Stockholm Research Reports in Demography | no 2023:25

Cultural proximity and migration

Evidence from multigenerational population registers on mother tongue

Rosa Weber and Jan Saarela

ISSN 2002-617X | Department of Sociology

Stockholm University
Demography Unit

Cultural proximity and migration: Evidence from Finnish multigenerational population registers on mother tongue

Rosa Weber[1,2,3] and Jan Saarela[3]

1: Stockholm University Demography Unit (SUDA), Department of Sociology, Stockholm University, SE-106 91 Stockholm, Sweden
2: Institut National d'Études Démographiques, 93300 Paris, France
3: Åbo Akademi Unviersity, FIN-65101 Vasa, Finland

Abstract

Using Finnish register data, this study assesses the role of cultural proximity in international migration. Finland has an ethnic minority of Swedish speakers, allowing us to proxy for different strengths of attachment to Finland across the population. We exploit information on individual and parental ethno-linguistic affiliation to identify Swedish and Finnish speakers with uniform and mixed backgrounds. Individuals with mixed backgrounds are particularly informative because they are bilingual but are raised in different communities. We find a gradient whereby Swedish speakers with a uniform background are the most likely to migrate, followed by individuals with mixed backgrounds. Finnish speakers with uniform backgrounds are the least likely to migrate. Among individuals with mixed backgrounds, international migration is more common for those living in predominantly Swedish-speaking municipalities. The patterns remain when controlling for parental and individual socioeconomic characteristics. These results underline that cultural proximity seems to play an important role in migration decisions.

Keywords: international migration; register data; return migration; ethno-linguistic background; cultural proximity

Stockholm Research Reports in Demography 2023:25
ISSN 2002-617X

1 Introduction

Migration intentions and decisions have been longstanding topics of research within the social sciences. This interest stems partially from the idea that migration flows can have important implications for economic, social and cultural development in both the origin and destination countries. However, the implications of migration—both at the country and individual levels—differ depending on the cultural proximity between migrants' origin and destination. Indeed, migration between culturally proximate countries can facilitate integration and boost returns on human capital (Docquier, Tansel, and Turati 2020). However, it can also prevent the full realisation of the potential benefits of international migration (Adserà 2015; Rapoport, Sardoschau, and Silve 2021). Therefore, gaining a better understanding of the role of cultural proximity in migration behaviour is important for guiding migration and integration policies.

A growing body of literature has assessed the role of cultural proximity or distance in migration decisions, but the findings remain heterogeneous (Wang, Graaff, and Nijkamp 2018). Some studies have indicated that culture has a strong impact on migration, and stronger than the differences in unemployment rates between countries (Adserà and Pytliková 2015; Arif 2020; Belot and Ederveen 2012; Belot and Hatton 2012; Docquier, Tansel, and Turati 2020), while others find that culture plays a minor role in migration decisions (Caragliu et al. 2013; Mayda 2010; Ruyssen and Salomone 2018). A common approach in previous empirical research concerned with the association between cultural proximity and migration has been to categorise migrants according to how "culturally close" the origin and destination countries are. These measures of cultural proximity generally refer to aspects such as similar gender and secular values, a common history and the use of the same language (Adserà and Pytliková 2015; Belot and Ederveen 2012; Ruyssen and Salomone 2018). Nevertheless, defining and measuring cultural proximity is problematic because it involves intertwined concepts such as affinity, language, norms and attitudes (Saarela and Scott 2019). The more heterogeneous the study population of migrants, the more difficult it is to isolate the concept from other contributing factors.

The present paper takes a novel approach to assessing the importance of cultural proximity for international migration. We exploit the fact that the Finnish population register includes information on the unique mother tongue of every citizen, that each person can be linked to their parents and that the register includes complete longitudinal information about emigrations, immigrations and the country of destination. Moreover, Finland has two official languages: Finnish and Swedish. The country's native population consists of Finnish speakers (Finnish-registered persons), who amount to about 87% of the total current population, and a numerical minority of Swedish speakers (Swedish-registered persons), who amount to about 5% of the population. Finnish is distinct from most other languages in the Nordic and European countries but is close to Estonian. Both Finnish and Estonian are part of the Finno-Ungarian group. Swedish, which is a Scandinavian language like Norwegian, Danish and Icelandic, is markedly different from Finnish.

Given that intermarriage between Swedish and Finnish speakers is common, a high share of individuals have one Finnish-speaking and one Swedish-speaking parent (Saarela 2021). Within mixed unions, parents need to choose one mother tongue to register for their child because they cannot record multilingualism. Prior evidence suggests that the mother tongue registered for individuals with a mixed background is somewhat arbitrary (Finnäs and O'Leary 2003; Obućina and Saarela 2020). Even though the decision has few binding and immediate consequences, it is indicative of a person's cultural proximity to either the Finnish- or Swedish-speaking community (Obućina and Saarela 2020). Behind the linguistic facet, cultural norms and values distinguish Finnish and Swedish speakers, and there are numerous identity markers related to group images, attitudes and social structure that separate the groups culturally (Hedberg 2004; McRae 1999; Saarela and Finnäs 2018). Besides its communicative function, a person's mother tongue also has a symbolic dimension (cf. Gans 1979).

Our empirical strategy is to investigate how a two-generational dimension of the mother tongue relates to international emigration and return migration. We distinguish between Swedish and Finnish speakers who have parents of the same ethno-linguistic affiliation (uniform background) and those with parents of different ethno-linguistic affiliations (mixed background). Individuals' attachment to Finland and to the Finnish culture can be assumed to be the highest among those with a Finnish uniform background (Rooth and Saarela 2007), followed by a Finnish mixed background and Swedish mixed background. It is expected to be the lowest among individuals with a Swedish uniform background, while this group has a stronger attachment to Sweden and the Swedish culture (Saarela and Scott 2019). The groups with a mixed background are particularly informative because they are exposed to the Finnish and Swedish languages at home but attend different schools (Finnish or Swedish ones) and are raised in different ethno-linguistic communities (Finnäs and O'Leary 2003; Saarela and Finnäs 2016). Although our measure of ethno-linguistic background provides a crude proxy of cultural proximity, it goes beyond the standard approach used in previous studies on the cultural antecedents of migration and fits well into the study context.

Beyond assessing differences in the likelihood of migrating and returning across ethno-linguistic groups, the empirical analysis studies potential mechanisms underlying the observed patterns. First, we control for parental and individual socioeconomic characteristics to assess the extent to which they account for the differences between the groups. Second, we differentiate between migration to different destination regions to study whether groups are more likely to migrate to countries that are culturally more proximate. In the Finnish context, the Swedish-speaking community can be assumed to be culturally most proximate to Sweden, followed by other Nordic countries (Hedberg 2004). Estonia, on the other hand, can be assumed to be culturally the most proximate to the Finnish-speaking community while not being culturally proximate to the Swedish-speaking community (Tammaru, Kumer-Haukanõmm, and Anniste 2010). We compare these regions to all other destinations (which we call the rest of the world), to which no group is expected to have a strong affinity. Third, we focus on individuals with a mixed background and assess whether the share of Swedish speakers in the municipality relates to migration behaviour independently of mother tongue registration. This allows us to assess the role of the residential context.

Our findings reveal a gradient whereby Swedish speakers with a uniform background are the most likely to migrate, followed by Swedish mixed background and Finnish mixed background. Finnish speakers with uniform backgrounds are the least likely group to migrate. Assessments of the potential mechanisms underlying the differences in the likelihood of migrating indicate that the gradient is attenuated but persists when we control for parental and individual socioeconomic characteristics. The gradient is strongest for migration to Sweden, followed by migration to the other Nordic countries. In contrast, we observe no gradient for migration to the rest of the world, and we find indications of a reverse gradient for migration to Estonia. We also find that living in a predominantly Swedish-speaking municipality increases the likelihood of migrating, independent of mother tongue registration. Still, Swedish speakers are more likely to migrate than Finnish speakers, even when the two groups live in similar municipalities. Return migration presents a mirror image. That is, Finnish speakers with a uniform background are the most likely group to return migrate, followed by Finnish mixed background and Swedish mixed background. Swedish speakers with a uniform background are the least likely group to return migrate. These patterns are consistent with the interpretation that cultural selectivity is important for the first and subsequent moves.

In the following, we discuss the context and literature in more detail and thereafter describe the data, empirical strategy and our findings. We conclude with a summary and interpretation of our results.

2 Context and previous literature

The Swedish-speaking minority in Finland can be traced back to the period when Finland was under Swedish rule (Tandefelt and Finnäs 2007). At this time, the main language spoken in the Finnish territory was Swedish. When Finland was ceded to Russian rule, Russia encouraged the development and use of the Finnish language as a counter to the influence of the Swedish language, hence underlining the importance of the Finnish language for national identity in Finland, even before the country gained its independence in 1917.

2.1 Mother tongue registration

In Finland, parents register the mother tongue of their child after birth (Saarela 2021). Parents can choose between Finnish, Swedish or other. As expected, Swedish-speaking couples record their children as Swedish speakers, while Finnish-speaking couples record their children as Finnish speakers (Obućina and Saarela 2020). In contrast, couples with mixed mother tongues have to make a more active decision because parents cannot choose multiple languages or record multilingualism (Tandefelt and Finnäs 2007). Hence, the children of intermarriages between Finnish or Swedish speakers are recorded as Finnish or Swedish speakers. In other words, the only way to distinguish individuals with a uniform ethno-linguistic background from those with a mixed background is through the linkage between children and parents, which is available in the analysed data. Considering that around 40% of Swedish speakers marry a Finnish speaker, this comprises a sizeable share of the population (Saarela 2021).

