CESifo WORKING PAPERS

10334
2023

March 2023

Decomposing Migrant Self-Selection: Education, Occupation, and Unobserved Abilities

Ilpo Kauppinen, Panu Poutvaara

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 10334*

Decomposing Migrant Self-Selection: Education, Occupation, and Unobserved Abilities

Abstract

We analyze self-selection and sorting of emigrants from Finland, using full-population administrative data from Statistics Finland. We analyze emigration events lasting at least five years and decompose migrant self-selection into education, occupation, and unobserved abilities. Our analysis focuses on Finnish citizens satisfying three criteria: they were between 25-54 years of age; they had no immigrant background; and they were employed. We find that emigrants from Finland are strongly positively self-selected in terms of education and earnings. We also find strong evidence of sorting: men who emigrate outside Nordic countries are considerably better educated and have higher earnings and residual earning than men who emigrate to Nordic countries. Most of the self-selection in terms of higher earnings can be explained by emigrants being more educated. Adding occupational controls increases the fraction of explained self-selection only marginally. While men are positively self-selected also with respect to residual earnings, women are not.

JEL-Codes: F220, I260, J310.

Keywords: international migration, self-selection, Roy model, education, residual earnings.

Ilpo Kauppinen
Vatt Institute for Economic Research
PL 1279
Finland – 00101 Helsinki
ilpo.kauppinen@vatt.fi

Panu Poutvaara
ifo Institute – Leibniz Institute for Economic Research at the University of Munich
Poschingerstr. 5
Germany – 81679 Munich
poutvaara@ifo.de

We are grateful to the participants at the IZA Annual Migration Meeting 2022 for valuable comments, and to Padmaja Kadambi for excellent research assistance.

1 Introduction

Migrant self-selection has important economic consequences for both the origin and the destination countries. Skill distribution of recent migrants affect not only the wages of citizens and earlier migrants in the destination countries, but also the wages of the people who stayed in the origin countries (Borjas 2003). Workers who are substitutes for migrants in the origin country and complements in the destination country tend to gain, whereas workers who are substitutes for migrants in the destination country and complements in the origin country tend to lose. Furthermore, both immigration of net recipients from income redistribution and emigration of net payers may pose a challenge to public finances (Wildasin 1991; Sinn 1997). Pay-as-you-go public pensions aggravate the challenge that emigration of young workers poses to public finances and may turn immigration of even low-skilled workers a fiscal gain in the net present value terms if they are young (Razin and Sadka 2000).

In a seminal contribution, Borjas (1987) presented migration decision as an extension of Roys's (1951) model of occupational self-selection. In the Roy-Borjas model, individuals differ in their productive capacities (or skills) and countries differ in the returns to skills. The key parameters in the Roy-Borjas model are to what extent skills are transferable between countries, and what are the returns to skills in the origin country and in potential destination countries. After comparing income opportunities in the origin country with income opportunities (net of migration costs) in potential destination countries, potential migrants would like to migrate if this increases their expected utility.

If skills are largely transferable and returns to skills in the potential destination country are higher (lower) than in the origin country, then migrants tend to come predominantly from the upper (lower) parts of the skill distribution. This implies that a country like the United States would attract predominantly high-skilled immigrants from Western and Northern European countries that have more equal income distribution, while migrants from countries with more unequal income distribution than the destination country can be expected to be negatively self-selected. In line with the prediction, Borjas (1987) finds that immigrants from more egalitarian Western European countries perform generally well in the United States.

In this paper, we test the predictions of the Roy-Borjas model using full-population administrative data from Finland. We analyze self-selection and sorting of emigrants in terms of education, occupation, and unobserved abilities. We decompose overall migrant self-selection separately for men and women into components that can be explained by education, other demographic characteristics like age and family situation, occupation, and unobservable abilities measured by residual wages from Mincerian wage regressions. We analyze emigration into other Nordic countries and to the rest of the world separately. Other Nordic countries are culturally and historically quite similar to Finland. Further, both have relatively small income differences in international comparison, as well as high levels of taxation and social protection. According to Borjas (1987) model, this implies that the emigrants from these countries to the rest of the world can be expected to come from the upper parts of skill distribution, assuming that skills from Finland can be expected to be largely applicable in the rest of the world.

We build on an extensive literature that has extended both the theoretical framework and the empirical analyses of Borjas (1987) in many aspects, summarized more widely than the space allows here and further extended in Borjas et al. (2019) and in Aksoy and Poutvaara (2021). Borjas and Bratsberg (1996) find that the selective return migration accentuates self-selection patterns, and we contribute to this question by analyzing how self-selection differs between short-term and long-term emigrants. Grogger and Hanson (2011) present a model in which migration cost has two components: a fixed monetary cost and a component that may increase or decrease with skills, instead of time cost in Borjas (1987) that was independent of skills and depends in monetary terms on net wage rate over the lost working time. Grogger and Hanson (2011) show that their model explains two stylized patterns of global migration flows: positive selection (migrants are generally more educated than non-migrants) and positive sorting (more educated migrants tend to choose destinations offering higher returns to skills). Inspired by this, we present evidence on how migrants from Finland to different destinations are self-selected and sorted in terms of their education. Borjas et al. (2019) show that the Roy model allows stronger predictions than previously considered: the conditions that Borjas (1987) shows to result in positive or negative selection in terms of expected earnings imply a stochastic dominance relationship between the earnings distributions of migrants and non-migrants.

An analysis of full-population administrative data from Denmark supports this: emigrants are better educated than non-migrants, and the earnings distribution of both men and women who emigrate almost stochastically dominates the earnings distribution of non-migrants. We compare the earnings distributions of migrants and stayers to see whether a similar pattern prevails in Finland.

In an important contribution, Chiquiar and Hanson (2005) show that the migrants from more unequal countries may be self-selected from lower, upper or intermediate parts of the skill distribution when migration costs are negatively correlated with skills. Chiquiar and Hanson (2005) combine information from American and Mexican censuses to predict what Mexican migrants in the United States would have earned in Mexico. The counterfactual estimated earnings based on observed characteristics suggest that the Mexican migrants would have been in the middle of Mexico's wage distribution. Subsequent improvements made in accessible datasets have allowed to analyze the migrant self-selection in terms of actual earnings and have directed attention to residual earnings. Fernandez-Huertas Moraga (2011) and Kaestner and Malamud (2014) use survey data that reported the actual pre-migration earnings of subsequent Mexican emigrants and find evidence of negative selection, as predictions from Borjas (1987) would suggest. Gould and Moav (2016) analyze the emigration of men from Israel to the United States and find an inverse U-shape of the probability of emigration in residual earnings. Borjas et al. (2019), instead, show that both male and female emigrants from Denmark are strongly positively self-selected in terms of their residual earnings. We contribute to the analysis of migrant self-selection by bringing new evidence from a different country, and, more importantly, by analyzing mechanisms behind self-selection and sorting in terms of residual earnings. To do the latter, we analyze self-selection in terms of residual earnings twice. First when controlling for education and demographic controls as in Borjas et al. (2019), and then by introducing occupational controls to study which fraction of self-selection are explained by these. This links our paper also to Patt et al. (2021), which is an analysis on the role of occupation for self-selection among Mexican migrants to the United States. They find that migrants have higher manual skills and lower cognitive skills than non-migrants.

Roy-Borjas model has also been tested under historical contexts. Abramitzky et al. (2012) analyze migrant self-selection from Norway to the United States during the age of mass migration (1850–1913). At that time, the United States had largely open borders to European migrants, allowing to analyze migrant self-selection in the absence of selective immigration policy. Furthermore, at that time Norway was poorer and more unequal than the United States. Abramitzky et al. (2012) find that migrants from urban areas in Norway were negatively self-selected, in line with the Borjas (1987) prediction. Wegge (2002) analyzes occupational self-selection of emigrants from nineteenth-century Hesse-Cassel and concludes that emigrants were positively self-selected in terms of skills, but negatively self-selected in terms of financial wealth. Spitzer and Zimran (2018) analyze how migrants from Italy to the United States between 1907 and 1925 were self-selected in terms of their height. They find evidence of negative self-selection at the national level, but positive self-selection at the local level.

We restrict our analysis to Finnish citizens with no immigrant background as return decisions of previous migrants are qualitatively different. When analyzing self-selection with respect to earnings or residual earnings, we restrict the analysis to prime working-age individuals who were at least 25 and at most 54 years old (as in Borjas et al. (2019)), and who were registered as employed for 12 months during the year. We also required their main type of activity to have been being employed and dropped individuals who were classified as entrepreneurs. This is because taxable income of entrepreneurs, owing to its tendency to fluctuate, makes it a noisier measure of earnings potential than the same for employees.

2 Data

Our main data source is the FOLK-modules, a high-quality registry-based full-population research data provided and maintained by Statistics Finland. The modules include rich information on socioeconomic characteristics of the population residing permanently in Finland. The data also has information on emigration and eventual return migration events

including information on the date of migration and the destination country.[1] We restrict the analysis to individuals who are of Finnish origin and are born in Finland.[2] For this study, an emigrant is an individual in (any of the cross-sections of) our dataset who emigrates from Finland in the given calendar year and stays abroad for at least 365 days. The rest of the observations are defined as non-migrants. As we are working in a full population panel setting, most individuals are included in the data multiple times.

The next section describes the migration flows from Finland in the period 1988-2019. When analyzing self-selection of emigrants from Finland (in the latter sections) we impose further restrictions on the sample. To focus on migration that is permanent in nature, we drop observations with emigration spells that are shorter than five years. We further aim to limit the self-selection analysis to individuals who worked full time and have likely completed their studies, as the observed annual income of part-time workers may not be indicative of their true earnings potential. Therefore, we restrict the analysis to prime working-age individuals who are at least 25 and at most 54 years old and who were registered as employed for 12 months during the year they were included in the data. As the variable for the number of months an individual has been employed during the year is only available from 1997 onwards, we restrict our analysis of self-selection of emigrants to emigration events taking place in 1998 or later. We also require that the included individuals have been employed and drop individuals whose socioeconomic group was either entrepreneurs or farmers and forestry entrepreneurs, as well as individuals with a missing socioeconomic group. We analyze emigration events taking place no later than 2015 and use data of immigration events until the end of year 2020 to identify return migration.

[1] It is possible to migrate without registering, but we expect the share of people migrating without registration to be small, as those who do not register emigration are obliged to pay taxes additionally on their income earned abroad. Consequently, especially those earning high income have a strong incentive to register their emigration. Return migration should also be registered, as one needs to be registered in Finland to be covered by the Finnish social insurance and public health care. Also, information on residence is needed when receiving earned income or capital income in Finland.

[2] Statistics Finland defines a person to have a "Finnish background" if at least one of the parents was born in Finland. In practice, those who are of Finnish background and are born in the country also hold a citizenship.

Taken together, these restrictions mean that in most of our analyses, we only include data consisting of cross-sections of the population from years 1997 to 2014 to analyze self-selection of emigrants who left from 1998 to 2015.

The income variable is total earned income. We drop individuals whose income is missing or zero. The income variable is top-coded so that incomes in the highest annual percentile are registered as a median within that group. This generates a peak at that income level.

3 Emigration Flows

In this section, we describe emigration flows from Finland among the native Finns during a three-decade time frame, from 1988 to 2019. We also show how emigration to different destinations has changed over time and discuss major changes that may have affected the level of emigration and changes in sorting to different destination countries.

After strong growth and an overheated economy in the late 1980s, Finland experienced an exceptionally severe recession and banking crisis. Finnish gross domestic product decreased by 13 percent from 1990 to 1993, and unemployment rose from 3% in 1990 to 20% in 1994 (Honkapohja and Koskela 1999). After devaluation and later floating of the Finnish currency, Finnish export sector regained its competitive edge, and the economy grew quickly in the remaining 1990s. Economic recovery was further boosted by Finland joining the European Union (EU) in 1995 and by Nokia's success for several years as the leading mobile phone manufacturer. The economy grew rapidly until the global financial crisis, and at a substantially lower speed thereafter. Like most developed countries, aging population put public finances under increasing stress. Additionally, in the later years Nokia lost its dominance in mobile phone market which contributed to the slow economic growth in 2010s.

Figures 1.1-1.3 show emigration from Finland from 1988 to 2019. The analyses of emigrants and underlying population are restricted to Finnish citizens with no immigrant background; without such restrictions, emigration would appear to have increased considerably more as a large share of persons emigrating from Finland in each year are foreign citizens who have lived in Finland temporarily. Figure 1.1 depicts the numbers of emigrants by year separately for all ages and for prime working-age individuals who are

at least 25 and at most 54 years old. Figure 1.2 is an extension of figure 1.1 where the rate of yearly emigration is expressed in percentages. Last, figure 1.3 shows the number of emigration events among the working age individuals separately to other Nordic countries, and to the rest of the world.

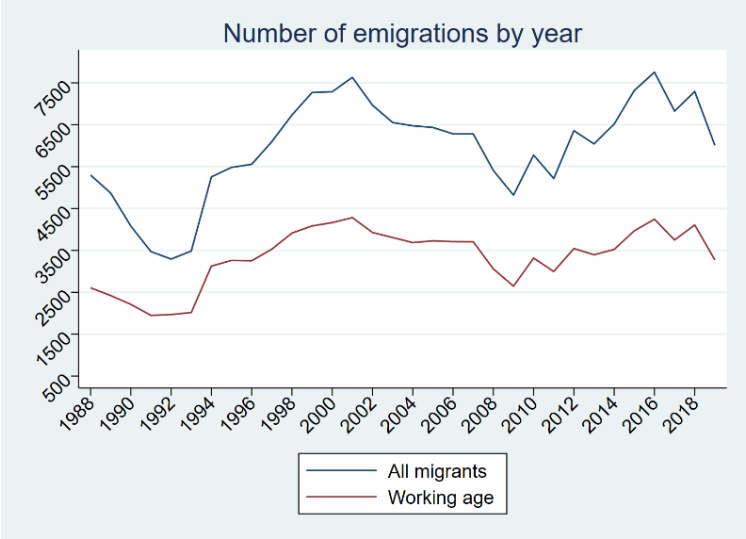

Figure 1.1. Emigration Flows

Notes: This figure shows for each year the number of Finnish citizens with at least one Finnish born parent, and who emigrated for at least 365 days, for two groups: without age restrictions and for those aged 25 to 54 in the year before emigration.

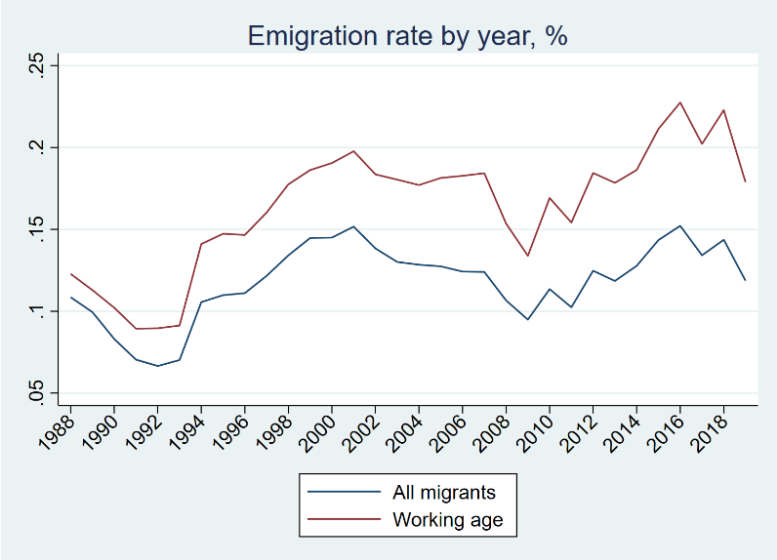

Figure 1.2. Emigration Rates

Notes: This figure shows for each year the percentage of Finnish citizens with at least one Finnish born parent, and who emigrated for at least 365 days, for two groups: without age restriction, and for working-age migrants including only those aged 25 to 54 in the year before emigration.

