

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

by Olof Ejermo, Kerstin Enflo, Björn Eriksson, and Erik Prawitz

Online Appendix A: Definitions and data

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 (Minnesota Population Center, 2020), a sample of the 1930 census provided by The Swedish National Archives (Swedish National Archives, 2022), the immigration registers provided by Statistics Sweden and Swedish Emigrant Institute (2014) 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 virtu-

ally 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 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 A.1 provides a summary of the distribution of the sample across regions.

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	17.0
Göteborgs och Bohus län	457,010	87,101	19.1
Älvsborgs 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	11.2
Jämtlands län	134,514	49,600	36.9
Västerbottens län	204,031	203,999	100.0
Norrbottnens län	199,826	14,159	7.1
Total	6,142,191	1,964,303	32.0

Note: This table reports the population size for the different counties and sample sizes and their shares for the different counties for the 1930 census which had been digitized and available for linking in the paper.
Source: Statistics Sweden (1935).

The emigration register contain the majority of all emigrants, but is not as complete as the censuses and the death index. The official statistics reported that 1,021,306 emigrants left for North America between 1860-1920, a number which is generally considered an underestimation (Statistics Sweden, 1860-1910 and Statistics Sweden, 1911-1920). The equivalent number of emigrants recorded in EMIBAS is 828,204. 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.

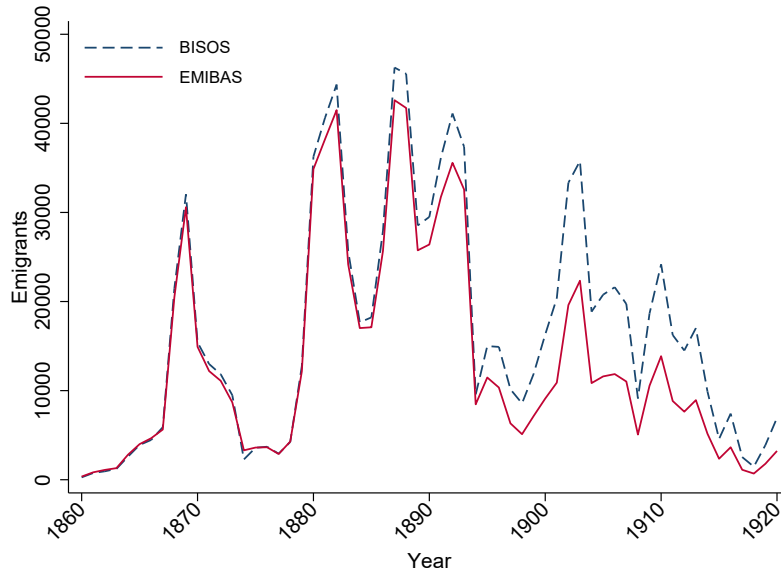


FIGURE A.1: EMIGRATION 1860-1920

Notes: This figure shows the number of emigrants by departure year recorded in the official statistics and the emigration registers (EMIBAS). Aggregate total emigration 1860-1920 from the official statistics published by Statistics Sweden (1860-1910) and Statistics Sweden (1911-1920). Emigration registers from Swedish Emigrant Institute and Federation of Swedish Genealogical Societies (2005)

A.2 Linking procedure

In this Online 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 (Ruggles, 2002).³⁷ Disqualified variables therefore include information such as current location of residence and civil 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 when available) 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. Table A.3 displays a summary of all index variables available across the different data sources.

Table A.2 shows emigrant and census linkage rates for our main sample of individuals aged below 15 years in 1880 (panel A) and the full population (panel B). Table A.3 displays the associated link variables.

³⁷Recent 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: EMIGRANT AND CENSUS LINKAGE RATES
(A) MAIN CHILDHOOD SAMPLE

Linked to						
Census	Obs.	Census 1890	Census 1900	Census 1910	Death index	Emigration register
Men						
1880	754,434	66.4%			50.6%	5.2%
1890	642,201		65.0%		58.2%	5.5%
1900	550,031			73.4%	71.1%	2.7%
1910	486,083				77.3%	0.3%
1930	126,297				87.0%	
Women						
1880	737,999	67.1%			55.5%	4.5%
1890	638,400		65.8%		63.2%	4.9%
1900	551,931			74.6%	74.6%	1.5%
1910	504,740				77.9%	0.2%
1930	132,963				85.7%	
(B) FULL POPULATION						
Linked to						
Census	Obs.	Census 1890	Census 1900	Census 1910	Death index	Emigration register
Men						
1880	4,624,825	62.9%			59.8%	4.2%
1890	4,843,782		67.8%		67.3%	2.2%
1900	5,200,111			71.5%	73.6%	1.6%
1910	5,586,360				75.4%	0.5%
Total	22,219,381				70.3%	1.8%

Note: This table reports in Panel (A) the linkage rate from one census to the next (Minnesota Population Center, 2020) and to the death index (Federation of Swedish Genealogical Societies, 2018) and emigration registers (Swedish Emigrant Institute and Federation of Swedish Genealogical Societies, 2005), respectively, by men and women for individuals aged below 15 years in 1880. Panel (B) shows the linkage rate from one census to the next and to the death index and emigration registers, respectively for the full sample.

TABLE A.3: 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	YES	YES	YES	YES	YES	YES	YES	YES
Observations	4,624,825	4,843,782	5,200,111	5,586,360	1,964,303	11,140,380	570,662	525,998

Note: This table gives an overview the variables available for linking among the different sources of data. Sources: 1880-1910 census (Minnesota Population Center, 2020), 1930 census (Swedish National Archives, 2022), death index (Federation of Swedish Genealogical Societies, 2018), emigration registers (Swedish Emigrant Institute and Federation of Swedish Genealogical Societies, 2005) and immigration registers (Statistics Sweden and Swedish Emigrant Institute, 2014)

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 (see Jaro, 1989; Winkler, 1990, for a more detailed discussion of the Jaro-Winkler algorithm).

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., 2020). 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. We complement the linked emigration data above with links between the censuses and the death index. We link each census to the subsequent census, and the death index, following the procedure described above. In terms of the share that are linked between each source the algorithm performs well. 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 description of the census and death index linking see Eriksson (2015); Dribe and Eriksson (2018).

A.3 Complementary linked data

We also use links from the list of Swedish emigrants to the US census in 1900 (Ruggles et al., 2021) as created in Castillo (2024). 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, 2024), which lowers the linking rate compared to those that rely on Swedish data.³⁸

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.

³⁸Roughly 16% of the Swedish-born men in the US Census of 1900 are linked to the emigrant register.

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 (see Figure A.2). In the years immediately following a census the linkage rates are about 70 per cent, and never falls below 50 per cent in a given year.

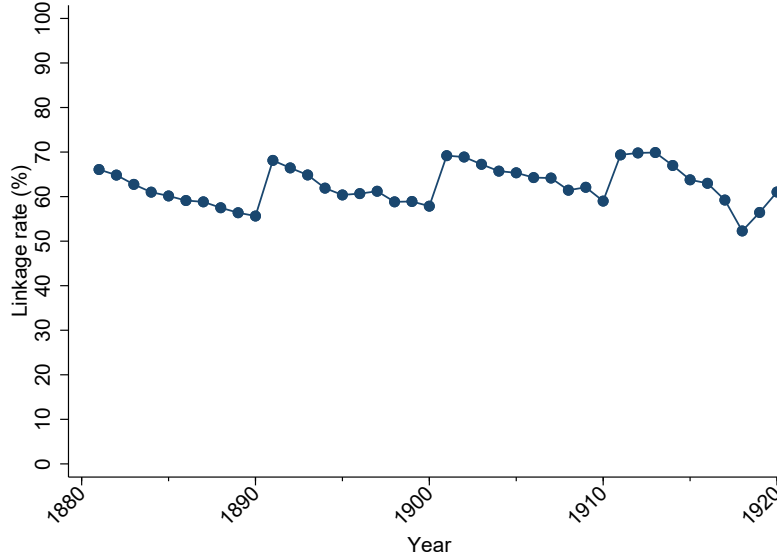


FIGURE A.2: EMIGRATION REGISTER LINKAGE RATES BY DEPARTURE YEAR, 1880-1920

Notes: This figure shows the linkage rate from EMIBAS by departure year. We only consider Swedish born individuals with information on birth parish and birth year. Individuals without such information are excluded.

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 by Lambert et al. (2013) in order to code occupations into distinct classes. HISCLASS is based on the HISCO coding scheme

(Van Leeuwen et al., 2002; Van Leeuwen, 2016) 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. Updated HISCO codes for the Swedish censuses based on revised coding of occupational strings were obtained by the SwedPop (2022) project for the Swedish historical censuses.