Approximately two-thirds of the children born in mixed unions are Swedish-registered (Saarela 2021), indicating a preference towards recording children as Swedish speakers while also pointing to considerable heterogeneity in parental decisions. Prior work shows that parental education and, to some extent, the share of Swedish speakers in the area are important in this decision (Finnäs and O'Leary 2003). The gender of the child plays a more modest role (Obućina and Saarela 2020), and there is limited within-family variation, meaning that the gender of the first-born child plays a slightly more important role than the gender of subsequent children. The mother's ethno-linguistic affiliation is the most influential factor, which may be related to the wording of the question, where parents are asked to record the child's *mother* tongue (Tandefelt and Finnäs 2007). Other proposed hypotheses are that mothers present a closer link to their children's culture or face greater restrictions from their family (Finnäs and O'Leary 2003). In short, there is considerable heterogeneity in the factors underlying parents' decisions. Information on two generations allows us to identify the groups for which mother tongue registration is nonsystematic.

2.2 Implications of ethno-linguistic affiliation

In Finland, all children learn Finnish and Swedish in school. However, there are two parallel school systems from kindergarten to university. In one school system, (all) instruction is in Finnish and the other (all) instruction in Swedish. Children tend to go to Swedish or Finnish schools, depending on their ethno-linguistic affiliation (Finnäs and O'Leary 2003; Tandefelt and Finnäs 2007). Still, children with a mixed background can be expected to have good knowledge of both Finnish and Swedish through their parents and extended families. In contrast, Finnish is the first language or mother tongue for Finnish speakers with a uniform background, while Swedish is the first language or mother tongue for Swedish speakers with a uniform background.

Beyond language, a large body of literature shows that Swedish and Finnish speakers differ across an array of sociodemographic outcomes, even though there is considerable regional variation in ethno-linguistic gradients. Swedish speakers tend to have lower unemployment (Saarela and Finnäs 2003), higher wages (Saarela and Finnäs 2004), greater wealth (Saarela 2006) and more social capital than Finnish speakers (Hyyppä and Mäki 2001; Nyqvist et al. 2008). Swedish speakers tend to live longer (Koskinen and Martelin 2003; Reini and Saarela 2021; Saarela and Finnäs 2005; Saarela and Finnäs 2016) and have lower alcohol consumption than Finnish speakers (Paljärvi et al. 2009; Saarela and Kolk 2020). Voting behaviour has also been found to differ across the groups. Two-thirds of Swedish speakers vote for the Swedish People's Party, a primarily ethnic party whose political agenda is to protect the interests of the Swedish-speaking community (Westinen 2015).

Prior studies also show that these groups differ in their migration and return migration behaviours. Swedish speakers have higher emigration rates to Sweden and lower return migration rates than Finnish speakers (Hedberg and Kepsu 2003; Saarela and Scott 2017; Weber and Saarela 2019). Swedish speakers have also been found to earn higher wages in Sweden, which partly accounts for their lower return migration rates (Rooth and Saarela 2007;

Saarela and Scott 2017). However, the ethno-linguistic differences in migration and return migration also seem to be related to the group’s cultural proximity to Sweden (Hedberg 2004). Swedish-speaking migrants have a much higher likelihood of naturalising in Sweden than Finnish-speaking migrants (Saarela and Scott 2019). Swedish speakers are also more likely to naturalise in Sweden than other Nordic migrants, even though the differences are smaller. The direct benefits of naturalisation are modest and the same for all Nordic-born migrants in Sweden. As a result, these differences have been considered, at least in part, to be a mode of cultural expression and to reflect a part of the identity of Swedish speakers in Finland (Hedberg 2004; Saarela and Scott 2019).

The present study contributes to this stream of the literature by incorporating information on parents’ mother tongue, which allows us to identify individuals with mixed backgrounds. These groups are especially interesting because they are similar, save for the fact that they are raised in different ethno-linguistic communities. In contrast to much of the literature that has focused on migration and return migration from Sweden, we also incorporate migration to different destination regions; this allows us to assess the cultural proximity argument by comparing group-level differences in migration to Sweden and the Nordic countries versus Estonia and the rest of the world.

2.3 The role of culture in migration decisions

A growing body of literature has been assessing the role of culture in migration decisions. However, country-level analyses have revealed mixed findings. On the one hand, some studies have found that cultural distance inhibits migration between countries (Adserà 2015; Adserà and Pytliková 2015; Arif 2020; Belot and Ederveen 2012; Belot and Hatton 2012; Docquier, Tansel, and Turati 2020). These studies suggest that cultural distance carries greater weight for migration decisions than differences in aggregate unemployment rates in the origin and destination countries but less weight than the pull effect of income and social networks abroad (Adserà and Pytliková 2015). On the other hand, other studies have found that cultural distance is not strongly related to migration (Caragliu et al. 2013; Mayda 2010; Ruyssen and Salomone 2018). Recent work shows that perceptions of gender discrimination positively relate to the migration intentions of women (Ruyssen and Salomone 2018). However, the likelihood that migration intentions are realised depends on more traditional determinants such as household income and network effects. Caragliu et al. (2013) further note that estimates are sensitive to the measure of the cultural distance used.

Prior studies have used an array of indicators to measure culture, which build on one or multiple indicators of values and attitudes, religion and language. Some studies rely on measures indicating that countries share a common border or colonial past (Docquier, Peri, and Ruyssen 2014; Mayda 2010). Others assess country-level differences in values and attitudes. Common measures include Inglehart and Baker's (2000) measure of cultural distance, which builds on value items collected in the World Values Survey (Arif 2020; Belot and Ederveen 2012; Caragliu et al. 2013). These items assess traditional versus secular rational values and postmaterialist values. Belot and Ederveen (2012) also assess a measure of cultural distance provided by Hofstede (1991, 2001) that relies on information from the 1970s. Measures of religious distance build on distance indicators between the official religion in the origin and destination countries (Belot and Ederveen 2012) or on individuals' level of religiosity (Docquier, Tansel, and Turati 2020). Measures of linguistic distance tend to be more consistent across studies. Many indicators are based on the number of common nodes on the linguistic tree that are shared between one language and another (Adserà 2015; Adserà and Pytliková 2015; Belot and Ederveen 2012; Belot and Hatton 2012), while others rely on an indicator of sharing a common language (Docquier, Peri, and Ruyssen 2014; Mayda 2010). This variety of measures may at least partly explain the different findings.

Although very interesting for their external validity, a concern is that the internal validity of this body of research is low. There might be other aspects related to migration flows that are not captured by culture measured at the country level. Disentangling culture from other related concepts, such as institutional arrangements, is intricate. The more heterogeneous the study population of migrants, the more difficult it is to isolate the concept from other contributing factors. For instance, Belgium, Canada, Spain and Finland have sizeable linguistic minorities, which may be important to incorporate to understand migration patterns. Likewise, other countries such as Bosnia-Herzegovina have an array of religions, and it is debatable whether one is more dominant than the others. This makes it difficult to assess the extent to which cross-country analyses capture culture or other differences between countries.

A set of recent studies has addressed some of these concerns with individual-level analyses. An assessment of cultural selectivity during the Age of Mass Migration indicates that individualism was strongly related to the likelihood to emigrate to the US from Scandinavian countries (Knudsen 2019). Another study found that moral values and civicness are related to the likelihood of moving from the south to the north of Italy (Casari et al. 2018). In other words, individuals who are less likely to be free riders and more conscientious persons are more likely to move to the north of Italy. In Germany, a link between risk attitudes and migration based on ex-post characteristics of migrants and stayers has been found (Jaeger et al. 2010). A number of studies on Asia have also assessed the role of culture in internal migration. Indonesia comprises many different ethnic groups that differ in migration propensities (Auwalin 2020). However, the characteristics of the context of residence play a role for all groups. In China and Nepal, cultural distance has been found to be negatively related to internal migration (Li 2018; Fafchamps and Shilpi 2013).

We contribute to the international literature by studying the Finnish context, which provides us with detailed individual-level register data; these include complete longitudinal information on migration, return migration and mother tongue registration, allowing us to proxy the different strengths of attachment to Finland across the population and investigate the extent to which this proxy captures concepts related to cultural proximity. Although the context is specific, it can provide insights into migration in comparable settings, such as Belgium, Canada and Spain.

3 Data

We use Finnish individual-level register data that cover the total population from 1987 to 2020.[1] The data include information on the month and year of emigration and immigration from and to Finland, information on the ethno-linguistic affiliation (mother tongue) and sociodemographic characteristics of children and their parents and a unique identifier of each person's municipality of residence. We analyse Finnish-born individuals from 32 entire birth cohorts (people born 1970–2001).

We restrict our analysis to native-born persons who had not moved abroad before age 19 to analyse individuals' first independent migration abroad. We also restrict the analysis to individuals for whom we have information about both parents' mother tongues and whose individual and parents' mother tongues are either Finnish or Swedish. Most individuals with another mother tongue or whose parents have another mother tongue comprise second-generation immigrants; they constitute a small share of the Finnish population (2%) and are a diverse group. More details on the sample restrictions are provided in Table A1 in the

[1] All data access to Finnish register data, data preparation and analyses are performed within Statistics Finland's remote access system FIONA (contract number is TK-52-694-18).

Appendix. Our final sample comprises 1,873,974 individuals, of whom 86,667 individuals, or 5%, migrated by 2020. In addition, 52,931 individuals return migrated, or 61% of the migrants.

3.1 Measurement of ethno-linguistic background

To measure ethno-linguistic background, we use information on the mother tongue of individuals and their mothers and fathers. This information is recorded recently after birth, and the vast majority of individuals do not change it over their life course (less than 0.5% do so). We exploit this information to define four main groups: 1) Finnish speakers with a uniform background (two Finnish-speaking parents), 2) Swedish speakers with a uniform background (two Swedish-speaking parents), 3) Swedish speakers with a mixed background (one Swedish and one Finnish-speaking parent) and 4) Finnish speakers with a mixed background (one Swedish and one Finnish-speaking parent).