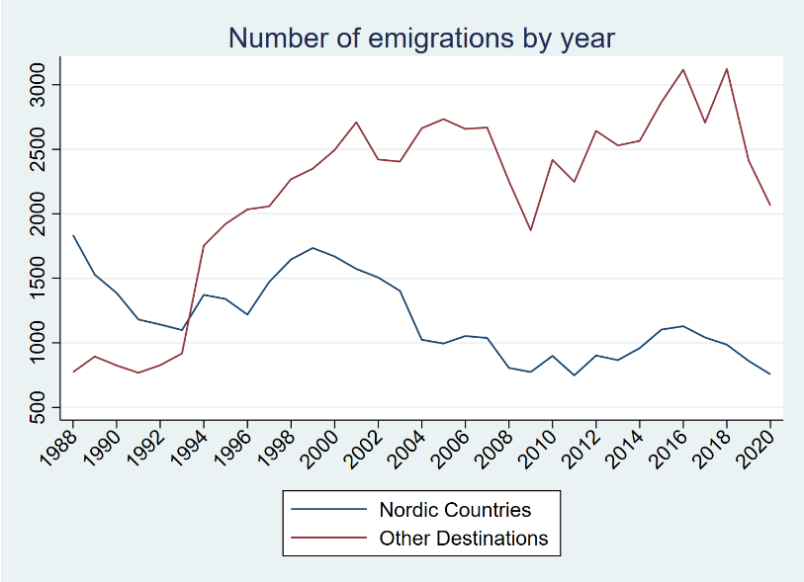

Figure 1.3. Working-Age Emigrant Flows by Destination

Notes: This figure shows for each year the number of Finnish citizens with at least one Finnish born parent and who emigrated for at least 365 days to Nordic countries and to other destinations.

Finnish emigrants' traditional main destination country is Sweden. As the economic crisis of early 1990s also affected the Swedish economy, Finland consequently experienced a drop in its emigration even though high unemployment could have been expected to be a powerful push factor to increase emigration. At the beginning of 1994, the European Economic Area that brought together the European Union and the European Free Trade Area (EFTA) came into force, allowing free labor mobility between Finland and the EU member states (even before Finland joined the EU at the beginning of 1995). Subsequent economic integration saw the number of native Finns emigrating annually increase from 3294 in 1992 to 7632 in 2001. Eventually, the number of people who emigrated annually decreased until the global financial crisis in 2008, after which it increased in the subsequent years. Nevertheless, it is noteworthy that in the last few years of our period of interest, the number of emigrants was at a similar level as in the late 1990s and early 2000s. During the period of our analysis, there is no trend of growing emigration of the native population from Finland.

Figure 1.3 shows that Nordic countries received most emigrants until 1993. In 1994 emigration to non-Nordic countries almost doubled from the previous year and grew rapidly until the end of the century. Emigration events to the 12 non-Nordic member states of the

European Union and Switzerland (which are part of the Common European Labor Market) increased from 562 in 1993 to 1166 in 1994.

Figure 2.1 depicts the numbers of emigrants by level of educational attainment. From 1993 onwards most of those who emigrated have higher education. This change is the joint effect of overall increase over time in the general level of education and change in self-selection of emigrants by education. It is worth noting that numbers of migrants with different levels of education give a misleading picture of the relationship between propensity to emigrate and education, as relatively small share of the population hold highest educational degrees. Nevertheless, it is useful for understanding the skill composition of emigrants.

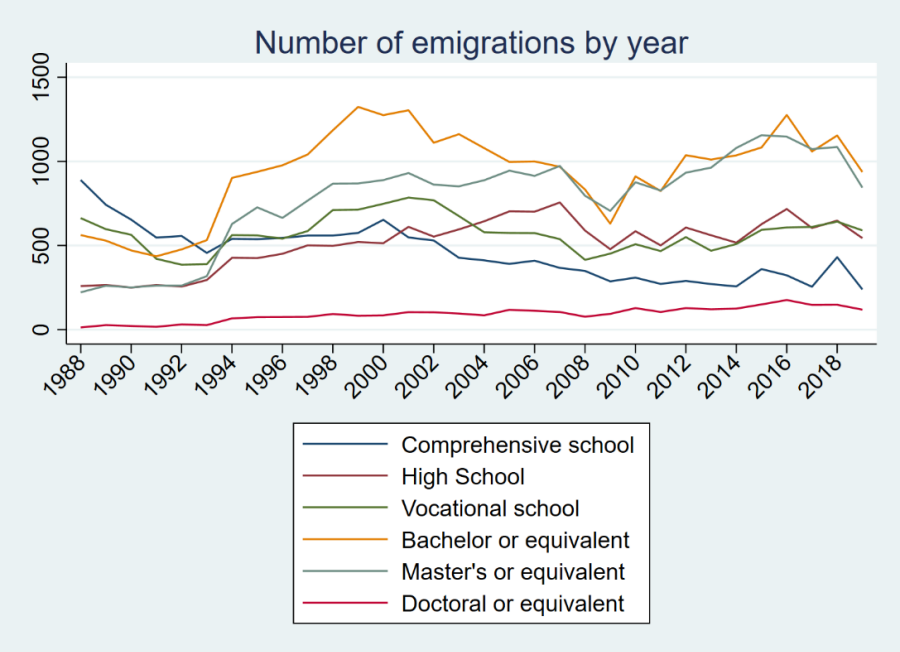

Figure 2.1. Emigration Flows by Educational Attainment

Notes: This figure shows for each year by educational attainment, the number of Finnish citizens with at least one Finnish born parent, and who emigrated for at least 365 days.

Figure 2.2 depicts emigration rates by educational attainment. Emigration rate is highest for those with doctoral degree, followed by those with master's degree or only high school degree. High emigration rate among those with doctoral or master's degrees implies that people with highest level of education are the most internationally mobile. The high emigration propensity for those with a high school degree can probably be explained by

those with a high school degree migrating to study abroad. In addition, the Finnish registry data does not include all degrees that are attained abroad, so some individuals who have high school as their highest attained education in the data, may have a higher degree from a foreign educational institution. Appendix figures A1.1-A1.3 present emigration rates separately by different age groups and show that those with doctoral degree are most mobile within those aged 25 to 30, 31 to 40, and 41 to 54.

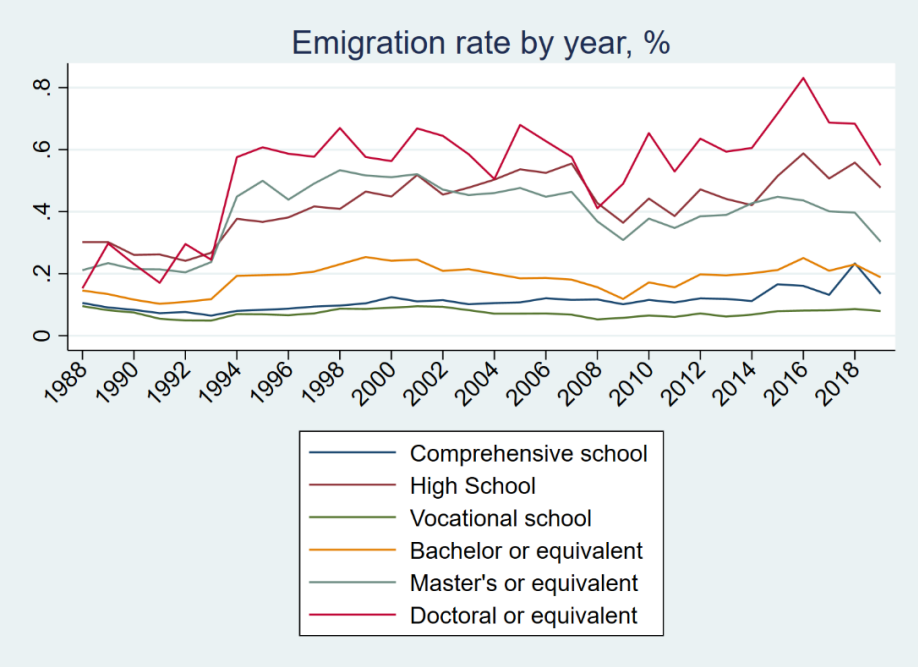

Figure 2.2. Emigration Rates by Educational Attainment

Notes: This figure shows for each year the percentage of Finnish citizens with at least one parent born in Finland, and with given educational attainment who emigrated for at least 365 days.

Figure 3 depicts the emigration rate by age for men and women separately. Young adults are most likely to emigrate, and among them, women are more likely to emigrate than men.[3] A possible reason for the difference could be that women are more likely to study or work as au pair abroad.

[3] In a Western-European comparison, the Finnish emigration rates are neither exceptionally low nor exceptionally high. Using harmonized data from European countries provided by Eurostat, Kalin et al. (2022) calculated emigration rates for in the age group 25 to 54 among the native-born. The calculated rate for Finland is 0.25%. Corresponding rates for the same year were 0.30% for Sweden, 0.31% for Denmark, 0.21% for Germany, 0.39% for the Netherlands, 0.18% for Spain and 0.20% for Italy. The rate that is calculated using the Eurostat data differs from the rate

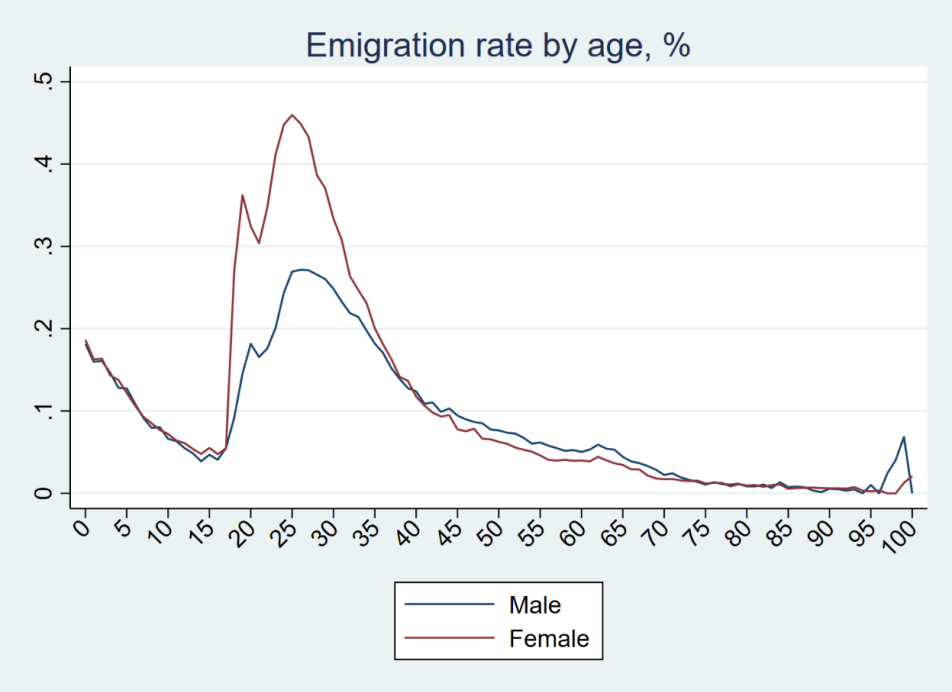

Figure 3. Emigration Rates by Age

Notes: This figure pools observations of Finnish citizens during 1987-2018, with at least one Finnish born parent and shows for each age measured in the end of the year the percentage of those who emigrated for at least 365 days during the following year.

Figures 4.1 and 4.2 describe the distribution of migration spell lengths. Figure 4.1 shows that almost 60% of migrants who reside abroad for a minimum of one year end up staying there for at least five years. Figure 4.2 presents the percentage of emigrants who stay abroad for at least five years according to the duration of the spell. Almost 80% of these long-term migrants are still abroad after ten years. When we restrict our self-selection analyses to migration spells of five years or more, we are essentially studying migration that is permanent in nature.

reported in Figure 1.2. because of the differences in restrictions that are impose on the data as well as how having migrated is defined.

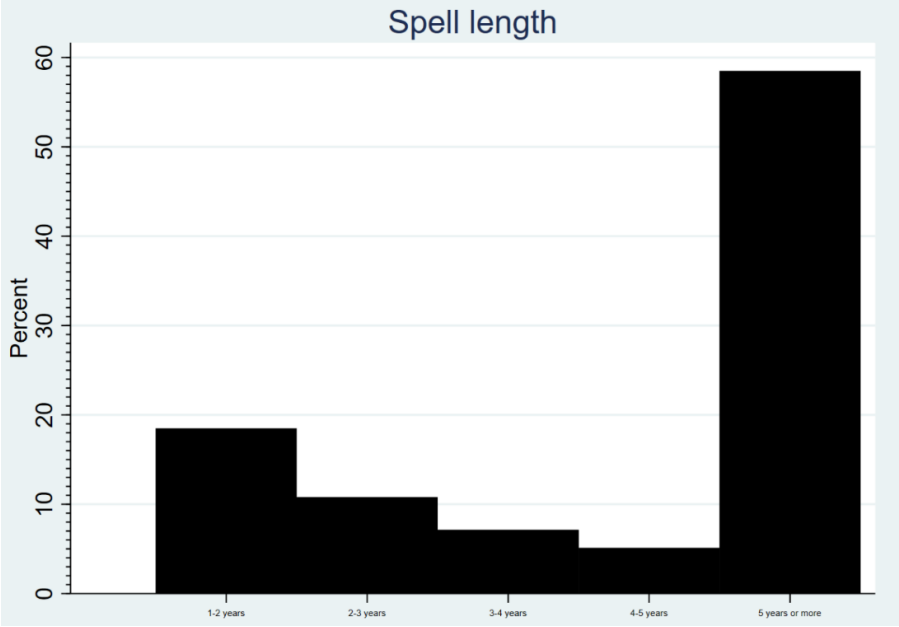

Figure 4.1. Spell Lengths of All Emigration Episodes

Notes: This figure pools observations of Finnish citizens with at least one parent born in Finland who emigrated for at least 365 days in the time span from year 1988 to year 2014 and shows the distribution of spell length.

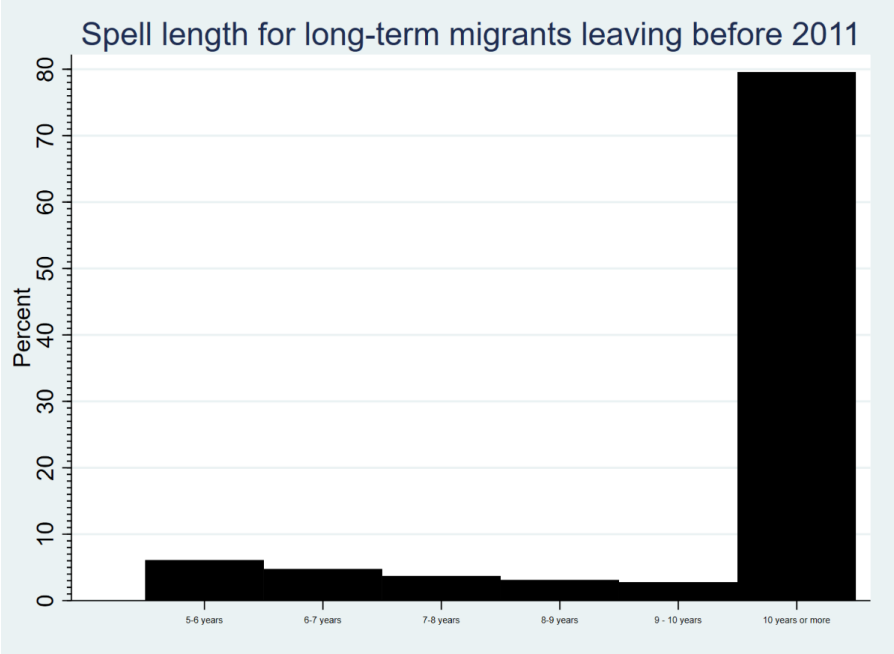

Figure 4.2. Spell Lengths for Spells of 5 Years or More

Notes: This figure pools observations of Finnish citizens with at least one parent born in Finland who emigrated for at least 5 years in the time span from year 1988 to year 2011 and shows the distribution of spell length.

