relationships between migrant status and a set of indicator variables capturing occupational class for men and women separately. All occupations are measured in the census prior to emigrating. The occupations of women are imputed using the occupation of the spouse, whenever it is missing for women. Medium-skilled manual workers are set as the reference group. Bars represent 95% confidence levels. Robust standard errors.



(A) Emigrant (own employment)

(B) Emigrant (spouse's employment)

(C) Returnee (own employment)

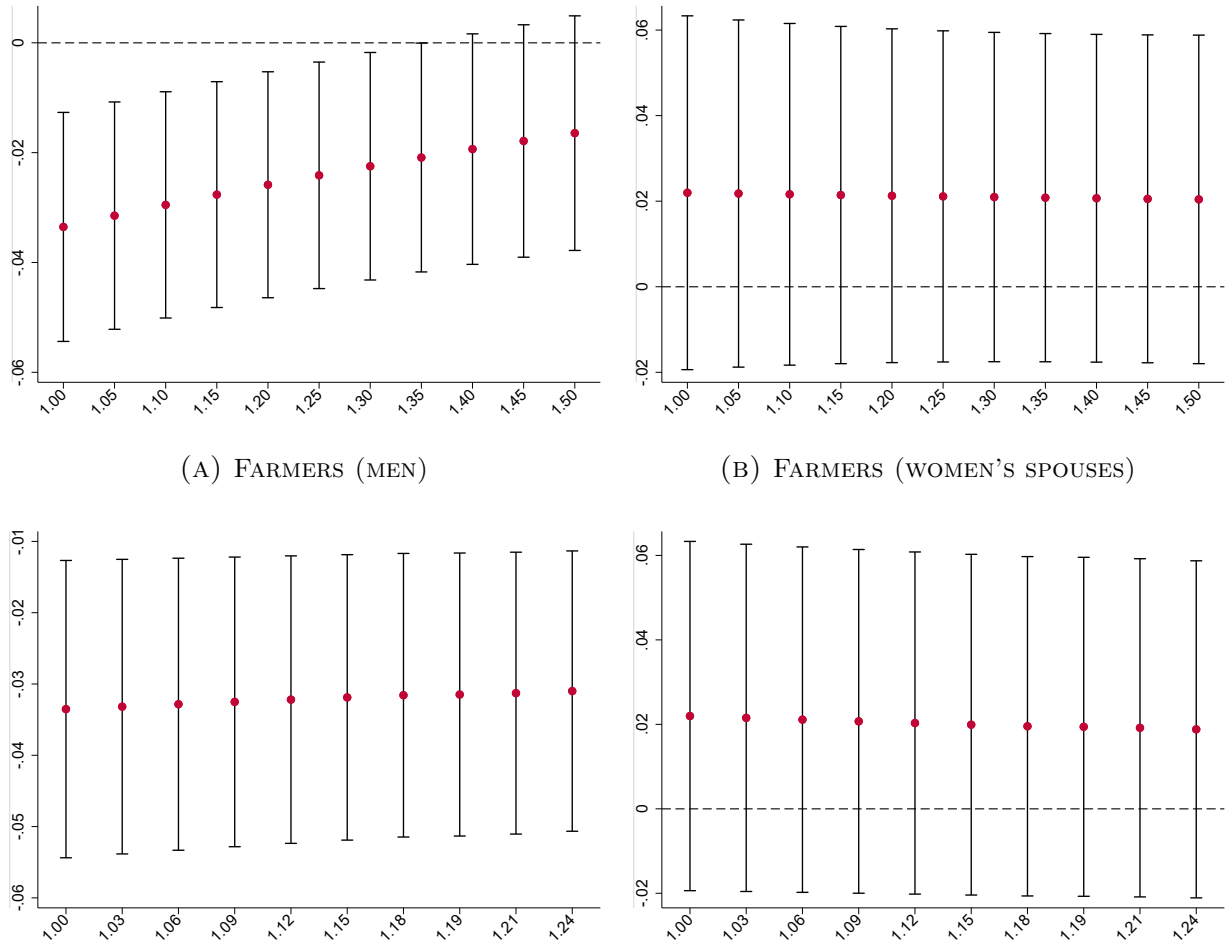
(D) Returnee (spouse's employment)