3.2 Descriptive statistics

Table 1 provides descriptive statistics. Swedish speakers and individuals with mixed backgrounds are overrepresented among migrants. Parental socioeconomic characteristics are measured when mothers and fathers are 35 years of age. For mothers and/or fathers who are not observed in the register at age 35, we rely on information collected in a random year between ages 25 and 65. On average, the parents of migrants have higher education, are more often employed and have higher income than parents of nonmigrants. Women and individuals with secondary education and the matriculation exam (a prerequisite for entering university) are more likely to migrate.

For return migration, Table 2 shows that the differences between returnees and nonreturnees are smaller but tend to go in the opposite direction. Finnish speakers with uniform backgrounds are overrepresented among returnees. On average, the parents of returnees have less education and lower incomes than those of nonreturnees. Men and less-educated individuals are more likely to return.

Table 1. Average measures for migrants and nonmigrants

	Migrants	Nonmigrants	Total
Background			
Finnish uniform background	0.80	0.93	0.93
Swedish uniform background	0.13	0.03	0.04
Swedish mixed background	0.04	0.02	0.02
Finnish mixed background	0.02	0.02	0.02
Parental characteristics			
Parental education			
Both parents have primary educ.	0.09	0.13	0.12
One parent has secondary educ.	0.16	0.22	0.21
Both parents have secondary educ.	0.15	0.22	0.21
One parent has tertiary educ.	0.30	0.26	0.26
Both parents have tertiary educ.	0.31	0.18	0.19
Parental employment status			
Neither parent is employed	0.03	0.04	0.04
One parent is employed	0.21	0.26	0.25
Both parents are employed	0.76	0.70	0.70
Parental income in quartiles			
Quartile 1 (bottom)	0.21	0.25	0.25
Quartile 2	0.23	0.25	0.25
Quartile 3	0.27	0.25	0.25
Quartile 4 (top)	0.29	0.25	0.25
Single parent	0.11	0.12	0.12
Individual characteristics			
Female	0.60	0.48	0.49
Educational attainment			
Primary educ.	0.27	0.32	0.32
Secondary educ. with*out* matric. exam	0.11	0.26	0.25
Secondary educ. with matric. exam	0.61	0.42	0.43
Birth cohort			
1970-1974	0.22	0.15	0.15
1975-1979	0.25	0.16	0.17
1980-1984	0.20	0.16	0.17
1985-1989	0.16	0.16	0.16
1990-1994	0.12	0.16	0.16
1995-2001	0.06	0.21	0.20

Continued

Table 1. Continued

	Migrants	Nonmigrants	Total
Region of residence			
Greater Helsinki	0.21	0.14	0.15
Uusimaa (Nyland)	0.10	0.09	0.09
Southwest Finland (Egentliga Finland)	0.08	0.08	0.08
Satakunta	0.03	0.05	0.05
Kanta-Häme (Egentliga Tavastland)	0.02	0.03	0.03
Pirkanmaa (Birkaland)	0.07	0.09	0.08
Päijät-Häme (Päijänne-Tavastland)	0.03	0.04	0.04
Kymenlaakso (Kymmenedalen)	0.03	0.04	0.04
South Karelia (Södra Karelen)	0.02	0.03	0.03
South Savo (Södra Savolax)	0.02	0.03	0.03
North Savo (Norra Savolax)	0.04	0.05	0.05
North Karelia (Norra Karelen)	0.03	0.04	0.03
Central Finland (Mellersta Finland)	0.04	0.05	0.05
South Ostrobothnia (Södra Österbotten)	0.02	0.04	0.04
Ostrobothnia (Österbotten)	0.08	0.03	0.03
Central Ostrobothnia (Mellersta Österbotten)	0.01	0.02	0.02
North Ostrobothnia (Norra Österbotten)	0.07	0.09	0.09
Kainuu (Kajanaland)	0.01	0.02	0.02
Lapland (Lappland)	0.04	0.04	0.04
Åland Islands	0.03	0.01	0.01
Number of individuals	86,667	1,787,307	1,873,974
	5%	95%	

Note. Proportion within each group reported. Regions of residence are provided in English with their Swedish names in parentheses when they differ. Greater Helsinki includes Helsinki, Espoo, Vantaa and Kaunianen.

Table 2. Average measures for returnees and nonreturnees

	Returnees	Nonreturnees	Total
Background			
Finnish uniform background	0.82	0.78	0.80
Swedish uniform background	0.12	0.15	0.13
Swedish mixed background	0.04	0.05	0.04
Finnish mixed background	0.02	0.02	0.02
Parental characteristics			
Parental education			
Both parents have primary educ.	0.10	0.08	0.09
One parent has secondary educ.	0.16	0.15	0.16
Both parents have secondary educ.	0.15	0.15	0.15
One parent has tertiary educ.	0.29	0.30	0.30
Both parents have tertiary educ.	0.30	0.32	0.31
Parental employment status			
Neither parent is employed	0.03	0.02	0.03
One parent is employed	0.21	0.21	0.21
Both parents are employed	0.76	0.76	0.76
Parental income in quartiles			
Quartile 1 (bottom)	0.21	0.20	0.21
Quartile 2	0.24	0.22	0.23
Quartile 3	0.27	0.27	0.27
Quartile 4 (top)	0.28	0.31	0.29
Single parent	0.11	0.11	0.11
Individual characteristics			
Female	0.57	0.66	0.60
Educational attainment			
Primary educ.	0.29	0.26	0.27
Secondary educ. with*out* matric. exam	0.12	0.10	0.11
Secondary educ. with matric. exam	0.60	0.65	0.61
Birth cohort			
1970-1974	0.24	0.19	0.22
1975-1979	0.26	0.22	0.25
1980-1984	0.20	0.19	0.20
1985-1989	0.15	0.17	0.16
1990-1994	0.11	0.15	0.12
1995-2001	0.04	0.08	0.06

Continued

Table 2. Continued

	Returnees	Nonreturnees	Total
Region of residence			
Greater Helsinki	0.20	0.23	0.21
Uusimaa (Nyland)	0.10	0.10	0.10
Southwest Finland (Egentliga Finland)	0.08	0.08	0.08
Satakunta	0.03	0.03	0.03
Kanta-Häme (Egentliga Tavastland)	0.02	0.02	0.02
Pirkanmaa (Birkaland)	0.07	0.07	0.07
Päijät-Häme (Päijänne-Tavastland)	0.03	0.03	0.03
Kymenlaakso (Kymmenedalen)	0.03	0.03	0.03
South Karelia (Södra Karelen)	0.02	0.02	0.02
South Savo (Södra Savolax)	0.02	0.02	0.02
North Savo (Norra Savolax)	0.04	0.04	0.04
North Karelia (Norra Karelen)	0.03	0.02	0.03
Central Finland (Mellersta Finland)	0.04	0.04	0.04
South Ostrobothnia (Södra Österbotten)	0.02	0.02	0.02
Ostrobothnia (Österbotten)	0.07	0.09	0.08
Central Ostrobothnia (Mellersta Österbotten)	0.02	0.01	0.01
North Ostrobothnia (Norra Österbotten)	0.08	0.06	0.07
Kainuu (Kajanaland)	0.01	0.01	0.01
Lapland (Lappland)	0.05	0.04	0.04
Åland Islands	0.04	0.03	0.03
Number of individuals	52,931	33,736	86,667
	61%	39%	

Note. Proportion within each group reported. Regions of residence are provided in English with their Swedish names in parentheses when they differ. Greater Helsinki includes Helsinki, Espoo, Vantaa and Kaunianen. We follow migrants until their return migration, death or the end of the observation period in 2020, whichever comes first. If individuals have returned during our observation period, they are classified as returnees and if they have not they are classified as non-returnees. We make no restrictions that returnees must have returned within a certain number of years.

Table 3 shows that nearly 30% of migrants from Finland move to Sweden. About 14% of migrants move to other Nordic countries (Norway, Denmark and Iceland). More than 50% of migrants move to the rest of the world and about 2% to Estonia. The rest of the world comprises a large share of destinations and is therefore sizeable. Still, among Swedish speakers with a uniform background, a relatively small share of individuals moves to the rest of the world, and most move to Sweden. Among the returnees, the patterns are strikingly similar. About 30% of return migrations are from Sweden, 18% from the other Nordic countries, 50% from the rest of the world and 2% from Estonia.

Table 3. Descriptive statistics of migration destinations by ethno-linguistic background

	Emigration					
	Finnish uniform	Swedish uniform	Swedish mixed	Finnish mixed	Total (N)	Total (%)
Destinations						
Sweden	14,405	7,889	1,863	665	24,822	29%
other Nordics	9,547	1,647	616	292	12,102	14%
rest of the world	43,762	1,825	1,239	1,010	47,836	55%
Estonia	1,772	42	46	4700%	1,907	2%
	Return migration					
	Finnish uniform	Swedish uniform	Swedish mixed	Finnish mixed	Total (N)	Total (%)
Destinations						
Sweden	10,226	4,247	1,055	414	15,942	30%
other Nordics	7,516	1,181	417	216	9,330	18%
rest of the world	24,365	875	612	536	26,388	50%
Estonia	1,180	29	28	34	1,271	2%

Note. Number of individuals reported in all columns apart from the last. The last column reports percentages.

4 Empirical strategy

We use an event history setup to assess differences in the likelihood of migrating by ethno-linguistic background. The population at risk includes the total Finnish population born in Finland in the birth cohorts 1970–2001 and observed in Finland at age 19 (N=1,873,974). Individuals enter the risk set in June of the calendar year of their nineteenth birthday. That is, individuals born in 1990 enter the risk set in June 2009, irrespective of whether they are born in April or December. This setup allows us to focus on independent migration decisions and analyse migration-prone ages while simultaneously assessing educational attainment in early adulthood. In particular, we measure whether individuals passed the matriculation exam, which is generally taken in May of the nineteenth birthday. The matriculation exam is a prerequisite for entering university studies and constitutes an important indicator of educational outcomes during early adulthood. Individuals are then followed until their first migration, death or the end of the observation period in 2020, whichever comes first. We specify the following model:

$$\lambda_i(t) = \lambda_0(t) \times \exp\{\beta backg_i\}, \tag{1}$$

where $\lambda_i(t)$ refers to the risk of migrating at time t. In this model, t refers to the time since June of the calendar year of their nineteenth birthday. The term $\lambda_0(t)$ indicates the baseline hazard, and $\lambda_o(t) = \lambda_j$ for t in each interval $[\tau_{J-1}, \tau_J)$. We partition duration into intervals with cut-points at the change of each calendar year for the first 15 years. The term $backg_i$ is a set of dummy variables that indicate whether individual i is a Swedish speaker with a uniform background, a Swedish speaker with a mixed background or a Finnish speaker with a mixed background. Finnish speakers with uniform backgrounds represent the omitted reference category. β represents the corresponding coefficient vector, which gives the difference in the risk of migrating between each group and the reference category (Finnish speakers with a uniform background).