4 Self-selection in Pre-migration Earnings and Observable Characteristics

In this section we present empirical evidence on the self-selection of emigrants from Finland in terms of educational attainment and standardized earnings from the year before emigration. The main empirical finding is that long-term emigrants were in general more productive prior to emigration than individuals who stayed in Finland.

As explained in the data section, we restrict the analysis to the cross section of individuals between 25 and 54 years of age who were employed, defined separately for each year from 1997 to 2014. Table 1 reports summary statistics of the obtained sample. The panel contains person-year observations for 11,132,212 male and 10,934,773 female non-migrants and 5450 male and 5307 female migrants covering years 1997 to 2014. By construction, these migrants are individuals who resided in Finland in the year for which self-selection is analyzed and emigrated in the subsequent year (between 1998 and 2015) and did not return to Finland within five years. These individuals are included as non-migrants in the years prior to the year in which they are included as persons emigrating in the following year, provided that they satisfied the age and employment conditions. As Table 1 shows, the Finnish emigrants, both men and women, are younger than the non-migrants. Despite the age difference, the emigrants' annual earnings were on average higher in the year prior to migration than those of the stayers.

To adjust for age, gender, and year effects, we construct a measure of "standardized earnings" following Borjas et al. (2019). Standardized earnings are defined by the ratio of a worker's annual gross earnings to the mean gross earnings of workers of the same age and gender during the same calendar year. Table 1 shows that those who emigrate earn more than non-migrants in terms of standardized earnings in the year before they emigrated.

Table 1. Summary Statistics

	Non-migrant men	Migrant Men	Non-migrant women	Migrant women
Observations	11,132,212	5450	10,934,773	5307
Age				
Average	40.0	35.4	41.2	34.2
Median	40.0	34.0	42.0	32.0
Annual earnings in 2020 euros				
Average	35311	46477	27193	30873
Median	34719	41061	28114	29481
Standardized annual earnings				
Average	1.0	1.3	1.0	1.2
Median	0.9	1.2	0.9	1.1

Notes: The data includes full time working individuals during 1997-2014 aged 25 to 54 and born in Finland with at least one Finnish born parent. Entrepreneurs and individuals with missing earnings information or non-positive earnings are excluded. Migration spells of five years or more. Standardized earnings are defined by the ratio of a worker's annual gross earnings to the mean gross earnings of workers of the same age and gender during the calendar year.

Appendix Table A1 reports the number of emigrants moving to different destinations. The most popular destination for both men and women is Sweden, followed by the United States, Germany, and the United Kingdom for men and the United Kingdom, the United States, and Germany for women. These four countries account for 53% of the total emigrations in our sample. The Appendix Table A2 shows that the shares of male and female migrants leaving for other Nordic countries and for other EU countries and Switzerland have decreased by one to three percentage points between the early period 1998-2007 and the late period 2008 to 2015, with the share emigrating to the rest of the world increasing correspondingly. Women are distinctly more likely than men to migrate to other Nordic

countries, whereas men are clearly more likely to migrate to the rest of the world than women.

Table 2 presents the highest education level attained by migrants and non-migrants, conditional upon migrants staying abroad for at least 5 years. Migrants tend to have higher educational qualifications than the non-migrants, for both men and women. The share of those with a master's degree or equivalent, or a PhD or equivalent, is 13 percent for men who stay in Finland, whereas the share is 30 percent for men who migrate to other Nordic countries and 48 percent for men who migrate to non-Nordic destinations. For women, the share of migrants to non-Nordic destinations who hold a master's degree or higher is 38 percent, whereas corresponding shares for non-migrants and migrants to other Nordic countries are 14 percent and 31 percent. Correspondingly, shares of those with vocational schooling and comprehensive schooling degrees are significantly lower among migrants than among non-migrants. Having a high school degree as the highest educational qualification is common among migrants to all destinations. This can be partly explained by those with high school degrees migrating to study abroad. It is also worth noticing, that some educational degrees that are attained abroad are not observed in the registry data. Hence, it is possible that some of the migrants who are registered as having only a high school degree may hold higher educational degrees from abroad.

Self-selection of emigrants in terms of education is even stronger in Finland than in Denmark among the same age groups. When analyzing self-selection of emigrants from Denmark from 1996 to 2005, Borjas et al. (2019) find that 8 percent of men and 5 percent of women who stay in Denmark have a master's degree or equivalent, or a PhD or equivalent, while this share is 12 percent for men and 13 percent for women who migrate to other Nordic countries, and 26 percent for men and 18 percent for women who migrate to other destinations.

Table 2. Education Levels of Non-migrants and Migrants

Education	Men			Women		
	Non-migrants	Nordic countries	Other destinations	Non-migrants	Nordic countries	Other destinations
Comprehensive school	14.7	12.4	4.5	11.6	6.8	4.3
High school	5.8	11.6	12.3	4.6	10.1	11.4
Vocational school	41.3	24.1	9.5	33.9	15.1	11.8
Bachelor or equivalent	25.0	22.1	26.1	35.8	37.4	34.6
Master's or equivalent	11.7	25.4	42.0	13.1	28.0	34.4
Doctoral or equivalent	1.4	4.4	5.6	1.0	2.6	3.5

Notes: The data includes full time working individuals during 1997-2014 aged 25 to 54 and born in Finland with at least one Finnish born parent and shows for each depicted group as column percentages the distribution of the highest education level when pooling observations from all years. Entrepreneurs and individuals with missing earnings information or non-positive earnings are excluded. Migration spells of five years or more.

Table 3 presents the Logit estimates for the probability of emigration by gender. The regressions confirm that for both men and women, those with higher education are more likely to emigrate. Furthermore, it shows that those with children are less likely to emigrate. Both these patterns are like what Borjas et al. (2019) find for Denmark. In terms of family situation, an interesting difference emerges. Borjas et al. (2019) find that marriage reduces the probability of emigration for both men and women. We find that in Finland, married women are less likely to emigrate, but married men are more likely to emigrate than single men when also other controls are included. Although having such a gender difference may appear puzzling on the first sight, it can be explained by our employment restriction if couples are more likely to emigrate when the female partner is not in full-time employment, and therefore not included in the target population of our analysis.

Table 3. Logit Estimates of the Probability of Emigration, by Gender

	Men		Women	
	Coefficient	Odds ratio	Coefficient	Odds ratio
	(1)	(2)	(3)	(4)
Married	0.1751***	1.1914***	-0.2350***	0.7905***
	(0.040)	(0.048)	(0.042)	(0.034)
Children	-0.6964***	0.4984***	-0.6912***	0.5010***
	(0.052)	(0.026)	(0.045)	(0.023)
Married*Children	-0.1118*	0.8942*	-0.1086*	0.8970*
	(0.066)	(0.059)	(0.064)	(0.058)
High school	1.0137***	2.7557***	0.8785***	2.4073***
	(0.067)	(0.184)	(0.075)	(0.181)
Vocational school	-0.4994***	0.6069***	0.4772***	0.6205***
	(0.064)	(0.039)	(0.072)	(0.045)
Bachelor or equivalent	0.6276***	1.8730***	0.3508***	1.4203***
	(0.059)	(0.111)	(0.066)	(0.094)
Master's or equivalent	1.7299***	5.6398***	1.1478***	3.1513***
	(0.058)	(0.326)	(0.067)	(0.213)
Doctoral or equivalent	2.2031***	9.0526***	1.8607***	6.4281***
	(0.080)	(0.727)	(0.100)	(0.644)
Age fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
N	11,144,864	11,144,864	10,945,028	10,945,028
Pseudo R-Squared	0.0705	0.0705	0.0741	0.0741

Notes: The table reports logistics regression results using regression coefficients (columns 1 and 3) and odds ratios (columns 2 and 4). The dependent variable is a dummy that gets value one if the individual emigrates the following year and does not return within five years. Individually clustered standard errors are in parentheses. Coefficients for the age and year fixed effects are not shown. The data includes full time working individuals during 1993-2014, aged 25 to 54 and born in Finland with at least one Finnish born parent. Entrepreneurs and individuals with missing earnings information or non-positive earnings are excluded.

* p<0.10, ** p<0.05, *** p<0.010.

Appendix Table A3 shows that if we control for education and log of standardized earnings simultaneously, self-selection with respect to university education remains positive, and is strongest for those with doctoral or master's degree. Interestingly, once education is controlled for, men are positively self-selected with respect to log standardized income, while women are negatively self-selected, even though women who emigrate have higher standardized earnings as shown in Table 1. A potential explanation for this is the co-location problem that couples face in deciding whether to migrate (Mincer 1978). Munk et al. (2022) show that couples that emigrate from Denmark emigrate much more often for male partner's than for female partner's work, which might explain weaker self-selection with respect to women's earnings. Our data does not allow testing this conjecture as a female who is registered as a single in Finland could also emigrate to live with a partner who already lives abroad.

The descriptive comparisons already suggest a positive selection of migrants. Differences in conditional averages could still hide substantial differences between the underlying probability distributions, so we proceed to compare cumulative distributions of standardized earnings between migrants and non-migrants. We also present kernel density estimates of the earnings density functions.

Figure 5 illustrates the cumulative earnings distributions by gender (male and female) for non-migrants, migrants to the Nordic countries, and migrants to non-Nordic countries. The values of the standardized earnings are truncated at -2 and 2 to make the graphs more tractable.[4] Men who emigrate are strongly positively self-selected in terms of their residual earnings. Men who emigrate outside Nordic countries have bigger residual earnings than men who emigrate to other Nordic countries, in line with migrant sorting predicted by Grogger and Hanson (2011). For women, selection and sorting patterns are weaker. Women who emigrate have higher average standardized earnings than women who do

[4] The shares of observations below the lower and above the upper truncation points are small. As the Finnish data does not have a reliable way to measure working full time, the data has individuals with very low incomes. This makes left tail of the earnings distributions long. The share of the population that falls outside the range is roughly 0.2 percent, all of which fall under the value -2. There are no observations above the value 2 due to the top-coding of high incomes.

not migrate, but there is no clear sorting pattern between women migrating to other Nordic countries and to other destinations.

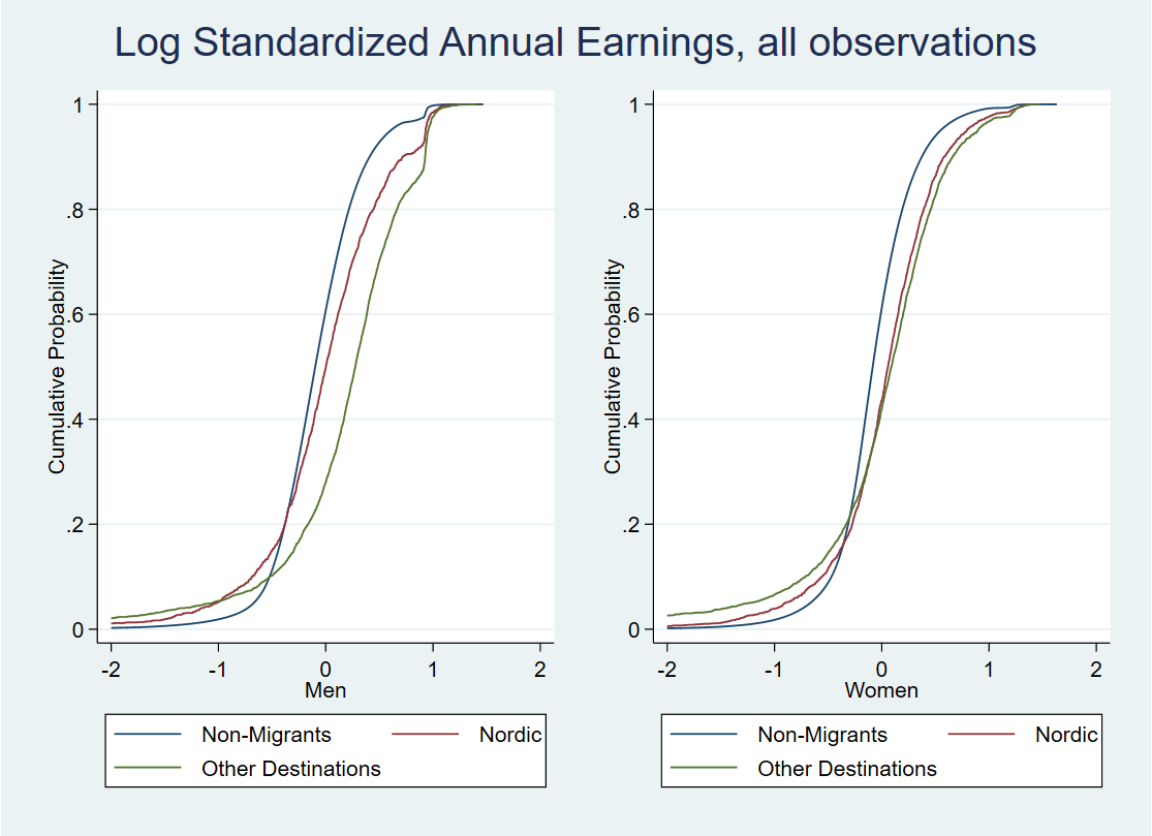

Figure 5. Cumulative Distribution Functions of Standardized Earnings

Notes: This figure pools observations of Finnish workers for the years 1997-2014, with at least one Finnish born parent, and who were aged 25 to 54 years and employed for 12 months. For each observation, standardized earnings are calculated by dividing the worker's annual gross earnings by the mean gross earnings of workers of the same age and gender during the same calendar year. The figure depicts the cumulative distribution function of log standardized earnings for three groups: non-migrants who did not emigrate during the following year, migrants to other Nordic countries during the following year who did not return in the next five years, and migrants to other destinations during the following year who did not return in the next five years.

The weaker selection to other Nordic countries than to the rest of the world among men can be explained by the relatively low rate of return to skills in Nordic countries, as well as by lower migration costs. There are several reasons why migration costs to other Nordic countries can be expected to be lower than to the rest of the world. First, all Nordic countries share cultural proximity that makes migration between them easier. Second, Sweden and Norway share border with Finland, and geographical proximity lowers both migration costs and costs of maintaining contacts with Finland after migration. Third,

Finland and Sweden have strong historic ties and large numbers of Finnish immigrants settled in Sweden before our period of analysis, providing a network for subsequent migrants. Such a network could reduce migration costs especially to those with less education, as shown by McKenzie and Rapoport (2010) in the context of migration from Mexico to the United States. Fourth, Finland has a Swedish-speaking minority of about 5 percent for whom migrating to Sweden is easy. Further, Swedish is also recognized as an official language in Finland, and it is mandatory to learn Swedish in schools. This might make migration to Sweden easier also for the Finnish-speaking majority.

Figure A2 from the appendix presents kernel estimates of the density functions of the logarithm of standardized pre-emigration earnings for men and women.[5] The density functions reveal the positive selection of migrants moving outside the Nordic countries along a major range of the income distribution for men, while the self-selection pattern is less clear for women. Kernel estimates for Finland are less clear than those for Denmark in Borjas et al. (2019), due to top-coding in the Finnish data.

Figure 6.1 presents emigration rates by decile of log standardized income for men and figure 6.2 for women. We find a U-shaped pattern, with emigration rate being highest in the top decile but also high (for women, second highest) in the lowest decile. A possible explanation for this is that at the upper part of the earnings distribution, migrants tend to be positively self-selected in terms of their transferable skills, while at the lowest decile of the earnings distribution there is also a considerable share of migrants with a bad job match.

[5] Following Leibbrandt et al. (2005), Fernandez-Huertas Moraga (2011) and Borjas et al. (2019), we use Silverman's reference bandwidth multiplied by 0.75 to prevent over-smoothing. The same bandwidth is used also in all reported density estimates.