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

Online Appendix B: Additional material

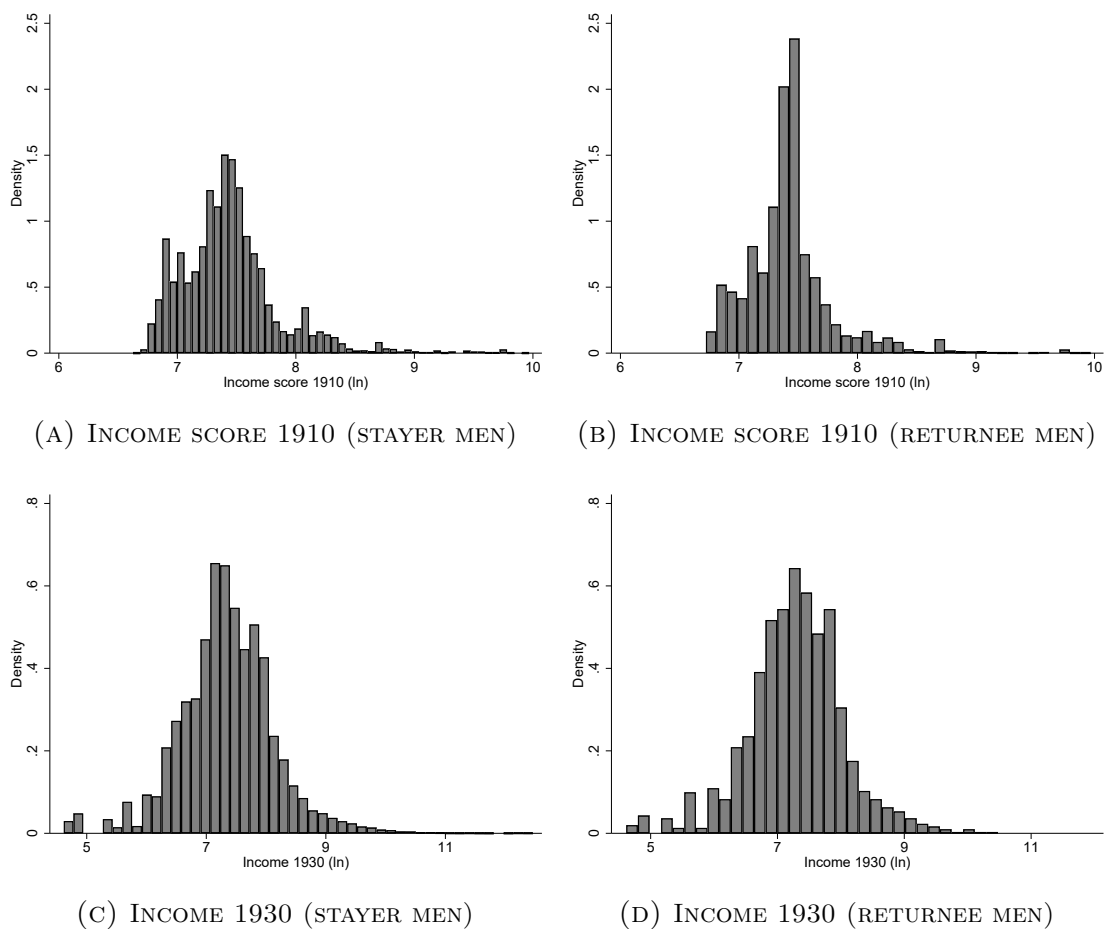
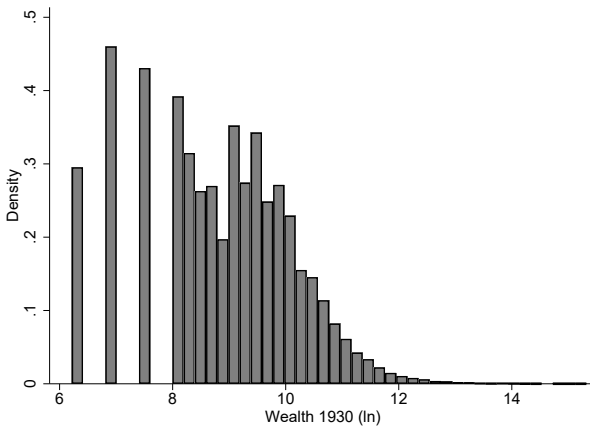
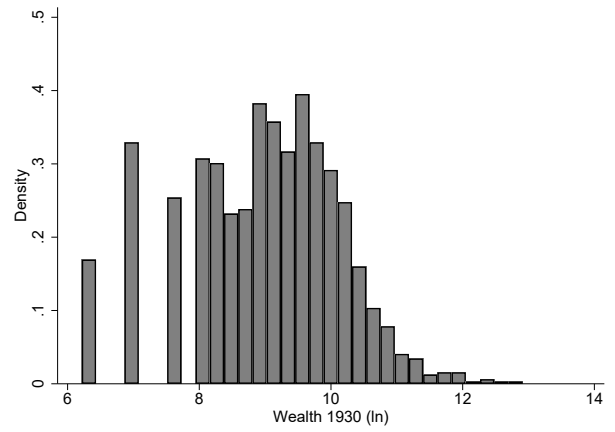


FIGURE B.1: DISTRIBUTION OF INCOME SCORE 1910 AND INCOME 1930

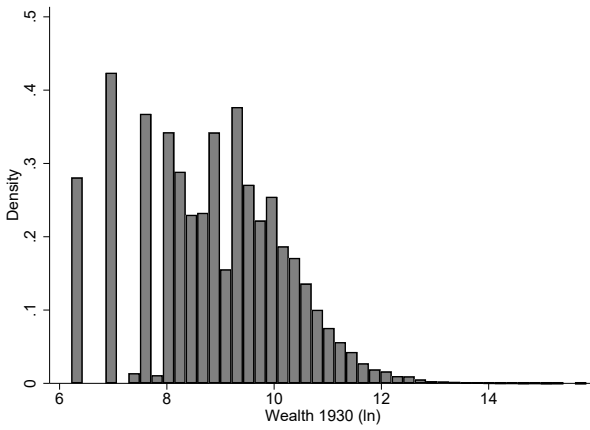
Notes: This figure displays the distribution of income scores in 1910 and income in 1930, both in natural logarithms, for men by returnee status in our main sample of individuals aged below 15 years in 1880. Income scores 1910 are only calculated for men. Individuals with zero income in 1930 are omitted. They represent 26.58% and 40.77% for male returnees and stayers, respectively.



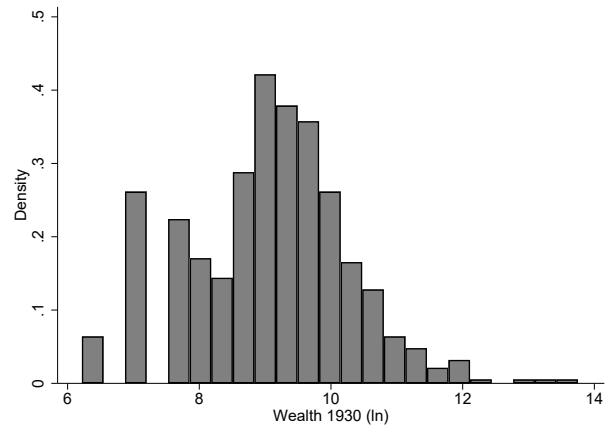
(A) STAYER MEN



(B) RETURNEE MEN



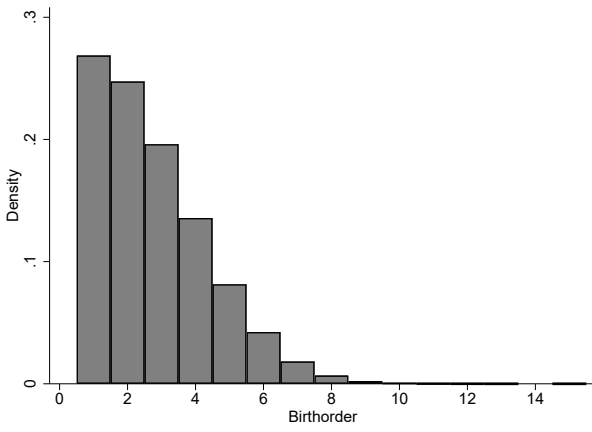
(C) STAYER WOMEN



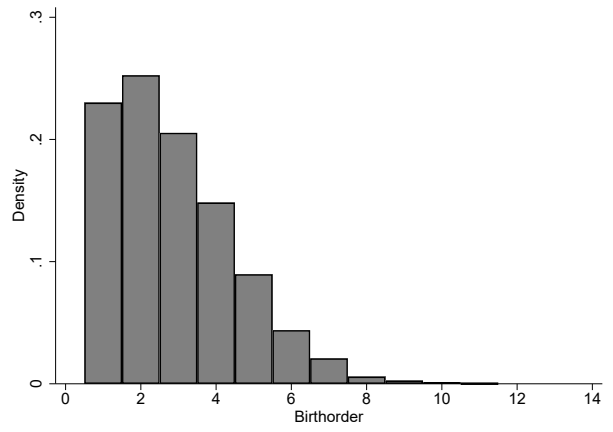
(D) RETURNEE WOMEN

FIGURE B.2: DISTRIBUTION OF WEALTH 1930

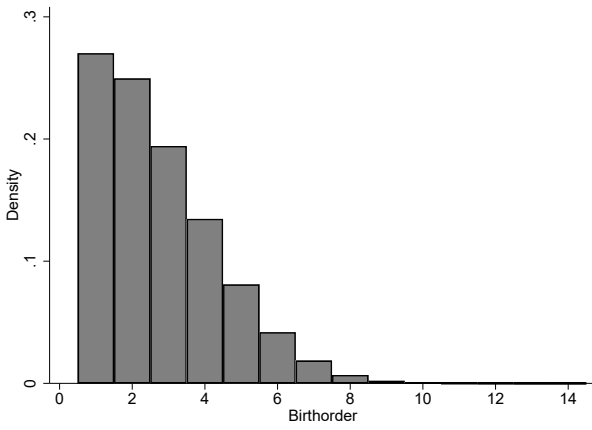
Notes: This figure displays the distribution of wealth in 1930, in natural logarithms, for men and women by returnee status in our main sample of individuals aged below 15 years in 1880. Individuals with zero wealth are omitted.



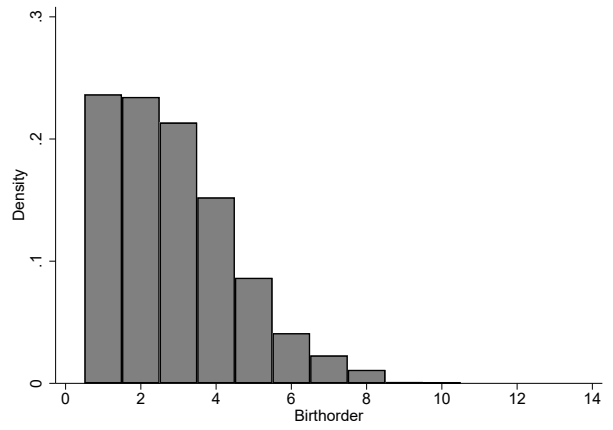
(A) STAYER MEN



(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 in our main sample of individuals aged below 15 years in 1880.