4.1 Parental and individual socioeconomic characteristics

Having established the overall gradient, we assess three potential mechanisms underlying the patterns. That is, we build on the detailed information provided in Finnish register data to assess whether ethno-linguistic background captures concepts related to cultural proximity.

First, we analyse the extent to which individuals' socioeconomic backgrounds shape the differences in migration behaviour across ethno-linguistic backgrounds. To address this point, we assess whether the gradient persists when we control for parental and individual socioeconomic characteristics in the following model:

$$\lambda_i(t) = \lambda_0(t) \times \exp\{\beta backg_i + \zeta par_i + \eta ind_i\}, \qquad (2)$$

where par_i refers to parental characteristics and ind_i to individual characteristics. Otherwise, the terms follow Equation 1. Parental socioeconomic characteristics refer to education, employment status, income and single parenthood. Parental education differentiates whether both parents have primary education, one parent has secondary education, both parents have secondary education, one parent has tertiary education or both parents have tertiary education. Parental employment status identifies whether neither parent is employed, one parent is employed or both parents are employed.[2] To measure parental income, we take the sum of the

[2] For individuals who grow up in a single parent household or who have missing information on the educational attainment or employment status of one parent, available information from the second parent is used. In this instance, a parent with primary education is classified as both parents having primary education.

mother's and father's income and rank it into quartiles. Single parenthood refers to a dummy variable and is measured when parents are age 35.

Individual socioeconomic characteristics refer to gender, educational attainment at age 19, birth cohort and region of residence at age 17 (when most still live in the parental home). These variables provide information before entering into the risk set. Educational attainment differentiates between primary education, secondary education with*out* a matriculation exam and secondary education with a matriculation exam. Secondary education with*out* a matriculation exam largely comprises vocational tracks in secondary education.

4.2 Distinctions by region of destination

Second, we assess whether the gradient varies across destinations. To the extent that high migration rates among Swedish speakers are driven by cultural proximity to Sweden, we expect that the gradient is accentuated for the likelihood of migrating to Sweden or the other Nordic countries where Scandinavian languages similar to Swedish are spoken. In contrast, Estonia is closer to Finland, leading us to expect a reverse gradient for migration to Estonia. In other words, Finnish speakers are expected to have a higher likelihood of moving to Estonia than Swedish speakers. Migration to the rest of the world includes a broad range of destinations, and no group is expected to have a strong affinity for these diverse destinations. To address this point, we use information on the country of destination to differentiate between migration to Sweden, the other Nordic countries, the rest of the world and Estonia. The event of interest is migration to a specific destination, and individuals are right censored at the time of migration to another destination. Otherwise, the setup is the same as that used for migration to any destination.

Specifically, we estimate Equation 2, where $\lambda_i(t)$ refers to the risk of migrating to Sweden, the other Nordic countries, the rest of the world or Estonia at time t. For simplicity, we assume that the risk of moving to different destinations is independent.

4.3 Role of the municipality of residence

Third, we study the role of the municipality of residence. We have argued that mother tongue registration impacts the community in which individuals grow up because Swedish and Finnish speakers attend different schools and are raised in different ethno-linguistic settings. Here, we investigate whether the composition of the municipality of residence is related to migration decisions independently of mother tongue registration. We incorporate information from the local context using information on the municipality composition at age 17. In these analyses, we focus on Swedish and Finnish speakers with a mixed background to assess differences in the migration behaviour of individuals for whom mother tongue registration is somewhat arbitrary. These two groups are similarly distributed across different municipalities (see Figure A1 in the Appendix). We use the following model:

$$\lambda_i(t) = \lambda_0(t) \times \exp\{\sigma mun_i + \beta backg_i + \gamma mun_i \times backg_i + \ \zeta par_i + \eta ind_i\}, (3)$$

where the term mun_i is a dummy variable that indicates whether individual i lives in a municipality with a high share of Swedish speakers at age 17 (50% or higher). Living in a municipality with a low share of Swedish speakers is the omitted reference category. σ represents the corresponding coefficient, which provides the difference in the risk of migrating between individuals living in areas with high versus low shares of Swedish speakers. The term $backg_i$ is a dummy variable that indicates whether individual i is a Swedish speaker with a mixed background. Finnish speakers with mixed backgrounds represent the omitted reference category. β represents the corresponding coefficient, which gives the difference in the risk of migrating between the two groups. The equation further includes an interaction term between $mun_i \times backg_i$. This allows for the association between the composition of the municipality and migration to vary between Swedish and Finnish speakers. The other terms follow Equation 2. We estimate models for both migration to all destinations and differentiate between the different destination regions.

4.4 Return migration

In the latter part of the analysis, we assess return migration. The interpretation of these analyses is informative for the longer-term implications of cultural selectivity. On the one hand, return migration may accentuate population changes if groups that were the most likely to migrate are also the least likely to return. On the other hand, population changes may be short-lived if the groups who were the most likely to migrate are also the most likely to return.

We follow a similar setup as for migration. We begin by establishing the overall gradient. The population at risk includes all individuals who made a first migration (N=86,667). Individuals enter the risk set when they migrate and are followed until their return migration, death or the end of the observation period in 2020, whichever comes first. We then assess the same set of mechanisms as above. First, we include controls for parental and individual socioeconomic characteristics. Second, we differentiate between return migration from different regions (Sweden, the other Nordic countries, the rest of the world and Estonia). Third, we focus on individuals with a mixed background, assessing the role of the municipality of residence in return migration.

5 Empirical findings

Figure 1 provides the hazard ratios from models estimating the likelihood of migrating across ethno-linguistic groups. All groups are compared with Finnish speakers with a uniform background, whose likelihood of migrating is indicated by the vertical line. Panel A provides raw differences based on Equation 1. Panel B provides estimates based on Equation 2, where controls for parental and, subsequently, individual socioeconomic characteristics are introduced. Corresponding estimates and standard errors are provided in Table A2 in the Appendix.

Figure 1A reveals a gradient in the likelihood of migrating by ethno-linguistic background. Swedish speakers with uniform backgrounds are the most likely group to migrate. Their likelihood of migrating is nearly five times as high as that of Finnish speakers with a uniform background (HR=4.5). Swedish speakers with a mixed background have a likelihood of migrating that is three times as high as that of Finnish speakers with a uniform background (HR=3.1), and Finnish speakers with a mixed background have a likelihood of migrating that is about 50% higher than that of Finnish speakers with a uniform background (HR=1.5).

The gradient indicates that the migration behaviour of individuals with a mixed background lies in between that of individuals with a uniform background. That is, Finnish (Swedish) speakers with a mixed background are more (less) likely to migrate than Finnish (Swedish) speakers with a uniform background. We also observe that Swedish speakers with a mixed background are more likely to migrate than Finnish speakers with a mixed background. This suggests that, even among mixed groups, which, in many respects, are similar and for whom mother tongue registration is somewhat arbitrary, factors captured by mother tongue registration play a role in migration behaviour. In the following, we analyse three potential mechanisms underlying these patterns to assess the extent to which ethno-linguistic background captures concepts related to cultural proximity.