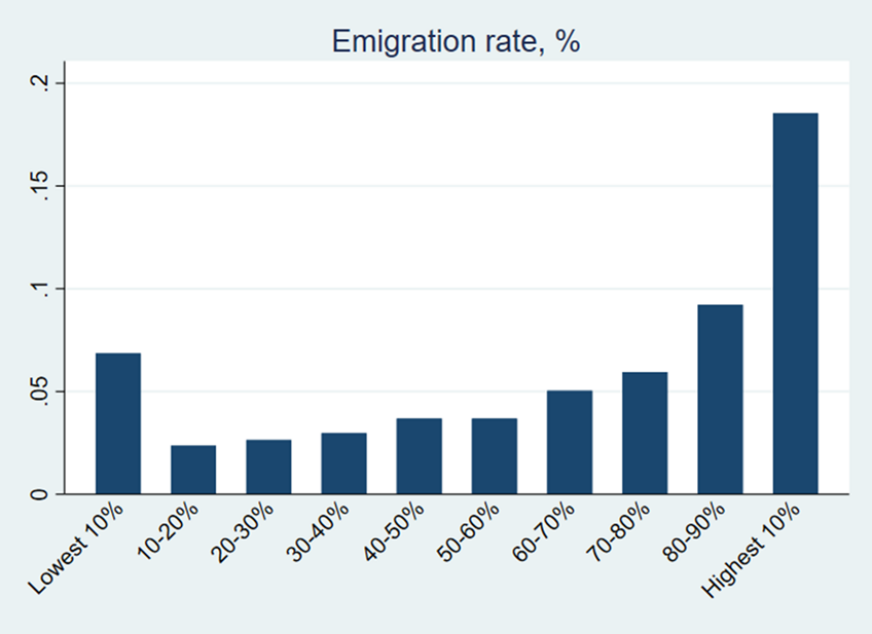

Figure 6.1. Emigration Rates by Decile of Log Standardized Income for Men

Notes: This figure depicts annual emigration rate of men satisfying the restrictions in Figure 5 by deciles of standardized earnings.

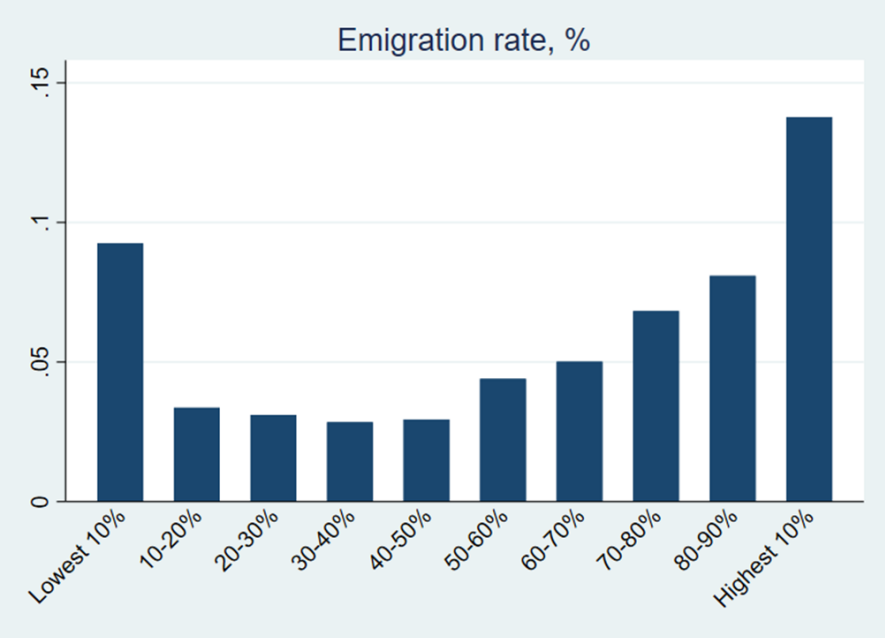

Figure 6.2. Emigration Rates by Decile of Log Standardized Income for Women

Notes: This figure depicts annual emigration rate of women satisfying the restrictions in Figure 5 by deciles of standardized earnings.

We also extend Borjas et al. (2019) by analyzing self-selection and sorting of emigrants according to their occupation. Table 4 shows 1-digit occupational classifications of non-migrants, migrants to other Nordic countries, and migrants to other destinations. We find that among both men and women, managers and professionals are significantly more likely to emigrate. We also find strong sorting between Nordic and non-Nordic destinations among men: managers and professionals form 27 percent of non-migrants, 47 percent of migrants to Nordic countries, and 66 percent of migrants to non-Nordic destinations. Among women, managers and professionals form 26 percent of non-migrants, 41 percent of migrants to Nordic countries, and 45 percent of migrants to non-Nordic destinations. Correspondingly, service and sales workers and manual workers in industry and workers in elementary occupations are considerably less likely to emigrate.

Table 4. Broad Occupational Groups of Non-migrants and Migrants

	Men			Women		
	Non-migrants	Nordic countries	Other destinations	Non - migrants	Nordic countries	Other destinations
Armed forces	1.3	n.a.	n.a.	0.1	n.a.	n.a.
Managers	5.1	10.3	16.8	2.2	3.9	5.5
Professionals	22.0	37.1	48.9	23.6	36.8	39.3
Technicians and associate professionals	18.5	19.1	18.2	25.5	28.4	22.6
Clerical support workers	4.0	4.9	4.3	11.1	10.8	12.3
Service and sales workers	10.5	10.3	4.2	26.3	15.9	15.8
Skilled agricultural, forestry and fishery workers	0.8	n.a.	n.a.	0.6	n.a.	n.a.
Craft and related trade workers	18.1	7.5	2.6	1.5	0.7	0.5
Plant and machine operators, and assemblers	15.0	8.3	2.9	3.3	0.1	1.7
Elementary occupations	4.8	2.2	1.4	5.9	2.5	2.0

Notes: This table pools observations of Finnish workers during 2010-2014, with at least one Finnish born parent, and who were aged at least 25 and at most 54 years and employed for 12 months. The table depicts column percentages of occupations at 1-digit level for three groups: non-migrants who did not emigrate during the following year, migrants to other Nordic countries during the following year who did not return in the next five years, and migrants to other destinations during the following year who did not return in the next five years. Cells with too few observations to report according to Statistics Finland privacy rules are marked with n.a.

We also checked the occupational distribution using 2-digit occupational codes. Migrant self-selection is especially pronounced for science and engineering professionals (in the population depicted in Table 4, this group includes 8 percent of non-migrant men and 2 percent of non-migrant women compared to 18 percent of men and 7 percent of women migrating to non-Nordic destinations), business and administration professionals (4 percent of non-migrant men and 5 percent of non-migrant women compared to 11 percent of men and 12 percent of women migrating to non-Nordic destinations) and information and communications technology professionals (4 percent of non-migrant men and 1 percent of non-migrant women compared to 11 percent of men and 3 percent of women migrating to non-Nordic destinations).

5 Self-selection in Unobserved Characteristics

In the previous section we showed how migrants are self-selected in terms of education and pre-migration earnings after adjusting for age and year effects. In this section, we examine self-selection in terms of the component of earnings that is associated with characteristics that are not observed. We adjust for differences in educational attainment and other observable variables between migrants and non-migrants by running Mincerian earnings regressions and study self-selection in terms of earnings regression residuals.[6] The residuals reflect the part of earnings that is uncorrelated with the observed measures of skill. The decomposition depends on the characteristics that are observed and can be included as regressors in the earnings regression model. As the set of characteristics depend on the dataset at hand, the selection is somewhat arbitrary, and the results depend on the choice of control variables. We begin our analysis with a specification which follows Borjas et al. (2019) as close as possible, following which we extend the analysis to include occupational categories.

Borjas et al (2019) argue that selection in terms of unobserved characteristics indicates how important the quality of job match is relative to the skill component that is internationally transferable. A pure random matching model would suggest that correlation of

[6] In the earnings regressions we use non-standardized annual earnings as the dependent variable. We include age and year fixed effects and run the regressions separately for men and women.

residuals related to match quality between potential jobs in the origin and destination would be zero. This would lead to negative selection in the residual earnings as only workers with a bad job match (corresponding to a negative residual) would find it optimal to emigrate. An alternative perspective is that the residual earnings reflect primarily the unobserved skills, and these are easily transferable across countries. If this is the case and the returns to unobserved skills are higher in the destination, emigrants would be positively self-selected in terms of unobservable characteristics (measured by residuals). Therefore, whether migrants from a country with smaller income differences are positively or negatively self-selected in terms of residuals would shed light on whether the residual earnings reflect primarily unobserved job match quality or unobserved abilities. If returns to unobserved skills are lower in destination, then migrants would be negatively self-selected even when the unobserved skills would be highly transferable. In that case, negative self-selection in residual earnings from a country with higher returns to skills to a destination country with lower returns is insufficient to conclude that this is because of unobserved job match quality.

Table 5 reports the Mincerian wage regressions used to calculate the residuals. The sample includes the whole population of prime-aged full-time workers pooled over the years between 1997 and 2014. The estimation equations include age and year fixed effects, dummies for educational attainment and dummies for being married and having children. Standard errors are clustered at the individual level and the model is estimated separately for men and women. The reported coefficients from the Mincerian wage regression for Finland are remarkably similar to those in Denmark as reported by Borjas et al. (2019). In both, the returns to bachelor's degree are higher for men than for women, and the returns to doctoral degree are higher for women than for men. The returns to master's degree are also higher for women, although the difference is very small in Finland. Another similarity between both is that being married and having children are associated with higher earnings for men and lower earnings for women than a single and childless person with same gender, age, and education could expect to earn.

Table 5. Mincerian Earnings Regressions, by Gender

	Men	Women
	Coefficient	Coefficient
	(1)	(2)
Married	0.0803***	-0.0249***
	(0.001)	(0.001)
Children	0.0312***	-0.0499***
	(0.001)	(0.001)
High school	0.1388***	0.1441***
	(0.002)	(0.002)
Vocational school	0.0679***	0.0349***
	(0.001)	(0.001)
Bachelor or equivalent	0.2859***	0.2285***
	(0.001)	(0.001)
Master's or equivalent	0.5424***	0.5518***
	(0.001)	(0.001)
Doctoral or equivalent	0.6253***	0.7409***
	(0.003)	(0.003)
Age fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
N	11,144,864	10,945,028
R-Squared	0.8447	0.8641

Notes: The table reports OLS results for earnings regressions. The dependent variable is log of annual earnings in 2020 euros. Coefficients for the age and year fixed effects are not shown. The data includes full time working individuals for the years 1997-2014, aged 25 to 54 and born in Finland with at least one Finnish born parent. Entrepreneurs and individuals with missing earnings information or non-positive earnings are excluded.

* p<0.10, ** p<0.05, *** p<0.010.

Figure 7 presents the cumulative distributions and Appendix Figure A3 Kernel densities of earnings regression residuals by gender for those who stay in Finland the following year, migrants to other Nordic countries, and migrants to non-Nordic countries. The values are truncated at -2 and 2, a range that covers most of the population. Emigrants from Finland are not as strongly positively self-selected with respect to residuals as emigrants from Denmark that Borjas et al. (2019) analyze. Instead, cumulative distribution functions

cross each other, so that migrants to non-Nordic countries are somewhat more likely to have high absolute values of negative residual earnings than migrants to Nordic countries, while migrants to non-Nordic countries are also more likely to have positive values of residual earnings than migrants to other Nordic countries and non-migrants. For women, the cumulative distribution function of residual earnings of migrants to non-Nordic countries is to the left of that of migrants to Nordic countries for negative residual earnings, and to the right for positive residual earnings.

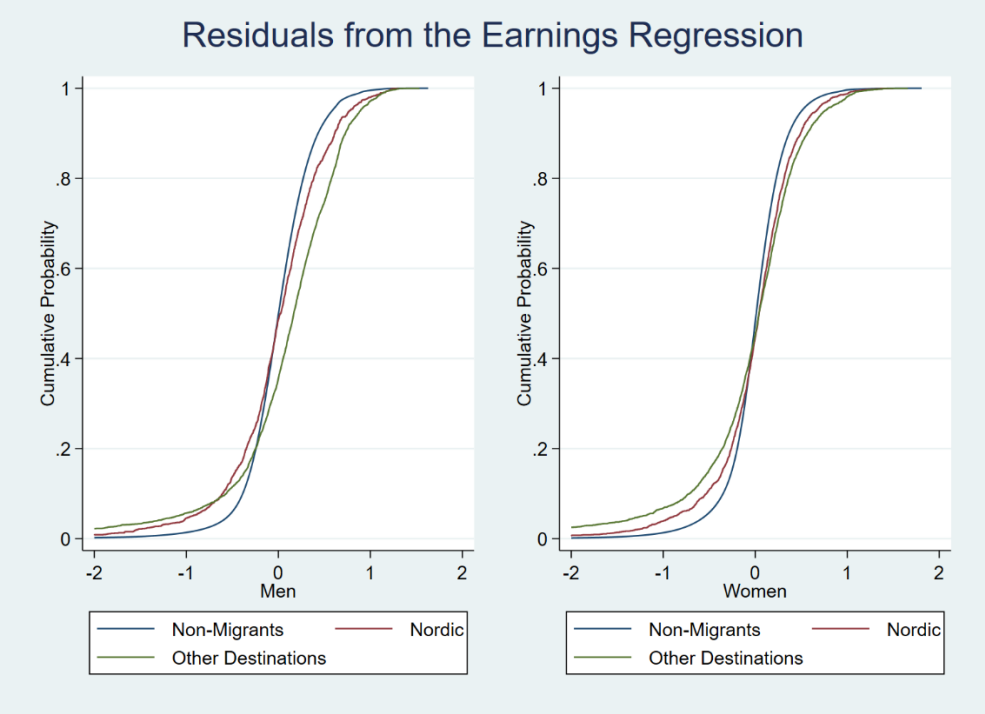

Figure 7. Cumulative Distribution Functions of Residual Earnings

Notes: This figure pools observations for the years 1997-2014 of Finnish workers with at least one Finnish born parent, and who were aged 25 to 54 years and employed for 12 months. For each observation, wage residuals are calculated based on the Mincerian wage regression in Table 5. The figure depicts the cumulative distribution function of residual earnings for three groups: non-migrants who did not emigrate during the following year, migrants to other Nordic countries during the following year who did not return in the next five years, and migrants to other destinations during the following year who did not return in the next five years.

Figure 8.1 aggregates self-selection patterns into emigration rates by deciles of residual earnings for men and figure 8.2 for women. We find a strong U-shaped pattern, with highest emigration rates in the lowest and highest decile of residual earnings. This suggests that bad job matches may be an important motivation to emigrate with low residual

earnings, while higher returns to unobserved abilities abroad are likely to play an important role with high residual earnings.

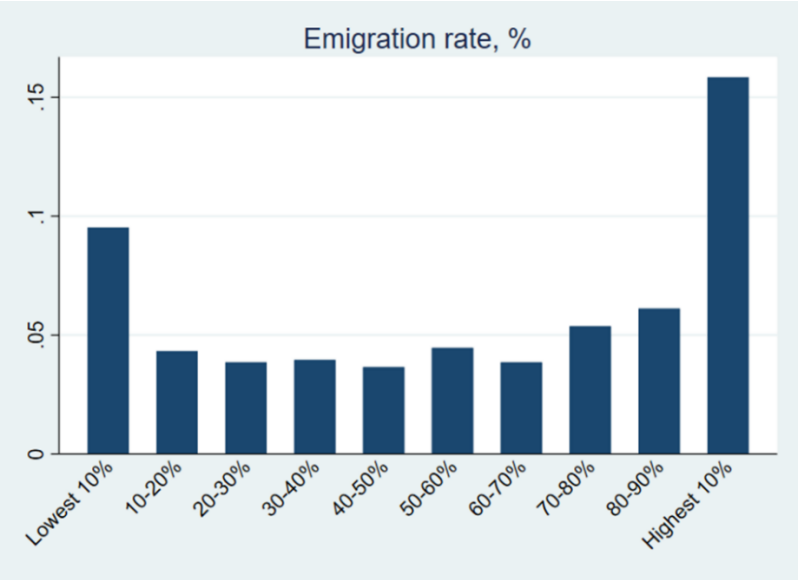

Figure 8.1. Emigration Rates by Decile of Earnings Regression Residuals for Men

Notes: This figure depicts annual emigration rate of men satisfying the restrictions in Figure 7 by deciles of residual earnings.