TABLE B.1: MOST COMMON OCCUPATIONS

	Freq.	Percent	Cum.
<i>Male returnees:</i>			
General farmer	2154	39.55	39.55
Farm worker	413	7.58	47.14
Small subsistence farmer	248	4.55	51.69
Carpenter	209	3.84	55.53
Working proprietor	193	3.54	59.07
Factory worker	106	1.95	61.02
Blacksmith	100	1.84	62.85
Logger	88	1.62	64.47
Worker	88	1.62	66.09
Labourer	79	1.45	67.54
<i>Male stayers:</i>			
General farmer	61100	21.29	21.29
Farm worker	29498	10.28	31.56
Small subsistence farmer	14241	4.96	36.52
Carpenter	10267	3.58	40.10
Labourer	8259	2.88	42.98
Working proprietor	7491	2.61	45.59
Factory worker	6686	2.33	47.92
Worker	6420	2.24	50.16
Blacksmith	5372	1.87	52.03
Seaman	5251	1.83	53.86
<i>Female returnees:</i>			
Hand and machine sewer	81	12.80	12.80
Housekeeper	70	11.06	23.85
General farmer	59	9.32	33.18
Working proprietor	51	8.06	41.23
House servant	50	7.90	49.13
Personal maid	23	3.63	52.76
Professional nurse	17	2.69	55.45
Small subsistence farmer	17	2.69	58.14
Worker	14	2.21	60.35
Professional midwife	13	2.05	62.40
<i>Female stayers:</i>			
House servant	11151	16.29	16.29
Hand and machine sewer	7208	10.53	26.82
Housekeeper	6136	8.96	35.78
Personal maid	4581	6.69	42.47
General farmer	3406	4.98	47.45
First-level education teacher	3131	4.57	52.02
Working proprietor	2991	4.37	56.39
Factory worker	1971	2.88	59.27
Retail trade salesperson	1751	2.56	61.83
Farm worker	1481	2.16	63.99

Notes: This table shows the ten most common occupational titles among returnees and stayers for both men and women separately in our main sample of individuals aged below 15 years in 1880. Individuals with no occupation are excluded.

TABLE B.2: MIGRANT SELECTION (MEN)

Dependent variable:	Emigrant		Returnee		Returnee	
Subsample:	All		Excluding permanent migrants		Excluding stayers	
	(1)	(2)	(3)	(4)	(5)	(6)
Unskilled	-0.011*** (0.002)	-0.016*** (0.002)	-0.004*** (0.001)	-0.004*** (0.001)	-0.021*** (0.007)	-0.016** (0.008)
Lower-skilled	0.005*** (0.002)	-0.012*** (0.002)	0.001 (0.001)	-0.002*** (0.001)	-0.003 (0.007)	0.010 (0.008)
Farmers	0.019*** (0.002)	-0.004** (0.002)	0.009*** (0.001)	0.006*** (0.001)	0.049*** (0.007)	0.069*** (0.007)
Unskilled	-0.025*** (0.002)	-0.012*** (0.002)	-0.005*** (0.001)	-0.004*** (0.001)	-0.006 (0.011)	-0.027** (0.011)
Lower-skilled	0.003 (0.002)	-0.005** (0.002)	-0.001 (0.001)	-0.002** (0.001)	-0.015* (0.008)	-0.007 (0.009)
Foremen	0.015** (0.007)	0.012* (0.007)	0.001 (0.003)	0.001 (0.003)	-0.016 (0.025)	-0.018 (0.026)
Lower-skilled	-0.027*** (0.005)	-0.012*** (0.004)	-0.002 (0.002)	0.001 (0.002)	0.065** (0.033)	0.046 (0.033)
Medium-skilled	-0.021*** (0.003)	-0.011*** (0.002)	-0.000 (0.001)	0.001 (0.001)	0.057*** (0.015)	0.047*** (0.015)
Medium-skilled managers	-0.016*** (0.003)	-0.011*** (0.002)	-0.003** (0.001)	-0.002 (0.001)	0.010 (0.013)	0.009 (0.014)
High-skilled	-0.037*** (0.003)	-0.035*** (0.003)	-0.002 (0.002)	-0.002 (0.002)	0.098*** (0.027)	0.101*** (0.027)
High-skilled managers	-0.057*** (0.004)	-0.055*** (0.004)	-0.009*** (0.002)	-0.007*** (0.002)	0.093* (0.049)	0.101** (0.051)
Birthyear FE	No	Yes	No	Yes	No	Yes
Family controls	No	Yes	No	Yes	No	Yes
Childhood municip. FE	No	Yes	No	Yes	No	Yes
Observations	598,953	598,952	598,953	598,952	53,064	52,907
Mean dep. var.	0.09	0.09	0.02	0.02	0.22	0.22

Notes: OLS regressions for male sample of individuals aged below 15 years in 1880 (including permanent emigrants). 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 municip. FE are fixed effects for the municipality of residence in 1880. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

TABLE B.3: MIGRANT SELECTION (WOMEN)

Dependent variable:	Emigrant		Returnee		Returnee	
Subsample:	All		Excluding permanent migrants		Excluding stayers	
	(1)	(2)	(3)	(4)	(5)	(6)
Unskilled	-0.011*** (0.002)	-0.015*** (0.002)	-0.003*** (0.001)	-0.003*** (0.001)	-0.021*** (0.008)	-0.005 (0.008)
Lower-skilled	-0.000 (0.002)	-0.014*** (0.002)	-0.002** (0.001)	-0.003*** (0.001)	-0.020*** (0.007)	0.003 (0.008)
Farmers	-0.003** (0.001)	-0.020*** (0.002)	0.002*** (0.001)	0.001* (0.001)	0.039*** (0.007)	0.069*** (0.008)
Unskilled	-0.018*** (0.002)	-0.007*** (0.002)	-0.002** (0.001)	-0.001 (0.001)	0.028** (0.011)	0.010 (0.012)
Lower-skilled	0.000 (0.002)	-0.005*** (0.002)	-0.002** (0.001)	-0.002*** (0.001)	-0.022*** (0.009)	-0.010 (0.009)
Foremen	0.010 (0.007)	0.011 (0.007)	0.005* (0.003)	0.005* (0.003)	0.040 (0.030)	0.042 (0.031)
Lower-skilled	-0.030*** (0.004)	-0.013*** (0.004)	-0.003 (0.002)	-0.000 (0.002)	0.059* (0.034)	0.025 (0.035)
Medium-skilled	-0.027*** (0.002)	-0.020*** (0.002)	0.000 (0.001)	0.001 (0.001)	0.110*** (0.018)	0.108*** (0.018)
Medium-skilled managers	-0.023*** (0.002)	-0.020*** (0.002)	-0.002** (0.001)	-0.002* (0.001)	0.036** (0.015)	0.043*** (0.016)
High-skilled	-0.049*** (0.002)	-0.046*** (0.003)	-0.006*** (0.001)	-0.005*** (0.001)	0.098*** (0.037)	0.118*** (0.038)
High-skilled managers	-0.060*** (0.003)	-0.058*** (0.003)	-0.009*** (0.001)	-0.009*** (0.001)	0.065 (0.077)	0.092 (0.081)
Birthyear FE	No	Yes	No	Yes	No	Yes
Family controls	No	Yes	No	Yes	No	Yes
Childhood municip. FE	No	Yes	No	Yes	No	Yes
Observations	585,871	585,870	585,871	585,870	39,182	38,989
Mean dep. var.	0.07	0.07	0.01	0.01	0.18	0.18

Notes: OLS regressions for female sample of individuals aged below 15 years in 1880 (including permanent emigrants). 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 municip. FE are fixed effects for the municipality of residence in 1880. Standard errors clustered at the level of the household head. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

TABLE B.4: PRE-MIGRATION CHARACTERISTICS

	Men			Women		
	(1)	(2)	(3)	(4)	(5)	(6)
Age	1.639*** (0.082)	1.644*** (0.083)	1.518*** (0.114)	1.154*** (0.107)	1.158*** (0.109)	0.995*** (0.150)
Observations	147,472	147,472	147,472	156,278	156,278	156,278
Birth order	-0.034 (0.036)	-0.111*** (0.036)	-0.470*** (0.040)	0.071 (0.047)	0.024 (0.046)	-0.328*** (0.055)
Observations	147,472	147,472	147,472	156,278	156,278	156,278
Eldest brother/sister	0.013 (0.009)	0.022** (0.009)	0.102*** (0.015)	-0.011 (0.012)	-0.004 (0.012)	0.057*** (0.019)
Observations	147,472	147,472	147,472	156,278	156,278	156,278
Father's income score (ln)	-0.008 (0.007)	-0.001 (0.004)	0.000 (.)	-0.002 (0.010)	-0.001 (0.006)	0.000 (.)
Childhood hh controls	No	Yes	No	No	Yes	No
Sibling FE	No	No	Yes	No	No	Yes
Observations	144,656	144,656	144,656	153,154	153,154	153,154

Notes: OLS regressions by sex for individuals aged below 15 years in 1880 (including permanent emigrants). Each entry displays the coefficient on *Returnee* from a separate regression. The dependent variable is displayed on the left-hand side for each row. 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.5: RETURNEE SELECTION IN THE US AND SWEDISH LABOR MARKETS (MEN)

Dependent variable:	US income score 1900 (ln)				Father's 1880 income score (ln)	
	(1)	(2)	(3)	(4)	(5)	(6)
Return migrant (=1)	-0.100 (0.061)	-0.127** (0.060)	-0.124** (0.059)	-0.132** (0.064)	0.017*** (0.003)	0.009 (0.014)
Father income score (ln)				0.110** (0.054)		
US state FE	No	Yes	Yes	Yes	No	No
Age controls	No	No	Yes	Yes	Yes	Yes
Childhood municip. FE	No	No	No	Yes	Yes	Yes
Observations	6,618	6,618	6,618	6,618	57,622	6,618
Mean dep. var.	3.73	3.73	3.73	3.73	7.40	7.41

Notes: OLS regressions. Columns 1–4 display results for male emigrants aged below 15 years in 1880 also observed in the US census 1900 with an occupation. Column 5–6 display results using the income scores of their fathers in Sweden for our main sample. Column 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.6: RETURNEES RETURNING HOME?