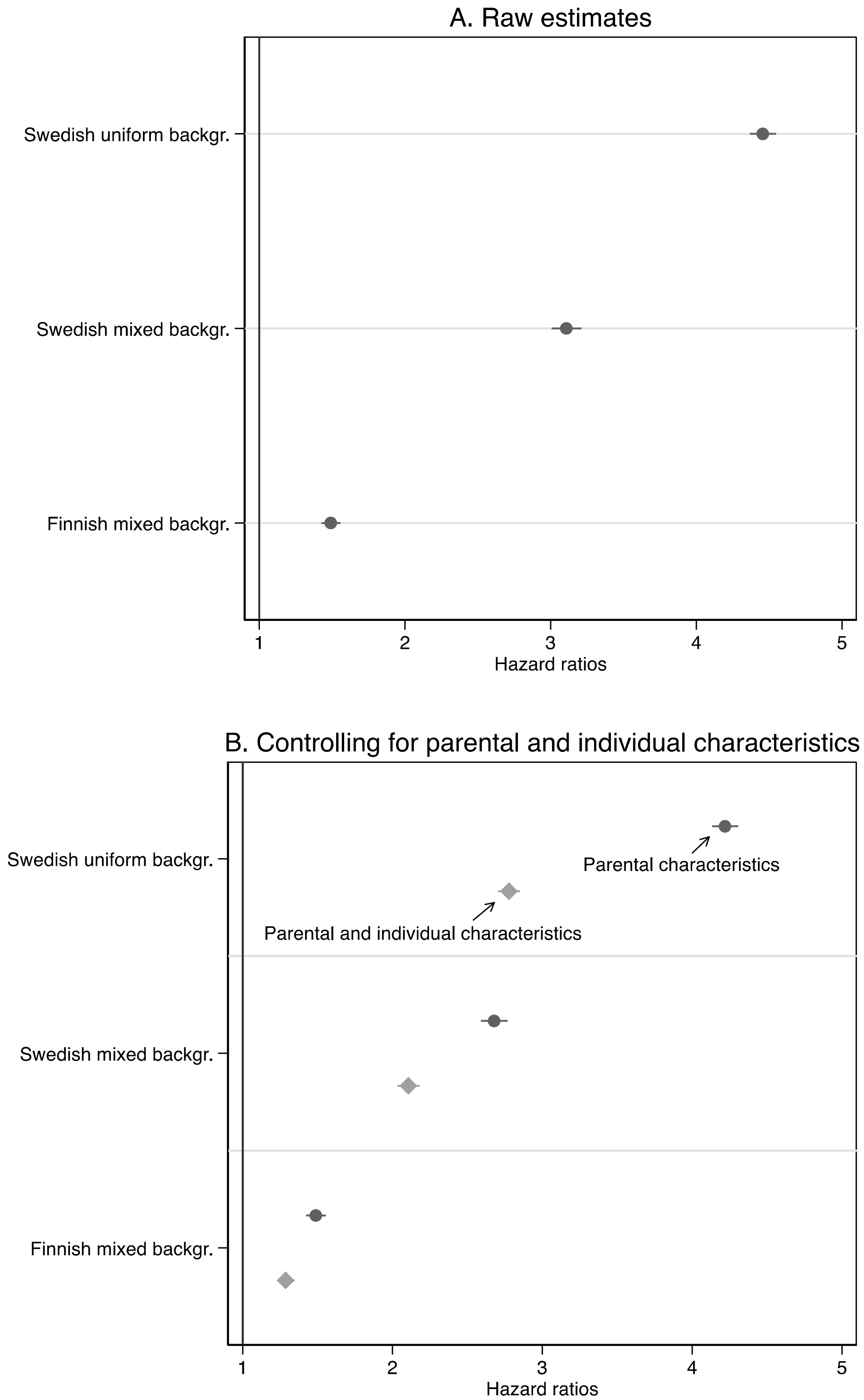

Figure 1. Hazard ratios from piecewise constant exponential models on the likelihood of migrating to all destinations. The vertical line indicates the hazard ratio of Finnish speakers with a uniform background, which is the reference group. Panel A provides raw estimates based on Equation 1. Panel B provides estimates based on Equation 2, where controls for parental socioeconomic characteristics (education, employment status, income and single parenthood) and, subsequently, individual socioeconomic characteristics (gender, educational attainment, birth cohort and region of residence) are introduced. The full set of estimates are provided in Table A2 in the Appendix.

5.1 Parental and individual socioeconomic characteristics

Figure 1B shows that the gradient is attenuated but persists when we control for parental and individual socioeconomic characteristics. Among Swedish speakers with a uniform background, controlling for individual (more so than parental) socioeconomic characteristics reduces the raw difference in the likelihood of migrating. Still, in the final model (that controls for both parental and individual socioeconomic characteristics), Swedish speakers with a uniform background continue to be about three times as likely to migrate as Finnish speakers with a uniform background (as compared with five times in panel A). For Swedish and Finnish speakers with mixed backgrounds, both parental and individual characteristics play a role. In the final model, Swedish speakers with a mixed background are about two times as likely to migrate as Finnish speakers with a uniform background (as compared with three times in panel A), and Finnish speakers with a mixed background are about 30% more likely to migrate than Finnish speakers with a uniform background (as compared with 50% in panel A).

The magnitude of the differences between ethno-linguistic groups is sizeable when compared with the hazard ratios of the control variables (see the forest plot in Figure A2 in the Appendix). Ethno-linguistic background carries greater weight for migration decisions than parental employment, income and individuals' education level. In contrast, parental education and living on the Åland Islands, which are close to Sweden, are similarly important as ethno-linguistic backgrounds. Specifically, the likelihood of migrating is about two times higher among individuals with two tertiary educated parents when compared with individuals with two compulsory educated parents. Individuals from the Åland Islands are nearly five times as likely to migrate as individuals from the capital region of Helsinki.

5.2 Distinctions by region of destination

Based on Equation 2, Figure 2 provides results from competing risk models on the likelihood of migrating to Sweden, the other Nordic countries, the rest of the world and Estonia. Corresponding estimates and standard errors are provided in Table A3 in the Appendix. We observe an accentuated gradient for the likelihood of migrating to Sweden and other Nordic countries. Swedish speakers with a uniform background are about seven times as likely to migrate to Sweden as Finnish speakers with a uniform background. Swedish speakers with a mixed background are nearly five times as likely to migrate to Sweden, and Finnish speakers with a mixed background are twice as likely to do so as Finnish speakers with a uniform background.

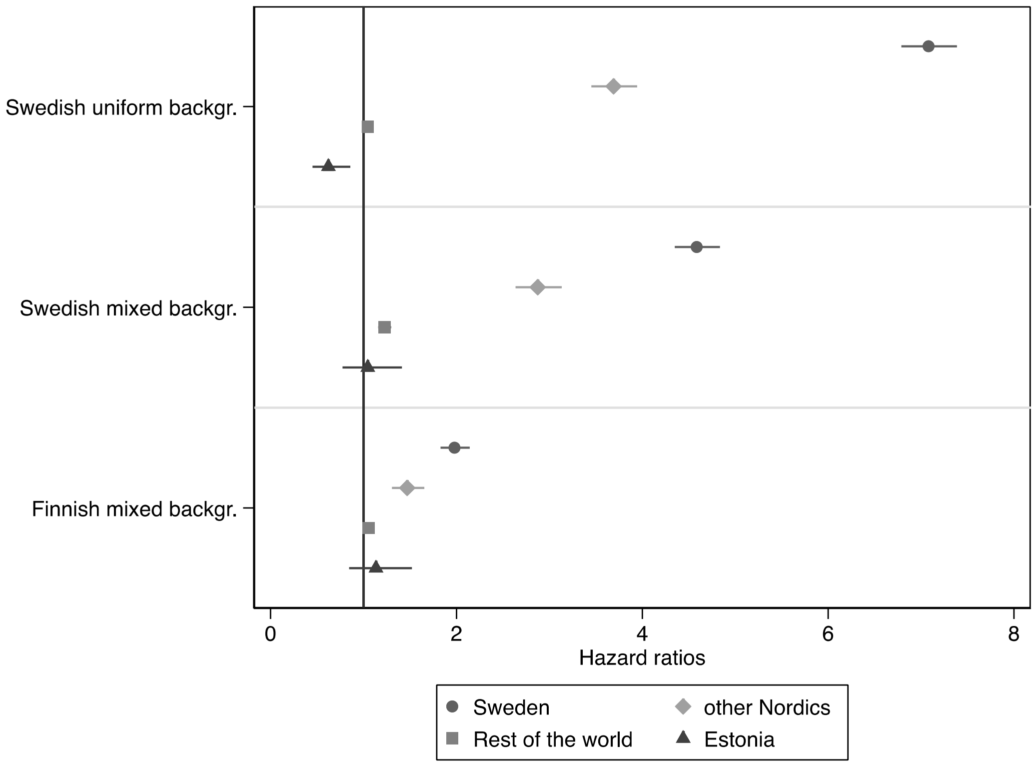

Figure 2. Hazard ratios from competing risk models on the likelihood of migrating to Sweden, the other Nordic countries, the rest of the world and Estonia, based on Equation 2. The vertical line indicates the hazard ratio of Finnish speakers with a uniform background, which is the reference group. Models control for parental socioeconomic characteristics (education, employment status, income and single parenthood) and individual socioeconomic characteristics (gender, educational attainment, birth cohort and region of residence). The full set of estimates are provided in Table A3 in the Appendix.

For moves to the other Nordic countries, the gradient is less pronounced, but it is also notable. Swedish speakers with a uniform background are roughly four times as likely to migrate to other Nordic countries as Finnish speakers with a uniform background. Swedish speakers with a mixed background are about three times as likely to migrate to other Nordic countries, and Finnish speakers with a mixed background are about 50% more likely to do so than Finnish speakers with a uniform background.

For moves to the rest of the world, we observe no gradient by ethno-linguistic background. The likelihood of migrating to the rest of the world is similar across all groups, but Swedish speakers with a mixed background have a somewhat higher likelihood of migrating to the rest of the world than the other groups. Still, the hazard ratio is small when compared with the estimates for Sweden and the other Nordic countries.

Regarding the likelihood of migrating to Estonia, Swedish speakers with a uniform background are nearly 40% less likely to migrate to Estonia than Finnish speakers with a uniform background. Among Swedish and Finnish speakers with mixed backgrounds, the likelihood of migrating to Estonia does not significantly differ from that of Finnish speakers with uniform backgrounds. This is likely because of the small number of moves in this direction.

5.3 Role of the municipality of residence

We further assess the differences in migration risks between Swedish and Finnish speakers with a mixed background. We provide estimates from models on the likelihood of migrating to all destinations, Sweden, the other Nordic countries, the rest of the world and Estonia based on Equation 3. The models include an interaction between the share of Swedish speakers in the municipality at age 17 and mother tongue. Table 4 presents the group-specific hazard ratios based on this model. The interaction coefficients and full set of hazard ratios are provided in Table A4 in the Appendix.

Table 4. Hazard ratios from models estimating the likelihood of migrating among individuals with a mixed background living in municipalities with low or high shares of Swedish speakers

	Finnish mixed background		Swedish mixed background	
	low share Swedish speakers	high share Swedish speakers	low share Swedish speakers	high share Swedish speakers
Destinations				
All destinations	1	1.317***	1.638***	1.858***
		(0.093)	(0.053)	(0.097)
Sweden	1	1.819***	2.248***	3.069***
		(0.190)	(0.124)	(0.242)
other Nordics	1	1.683**	1.850***	2.390***
		(0.291)	(0.153)	(0.313)
rest of the world	1	0.931	1.262***	0.915
		(0.113)	(0.059)	(0.088)
Estonia	1	1.492	1.131	0.534
		(0.858)	(0.255)	(0.317)

Note. The hazard ratios and standard errors are obtained using the lincom post-estimation command in Stata. Models control for parental socioeconomic characteristics (education, employment status, income and single parenthood) and individual socioeconomic characteristics (gender, educational attainment, birth cohort and region of residence). The full set of estimates are provided in Table A4 in the Appendix. Clustered standard errors at the individual level in parentheses. * p < 0.05; ** p < 0.01; *** p < 0.001 (two-tailed).