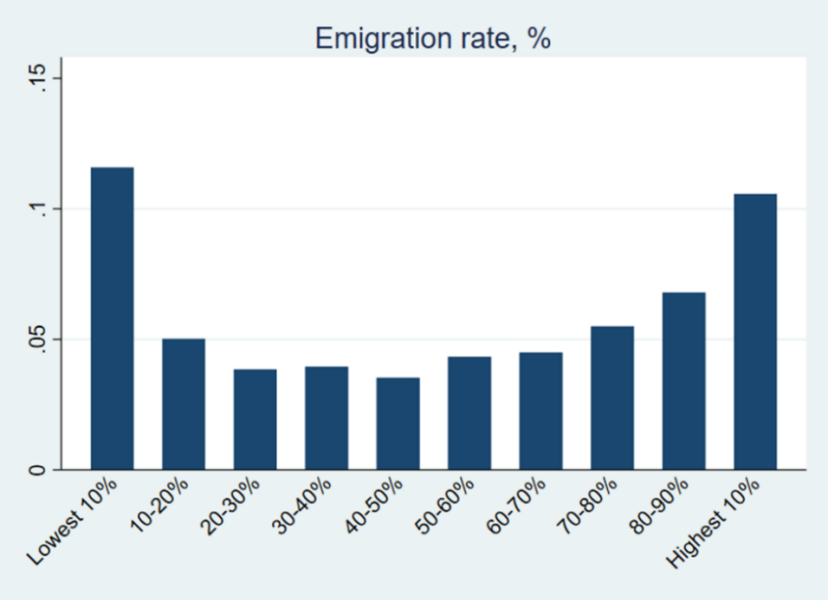

Figure 8.2. Emigration Rates by Decile of Earnings Regression Residuals for Women

Notes: This figure depicts annual emigration rate of women satisfying the restrictions in Figure 7 by deciles of residual earnings.

We next extend the analysis to control for occupational classification. We restrict the analysis to emigration from years 2011 to 2015, analyzing self-selection based on characteristics for the years 2010 to 2014. This restriction is needed as there had been changes in occupational classification between 2009 and 2010. Appendix Table A4 presents wage regressions for men, first without educational and occupational controls, then with educational controls, and 1-digit and 2-digit occupational controls that are used in the calculation of these residuals. Appendix Table A5 present corresponding regressions for women. Figure 9 presents the cumulative distributions of wage residuals for male and female migrants moving to other Nordic countries, non-Nordic countries and for those who stayed in Finland the following year when controlling for occupation at 1-digit level.[7] These patterns remain the same after controlling for occupations at two-digit level (Figure A5). U-shaped self-selection pattern in terms of deciles of residual earnings also prevails after controlling for occupations (Appendix Figures A6.1-A7.2).

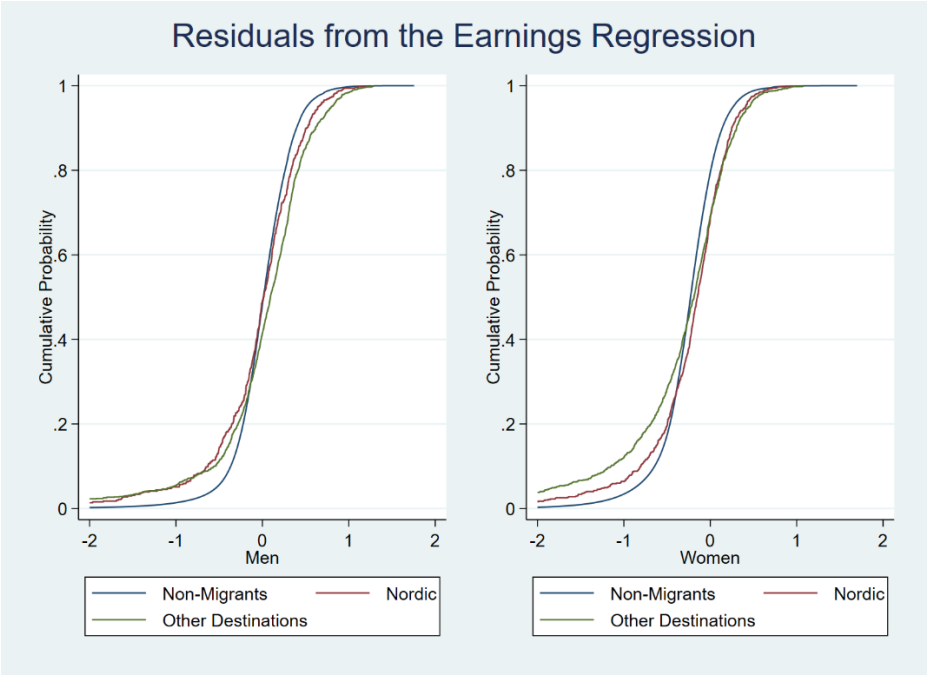

Figure 9. Cumulative Distribution Functions of Residual Earnings with Broad Occupational Controls

Notes: This figure pools observations for the years 2010-2014 of Finnish workers with at least one Finnish born parent, and who were aged 25 to 54 years and employed for 12 months. For each

[7] As Figure 9 uses a shorter time than Figure 7, we present as Figure A4 also the distribution of residual earning without occupational controls for years 2010 to 2015.

observation, wage residuals are calculated based on the Mincerian wage regression in Tables A4 and A5. The figure depicts the cumulative distribution function of residual earnings for three groups: non-migrants who did not emigrate during the following year, migrants to other Nordic countries during the following year who did not return in the next five years, and migrants to other destinations during the following year who did not return in the next five years.

Borjas et al. (2019) also estimated which fraction of emigrant self-selection from Denmark in terms of standardized earnings can be explained by observed characteristics, and which fraction by unobserved characteristics. To do so, they followed Chiquiar and Hanson (2005) and calculated predicted counterfactual wage distribution of emigrants. The counterfactual earnings distribution for migrants is estimated by weighting the data for non-migrants so that the distribution of observable characteristics mimics that for the migrants. The weights are based on logit models explaining migration decisions with observed individual characteristics.[8] They then compared the estimated distribution to the actual pre-migration earnings distributions for migrants and non-migrants. The difference between the density for non-migrants and the counterfactual density reflects the part of self-selection that is due to observed characteristics whereas the difference between the counterfactual and actual densities for migrants reflects the part of selection that is due to individual characteristics of that are not included in the logit model. In addition to presenting Kernel densities, Borjas et al. (2019) also quantified the differences by comparing the averages of the actual and counterfactual distributions. Deducting the average standardized earnings of non-migrants from the average standardized counterfactual earnings of migrants and dividing the difference by the actual difference in standardized earnings between migrants and non-migrants tell us, which fraction of migrant self-selection in terms of standardized earnings observable characteristics can explain. It turned out that observed characteristics could explain only 30% of the self-selection of male emigrants, and about half of the self-selection of female migrants. We follow the same process for self-selection of Finnish emigrants, but we include additional steps.

Figure 10 presents log standardized earnings of migrants and non-migrants, as well as four predicted counterfactual wage distributions of migrants, each of which is based on different logit model explaining migration decisions: model (1) controls only for age,

[8] The details of the estimation procedure are described in the appendix.

marital status and having children; model (2) controls also for education; and models (3) and (4) add indicators for occupational group at 1-digit and 2-digit level respectively. In addition, all models include a vector of year fixed effects. The regression results are reported in Tables A6 to A9 in the appendix, for regression coefficients from logit regressions and odds ratios, separately by gender.

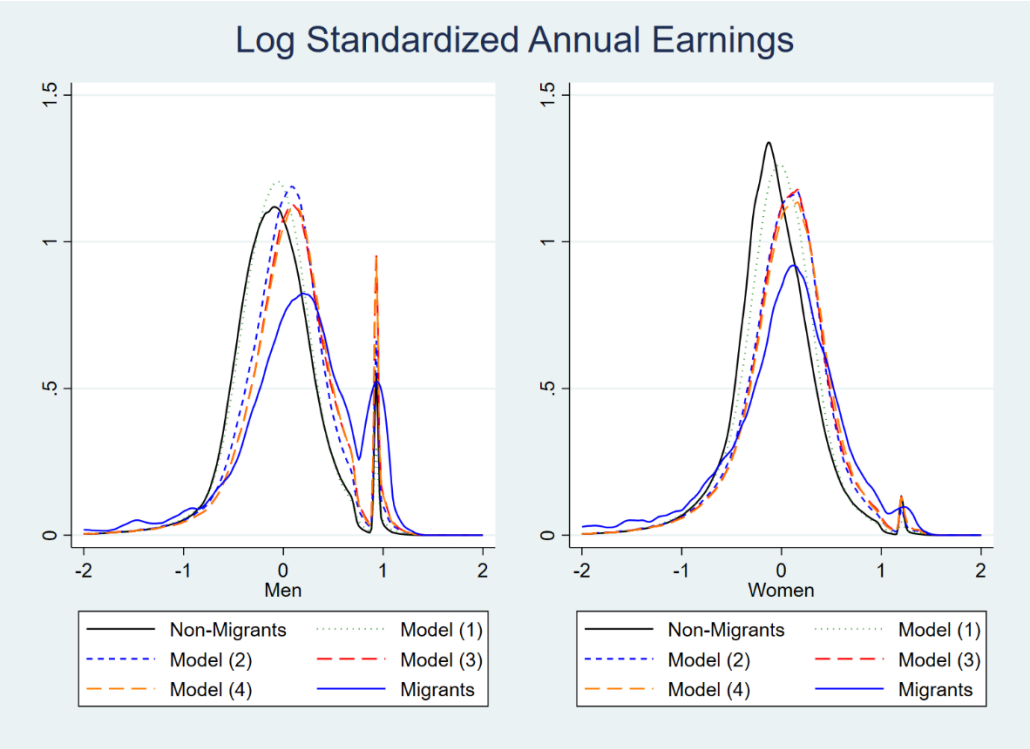

Figure 10. Counterfactual and Actual Densities of Log Standardized Earnings

Notes: This figure pools observations for the years 2010-2014 of male and female Finnish workers (separately) with at least one Finnish born parent, and who were aged 25 to 54, and employed for 12 months. Non-migrants did not emigrate during the following year, while migrants emigrated during the following year and did not return to Finland during the following five years. Non-Migrants and Migrants depict actual distributions of log standardized earnings. Models 1-4 present counterfactual earnings distributions for migrants, estimated based on coefficients in the corresponding column in Table A6 for men and A8 for women.

Table 6 presents actual and counterfactual differences between the average log standardized earnings of migrants and non-migrants, and which fraction of migrant self-selection various observable characteristics included in models (1) to (4) can explain, separately for men and women.

Table 6. Actual and Counterfactual Differences between the Average Log Standardized Earnings of Migrants and Non-migrants

	Men	Women
Non-migrant average	-0.0884	-0.0834
True average for migrants	0.0846	-0.0389
True difference	0.173	0.045
Model 1 (age and family controls)		
Estimated average for migrants	-0.0875	-0.0523
Counterfactual difference	0.001	0.031
Share of the actual difference explained by observable characteristics, %	0.5	69.9
Model 2 (adding education)		
Estimated average for migrants	0.027	0.026
Counterfactual difference	0.116	0.109
Share of the actual difference explained by observable characteristics, %	66.8	245.8
Model 3 (adding 1-digit occupational code)		
Estimated average for migrants	0.081	0.0479
Counterfactual difference	0.170	0.131
Share of the actual difference explained by observable characteristics, %	98.0	295.1
Model 4 (adding 2-digit occupational code)		
Estimated average for migrants	0.083	0.0554
Counterfactual difference	0.171	0.139
Share of the actual difference explained by observable characteristics, %	99.0	311.9

Notes: This table is based on actual and counterfactual distributions presented in Figure 10, separately for men and women. The first panel shows average log standardized earnings of migrants and non-migrants and the true difference in these. Subsequent panels show estimated averages based on counterfactual distributions in models 1-4, as well as which share of actual difference in the first panel each model explains.

We find major gender difference in the share of actual difference in log standardized earnings explained by observable characteristics. Age and family controls explain less than one percent of average earnings difference between migrants and non-migrants for men but 70 percent for women, while education explains 66 percent of earnings difference for men and 176 percent for women. Adding occupation boosts the share of earnings difference explained for men by 32 to 33 percent, leaving only 1-2 percent of average difference for unobserved characteristics. For women, education and occupation explain more than 100 percent of average wage differences, implying negative self-selection in

unobserved characteristics. This could reflect women with a low-paying job match either searching for a better job match abroad or being more likely to emigrate for family-related reasons because their opportunity cost is lower.

6 Conclusion

In this article, we analyzed migrant self-selection and sorting in terms of education, occupation, earnings, and residual earnings. Our analysis using Finnish data extended the analysis of self-selection of emigrants from Denmark in Borjas (2019). We included an analysis of self-selection in terms of occupations, using in some specifications one-digit or two-digit occupational classifications, in addition to education and demographic controls. Similar to Borjas et al. (2019), we focused our analysis of self-selection on employees aged between 25 and 54 in the year before eventual emigration. We analyzed only Finns without immigration background, as for those with immigrant background emigration from Finland might correspond to returning to their home country, or country from which their parents came from.

We find that emigrants from Finland are strongly positively self-selected in terms of education and earnings. Furthermore, there is a clear sorting pattern with respect to education in line with what Grogger and Hanson (2011) predict. More educated migrants tend to choose destinations offering higher returns to skills, which corresponds to non-Nordic countries, while less educated migrants are more likely to migrate to other Nordic countries, which are culturally and geographically closer and have relatively small income differences in international comparison, just as Finland does. This gender difference in sorting is more pronounced for men, which could reflect women migrating more often due to partner or other family considerations, as documented by Munk et al. (2022) for emigration from Denmark.

We also analyzed sorting with respect to standardized earnings and residual earnings. Men migrating outside Nordic countries have clearly higher standardized earnings than men migrating to other Nordic countries. For women, the sorting pattern with respect to standardized earnings is more mixed. The share of women with low standardized earnings is higher among women migrating outside Nordic countries than among women

migrating to other Nordic countries. At the same time, the share of women with high standardized earnings is slightly higher among women migrating outside Nordic countries than among women migrating to other Nordic countries. We also find that men migrating outside Nordic countries have higher residual earnings than men migrating to other Nordic countries, although this difference is smaller than what Borjas et al. (2019) find for men emigrating from Finland.

The main difference between self-selection from Finland and self-selection from Denmark is how women are self-selected in terms of residual earnings. Borjas et al. (2019) concluded that about 30% of the positive selection of Danish men who emigrate in pre-migration earnings can be attributed to the observable characteristics included in the logit model and about 70% is attributable to unobservable determinants of productivity, while for Danish women who emigrate, observable and unobservable characteristics are about equally important. We find that about 66 percent of self-selection of Finnish men who emigrate in terms of average log earnings can be explained by their education, less than one percent by age and family controls, and 32-33 percent by occupation once education is controlled for. For women, age and family controls alone explain 70 percent and education explains clearly more than 100 percent of positive self-selection, implying that on average, women are negatively self-selected in residual earnings. This is a major difference compared with the self-selection from Denmark. We hope that the finding of different self-selection patterns with respect to residual earnings inspires further studies on the topic, using data from other countries and periods to establish the relative prevalence of different self-selection and sorting patterns, and which factors predict which patterns are dominant in different countries.

7 References

Abramitzky Ran, Boustan Leah Platt, and Eriksson Katherine. 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.

Aksoy Cevat G. and Poutvaara Panu. 2021. Refugees' and irregular migrants' self-selection into Europe. *Journal of Development Economics*, 151: 102681.

Borjas George J. 1987. Self-selection and the earnings of immigrants. *American Economic Review*, 77(4): 531–53.

Borjas George J. 2003. The labor demand curve is downward sloping: reexamining the impact of immigration on the labor market. *The Quarterly Journal of Economics*, 118(4): 1335–1374.

Borjas George J., and Bratsberg Bernd.1996. Who leaves? The outmigration of the foreign-born. *The Review of Economics and Statistics*, 78(1):165–76.

Borjas George J., Kauppinen Ilpo, and Poutvaara Panu. 2019. Self-selection of Emigrants: Theory and evidence on stochastic dominance in observable and unobservable characteristics. *The Economic Journal*, 129(167): 143–171.