Dependent variable:	In home muni			In home region		
<i>Panel A: Men</i>	(1)	(2)	(3)	(4)	(5)	(6)
Returnee	0.037*** (0.010)	-0.003 (0.012)	0.004 (0.014)	0.049*** (0.008)	0.013 (0.010)	0.016 (0.011)
Returnee x Eldest brother			-0.024 (0.024)			-0.010 (0.019)
Observations	147,472	147,472	147,472	147,472	147,472	147,472
Mean dep. var.	0.44	0.44	0.44	0.77	0.77	0.77
<i>Panel B: Women</i>	(1)	(2)	(3)	(4)	(5)	(6)
Returnee	0.002 (0.013)	-0.004 (0.016)	-0.003 (0.019)	-0.001 (0.012)	-0.015 (0.013)	-0.012 (0.016)
Returnee x Eldest sister			-0.003 (0.033)			-0.012 (0.027)
Observations	156,278	156,278	156,278	156,278	156,278	156,278
Mean dep. var.	0.38	0.38	0.38	0.75	0.75	0.75
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

Notes: OLS regressions by sex for individuals aged below 15 years in 1880. 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. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

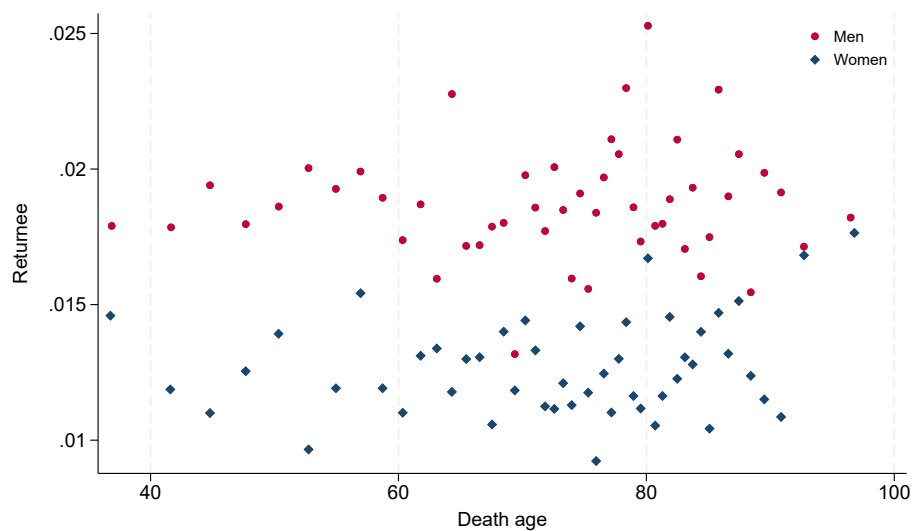


FIGURE B.4:
RETURNEES AND DEATH AGE BY SEX

Notes: Binned scatter plots by sex for individuals aged below 15 years in 1880. This figure displays the correlation between an indicator for if the individual is a returnee and the death age of the individual as observed in the death index. Controls include childhood municipality, age and birth-order fixed effects, and indicators for being eldest brother and eldest sister, respectively.

TABLE B.7: EXTENSIVE MARGIN FOR LABOR INCOME AND WEALTH IN 1910 AND 1930

Dependent variable:	Employed 1910 (=1)	Employed 1930 (=1)	Any wealth 1930 (=1)			
<i>Panel A: Men</i>	(1)	(2)	(3)	(4)	(5)	(6)
Returnee	0.004 (0.006)	0.008 (0.007)	0.016 (0.017)	0.004 (0.022)	0.097*** (0.021)	0.048* (0.025)
Observations	147,472	147,472	22,354	22,354	22,354	22,354
Mean dep. var.	0.92	0.92	0.84	0.84	0.62	0.62
<i>Panel B: Women</i>	(1)	(2)	(3)	(4)	(5)	(6)
Returnee	-0.028*** (0.010)	-0.003 (0.014)	-0.013 (0.024)	0.000 (0.031)	0.078** (0.032)	0.023 (0.042)
Observations	156,278	156,278	23,546	23,546	23,546	23,546
Mean dep. var.	0.20	0.20	0.19	0.19	0.48	0.48
<i>Panel C: Women's spouses</i>	(1)	(2)	(3)	(4)	(5)	(6)
Returnee	0.017 (0.013)	-0.018 (0.017)	-0.001 (0.046)	-0.031 (0.054)	0.116** (0.046)	0.090 (0.055)
Observations	156,278	156,278	9,201	9,201	9,201	9,201
Mean dep. var.	0.64	0.64	0.83	0.83	0.63	0.63
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

Notes: OLS regressions by sex for individuals aged below 15 years in 1880. 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.8: INTERGENERATIONAL OCCUPATIONAL MOBILITY IN FULL CHILDHOOD SAMPLE

Dependent variable:	Mobile		Downward		Upward	
<i>Panel A: Men</i>	(1)	(2)	(3)	(4)	(5)	(6)
Returnee	-0.054*** (0.011)	-0.015 (0.013)	-0.040*** (0.010)	-0.002 (0.013)	-0.013 (0.010)	-0.012 (0.013)
Observations	119,210	119,210	119,210	119,210	119,210	119,210
Mean dep. var.	0.69	0.69	0.29	0.29	0.40	0.40
<i>Panel B: Women's spouses</i>	(1)	(2)	(3)	(4)	(5)	(6)
Returnee	0.017 (0.017)	0.033 (0.023)	-0.045*** (0.015)	-0.012 (0.020)	0.062*** (0.018)	0.044* (0.023)
Observations	69,967	69,967	69,967	69,967	69,967	69,967
Mean dep. var.	0.69	0.69	0.26	0.26	0.43	0.43
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

Notes: OLS regressions by sex for individuals aged below 15 years in 1880. 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:

$$Emigrate_{i\tau} = \sum_{j=1}^{12} \beta_j HC_{j,it} + \phi_t + v_{it}, \quad (4)$$

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 $t, \dots, t + 10$, and zero otherwise.³⁹ $HC_{j,it}$ are a set of indicator variables capturing how emigrants were selected in terms of their occupational status in t , measured by $j = 1, \dots, 12$ 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 .

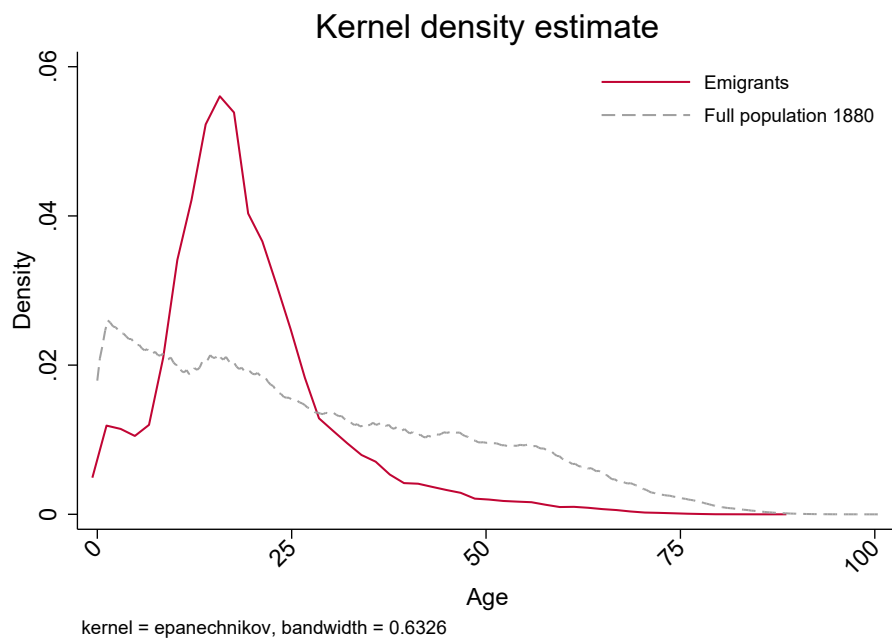
Similarly, we study return migrant selection by estimating the corresponding model:

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

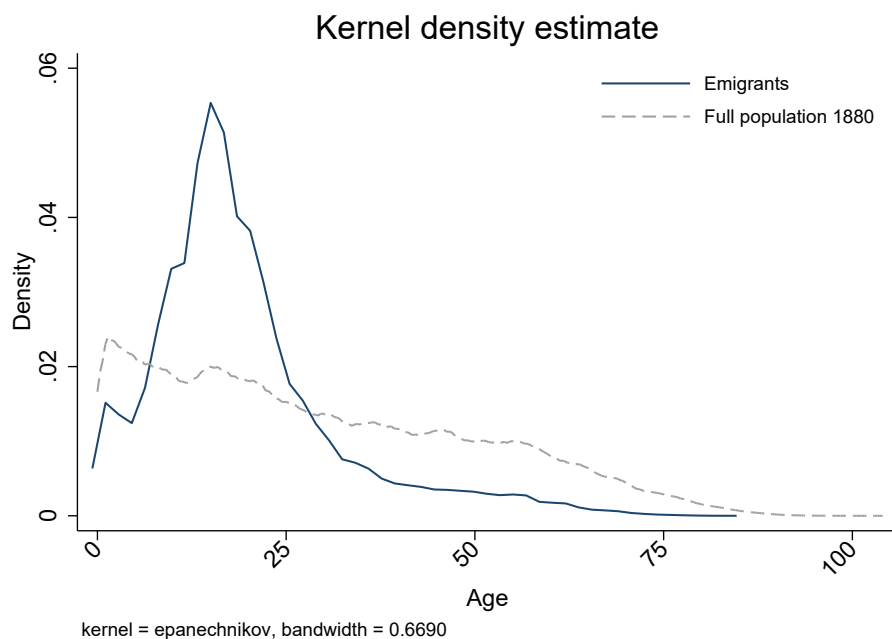
where $Return_{i\tau}$ is a dummy variable that takes the value 1 if a return migrant i emigrated in τ , with $t < \tau < t + 10$, i.e., between two censuses $t, \dots, t + 10$, and 0 otherwise. Thus, we compare return migrants with never-returning emigrants in the census year prior to emigration.