Figure B.11:
Female migrant selection: women's own employment vs women's spouses

Notes: OLS regressions for full female working-age sample by sex. These figures display the relationships between migrant status and a set of indicator variables capturing occupational class for men and women separately. Figure A–D display the probability of emigrating, while figure E–H display the probability of returning, conditional on emigrating. All occupations are measured in the census prior to emigrating. The occupations of women are imputed using the occupation of the spouse, whenever it is missing for women. Medium-skilled manual workers are set as the reference group. Bars represent 95% confidence levels. Robust standard errors.

Appendix C: Robustness

C.1 Figures



(A) Farmers (men)

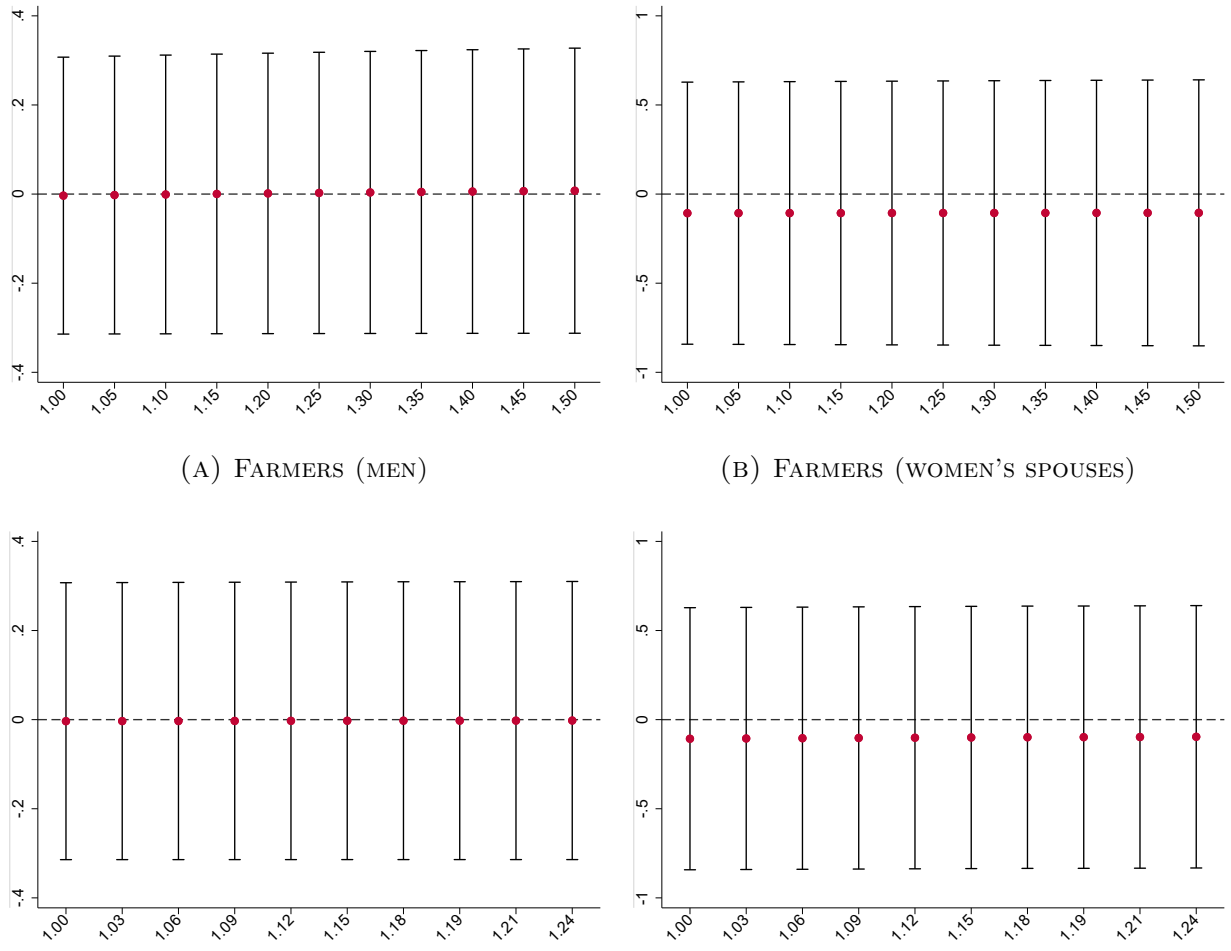
(B) Farmers (women's spouses)