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

Fernandez-Huertas Moraga Jesus. 2011. New evidence on emigrant selection. *The Review of Economics and Statistics*, 93(1): 72-96.

Gould Eric D., and Moav Omer.2016. Does high inequality attract high skilled immigrants?. *The Economic Journal*, 126(593): 1055-1091.

Grogger Jeffrey, and Hanson Gordon H..2011. Income maximization and the selection and sorting of international migrants. *Journal of Development Economics*, 95(1): 42-57.

Honkapohja Seppo, and Koskela Erkki. 1999. The economic crisis of the 1990s in Finland. *Economic Policy*, 14(29): 400-436.

Kalin Salla, Kauppinen Ilpo, Kotakorpi Kaisa, Pirttilä Jukka. 2022. Migration and tax policy: Evidence from Finnish full population data. Vatt Working papers 150.

Kaestner Robert, and Malamud Ofer. 2014. Self-selection and international migration: new evidence from Mexico. *Review of Economics and Statistics*, 96(1): 78-91.

Leibbrandt Murray, Levinsohn James, and McCrary Justin. 2005. Incomes in South Africa since the fall of apartheid. NBER Working Paper No. 11384.

Mincer Jacob (1978) Family migration decisions. *Journal of Political Economy*, 86: 749–773.

Munk Martin D., Poutvaara Panu, and Nikolka Till. 2022. International Family Migration and the Dual-Earner Model. *Journal of Economic Geography* 22(2): 263–287.

Patt Alexander, Ruhose Jens, Wiederhold Simon, and Flores Miguel.2021. International emigrant selection on occupational skills. *Journal of the European Economic Association*, 19(2): 1249-1298.

Razin Assaf and Sadka Efraim. 2000. Unskilled migration: a burden or a boon for the welfare state? *Scandinavian Journal of Economics*, 102(3): 463–479.

Roy Andrew Donald.1951. Some thoughts on the distribution of earnings. *Oxford Economic Papers*, 3(2):135-146.

Sinn Hans-Werner. 1997. The selection principle and market failure in systems competition, *Journal of Public Economics*, 66(2): 247–74.

Spitzer Yannay, and Zimran Ariell. 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.

Wegge Simone A.2002. Occupational self-selection of European emigrants: evidence from nineteenth-century Hesse-Cassel. *European Review of Economic History*, 6(3): 365-394.

Wildasin David.E. 1991. Income redistribution in a common labor market, *American Economic Review*, 81(4): 757–74.

Appendix

Details Related to Counterfactual Earnings Distributions

This presentation follows closely the technical appendix in Borjas et al. (2019). Let w represent the logarithm of standardized annual earnings as defined earlier (i.e. earnings adjusted for age, gender, and year effects). Let $f(w|x)$ be the earnings density in Finland, conditional on observable characteristics x. Let I be an indicator variable equal to one if the individual migrates the following year and equal to zero otherwise.

Define $h(x|I=0)$ as the conditional density of observed characteristics among workers in Finland who choose not to migrate, and $h(x|I=1)$ is the corresponding conditional density among migrants.

The observed wage density for the non-migrants is

$$g(w|I=0) = \int f(w|x, I=0)h(x|I=0)\,dx. \tag{1}$$

Similarly, the observed wage density for the migrants is

$$g(w|I=1) = \int f(w|x, I=1)h(x|I=1)\,dx. \tag{2}$$

The self-selection results that we have reported so far consist of comparing the distribution functions associated with the densities in equations (1) and (2). If the pre-migration earnings density for non-migrants were not available and we would estimate it based on the observable characteristics of the migrants, the obtained counterfactual distribution would be:

$$\hat{g}(w|I=1) = \int f(w|x, I=0)h(x|I=1)\,dx. \tag{3}$$

The density in equation (3) corresponds to the density of income for non-migrants but integrated over the density of observable characteristics for migrants. The density in (3) can be written as

(4)

$$\hat{g}(w|I=1) = \int f(w|x, I=0)h(x|I=0)\frac{h(x|I=1)}{h(x|I=0)}dx$$

$$= \int \theta(x)f(w|x, I=0)h(x|I=0)\,dx,$$

where $\theta(x) = \frac{h(x|I=1)}{h(x|I=0)}$.

We use Bayes' law to write:

(5) $$h(x) = \frac{h(x|I=0)Pr(I=0)}{\Pr(I=0|x)} \text{ and } h(x) = \frac{h(x|I=1)Pr(I=1)}{\Pr(I=1|x)},$$

where $h(x)$ is the unconditional density of observed characteristics.

We combine (4) and (5) to solve for $\theta(x)$:

(6)

$$\theta(x) = \frac{Pr(I=1|x)}{1-Pr(I=1|x)}\frac{Pr(I=0)}{Pr(I=1)}.$$

The proportion $Pr(I = 0)/Pr(I = 1)$ is a constant related to the proportion of migrants in the data, so we can set it to one in kernel density estimation. That is, we replace the weight $\theta(x)$ by:

(7)

$$\theta^e(x) = \frac{Pr(I = 1|x)}{1 - Pr(I = 1|x)}.$$

As in Chiquiar and Hanson (2005), the individual weights $\theta^e(x)$ are calculated by estimating a logit model where the dependent variable indicates if a person emigrated. We estimate four different models, separately for men and women. Model (1) controls only for age, marital status and having children; model (2) controls also for education; and models (3) and (4) add indicators for occupational group at 1-digit and 2-digit level respectively. In addition, all models include a vector of year fixed effects. The regression results are reported in Tables A6 and A7.

We use the regression for computing a predicted migration probability for each non-migrant individual in the data. Based on the predicted probabilities we then compute the weights θ^e that we use in counterfactual density estimation. As earlier, we use Silverman's reference bandwidth multiplied by 0.75. Figure 10 presents the resulting counterfactual density functions of the logarithm of standardized earnings as well as the actual distributions for migrants and non-migrants. To conduct the counterfactual analysis we pool the sample of all migrants (regardless of whether they moved to Nordic countries or not).

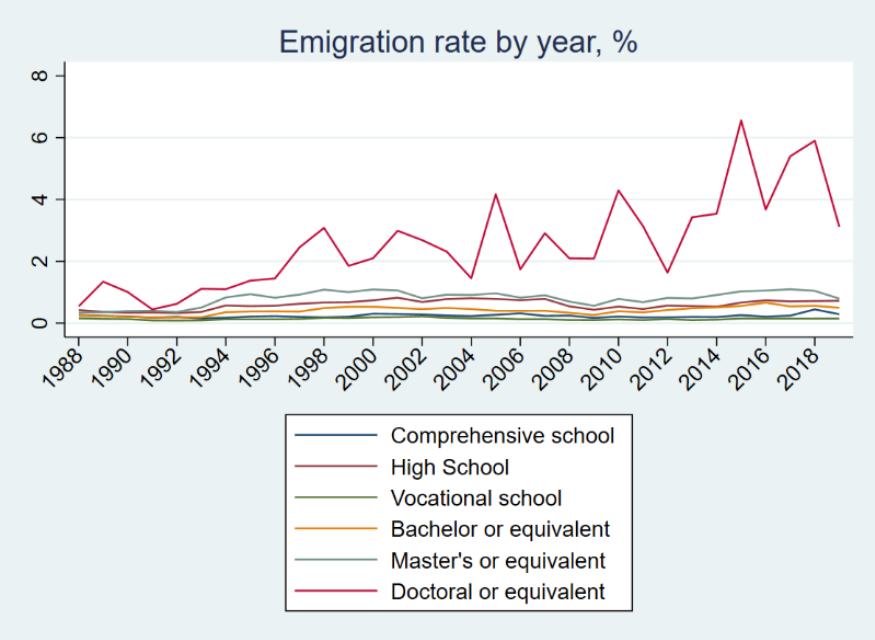

Figure A1.1. Emigration Rates by Educational Attainment for the Population Aged 25-30

Notes: This figure shows for each year the percentage of Finnish citizens aged 25-30 years, with at least one parent born in Finland , and who emigrated for at least 365 days.

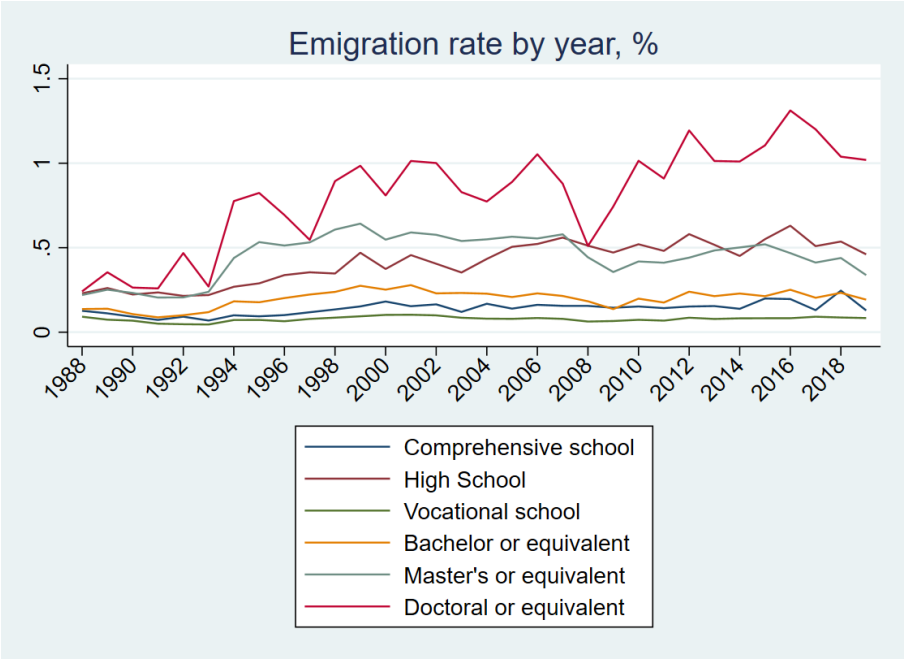

Figure A1.2. Emigration Rates by Educational Attainment for the Population Aged 31-40

Notes: This figure shows for each year the percentage of Finnish citizens aged 31-40 years, with at least one parent born in Finland, and who emigrated for at least 365 days.

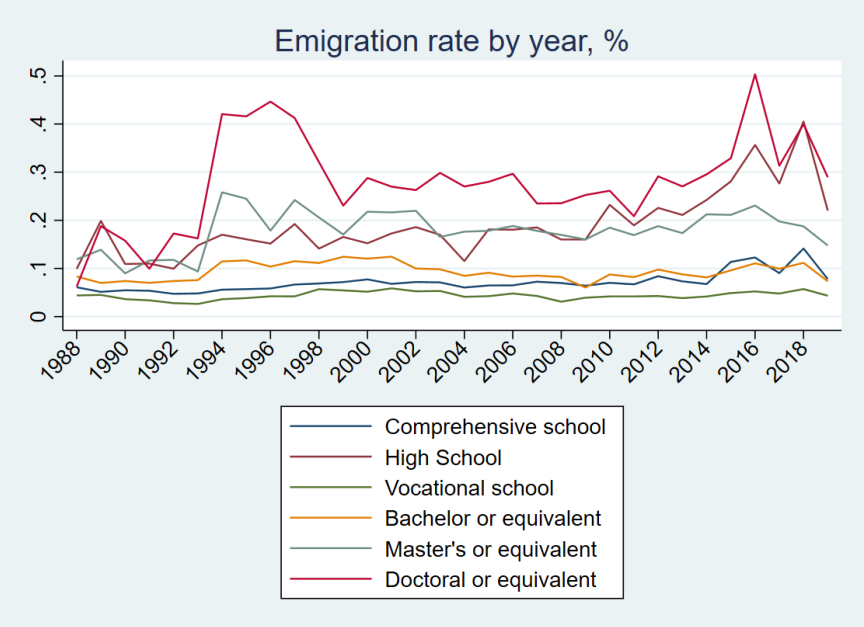

Figure A1.3. Emigration Rates by Educational Attainment for the Population Aged 41-54

Notes: This figure shows for each year the percentage of Finnish citizens aged 41-54 years, with at least one parent born in Finland, and who emigrated for at least 365 days.

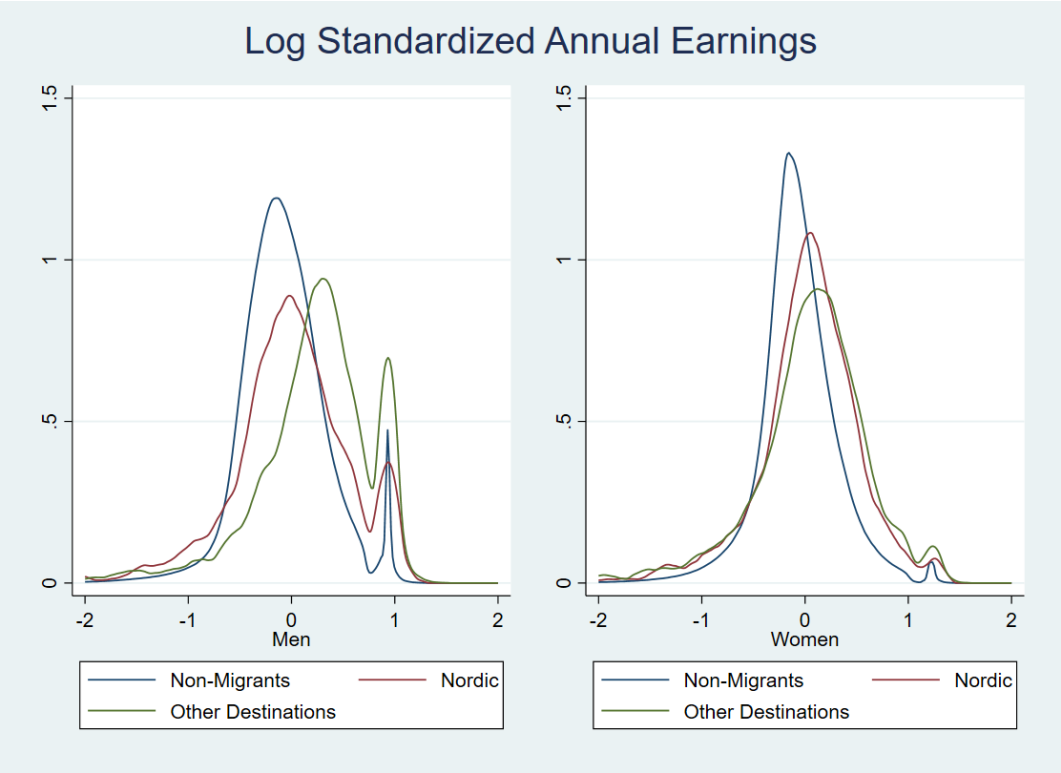

Figure A2. Density Functions for Standardized Earnings

Notes: This figure pools observations for the years 1997-2014 of Finnish workers with at least one parent born in Finland, who were aged at least 25 and at most 54 years and employed for 12 months. For each observation, standardized earnings are calculated by dividing the worker's annual gross earnings by the mean gross earnings of workers of the same age and gender during the same calendar year. The figure depicts the density function of log standardized earnings for three groups: non-migrants who did not emigrate during the following year, migrants to other Nordic countries during the following year who did not return in the next five years, and migrants to other destinations during the following year who did not return during the next five years. The income variable is top-coded so that incomes in the highest annual percentile are registered as a median within that group.