We start by documenting that emigrants were younger than the rest of the population (see Figure B.5) before we proceed to estimating the above equations (4 and 5) with and without birth-year and birth-parish fixed effects.

³⁹Repeat migrants are only treated as emigrants at his or her first-time emigration event.



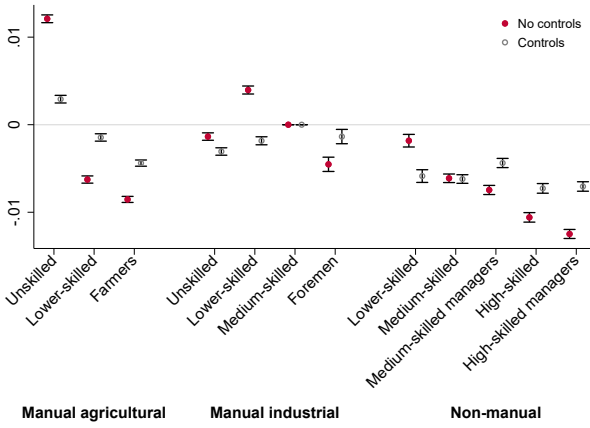
(A) MEN



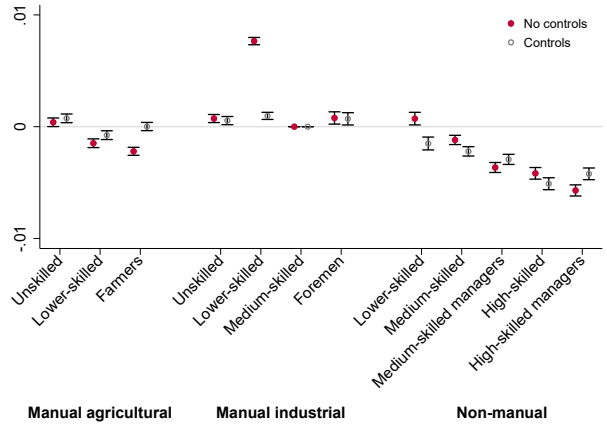
(B) WOMEN

FIGURE B.5: AGE DENSITY OF EMIGRANTS BY SEX

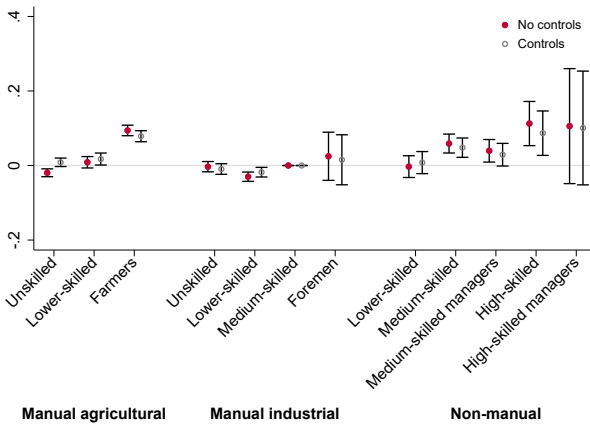
Notes: This figure displays kernel density plots (using the Epanechnikov kernel) for the age of emigrants in the full population of emigrants (by sex) alongside the age distribution in the 1880 census.



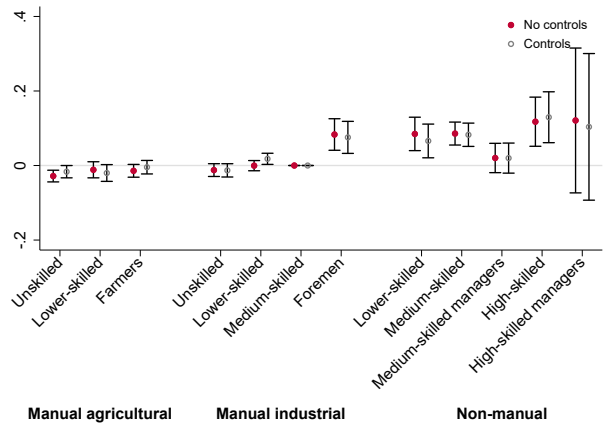
(A) EMIGRANT MEN



(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.

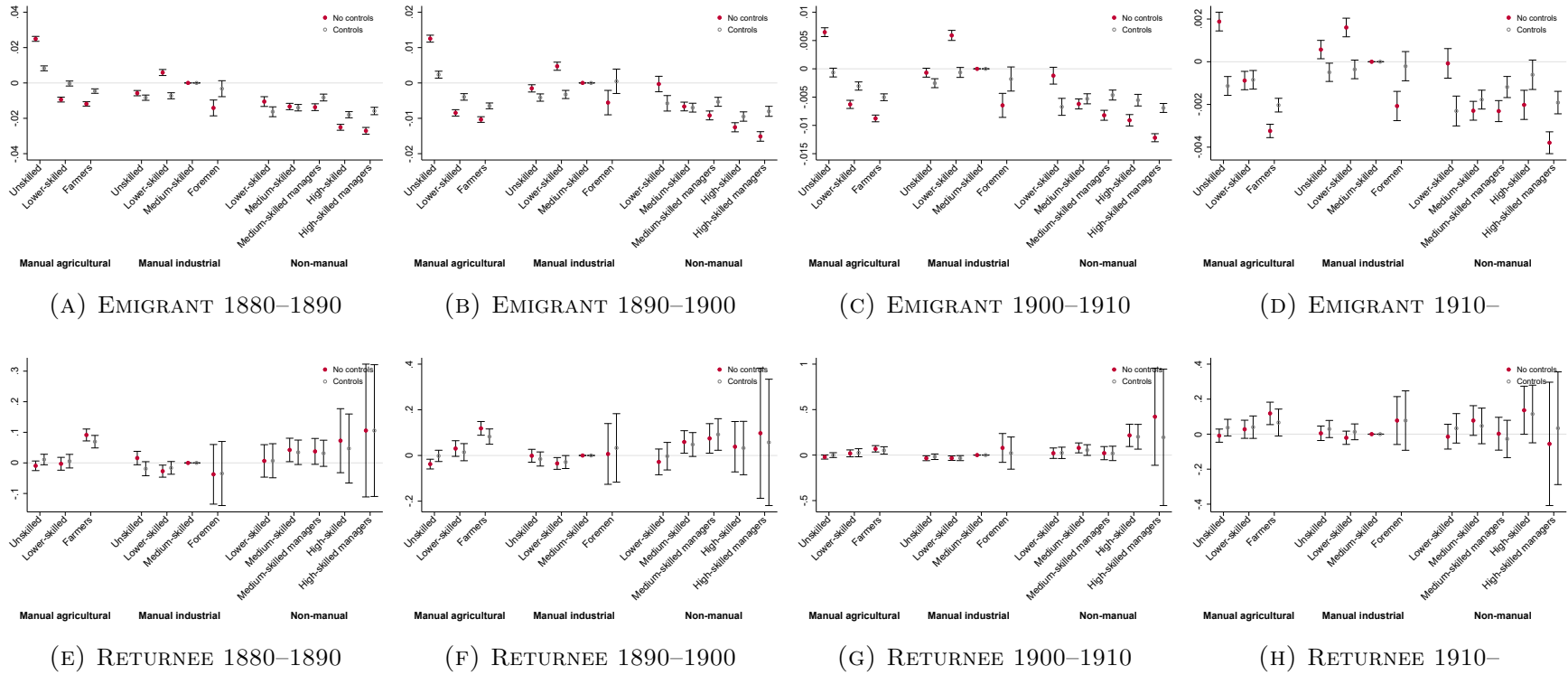


FIGURE B.7:
DECADEL MALE MIGRANT SELECTION BY OCCUPATIONAL GROUP

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.