(C) Farm hands (men)

(D) Farm hands (women's spouses)

Figure C.1:
Robustness of agricultural income score adjustment (1910)

Notes: OLS regressions for the linked childhood sample by sex. This figure displays separate coefficients for the regression in Table 2, column 2, using different adjustments for the income of farmers. Adjustments of 1.00 represent no adjustment. The adjustments used in the main results are 1.35 in panels A–B and 1.19 in panels C–D. All regressions include birth-parish, age and birth-order fixed effects, indicators for being eldest brother and eldest sister, respectively as well as sibling fixed effects. Standard errors clustered at the level of the household head. Bars represent 95% confidence intervals.



(A) Farmers (men)

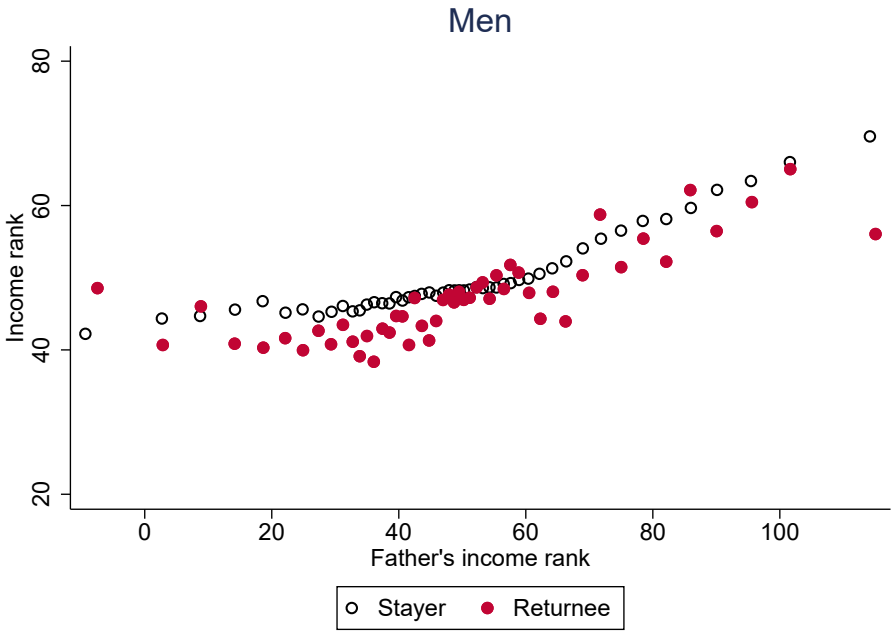
(B) Farmers (women's spouses)

(C) Farm hands (men)

(D) Farm hands (women's spouses)

Figure C.2:
Robustness of agricultural income adjustment (1930)

Notes: OLS regressions for the linked childhood sample by sex. This figure displays separate coefficients for the regression in Table 2, column 2, using different adjustments for the income of farmers. Adjustments of 1.00 represent no adjustment. The adjustments used in the main results are 1.35 in panels A–B and 1.19 in panels C–D. All regressions include birth-parish, age and birth-order fixed effects, indicators for being eldest brother and eldest sister, respectively as well as sibling fixed effects. Standard errors clustered at the level of the household head. Bars represent 95% confidence intervals.



(A) MEN

(B) WOMEN

FIGURE C.3:
INTERGENERATIONAL MOBILITY: RANK-RANK CORRELATIONS WITH UNADJUSTED FARM INCOME

Notes: Binscatter plots for the linked childhood sample by sex. Controls include birth-parish, age and birth-order fixed effects, and indicators for being eldest brother and eldest sister, respectively. Returnees in solid circles.