Among Finnish speakers with a mixed background, individuals who live in municipalities with a high share of Swedish speakers are about 30% more likely to migrate than their counterparts who live in municipalities with a low share of Swedish speakers (see Table 4). However, their higher migration propensity is largely explained by migration to Sweden and the other Nordic countries. Specifically, Finnish speakers in municipalities with high shares of Swedish speakers are 80% more likely to migrate to Sweden and 70% more likely to migrate to other Nordic countries than Finnish speakers in municipalities with low shares of Swedish speakers.

Swedish speakers in municipalities with low and high shares of Swedish speakers are also more likely to migrate to Sweden and the other Nordic countries when compared with Finnish speakers in municipalities with low shares. Swedish speakers in municipalities with high shares are particularly likely to migrate to Sweden; they are three times as likely to do so as Finnish speakers in areas with low shares. Still, Swedish speakers in areas with low shares are also two times more likely to migrate than Finnish speakers in similar areas.

The interpretation of these results is twofold. Differences in migration across ethno-linguistic groups vary both by residential context and individual factors. On the one hand, we observe a strong impact of living in predominantly Swedish-speaking municipalities among Finnish speakers with a mixed background who attend Finnish schools. That is, the residential context relates to migration independent of mother tongue registration. On the other hand, we continue to observe differences between Swedish and Finnish speakers with a mixed background who live in similar municipalities. This suggests that, among individuals for whom mother tongue registration is somewhat arbitrary, ethno-linguistic affiliation relates to migration behaviour, even when they live in similar areas.

5.4 Return migration

In the last step, we assess the extent to which the above gradients are reversed or remain the same for return migration. Figure 3 provides hazard ratios for return migration, revealing a reverse gradient by ethno-linguistic background (see panel A). That is, Finnish speakers with a uniform background are the most likely group to return migrate, followed by Finnish mixed background and Swedish mixed background, respectively. Swedish speakers with a uniform background are the least likely group to return migrate. Figure 3B indicates that the estimates do not substantially change when we control for socioeconomic characteristics, likely because they are measured prior to the first migration, hence providing less precise information on the

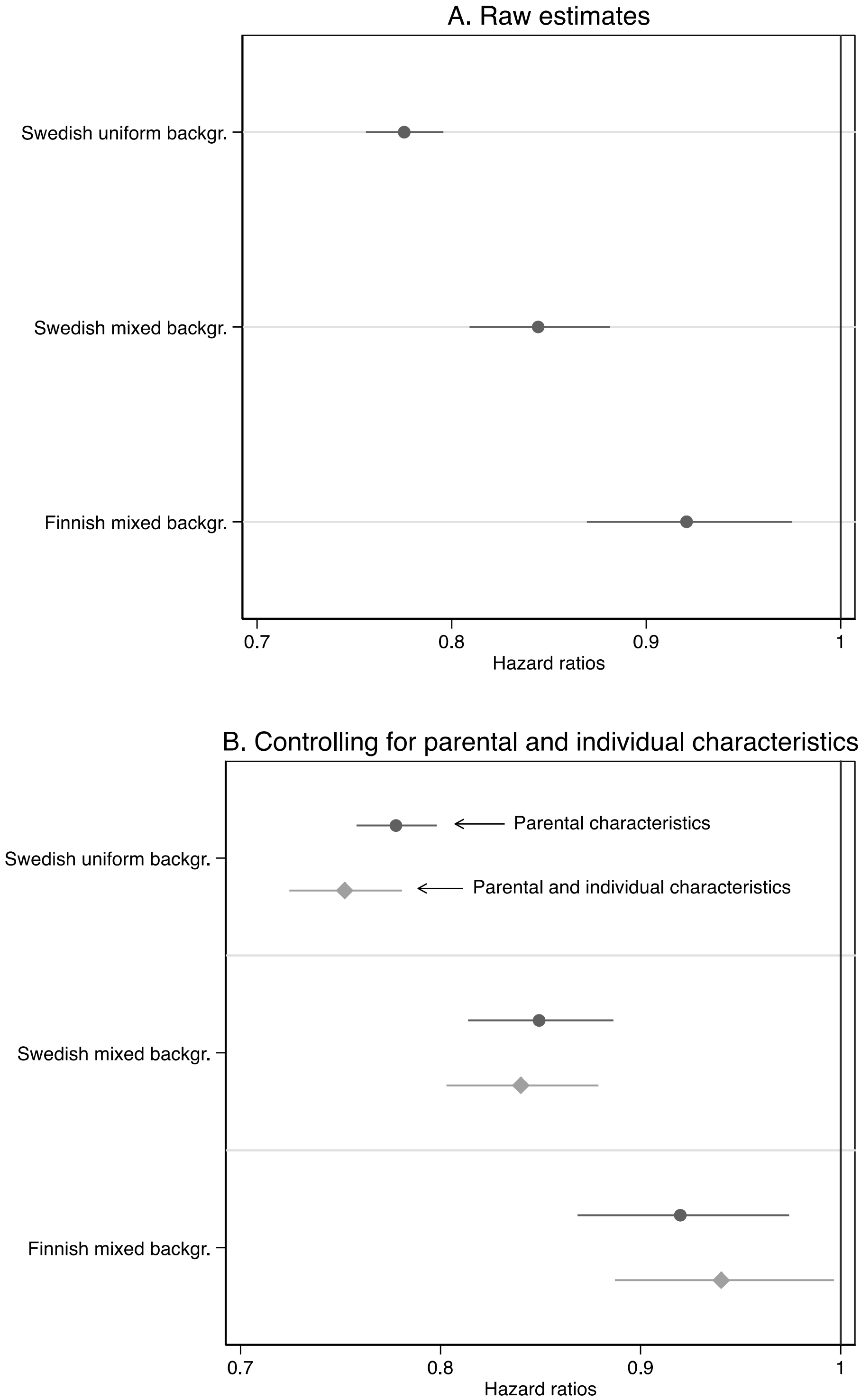

Figure 3. Hazard ratios from piecewise constant exponential models on the likelihood of return migrating from all countries. The vertical line indicates the hazard ratio of Finnish speakers with a uniform background, which is the reference group. Panel A provides raw estimates based on Equation 1. Panel B provides estimates based on Equation 2, where controls for parental socioeconomic characteristics (education, employment status, income and single parenthood) and, subsequently, individual socioeconomic characteristics (gender, educational attainment, birth cohort and region of residence) are introduced. The full set of estimates are provided in Table A5 in the Appendix.

factors influencing return decisions. We also observe that the magnitude of differences between ethno-linguistic groups is sizeable and larger than that of parental socioeconomic characteristics. The effects are comparable to differences in individuals' education level, gender and cohort (see Figure A3 in the Appendix).

Figure 4 provides the results from competing risk models on the likelihood of return migrating from different regions. The gradient is strongest for Sweden, followed by the other Nordic countries. We also find that Swedish speakers with uniform and mixed backgrounds are less likely to return from the rest of the world than Finnish speakers with a uniform background. Table 5 shows that, among Finnish speakers with a mixed background, individuals in municipalities with high shares of Swedish speakers are less likely to return migrate than those in municipalities with low shares of Swedish speakers. Swedish speakers in municipalities with low shares of Swedish speakers are also less likely to return migrate from all destinations than Finnish speakers in municipalities with low shares of Swedish speakers. The other estimates are nonsignificant. However, Swedish speakers in municipalities with high shares are significantly less likely to return from Sweden and other Nordic countries than Finnish speakers in municipalities with low shares. This suggests that the residential context (in Finland) relates to return migration independent of mother tongue registration (indicated by differences in return migration propensities between the two groups of Finnish speakers). Still, Swedish speakers are somewhat less likely to return than Finnish speakers who live in similar areas. The estimates for Estonia are excluded from these analyses due to the small number of observations (see Tables A6 and A7 in the Appendix for the results).

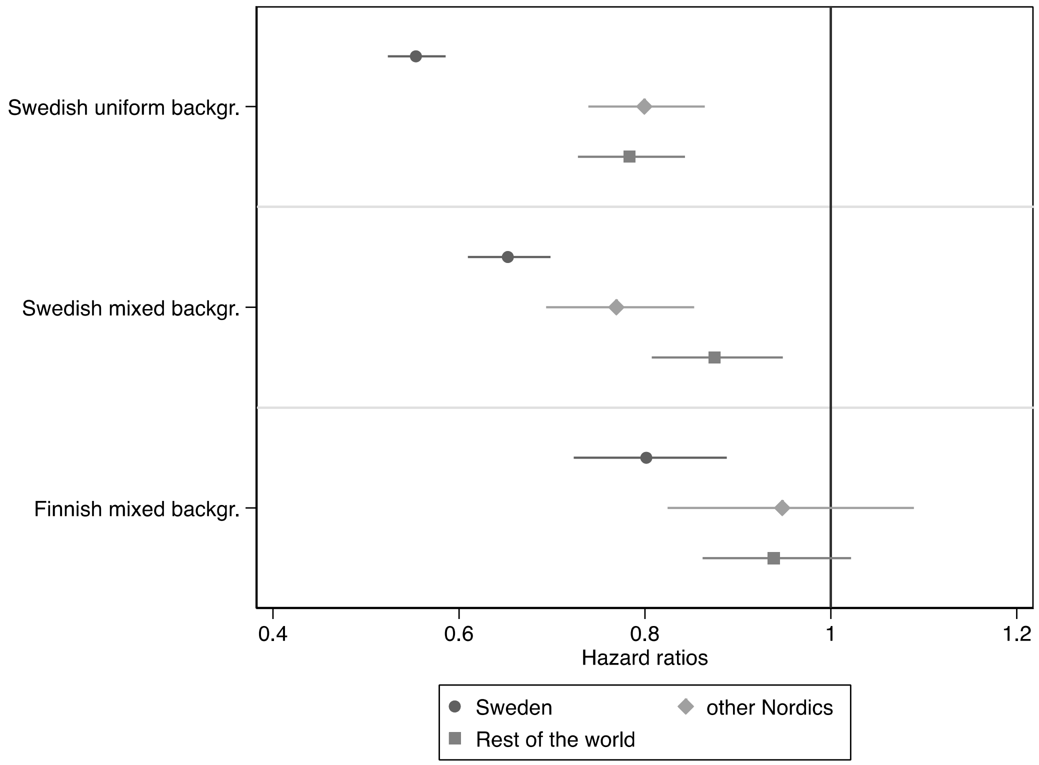

Figure 4. Hazard ratios from competing risk models on the likelihood of return migrating from Sweden, the other Nordic countries, the rest of the world and Estonia, based on Equation 2. The vertical line indicates the hazard ratio of Finnish speakers with a uniform background, which is the reference group. Models control for parental socioeconomic characteristics (education, employment status, income and single parenthood) and individual socioeconomic characteristics (gender, educational attainment, birth cohort and region of residence). The full set of estimates are provided in Table A6 in the Appendix.