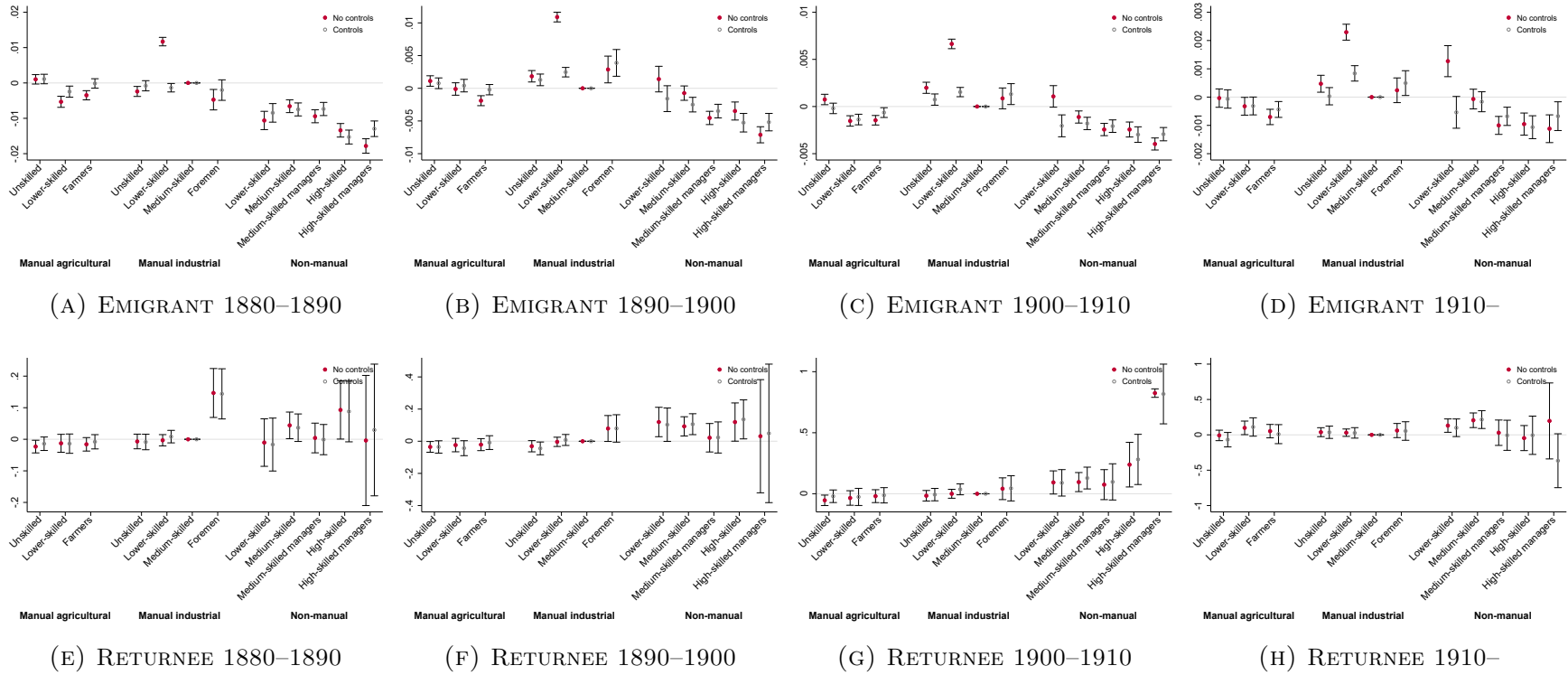
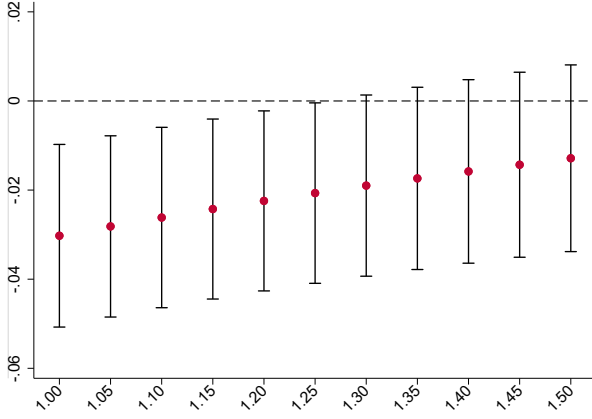


FIGURE B.8:
DECADEL FEMALE MIGRANT SELECTION BY OCCUPATIONAL GROUP

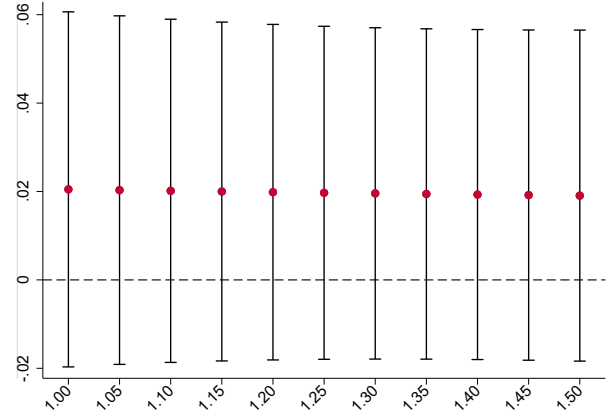
Notes: OLS regressions for full female working-age sample by census year. These figures display the relationships between migrant status and a set of indicator variables capturing occupational class for men and women separately. Figure A–D display the probability of emigrating, while figure E–H display 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.

Online Appendix C: Robustness

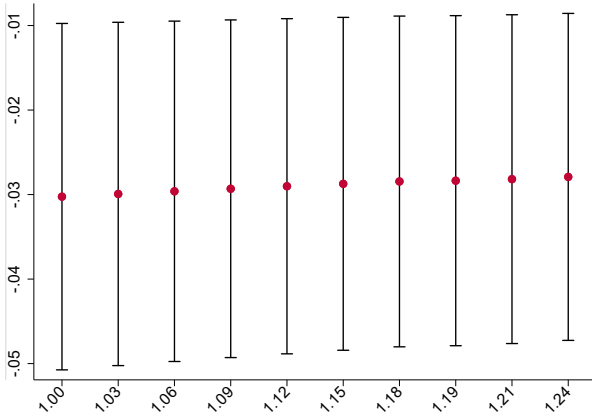
C.1 Figures



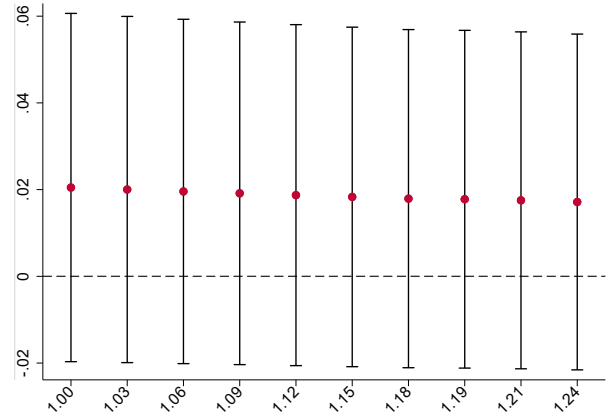
(A) FARMERS (MEN)



(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 by sex for individuals aged below 15 years in 1880. This figure displays separate coefficients for the regression in Table 3, column 5, 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.

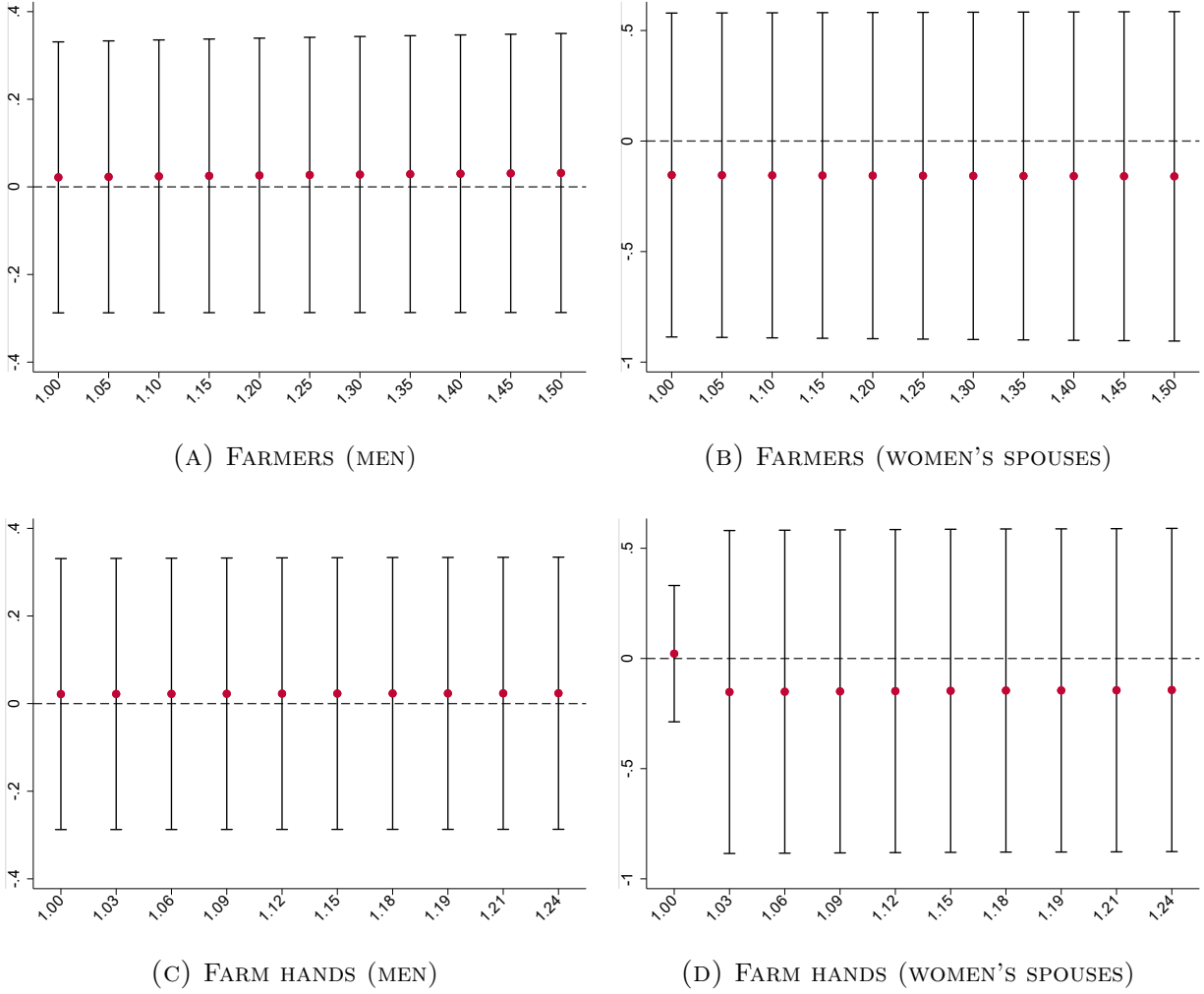
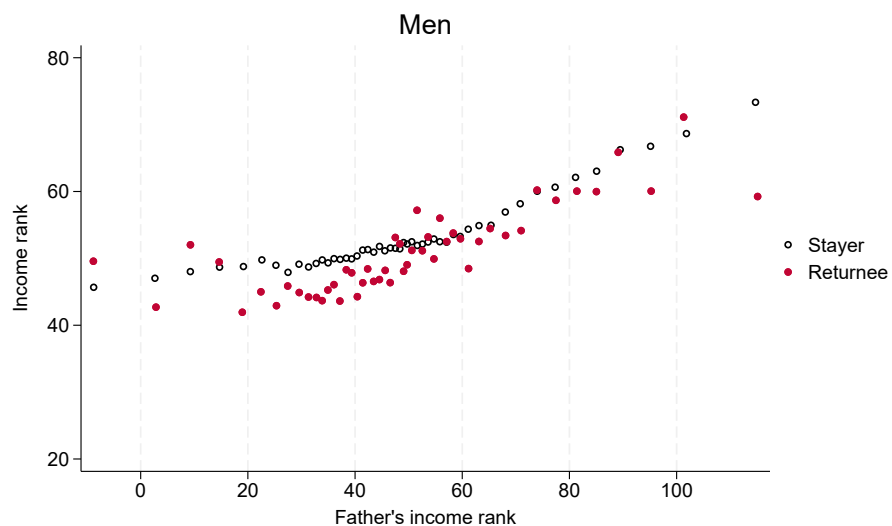
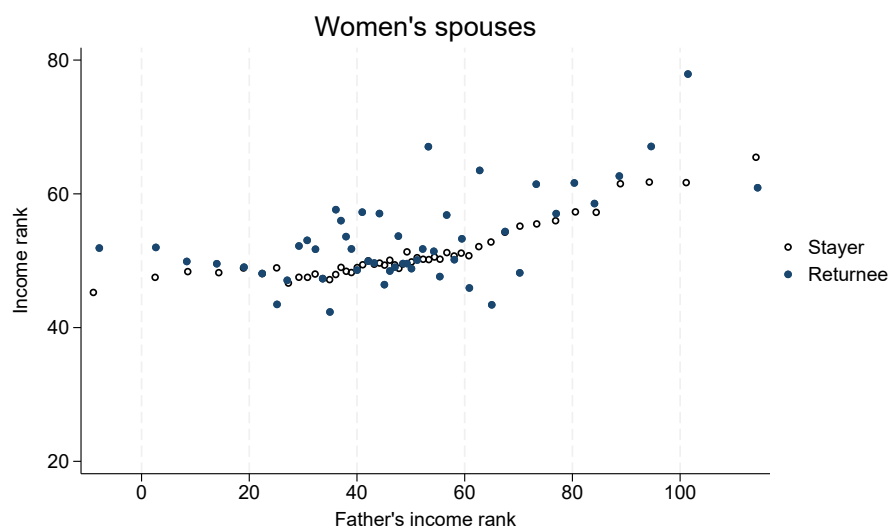