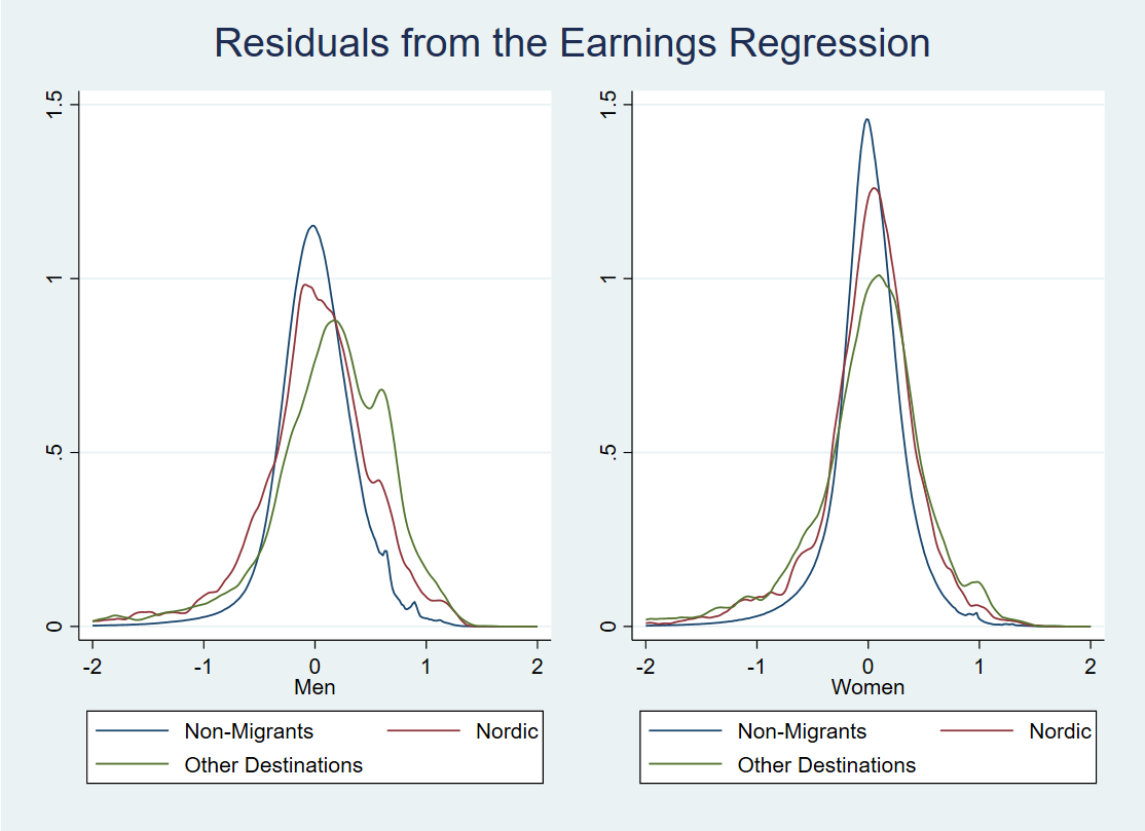

Figure A3. Density Functions of Residuals from Earnings Regression Explaining Standardized Earnings

Notes: This figure pools observations of Finnish workers with at least one parent born in Finland from year 1997 to year 2014 who were aged at least 25 and at most 54 years and employed for 12 months. For each observation, wage residuals are calculated based on the Mincerian wage regression in Table 5. The figure depicts the density function of residual earnings for three groups: non-migrants who did not emigrate during the following year, migrants to other Nordic countries during the following year who did not return in the next five years, and migrants to other destinations during the following year who did not return during the next five years. The income variable is top-coded so that incomes in the highest annual percentile are registered as a median within that group.

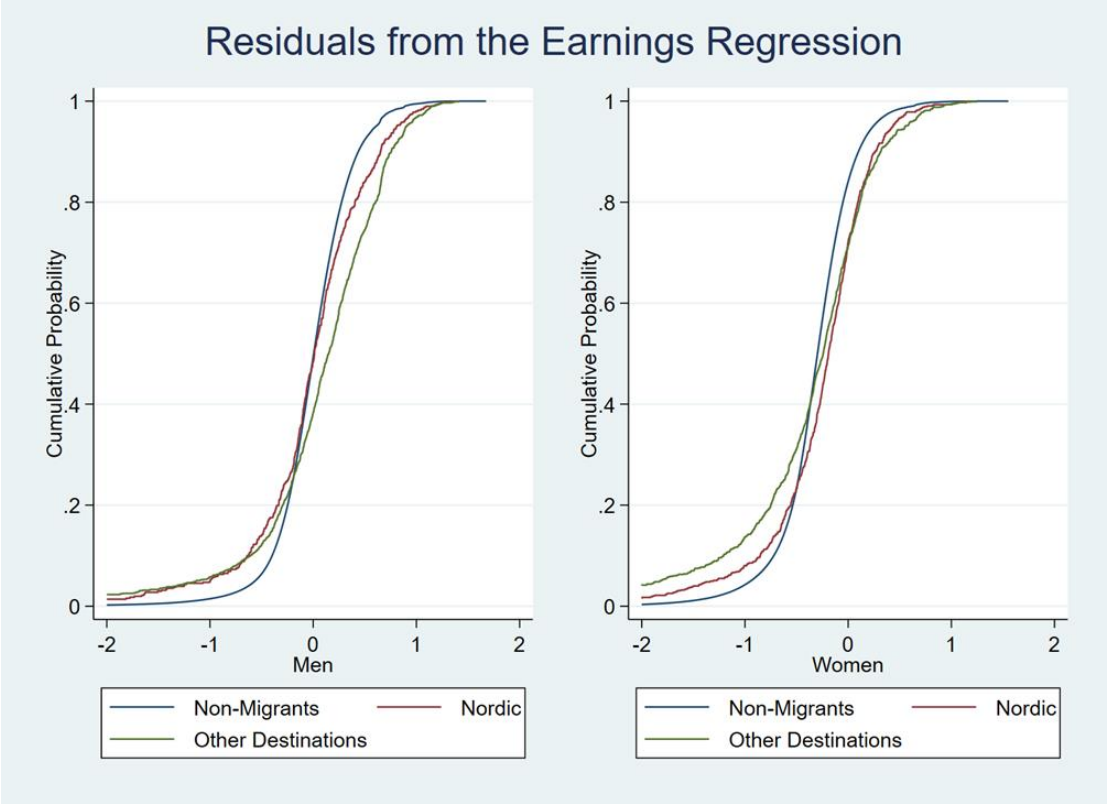

Figure A4. Cumulative Distribution Functions of Residual Earnings without Occupational Controls for Years 2010 to 2014.

Notes: This figure pools observations of Finnish workers with at least one parent born in Finland from year 2010 to year 2014 who were aged at least 25 and at most 54 years and employed for 12 months. For each observation, wage residuals are calculated based on the Mincerian wage regression in Tables A4 and A5. The figure depicts the density function of residual earnings for three groups: non-migrants who did not emigrate during the following year, migrants to other Nordic countries during the following year who did not return in the next five years, and migrants to other destinations during the following year who did not return during the next five years. The income variable is top-coded so that incomes in the highest annual percentile are registered as a median within that group.

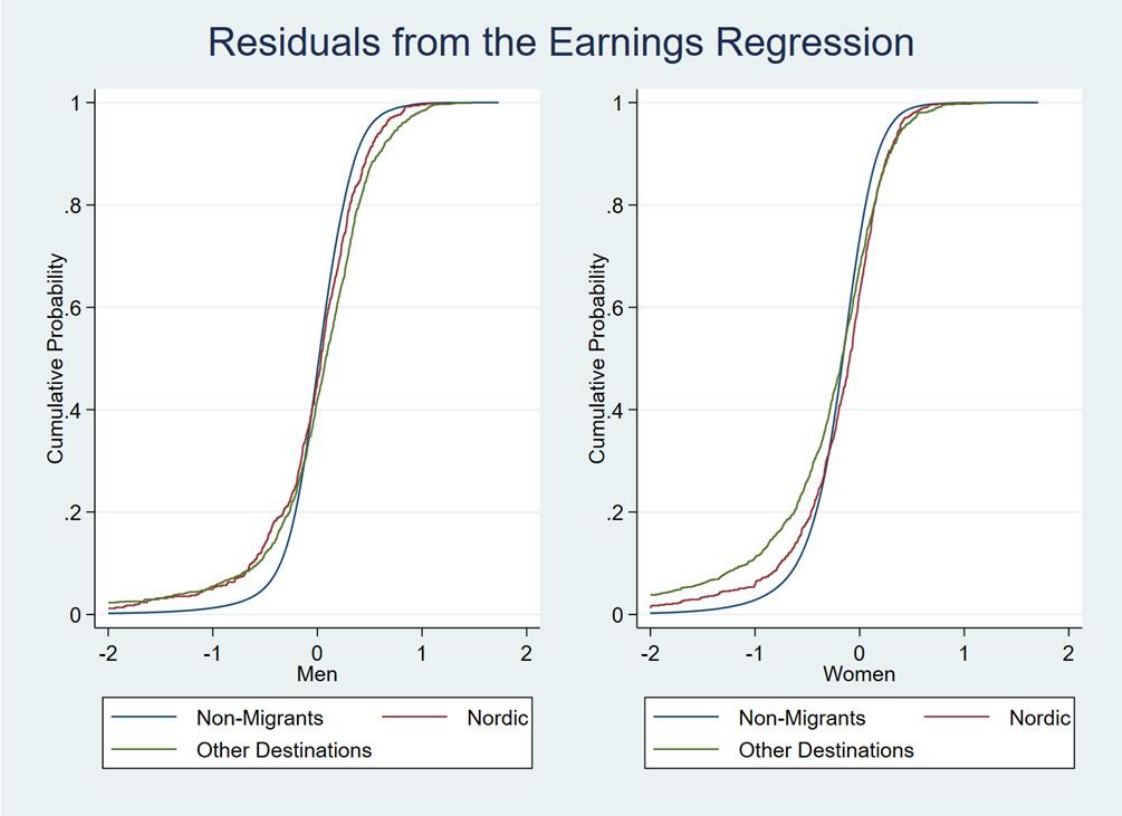

Figure A5. Cumulative Distribution Functions of Residual Earnings with 2-Digit Occupational Controls for Years 2010 to 2014

Notes: This figure pools observations of Finnish workers with at least one parent born in Finland from year 2010 to year 2014 who were aged at least 25 and at most 54 years and employed for 12 months. For each observation, wage residuals are calculated based on the Mincerian wage regression that includes 2-digit occupational controls in Tables A4 and A5. The figure depicts the density function of residual earnings for three groups: non-migrants who did not emigrate during the following year, migrants to other Nordic countries during the following year who did not return in the next five years, and migrants to other destinations during the following year who did not return during the next five years. The income variable is top-coded so that incomes in the highest annual percentile are registered as a median within that group.

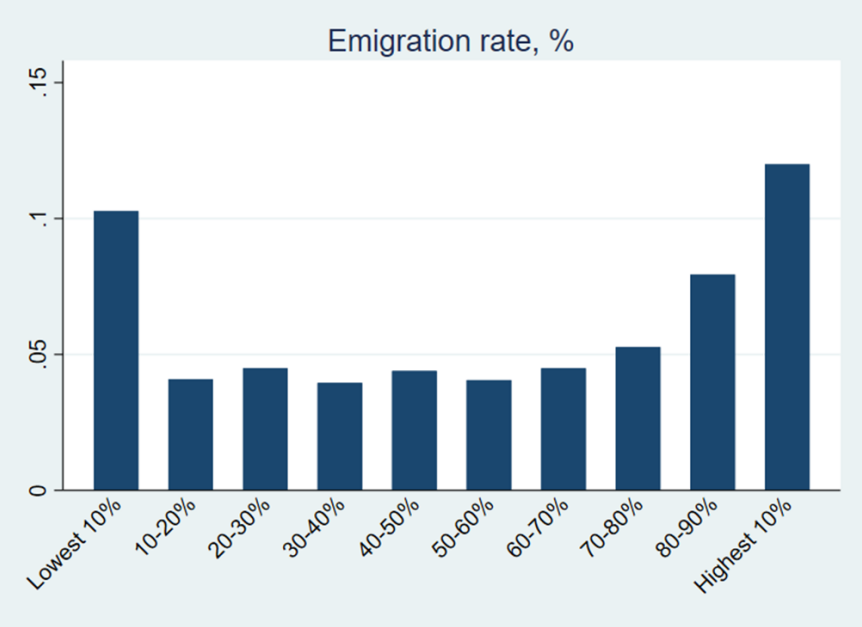

Figure A6.1. Emigration Rates by Decile of Residuals from an Earnings Regression with 1-Digit Occupational Controls for Men.

Notes: This figure depicts annual emigration rate of men by deciles of residual earnings based on column 3 of Table A4.

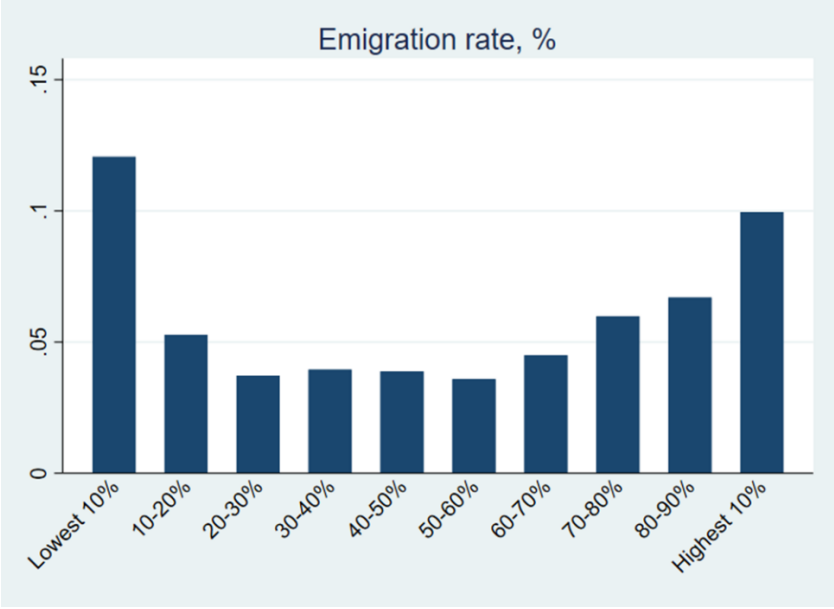

Figure A6.2. Emigration Rates by Decile of Residuals from an Earnings Regression with 1-Digit Occupational Controls for Women

Notes: This figure depicts annual emigration rate of women by deciles of residual earnings based on column 3 of Table A5.

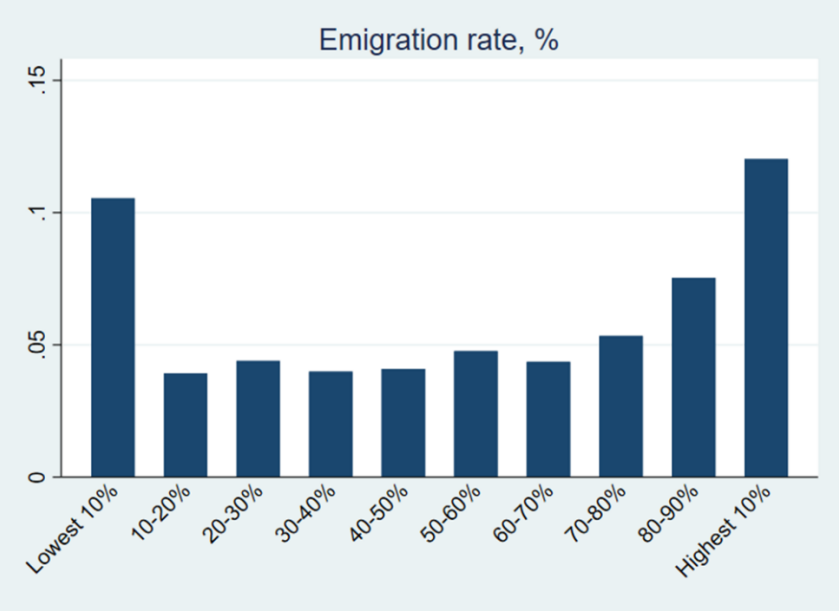

Figure A7.1. Emigration Rates by Decile of Residuals from an Earnings Regression with 2-Digit Occupational Controls for Men

Notes: This figure depicts annual emigration rate of men by deciles of residual earnings based on column 4 of Table A4.