Table 5. Hazard ratios from models estimating the likelihood of return migrating among individuals with a mixed background living in municipalities with low or high shares of Swedish speakers

	Finnish mixed background		Swedish mixed background	
	low share Swedish speakers	high share Swedish speakers	low share Swedish speakers	high share Swedish speakers
Destinations				
All destinations	1	0.747**	0.907*	0.770***
		(0.075)	(0.037)	(0.050)
Sweden	1	0.745*	0.875	0.649***
		(0.106)	(0.062)	(0.061)
other Nordics	1	0.616*	0.770**	0.727*
		(0.144)	(0.071)	(0.097)
rest of the world	1	0.685	0.910	0.872
		(0.135)	(0.059)	(0.119)

Note. The hazard ratios and standard errors are based on the lincom post-estimation command in Stata. Models control for parental socioeconomic characteristics (education, employment status, income and single parenthood) and individual socioeconomic characteristics (gender, educational attainment, birth cohort and region of residence). The full set of estimates are provided in Table A7 in the Appendix. Clustered standard errors at the individual level in parentheses. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ (two-tailed).

6 Conclusion and discussion

The present study has assessed the role of ethno-linguistic background in migration behaviour. Exploiting Finnish register data, we identify native persons' ethno-linguistic affiliation and background using information on individual and parental mother tongues. These measures allow us to proxy individuals' attachment to Finland and to Finnish culture, which can be assumed to differ across Finnish and Swedish speakers, as well as across those with a uniform and mixed background (Gans 1979; Hedberg 2004; McRae 1999; Saarela and Finnäs 2018). Individuals with a mixed background are especially interesting because they are, in many respects, similar but attend different schools and are raised in different ethno-linguistic communities (Finnäs and O'Leary 2003; Saarela and Finnäs 2016).

Our results reveal an ethno-linguistic gradient in the likelihood of migrating. Swedish speakers with a uniform background are the most likely to migrate, followed by Swedish mixed background and Finnish mixed background. Finnish speakers with uniform backgrounds are the least likely group to migrate. For return migration, we observe the opposite pattern, which is consistent with the interpretation that cultural selectivity is accentuated by return migration. Our finding that Swedish speakers are more likely to emigrate and less likely to return migrate than Finnish speakers goes in line with previous findings (Hedberg and Kepsu 2003; Rooth and Saarela 2007; Saarela and Scott 2017; Weber and Saarela 2019). However, the novelty of our results lies in the fact that we also observe differences in the migration behaviour of the two mixed groups. That is, Swedish speakers with a mixed background are more likely to emigrate and are less likely to return migrate than Finnish speakers with a mixed background. This indicates that factors captured by mother tongue registration relate to migration behaviour, even among groups that are similar in other respects.

Having established the overall gradient, the remainder of the analyses investigate the extent to which ethno-linguistic background captures concepts related to cultural proximity. First, we assess whether the gradient is driven by differences in individuals' socioeconomic backgrounds. The results reveal that the gradient is attenuated, but it persists when we control for parental and individual socioeconomic characteristics. We further find that ethno-linguistic background carries greater weight for migration decisions than parental employment, income and individuals' education level, while parental education is similarly important as ethno-linguistic background. For return migration, we observe that ethno-linguistic background is more important than parental characteristics, while individuals' education level, gender and cohort are similarly important as their ethno-linguistic background. This indicates that differences across ethno-linguistic groups are not accounted for by the differences in socioeconomic background.

Second, we assess whether the high migration rates among Swedish speakers are driven by migration to Sweden and the other Nordic countries, here following the argument that this group has a strong attachment to Sweden and Swedish culture (Saarela and Scott 2019). In line with this reasoning, our results indicate that the gradient is strongest for migration to Sweden, followed by migration to other Nordic countries. In contrast, we observe no gradient for migration to the rest of the world, and we find indications of a reverse gradient for migration to Estonia. Indeed, Estonia is linguistically and culturally closer to Finland (Tammaru, Kumer-Haukanõmm, and Anniste 2010).

Third, we focus on individuals with a mixed background and investigate how the residential context shapes migration. These analyses reveal that both the context in which individuals live and their mother tongue play an important role. On the one hand, individuals with the same mother tongue have different migration behaviours depending on the area in which they live. That is, Finnish speakers in municipalities with a high share of Swedish speakers are more likely to migrate than Finnish speakers in municipalities with a low share of Swedish speakers. On the other hand, among individuals who live in similar municipalities, Swedish speakers are more prone to migrate than Finnish speakers. That is, we observe differences between Swedish and Finnish speakers in municipalities with low shares of Swedish speakers, as well as between the two groups in municipalities with high shares of Swedish speakers. In short, both the context and factors captured by mother tongue registration, including but not limited to values and cultural proximity, are associated with individuals' migration behaviour (Gans 1979; Hedberg 2004; McRae 1999; Saarela and Finnäs 2018).

In line with previous studies, we find a strong relationship between our proxy of cultural proximity and individuals' migration behaviour (Casari et al. 2018; Jaeger et al. 2010; Knudsen 2019). Our results also reveal a stronger gradient for migration to Sweden and the other Nordic countries than for the rest of the world and Estonia, which is in line with previous country-level analyses showing that cultural proximity promotes migration between countries (Adserà and Pytliková 2015; Arif 2020; Belot and Ederveen 2012; Belot and Hatton 2012; Docquier, Tansel, and Turati 2020). We contribute to this stream of the literature through analyses of individual-level register data that allow us to study both emigration and return migration. Information on individuals' ethno-linguistic background further provides us with a proxy for different strengths of attachment to Finland and Sweden across the population. Even though we lack direct information on individuals' affinities, cultural practices or language fluency, we have probed our proxy measure in multiple ways. This analysis has allowed us to substantiate how ethno-linguistic background captures concepts related to cultural proximity.

Our results suggest that differences in culture can have important implications for migration. Cultural proximity can promote migration by reducing the obstacles individuals face when relocating. On the contrary, cultural distance can impede migration when individuals encounter difficulties in getting their qualifications accredited or other hurdles when organising the relocation (Docquier, Peri, and Ruyssen 2014). This entails both positive and negative aspects. On the one hand, cultural proximity can boost returns to human capital and facilitate integration because migrants already speak the language or have affinities similar to the majority in the destination country (Docquier, Tansel, and Turati 2020). On the other hand, it can prevent the realisation of potential gains from international migration as individuals forego opportunities in countries with which they are less familiar (Adserà 2015; Rapoport, Sardoschau, and Silve 2021).

Our study contributes to the literature by assessing a two-generational dimension of mother tongue to proxy cultural proximity. We find an ethno-linguistic gradient in the likelihood of both migrating and return migrating, which is the strongest for migration to Sweden and the other Nordic countries. We also find that both the residential context and factors captured by mother tongue registration play a role in the differences in migration behaviour. In future research, it would be important to probe the concept of cultural proximity using a combination of register and survey data to gain a comprehensive understanding of its role in migration both in open and closed border settings.

Acknowledgments

We would like to thank participants at the European Society for Population Economics Conference in Belgrade and a workshop at the DIW in Berlin for their helpful comments and input on earlier versions of the paper. This work was supported by the Åbo Akademi University Foundation's funding of the DemSwed Internal Centre of Excellence 2019-2024.

References

Adserà, Alicia. 2015. “Language and Culture as Drivers of Migration.” *IZA World of Labor* 164 (July): 1–10. doi:10.15185/izawol.164.

Adserà, Alicia, and Mariola Pytliková. 2015. “The Role of Language in Shaping International Migration.” *The Economic Journal* 125 (586): F49–F81. doi:10.1111/ecoj.12231.

Arif, Imran. 2020. “The Determinants of International Migration: Unbundling the Role of Economic, Political and Social Institutions.” *The World Economy* 43 (6): 1699–1729. doi:10.1111/twec.12889.

Auwalin, Ilmiawan. 2020. “Ethnic Identity and Internal Migration Decision in Indonesia.” *Journal of Ethnic and Migration Studies* 46 (13): 2841–2861. doi:10.1080/1369183X.2018.1561252.

Belot, Michèle, and Sjef Ederveen. 2012. “Cultural Barriers in Migration between OECD Countries.” *Journal of Population Economics* 25 (3): 1077–1105. doi:10.1007/s00148-011-0356-x.

Belot, Michèle, and Timothy Hatton. 2012. “Immigrant Selection in the OECD.” *The Scandinavian Journal of Economics* 114 (4): 1105–1128. doi:10.1111/j.1467-9442.2012.01721.x.

Caragliu, Andrea, Chiara Del Bo, Henri L. F. de Groot, and Gert-Jan M. Linders. 2013. “Cultural Determinants of Migration.” *The Annals of Regional Science* 51 (1): 7–32. doi:10.1007/s00168-012-0531-3.