FIGURE C.2:
ROBUSTNESS OF AGRICULTURAL INCOME ADJUSTMENT (1930)

Notes: OLS regressions by sex for individuals aged below 15 years in 1880. This figure displays separate coefficients for the regression in Table 3, column 7, 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 by sex for individuals aged below 15 years in 1880. Controls include birth-parish, age and birth-order fixed effects, and indicators for being eldest brother and eldest sister, respectively.

C.2 Tables

TABLE C.1: ROBUSTNESS: ALTERING CUTOFF FOR SHORT AND LONG STAYS ABROAD

Dependent variable:	In home county		In rural location		Farmer		Wealth (ln)	
<i>Panel A: Men</i>	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Returnee <7 Years Abroad	-0.025 (0.017)	-0.025 (0.017)	0.037** (0.017)	0.033* (0.017)	0.029 (0.019)	0.031 (0.019)	0.502 (0.379)	0.580 (0.392)
Returnee ≥ 7 Years Abroad	0.088*** (0.022)	0.089*** (0.024)	0.074*** (0.023)	0.063*** (0.023)	0.116*** (0.034)	0.122*** (0.036)	1.113** (0.563)	1.256** (0.580)
Observations	145,327	145,327	145,327	145,327	128,497	128,497	21,874	21,874
Mean dep. var.	0.77	0.77	0.55	0.55	0.19	0.19	5.48	5.48
<i>Panel B: Women</i>	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Returnee <7 Years Abroad	0.003 (0.023)	0.004 (0.023)	0.022 (0.026)	0.036 (0.026)	0.037 (0.029)	0.045 (0.030)	1.329* (0.763)	1.509** (0.759)
Returnee ≥ 7 Years Abroad	-0.043 (0.032)	-0.040 (0.035)	0.028 (0.033)	0.066* (0.037)	0.028 (0.042)	0.054 (0.048)	0.659 (0.793)	1.269 (0.954)
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	155,112	155,112	155,112	155,112	115,126	115,126	23,266	23,266
Mean dep. var.	0.75	0.75	0.53	0.53	0.18	0.18	4.26	4.26

Notes: OLS regressions by sex for individuals aged below 15 years in 1880. 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.2: ROBUSTNESS: WINSORIZING WEALTH

Dependent variable:	Wealth 1930 (ln)					
	None		5th percentile		10th percentile	
<i>Panel A: Men</i>	(1)	(2)	(3)	(4)	(5)	(6)
Returnee	0.916*** (0.190)	0.407* (0.226)	0.930*** (0.187)	0.420* (0.225)	0.924*** (0.184)	0.417* (0.222)
Observations	22,354	22,354	22,354	22,354	22,354	22,354
Mean dep. var.	5.51	5.51	5.46	5.46	5.38	5.38
<i>Panel B: Women's spouses</i>	(1)	(2)	(3)	(4)	(5)	(6)
Returnee	1.221*** (0.421)	1.003** (0.492)	1.209*** (0.413)	0.982** (0.487)	1.172*** (0.404)	0.956** (0.479)
Observations	9,201	9,201	9,201	9,201	9,201	9,201
Mean dep. var.	5.57	5.57	5.52	5.52	5.44	5.44
<i>Panel C: Women</i>	(1)	(2)	(3)	(4)	(5)	(6)
Returnee	0.758*** (0.292)	0.331 (0.374)	0.769*** (0.289)	0.335 (0.371)	0.778*** (0.286)	0.340 (0.367)
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	23,546	23,546	23,546	23,546	23,546	23,546
Mean dep. var.	4.27	4.27	4.24	4.24	4.20	4.20

Notes: OLS regressions by sex for individuals aged below 15 years in 1880. 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.3: ROBUSTNESS: WEIGHTING REGRESSIONS WITH THE INVERSE PROBABILITY OF BEING LINKED ACROSS CENSUSES

Dependent variable:	Income score (ln)		Income rank		Income 1930 (ln)		Wealth 1930 (ln)	
<i>Panel A: Men</i>	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Returnee	-0.009 (0.008)	-0.018* (0.010)	0.948 (0.597)	0.043 (0.787)	0.154 (0.128)	0.037 (0.163)	0.993*** (0.191)	0.452** (0.226)
Father's income rank			0.109*** (0.005)					
Returnee x Father's income rank			0.008 (0.024)	-0.019 (0.031)				
Observations	147,454	147,454	144,638	144,638	22,354	22,354	22,354	22,354
Mean dep. var.	7.44	7.44	63.08	63.08	6.11	6.11	5.41	5.41
<i>Panel B: Women's spouses</i>	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Returnee	0.022 (0.015)	0.017 (0.019)	2.376** (1.031)	1.570 (1.418)	0.050 (0.335)	-0.133 (0.379)	1.197*** (0.429)	1.023** (0.494)
Father's income rank			0.075*** (0.006)					
Returnee x Father's income rank			-0.063 (0.044)	-0.126** (0.063)				
Observations	79,500	79,500	78,162	78,162	9,201	9,201	9,201	9,201
Mean dep. var.	7.47	7.47	58.14	58.14	6.07	6.07	5.49	5.49
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Childhood hh controls	Yes	No	Yes	No	Yes	No	Yes	No
Sibling FE	No	Yes	No	Yes	No	Yes	No	Yes

Notes: OLS regressions by sex for individuals aged below 15 years in 1880 with inverse probability weights for if 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.4: ROBUSTNESS: WEIGHTING REGRESSIONS WITH THE INVERSE PROBABILITY OF BEING LINKED ACROSS CENSUSES

Dependent variable:	Married (=1)					
	All		Dropping Returnees after 1900		Dropping +married 1900	
<i>Panel A: Men</i>	(1)	(2)	(3)	(4)	(5)	(6)
Returnee	0.015 (0.010)	0.017 (0.013)	0.059*** (0.012)	0.063*** (0.017)	0.115*** (0.023)	0.099*** (0.032)
Observations	147,472	147,472	145,479	145,479	71,598	71,598
Mean dep. var.	0.68	0.68	0.68	0.68	0.52	0.52
<i>Panel B: Women</i>	(1)	(2)	(3)	(4)	(5)	(6)
Returnee	0.039*** (0.013)	0.007 (0.017)	0.105*** (0.016)	0.052** (0.022)	0.133*** (0.037)	0.104** (0.049)
Observations	156,278	156,278	155,224	155,224	60,190	60,190
Mean dep. var.	0.66	0.66	0.66	0.66	0.43	0.43
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

Notes: OLS regressions by sex for individuals aged below 15 years in 1880 with inverse probability weights for if 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: WEIGHTING REGRESSIONS WITH THE INVERSE PROBABILITY OF BEING LINKED ACROSS CENSUSES

Dependent variable:	Mobile		Downward		Upward				
					Any	Farmer	Skilled	Non-manual	
<i>Panel A: Men</i>	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Returnee	0.008 (0.017)	-0.017 (0.020)	-0.013 (0.014)	-0.016 (0.020)	0.020 (0.019)	-0.001 (0.024)	0.069*** (0.019)	-0.021 (0.020)	-0.014 (0.015)
Observations	54,132	54,132	54,132	54,132	54,132	54,132	54,132	54,132	54,132
Mean dep. var.	0.79	0.79	0.16	0.16	0.63	0.63	0.09	0.18	0.09
<i>Panel B: Women's spouses</i>	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Returnee	0.068*** (0.023)	0.068** (0.031)	-0.049** (0.021)	-0.026 (0.029)	0.117*** (0.029)	0.095** (0.038)	0.050 (0.035)	0.038 (0.042)	0.071** (0.035)
Observations	32,600	32,600	32,600	32,600	32,600	32,600	32,600	32,600	32,600
Mean dep. var.	0.82	0.82	0.15	0.15	0.68	0.68	0.13	0.19	0.09
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