C.2 Tables

Table C.1: Robustness: Including only individuals with siblings of the same sex observed in 1910

Dependent variable:	In home county (=1)		Married (=1)		Farmer (=1)		Income score (ln)		Income 1930 (ln)		Wealth 1930 (ln)	
Panel A: Men	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Returnee	0.050***	0.053***	0.072***	0.277***	0.114***	0.144***	-0.012**	-0.013*	0.142**	0.048	0.973***	0.881***
	(0.005)	(0.008)	(0.008)	(0.005)	(0.006)	(0.011)	(0.005)	(0.008)	(0.067)	(0.103)	(0.094)	(0.151)
Observations	314,111	147,674	311,502	146,569	289,256	136,297	314,092	147,664	78,511	37,022	78,511	37,022
Mean dep. var.	0.77	0.77	0.68	0.69	0.20	0.19	7.44	7.44	6.08	6.16	5.24	5.27
Panel B: Women	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Returnee	-0.010	0.000	0.098***	0.328***	0.053***	0.098***	0.026***	0.023*	0.277**	0.311	0.950***	0.999***
	(0.008)	(0.012)	(0.010)	(0.006)	(0.008)	(0.015)	(0.009)	(0.013)	(0.116)	(0.189)	(0.174)	(0.272)
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Childhood hh controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	328,701	156,676	327,237	156,244	272,447	130,392	218,022	102,684	49,267	23,169	49,267	23,169
Mean dep. var.	0.76	0.75	0.67	0.67	0.19	0.19	7.48	7.49	6.10	6.15	5.39	5.41

Notes: OLS regressions for the linked childhood sample by sex. Returnee is an indicator variable equal to one if the individual is linked to the emigrant lists 1880-1910 and observed in the 1910 census, and zero otherwise. Individual controls include age and birth-order fixed effects, and indicators for being eldest brother and eldest sister, respectively. Childhood household controls include fixed effects for the childhood municipality and the social class in 1880. Sibling FE are fixed effects for the household head in 1880. Standard errors clustered at the level of the household head. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

Table C.2: Robustness: Altering cutoff for short and long stays abroad

Dependent variable:	In home county		In rural location		Farmer		Wealth (ln)	
Panel A: Men	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Returnee <7 Years Abroad	-0.020	-0.020	0.038**	0.034*	0.021	0.024	0.544	0.628
	(0.018)	(0.018)	(0.018)	(0.018)	(0.020)	(0.020)	(0.393)	(0.408)
Returnee ≥ 7 Years Abroad	0.089***	0.090***	0.078***	0.069***	0.116***	0.125***	1.115**	1.263**
	(0.023)	(0.024)	(0.023)	(0.023)	(0.034)	(0.037)	(0.564)	(0.581)
Observations	145,201	145,201	145,201	145,201	128,394	128,394	21,859	21,859
Mean dep. var.	0.77	0.77	0.55	0.55	0.19	0.19	5.49	5.49
Panel B: Women	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Returnee <7 Years Abroad	0.001	0.002	0.013	0.027	0.040	0.049	1.476**	1.595**
	(0.023)	(0.024)	(0.026)	(0.027)	(0.030)	(0.031)	(0.706)	(0.696)
Returnee ≥ 7 Years Abroad	-0.043	-0.041	0.028	0.065*	0.029	0.055	0.843	1.254
	(0.032)	(0.035)	(0.033)	(0.037)	(0.042)	(0.048)	(0.788)	(0.939)
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sibling FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Emigration period FE	No	Yes	No	Yes	No	Yes	No	Yes
Observations	155,061	155,061	155,061	155,061	115,086	115,086	23,264	23,264
Mean dep. var.	0.75	0.75	0.53	0.53	0.18	0.18	4.20	4.20

Notes: OLS regressions for linked childhood sample. Returnee < 5 Years Abroad is an indicator variable equal to one if the individual has emigrated and returned by 1910 and has spent less than 5 years abroad, and zero otherwise. Returnee ≥ 5 Years Abroad is an indicator variable equal to one if the individual has emigrated and returned by 1910 and has more than 5 years abroad, and zero otherwise. The omitted category is permanent stayers. Individual controls include age and birth-order fixed effects, and indicators for being eldest brother and eldest sister, respectively. Childhood household controls include fixed effects for the childhood municipality and the social class in 1880. Sibling FE are fixed effects for the household head in 1880. Standard errors clustered at the level of the household head. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$. Sample of individuals born 1866-80.