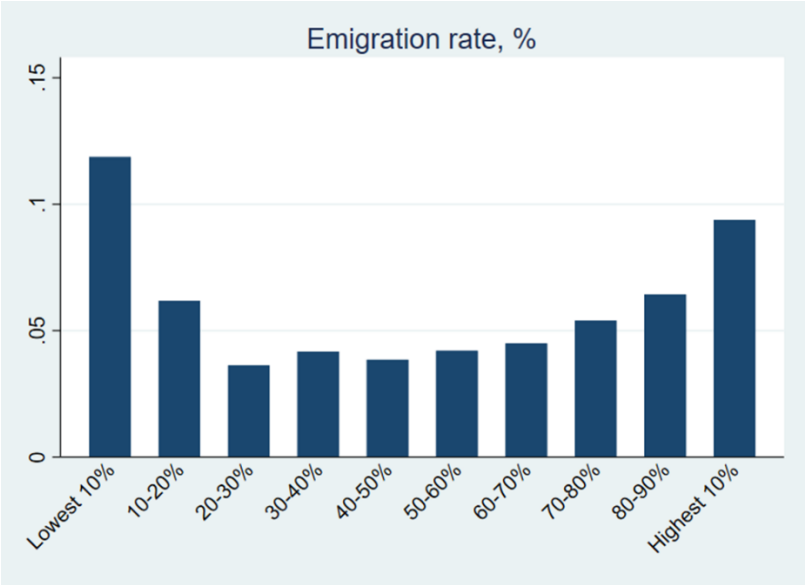

Figure A7.2. Emigration Rates by Decile of Residuals from an Earnings Regression with 2-Digit Occupational Controls for Women

Notes: This figure depicts annual emigration rate of women by deciles of residual earnings based on column 4 of Table A5.

Table A1. Number of Emigrants, by Destination

	Men	Women
Sweden	1207	1510
United States	625	414
United Kingdom	417	584
Germany	485	410
Norway	256	375
Switzerland	310	235
Denmark	139	250
Belgium	154	193
Spain	163	158
Netherlands	126	185
Estonia	196	47
France	108	127
China	158	38
Australia	94	93
Luxembourg	94	78
Singapore	94	44
Canada	71	54
Ireland	55	67
Italy	42	68
Austria	46	59
Russia	47	16
Japan	41	13
Portugal	12	12
New Zealand	13	10
Greece	6	14
Other	491	253
Total	5450	5307

Notes: The data includes full time working individuals for the period 1997-2014, aged 25 to 54 years and born in Finland with no immigration background. Entrepreneurs and individuals with missing earnings information or non-positive earnings are excluded. Migration spells of five years or more.

Table A2. Percentages of Migrants by Destination Country Groups

	1998-2007	2008- 2015
	(1)	(2)
Men		
Other Nordic Countries	30.0	29.0
Other EU-15 and Switzerland	36.8	35.7
Other	33.2	35.3
Women		
Other Nordic Countries	41.0	39.8
Other EU-15 and Switzerland	41.5	38.5
Other	17.5	21.7
Total		
Other Nordic Countries	35.3	34.4
Other EU-15 and Switzerland	39.1	37.1
Other	25.6	28.5

Notes: The data includes full time working individuals for the period 1997-2014, aged 25 to 54 years and born in Finland with no immigration background. Entrepreneurs and individuals with missing earnings information or non-positive earnings are excluded. Migration spells of five years or more.

Table A3. Logit Estimates of the Probability of Emigration, by Gender

	Men		Women	
	Coefficient	Odds ratio	Coefficient	Odds ratio
	(1)	(2)	(3)	(4)
Log of standardized income	0.4464***	1.5627***	-0.2149***	0.8067***
	(0.066)	(0.103)	(0.050)	(0.041)
Married	0.1479***	1.1594***	-0.2342***	0.7912***
	(0.041)	(0.047)	(0.042)	(0.034)
Children	-0.7011***	0.4960***	-0.7121***	0.4906***
	(0.052)	(0.026)	(0.045)	(0.022)
Married*Children	-0.1347**	0.8740**	-0.1114*	0.8946*
	(0.066)	(0.058)	(0.064)	(0.058)
High school	0.9545***	2.5973***	0.8920***	2.4399***
	(0.066)	(0.172)	(0.075)	(0.184)
Vocational school	-0.5348***	0.5858***	-0.4671***	0.6268***
	(0.064)	(0.038)	(0.072)	(0.045)
Bachelor or equivalent	0.5147***	1.6732***	0.3965***	1.4866***
	(0.061)	(0.102)	(0.067)	(0.100)
Master's or equivalent	1.5147***	4.5479***	1.2510***	3.4938***
	(0.065)	(0.297)	(0.072)	(0.251)
Doctoral or equivalent	1.9676***	7.1536***	1.9923***	7.3321***
	(0.089)	(0.633)	(0.106)	(0.775)
Age fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
N	11,144,864	11,144,864	10,945,028	10,945,028
Pseudo R-Squared	0.0722	0.0722	0.0745	0.0745

Notes: The table reports logistics regression results using regression coefficients (columns 1 and 3) and odds ratios (columns 2 and 4). The dependent variable is a dummy that gets value one if the if the individual is a long-term migrant migrating to other Nordic countries or other destinations the following year. Individually clustered standard errors are in parentheses. Coefficients for the age and year fixed effects are not shown. The data includes full time working individuals for the period 1993-2014, aged 25 to 54 years and born in Finland with no immigration background. Entrepreneurs and individuals with missing earnings information or non-positive earnings are excluded. Migration spells of five years or more.

* p<0.10, ** p<0.05, *** p<0.010.

Table A4. Earnings Regressions for Men

	Coefficient (1)	Coefficient (2)	Coefficient (3)	Coefficient (4)
Married	0.1401***	0.0841***	0.0664***	0.0606***
	(0.001)	(0.001)	(0.001)	(0.001)
Children	0.0465***	0.0507***	0.0406***	0.0383***
	(0.001)	(0.001)	(0.001)	(0.001)
High school		0.1156***	0.0493***	0.0376***
		(0.003)	(0.002)	(0.002)
Vocational school		0.0775***	0.0734***	0.0616***
		(0.001)	(0.001)	(0.001)
Bachelor or equivalent		0.2918***	0.1584***	0.1219***
		(0.002)	(0.002)	(0.002)
Master's or equivalent		0.5254***	0.3007***	0.2654***
		(0.002)	(0.002)	(0.002)
Doctoral or equivalent		0.6072***	0.3815***	0.3622***
		(0.004)	(0.004)	(0.004)
Age fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Occupation 1 digit fixed effects	No	No	Yes	No
Occupation 2 digit fixed effects	No	No	No	Yes
N	2,962,695	2,962,695	2,962,695	2,962,695
R-Squared	0.1152	0.2574	0.3489	0.3915

Notes: The table reports OLS results for earnings regressions. The dependent variable is log of annual earnings in 2020 euros. Coefficients for the age and year fixed effects are not shown. The data includes full time working men aged 25 to 54 years and born in Finland with no immigration background in 2010-2014. Entrepreneurs and individuals with missing earnings information or non-positive earnings or missing occupation information are excluded.

* p<0.10, ** p<0.05, *** p<0.010.

Table A5. Earnings Regressions for Women

	Coefficient (1)	Coefficient (2)	Coefficient (3)	Coefficient (4)
Married	0.0217***	-0.0189***	-0.0258***	-0.0260***
	(0.001)	(0.001)	(0.001)	(0.001)
Children	-0.0736***	-0.0622***	-0.0631***	-0.0607***
	(0.001)	(0.001)	(0.001)	(0.001)
High school		0.1319***	0.0554***	0.0508***
		(0.003)	(0.003)	(0.003)
Vocational school		0.0553***	0.0538***	0.0426***
		(0.002)	(0.002)	(0.002)
Bachelor or equivalent		0.2477***	0.1033***	0.0976***
		(0.002)	(0.002)	(0.002)
Master's or equivalent		0.5546***	0.2979***	0.2975***
		(0.002)	(0.002)	(0.002)
Doctoral or equivalent		0.7586***	0.4837***	0.4525***
		(0.004)	(0.004)	(0.004)
Age fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Occupation 1 digit fixed effects	No	No	Yes	No
Occupation 2 digit fixed effects	No	No	No	Yes
N	3,094,422	3,094,422	3,094,422	3,094,422
R-Squared	0.0727	0.2570	0.3272	0.3643

Notes: The table reports OLS results for earnings regressions. The dependent variable is log of annual earnings in 2020 euros. Coefficients for the age and year fixed effects are not shown. The data includes full time working women aged 25 to 54 years and born in Finland with no immigration background in 2010-2014. Entrepreneurs and individuals with missing earnings information or non-positive earnings or missing occupation information are excluded.

* p<0.10, ** p<0.05, *** p<0.010.

Table A6. Logit Estimates of the Probability of Emigration for Men

	Coefficient (1)	Coefficient (2)	Coefficient (3)	Coefficient (4)
Married	0.4311***	0.2796***	0.2242***	0.2272***
	(0.068)	(0.068)	(0.069)	(0.069)
Children	-0.9558***	-0.7228***	-0.7093***	-0.6947***
	(0.091)	(0.093)	(0.093)	(0.093)
Married*Children	0.0282	-0.2229*	-0.2713**	-0.2724**
	(0.116)	(0.116)	(0.116)	(0.116)
High school		0.9369***	0.4448***	0.4273***
		(0.123)	(0.126)	(0.127)
Vocational school		-0.6322***	-0.6221***	-0.6233***
		(0.121)	(0.121)	(0.123)
Bachelor or equivalent		0.5669***	-0.1675	-0.1768
		(0.113)	(0.124)	(0.126)
Master's or equivalent		1.6282***	0.6431***	0.6983***
		(0.109)	(0.128)	(0.129)
Doctoral or equivalent		2.1575***	1.1552***	1.3418***
		(0.146)	(0.160)	(0.162)
Age fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Occupation 1 digit fixed effects	No	No	Yes	No
Occupation 2 digit fixed effects	No	No	No	Yes
N	2,962,695	2,962,695	2,962,695	2,955,018
Pseudo R-Squared	0.0278	0.0701	0.0835	0.0876

Notes: The table reports logit regression results using regression coefficients. The dependent variable is a dummy that gets value one if the individual is a long-term migrant migrating to other Nordic countries or other destinations the following year. Individually clustered standard errors are in parentheses. Coefficients for the age and year fixed effects and occupation fixed effects are not shown. The data includes full time working individuals aged 25 to 54 and born in Finland with no immigration background in 2010-2014. Entrepreneurs and individuals with missing earnings information or non-positive earnings or missing occupation information are excluded. Migration spells of five years or more.

* p<0.10, ** p<0.05, *** p<0.010.

Table A7. Odds Ratio Estimates of the Probability of Emigration for Men

	Odds ratio (1)	Odds ratio (2)	Odds ratio (3)	Odds ratio (4)
Married	1.5389***	1.3227***	1.2513***	1.2550***
	(0.105)	(0.090)	(0.086)	(0.086)
Children	0.3845***	0.4854***	0.4920***	0.4992***
	(0.035)	(0.045)	(0.046)	(0.047)
Married*Children	1.0286	0.8002*	0.7624**	0.7616**
	(0.119)	(0.093)	(0.089)	(0.089)
High school		2.5520***	1.5601***	1.5331***
		(0.314)	(0.196)	(0.195)
Vocational school		0.5314***	0.5368***	0.5362***
		(0.064)	(0.065)	(0.066)
Bachelor or equivalent		1.7628***	0.8458	0.8379
		(0.199)	(0.105)	(0.106)
Master's or equivalent		5.0949***	1.9023***	2.0102***
		(0.558)	(0.243)	(0.260)
Doctoral or equivalent		8.6493***	3.1745***	3.8261***
		(1.262)	(0.509)	(0.619)
Age fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Occupation 1 digit fixed effects	No	No	Yes	No
Occupation 2 digit fixed effects	No	No	No	Yes
N	2,962,695	2,962,695	2,962,695	2,955,018
Pseudo R-Squared	0.0278	0.0701	0.0835	0.0876

Notes: The table reports logit regression results using odds ratios. The dependent variable is a dummy that gets value one if the individual is a long-term migrant migrating to other Nordic countries or other destinations the following year. Individually clustered standard errors are in parentheses. Coefficients for the age and year fixed effects and occupation fixed effects are not shown. The data includes full time working individuals aged 25 to 54 and born in Finland with no immigration background in 2010-2014. Entrepreneurs and individuals with missing earnings information or non-positive earnings or missing occupation information are excluded. Migration spells of five years or more.

* p<0.10, ** p<0.05, *** p<0.010.

Table A8. Logit Estimates of the Probability of Emigration for Women

	Coefficient (1)	Coefficient (2)	Coefficient (3)	Coefficient (4)
Married	-0.0348	-0.0757	-0.0895	-0.0783
	(0.071)	(0.071)	(0.071)	(0.071)
Children	-0.9514***	-0.7370***	-0.7275***	-0.6844***
	(0.076)	(0.079)	(0.079)	(0.079)
Married*Children	-0.0059	-0.1725	-0.1780	-0.1713
	(0.108)	(0.109)	(0.109)	(0.109)
High school		0.7217***	0.5504***	0.5390***
		(0.142)	(0.146)	(0.147)
Vocational school		-0.8362***	-0.8754***	-0.8470***
		(0.139)	(0.140)	(0.142)
Bachelor or equivalent		0.0987	-0.1912	-0.1467
		(0.128)	(0.140)	(0.142)
Master's or equivalent		0.8625***	0.4400***	0.5048***
		(0.129)	(0.154)	(0.153)
Doctoral or equivalent		1.3832***	0.9398***	1.0170***
		(0.185)	(0.204)	(0.203)
Age fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Occupation 1 digit fixed effects	No	No	Yes	No
Occupation 2 digit fixed effects	No	No	No	Yes
N	3,094,422	3,094,422	3,094,422	3,070,910
Pseudo R-Squared	0.0514	0.0727	0.0759	0.0816

Notes: The table reports logit regression results using regression coefficients. The dependent variable is a dummy that gets value one if the individual is a long-term migrant migrating to other Nordic countries or other destinations the following year. Individually clustered standard errors are in parentheses. Coefficients for the age and year fixed effects and occupation fixed effects are not shown. The data includes full time working individuals aged 25 to 54 and born in Finland with no immigration background in 2010-2014. Entrepreneurs and individuals with missing earnings information or non-positive earnings or missing occupation information are excluded. Migration spells of five years or more.

* p<0.10, ** p<0.05, *** p<0.010.

Table A9. Odds Ratio Estimates of the Probability of Emigration for Women

	Odds ratio (1)	Odds ratio (2)	Odds ratio (3)	Odds ratio (4)
Married	0.9658	0.9271	0.9144	0.9247
	(0.069)	(0.066)	(0.065)	(0.066)
Children	0.3862***	0.4785***	0.4831***	0.5044***
	(0.029)	(0.038)	(0.038)	(0.040)
Married*Children	0.9941	0.8416	0.8370	0.8425
	(0.108)	(0.092)	(0.091)	(0.092)
High school		2.0579***	1.7339***	1.7142***
		(0.293)	(0.253)	(0.251)
Vocational school		0.4334***	0.4167***	0.4287***
		(0.060)	(0.058)	(0.061)
Bachelor or equivalent		1.1037	0.8259	0.8636
		(0.141)	(0.116)	(0.122)
Master's or equivalent		2.3690***	1.5527***	1.6566***
		(0.306)	(0.239)	(0.254)
Doctoral or equivalent		3.9876***	2.5594***	2.7648***
		(0.739)	(0.523)	(0.562)
Age fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Occupation 1 digit fixed effects	No	No	Yes	No
Occupation 2 digit fixed effects	No	No	No	Yes
N	3,094,422	3,094,422	3,094,422	3,070,910
Pseudo R-Squared	0.0514	0.0727	0.0759	0.0816

Notes: The table reports logit regression results using odds ratios. The dependent variable is a dummy that gets value one if the individual is a long-term migrant migrating to other Nordic countries or other destinations the following year. Individually clustered standard errors are in parentheses. Coefficients for the age and year fixed effects and occupation fixed effects are not shown. The data includes full time working individuals aged 25 to 54 and born in Finland with no immigration background in 2010-2014. Entrepreneurs and individuals with missing earnings information or non-positive earnings or missing occupation information are excluded. Migration spells of five years or more.

* p<0.10, ** p<0.05, *** p<0.010.