Notes: OLS regressions by sex for individuals aged below 15 years in 1880 with inverse probability weights for if 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.6: ROBUSTNESS: WEIGHTING REGRESSIONS WITH THE INVERSE PROBABILITY OF BEING LINKED ACROSS CENSUSES

Dependent variable:	Landowning farmer			Subsistence	Worker	In home muni	Wealth (ln)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Returnee	0.101*** (0.009)	0.052*** (0.012)	0.056*** (0.014)	-0.003 (0.004)	0.002 (0.013)	0.013 (0.015)	0.564** (0.257)
Returnee x Eldest brother			-0.014 (0.022)	0.008 (0.008)	0.007 (0.021)	-0.020 (0.026)	-0.352 (0.439)
Eldest brother	0.015*** (0.003)	0.014*** (0.004)	0.014*** (0.004)	0.003** (0.002)	-0.001 (0.005)	0.016*** (0.005)	0.062 (0.125)
Observations	130,363	130,363	130,363	130,363	130,363	130,363	22,354
Mean dep. var.	0.18	0.18	0.18	0.02	0.20	0.41	5.41
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

Notes: OLS regressions by sex for individuals aged below 15 years in 1880 with inverse probability weights for if 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.7: ROBUSTNESS: WEIGHTING REGRESSIONS WITH THE INVERSE PROBABILITY OF BEING LINKED ACROSS CENSUSES

Dependent variable:	Landowning farmer		In rural location		Wealth (ln)	
<i>Panel A: Men</i>	(1)	(2)	(3)	(4)	(5)	(6)
Returnee <5 Years Abroad	0.032 (0.020)	0.034 (0.020)	0.040* (0.022)	0.037* (0.022)	0.399 (0.402)	0.462 (0.413)
Returnee ≥ 5 Years Abroad	0.093*** (0.027)	0.098*** (0.029)	0.064*** (0.022)	0.058** (0.023)	1.098** (0.499)	1.226** (0.526)
Observations	128,497	128,497	128,497	128,497	21,874	21,874
Mean dep. var.	0.18	0.18	0.52	0.52	5.38	5.38
<i>Panel B: Women</i>	(1)	(2)	(3)	(4)	(5)	(6)
Returnee <5 Years Abroad	0.021 (0.031)	0.029 (0.032)	0.020 (0.037)	0.029 (0.037)	1.636* (0.866)	1.793** (0.859)
Returnee ≥ 5 Years Abroad	0.043 (0.033)	0.063* (0.036)	0.006 (0.033)	0.029 (0.037)	0.250 (0.762)	0.715 (0.882)
Observations	115,126	115,126	115,126	115,126	23,266	23,266
Mean dep. var.	0.17	0.17	0.50	0.50	4.18	4.18
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

Notes: OLS regressions by sex for individuals aged below 15 years in 1880 with inverse probability weights for if 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.8: ROBUSTNESS: ALTERNATIVE STANDARD ERROR CALCULATIONS

Dependent variable:	In home county (=1)		Married (=1)		Farmer (=1)		Income score (ln)		Income 1930 (ln)		Wealth 1930 (ln)	
<i>Panel A: Men</i>	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Returnee	0.048	0.013	0.072	0.062	0.110	0.049	-0.010	-0.015	0.137	0.031	0.946	0.407
<i>Robust</i>	(0.005)***	(0.010)	(0.007)***	(0.017)***	(0.006)***	(0.012)***	(0.005)**	(0.010)	(0.066)**	(0.163)	(0.092)***	(0.229)*
<i>Cluster-robust (Father)*</i>	(0.005)***	(0.010)	(0.007)***	(0.017)***	(0.006)***	(0.012)***	(0.005)**	(0.010)	(0.066)**	(0.162)	(0.093)***	(0.226)*
<i>Cluster-robust (Father hisclass)</i>	(0.008)***	(0.008)	(0.011)***	(0.014)***	(0.007)***	(0.007)***	(0.007)	(0.010)	(0.060)**	(0.111)	(0.254)***	(0.173)**
<i>Cluster-robust (Municipality)</i>	(0.005)***	(0.010)	(0.008)***	(0.017)***	(0.007)***	(0.012)***	(0.005)**	(0.010)	(0.068)**	(0.165)	(0.091)***	(0.223)*
Observations	314,331	147,472	311,632	145,479	289,460	130,363	314,312	147,454	78,567	22,354	78,567	22,354
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
<i>Panel B: Women</i>	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Returnee	-0.013	-0.015	0.099	0.048	0.048	0.022	0.027	0.017	0.271	-0.149	0.951	1.003
<i>Robust</i>	(0.008)*	(0.014)	(0.010)***	(0.022)**	(0.008)***	(0.016)	(0.008)***	(0.018)	(0.114)**	(0.382)	(0.169)***	(0.502)**
<i>Cluster-robust (Father)*</i>	(0.008)*	(0.013)	(0.010)***	(0.021)**	(0.008)***	(0.017)	(0.008)***	(0.018)	(0.114)**	(0.378)	(0.170)***	(0.492)**
<i>Cluster-robust (Father hisclass)</i>	(0.010)	(0.013)	(0.009)***	(0.015)***	(0.016)***	(0.009)**	(0.009)**	(0.024)	(0.131)*	(0.326)	(0.272)***	(0.324)***
<i>Cluster-robust (Municipality)</i>	(0.008)	(0.013)	(0.010)***	(0.021)**	(0.009)***	(0.016)	(0.009)***	(0.019)	(0.112)**	(0.410)	(0.171)***	(0.500)**
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	328,818	156,278	327,324	155,224	272,554	115,956	218,110	79,500	49,296	9,201	49,296	9,201
Mean dep. var.	0.76	0.75	0.67	0.66	0.19	0.18	7.48	7.47	6.10	6.08	5.39	5.57

Notes: OLS regressions by sex for individuals aged below 15 years in 1880. 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. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

TABLE C.9: RETURNEES AND PERCENTILE RANKED INCOME

Dependent variable:	Income score rank 1910					Income rank 1930	
<i>Panel A: Men</i>	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Returnee	0.533 (0.387)	-0.241 (0.387)	1.217*** (0.369)	0.994* (0.586)	0.229 (0.769)	2.088 (1.421)	0.513 (1.762)
Observations	314,314	314,313	314,312	147,454	147,454	22,354	22,354
Mean dep. var.	63.05	63.05	63.05	63.16	63.16	65.79	65.79
<i>Panel B: Women's spouses</i>	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Returnee	1.464** (0.590)	1.383** (0.590)	2.176*** (0.568)	2.006** (1.010)	1.398 (1.350)	0.749 (3.632)	-0.380 (3.956)
Individual controls	No	Yes	Yes	Yes	Yes	Yes	Yes
Childhood hh controls	No	No	Yes	Yes	No	Yes	No
Sibling FE	No	No	No	No	Yes	No	Yes
Observations	218,111	218,111	218,110	79,500	79,500	9,201	9,201
Mean dep. var.	58.40	58.40	58.40	58.28	58.28	64.93	64.93

Notes: OLS regressions by sex for individuals aged below 15 years in 1880. 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 C.10: RETURNEES AND PERCENTILE RANKED WEALTH 1930

Dependent variable:	Wealth rank 1930			
<i>Panel A: Men</i>	(1)	(2)	(3)	(4)
Returnee	8.509*** (1.739)	3.910* (2.082)	5.853*** (1.602)	2.965 (1.970)
Observations	22,354	22,354	22,354	22,354
Mean dep. var.	51.70	51.70	51.70	51.70
<i>Panel B: Women's spouses</i>	(1)	(2)	(3)	(4)
Returnee	11.179*** (3.843)	9.013** (4.550)	8.628** (3.699)	7.941* (4.574)
Observations	9,201	9,201	9,201	9,201
Mean dep. var.	52.58	52.58	52.58	52.58
<i>Panel C: Women's household</i>	(1)	(2)	(3)	(4)
Returnee	6.165*** (2.381)	2.753 (3.049)	-5.411 (6.290)	-0.319 (8.517)
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	23,546	23,546	3,606	1,450
Mean dep. var.	35.17	35.17	26.49	26.79

Notes: OLS regressions by sex for individuals aged below 15 years in 1880. 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 C.11: DROPPING SHORTER STAYS ABROAD

Dependent variable:	In home county (=1)	Married (=1)	Farmer (=1)	Income score (ln)	Income 1930 (ln)	Wealth 1930 (ln)
<i>Panel A: Men</i>	(1)	(2)	(3)	(4)	(5)	(6)
Returnee	0.015 (0.010)	0.070*** (0.018)	0.049*** (0.012)	-0.015 (0.010)	0.045 (0.165)	0.448* (0.236)
Observations	147,128	145,282	130,068	147,110	22,298	22,298
Mean dep. var.	0.77	0.68	0.19	7.44	6.13	5.51
<i>Panel B: Women</i>	(1)	(2)	(3)	(4)	(5)	(6)
Returnee	-0.015 (0.014)	0.046** (0.022)	0.022 (0.017)	0.000 (0.012)	-0.057 (0.213)	0.055 (0.274)
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes
Sibling FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	156,161	155,154	115,873	155,521	23,538	23,538
Mean dep. var.	0.75	0.66	0.18	7.33	1.31	1.28

Notes: OLS regressions by sex for individuals aged below 15 years in 1880, excluding emigrants staying less than three years abroad before returning. 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. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.