Table C.3: Robustness: Winsorizing wealth

Dependent variable:	Wealth 1930 (ln)					
	None		5th percentile		10th percentile	
Panel A: Men	(1)	(2)	(3)	(4)	(5)	(6)
Returnee	0.973***	0.404*	0.981***	0.417*	0.973***	0.415*
	(0.094)	(0.230)	(0.093)	(0.228)	(0.091)	(0.225)
Observations	78,511	22,334	78,511	22,334	78,511	22,334
Mean dep. var.	5.24	5.51	5.19	5.46	5.12	5.38
Panel B: Women's spouses	(1)	(2)	(3)	(4)	(5)	(6)
Returnee	0.950***	0.984**	0.923***	0.962**	0.902***	0.935**
	(0.174)	(0.486)	(0.169)	(0.480)	(0.165)	(0.472)
Observations	49,267	9,196	49,267	9,196	49,267	9,196
Mean dep. var.	5.39	5.57	5.33	5.52	5.25	5.44
Panel C: Women	(1)	(2)	(3)	(4)	(5)	(6)
Returnee	0.010	0.082	0.817***	0.549	0.812***	0.552
	(0.101)	(0.275)	(0.143)	(0.364)	(0.142)	(0.361)
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes
Childhood hh controls	Yes	No	Yes	No	Yes	No
Sibling FE	No	Yes	No	Yes	No	Yes
Observations	83,379	23,538	83,379	23,538	83,379	23,538
Mean dep. var.	1.18	1.29	4.06	4.18	4.02	4.15

Notes: OLS regressions for the linked childhood sample by sex. Returnee is an indicator variable equal to one if the individual is linked to the emigrant lists 1880-1910 and observed in the 1910 census, and zero otherwise. Individual controls include age and birth-order fixed effects, and indicators for being eldest brother and eldest sister, respectively. Childhood household controls include fixed effects for the childhood municipality and the social class in 1880. Sibling FE are fixed effects for the household head in 1880. Standard errors clustered at the level of the household head. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

Table C.4: Robustness: Weighting regressions with the inverse probability of being linked across censuses

Dependent variable:	In home county (=1)		Married (=1)		Farmer (=1)		Income score (ln)		Income 1930 (ln)		Wealth 1930 (ln)	
Panel A: Men	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Returnee	0.053***	0.016	0.072***	0.064***	0.117***	0.050***	-0.010**	-0.021**	0.150**	0.011	1.041***	0.449*
	(0.005)	(0.010)	(0.008)	(0.018)	(0.006)	(0.012)	(0.005)	(0.010)	(0.068)	(0.164)	(0.095)	(0.229)
Observations	314,111	147,308	311,502	145,384	289,256	130,228	314,092	147,290	78,511	22,334	78,511	22,334
Mean dep. var.	0.77	0.77	0.68	0.68	0.18	0.18	7.44	7.44	6.06	6.11	5.14	5.41
Panel B: Women	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Returnee	-0.008	-0.014	0.097***	0.051**	0.056***	0.024	0.026***	0.017	0.300***	-0.120	0.956***	0.981**
	(0.008)	(0.014)	(0.010)	(0.022)	(0.008)	(0.016)	(0.009)	(0.019)	(0.115)	(0.374)	(0.179)	(0.482)
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Childhood hh controls	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
Sibling FE	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Observations	328,701	156,195	327,237	155,167	272,447	115,887	218,022	79,446	49,267	9,196	49,267	9,196
Mean dep. var.	0.76	0.75	0.67	0.66	0.19	0.18	7.48	7.47	6.08	6.07	5.30	5.49

Notes: OLS regressions for the linked childhood sample by sex. Weighted regressions using the inverse probability that the individual is included in the 1910 census. Returnee is an indicator variable equal to one if the individual is linked to the emigrant lists 1880-1910 and observed in the 1910 census, and zero otherwise. Individual controls include age and birth-order fixed effects, and indicators for being eldest brother and eldest sister, respectively. Childhood household controls include fixed effects for the childhood municipality and the social class in 1880. Sibling FE are fixed effects for the household head in 1880. Standard errors clustered at the level of the household head. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

TABLE C.5: ROBUSTNESS: ALTERNATIVE STANDARD ERROR CALCULATIONS

Dependent variable:	In home county (=1)		Married (=1)		Farmer (=1)		Income score (ln)		Income 1930 (ln)		Wealth 1930 (ln)	
Panel A: Men	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Returnee	0.050	0.016	0.072	0.062	0.114	0.047	-0.012	-0.019	0.142	0.006	0.973	0.404
Robust	(0.005)***	(0.010)	(0.008)***	(0.017)***	(0.006)***	(0.012)***	(0.005)**	(0.010)*	(0.067)**	(0.164)	(0.093)***	(0.232)*
*Cluster-robust (Household head)**	(0.005)***	(0.010)	(0.008)***	(0.017)***	(0.006)***	(0.012)***	(0.005)**	(0.010)*	(0.067)**	(0.162)	(0.094)***	(0.230)*
Cluster-robust (hh hisclass)	(0.008)***	(0.009)*	(0.011)***	(0.014)***	(0.007)***	(0.007)***	(0.007)	(0.010)*	(0.058)**	(0.111)	(0.268)***	(0.186)*
Cluster-robust (Municipality)	(0.005)***	(0.010)*	(0.008)***	(0.017)***	(0.007)***	(0.012)***	(0.005)**	(0.010)*	(0.068)**	(0.166)	(0.091)***	(0.229)*
Observations	314,111	147,308	311,502	145,384	289,256	130,228	314,092	147,290	78,511	22,334	78,511	22,334
Mean dep. var.	0.77	0.77	0.68	0.68	0.20	0.19	7.44	7.44	6.08	6.13	5.24	5.51
Panel B: Women	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Returnee	-0.010	-0.017	0.098	0.048	0.053	0.024	0.026	0.018	0.277	-0.097	0.950	0.984
Robust	(0.008)	(0.014)	(0.010)***	(0.022)**	(0.008)***	(0.017)	(0.009)***	(0.019)	(0.116)**	(0.380)	(0.174)***	(0.494)**
*Cluster-robust (Household head)**	(0.008)	(0.014)	(0.010)***	(0.022)**	(0.008)***	(0.017)	(0.009)***	(0.019)	(0.116)**	(0.379)	(0.174)***	(0.486)**
Cluster-robust (hh hisclass)	(0.010)	(0.013)	(0.009)***	(0.016)**	(0.015)***	(0.008)**	(0.010)**	(0.023)	(0.125)**	(0.265)	(0.285)***	(0.375)**
Cluster-robust (Municipality)	(0.008)	(0.013)	(0.010)***	(0.022)**	(0.009)***	(0.017)	(0.009)***	(0.019)	(0.113)**	(0.414)	(0.175)***	(0.498)**
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Childhood hh controls	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
Sibling FE	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Observations	328,701	156,195	327,237	155,167	272,447	115,887	218,022	79,446	49,267	9,196	49,267	9,196
Mean dep. var.	0.76	0.75	0.67	0.66	0.19	0.18	7.48	7.48	6.10	6.09	5.39	5.57

Notes: OLS regressions for the linked childhood sample by sex. Standard errors calculated with different methods within parenthesis. *Robust* are heteroskedasticity-robust standard errors. *Cluster-robust (Household head)* are cluster-robust standard errors at the level of the household head in childhood 1880 (this is the default). *Cluster-robust (hh hisclass)* are cluster-robust standard errors at the level of the hisclass of the household head. *Cluster-robust (Municipality)* are cluster-robust standard errors at the level of the childhood municipality. Income score, income, and wealth are shown for the spouse for women. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.