Casari, Marco, Andrea Ichino, Moti Michaeli, Maria De Paola, Ginevra Marandola, and Vincenzo Scoppa. 2018. “Civicness Drain.” IZA Discussion Paper No. 11955. doi:10.2139/ssrn.3301707.

Docquier, Frédéric, Giovanni Peri, and Ilse Ruyssen. 2014. “The Cross-Country Determinants of Potential and Actual Migration.” *International Migration Review* 48 (s1): S37–S99. doi:10.1111/imre.12137.

Docquier, Frédéric, Aysit Tansel, and Riccardo Turati. 2020. “Do Emigrants Self-Select Along Cultural Traits? Evidence from the MENA Countries.” *International Migration Review* 54 (2). SAGE Publications Inc: 388–422. doi:10.1177/0197918319849011.

Fafchamps, Marcel, and Forhad Shilpi. 2013. “Determinants of the Choice of Migration Destination.” *Oxford Bulletin of Economics and Statistics* 75 (3): 388–409. doi:10.1111/j.1468-0084.2012.00706.x.

Finnäs, Fjalar, and Richard O’Leary. 2003. “Choosing for the Children: The Affiliation of the Children of Minority–Majority Group Intermarriages.” *European Sociological Review* 19 (5): 483–499. doi:10.1093/esr/19.5.483.

Gans, Herbert J. 1979. “Symbolic Ethnicity: The Future of Ethnic Groups and Cultures in America.” *Ethnic and Racial Studies* 2 (1): 1–20. doi:10.1080/01419870.1979.9993248.

Hedberg, Charlotta. 2004. *The Finland-Swedish Wheel of Migration Identity, Networks and Integration 1976-2000*. Uppsala: Kulturgeografiska institutionen.

Hedberg, Charlotta, and Kaisa Kepsu. 2003. “Migration as a Mode of Cultural Expression? The Case of the Finland-Swedish Minority’s Migration to Sweden.” *Geografiska Annaler: Series B, Human Geography* 85 (2): 67–84. doi:10.1111/1468-0467.00132.

Hofstede, G. 1991. *Cultures and Organizations: Software of the Mind*. London: McGraw-Hill.

Hofstede, G. 2001. *Culture’s Consequences: Comparing Values, Behaviors, Institutions, and Organizations across Nations*. 2nd edn. London: Sage.

Hyyppä, Markku T., and Juhani Mäki. 2001. “Individual-Level Relationships between Social Capital and Self-Rated Health in a Bilingual Community.” *Preventive Medicine* 32 (2): 148–155. doi:10.1006/pmed.2000.0782.

Inglehart, Ronald, and Wayne E. Baker. 2000. “Modernization, Cultural Change, and the Persistence of Traditional Values.” *American Sociological Review* 65 (1): 19–51. doi:10.2307/2657288.

Jaeger, David A, Thomas Dohmen, Armin Falk, David Huffman, Uwe Sunde, and Holger Bonin. 2010. “Direct Evidence on Risk Attitudes and Migration.” *The Review of Economics and Statistics* 92 (3): 684–689. doi:10.1162/REST_a_00020.

Knudsen, Anne Sofie Beck. 2019. “Those Who Stayed: Individualism, Self-Selection and Cultural Change During the Age of Mass Migration.” Working Paper. doi:10.2139/ssrn.3321790.

Koskinen, Seppo, and Tuija Martelin. 2003. “Why Is Mortality Low among the Swedish-Speaking Minority in Finland?” *Finnish Yearbook of Population Research* 39: 15–31.

Li, Nan. 2018. “The Long-Term Consequences of Cultural Distance on Migration: Historical Evidence from China.” *Australian Economic History Review* 58 (1): 2–35. doi:10.1111/aehr.12134.

Mayda, Anna Maria. 2010. “International Migration: A Panel Data Analysis of the Determinants of Bilateral Flows.” *Journal of Population Economics* 23 (4): 1249–1274. doi:10.1007/s00148-009-0251-x.

McRae, Susan. 1999. *Changing Britain: Families and Households in the 1990s*. Oxford University Press.

Nyqvist, Fredrica, Fjalar Finnäs, Gunborg Jakobsson, and Seppo Koskinen. 2008. “The Effect of Social Capital on Health: The Case of Two Language Groups in Finland.” *Health & Place* 14 (2): 347–360. doi:10.1016/j.healthplace.2007.09.001.

Obućina, Ognjen, and Jan Saarela. 2020. “Intergenerational Transmission of Ethnic Identity in Mixed Native Couples in Finland.” *Acta Sociologica* 63 (1): 103–123. doi:10.1177/0001699319847507.

Paljärvi, Tapio, Sakari Suominen, Markku Koskenvuo, Torsten Winter, and Jussi Kauhanen. 2009. “The Differences in Drinking Patterns between Finnish-Speaking Majority and Swedish-Speaking Minority in Finland.” *European Journal of Public Health* 19 (3): 278–284. doi:10.1093/eurpub/ckp007.

Rapoport, Hillel, Sulin Sardoschau, and Arthur Silve. 2021. “Migration and Cultural Change.” IZA Working Paper No.14772. doi:10.2139/ssrn.4114309.

Reini, Kaarina, and Jan Saarela. 2021. “Life Expectancy of the Ethnically Mixed: Register-Based Evidence from Native Finns.” *International Journal of Environmental Research and Public Health* 18 (7): 3415. doi:10.3390/ijerph18073415.

Rooth, Dan-Olof, and Jan Saarela. 2007. “Native Language and Immigrant Labour Market Outcomes: An Alternative Approach to Measuring the Returns for Language Skills.” *Journal of International Migration and Integration* 8 (2): 207–221. doi:10.1007/s12134-007-0014-z.

Ruyssen, Ilse, and Sara Salomone. 2018. “Female Migration: A Way out of Discrimination?” *Journal of Development Economics* 130 (January): 224–241. doi:10.1016/j.jdeveco.2017.10.010.

Saarela, Jan. 2006. “Wealth in Two Ethnic Groups: The Role of Internal Migration Background.” *Finnish Yearbook of Population Research*, 43–64.

Saarela, Jan. 2021. *Finlandssvenskarna 2021 - En Statistisk Rapport*. Folktinget.

Saarela, Jan, and Fjalar Finnäs. 2003. “Unemployment and Native Language: The Finnish Case.” *The Journal of Socio-Economics* 32 (1): 59–80. doi:10.1016/S1053-5357(03)00007-6.

Saarela, Jan, and Fjalar Finnäs. 2004. “Interethnic Wage Variation in the Helsinki Area.” *Finnish Economic Papers* 17 (1): 35–48.

Saarela, Jan, and Fjalar Finnäs. 2005. “Mortality Inequality in Two Native Population Groups.” *Population Studies* 59 (3): 313–320. doi:10.1080/00324720500249372.

Saarela, Jan, and Fjalar Finnäs. 2016. “The Ethno-Linguistic Community and Premature Death: A Register-Based Study of Working-Aged Men in Finland.” *Journal of Racial and Ethnic Health Disparities* 3 (2): 373–380. doi:10.1007/s40615-015-0163-2.

Saarela, Jan, and Fjalar Finnäs. 2018. “Ethno-Linguistic Exogamy and Divorce: Does Marital Duration Matter?” *Sociological Focus* 51 (4): 279–303. doi:10.1080/00380237.2018.1431506.

Saarela, Jan, and Martin Kolk. 2020. “Alcohol-Related Mortality by Ethnic Origin of Natives: A Prospective Cohort Study Based on Multigenerational Population Register Data from Finland and Sweden.” *BMJ Open* 10 (11). British Medical Journal Publishing Group: e042234. doi:10.1136/bmjopen-2020-042234.

Saarela, Jan, and Kirk Scott. 2017. “Mother Tongue, Host Country Earnings, and Return Migration: Evidence from Cross-National Administrative Records.” *International Migration Review* 51 (2): 542–564. doi:10.1111/imre.12230.

Saarela, Jan, and Kirk Scott. 2019. “Naturalization in a Context of Free Mobility: Evidence from Cross-National Data on Finnish Immigrants in Sweden.” *European Journal of Population* 36 (June): 317–335. doi:10.1007/s10680-019-09530-3.

Tammaru, Tiit, Kaja Kumer-Haukanõmm, and Kristi Anniste. 2010. “The Formation and Development of the Estonian Diaspora.” *Journal of Ethnic and Migration Studies* 36 (7): 1157–1174. doi:10.1080/1369183X.2010.481614.

Tandefelt, Marika, and Fjalar Finnäs. 2007. “Language and Demography: Historical Development” 2007 (187–188). De Gruyter Mouton: 35–54. doi:10.1515/IJSL.2007.049.

Wang, Zhiling, Thomas de Graaff, and Peter Nijkamp. 2018. “Barriers of Culture, Networks, and Language in International Migration: A Review.” *REGION* 5 (1): 73–89. doi:10.18335/region.v5i1.203.

Weber, Rosa, and Jan Saarela. 2019. “Circular Migration in a Context of Free Mobility: Evidence from Linked Population Register Data from Finland and Sweden.” *Population, Space and Place* 25 (e2230): 1–12. doi:https://doi.org/10.1002/psp.2230.

Westinen, Jussi. 2015. “Cleavages – Dead and Gone? An Analysis of Cleavage Structure and Party Choice in Contemporary Finland.” *Scandinavian Political Studies* 38 (3): 277–300. doi:10.1111/1467-9477.12046.

Appendix

Table A1. Sample restrictions

	N	Percentage excluded
Finnish register data (masterfile)	8,290,911	
Merge to yearly file for individuals observed at age 19	2,181,627	73.69%
1. Restrict to cohorts 1970-2001	2,046,824	6.18%
2. Restrict to Finnish-born population	1,955,907	4.44%
3. Exclude individuals with missing information on background	1,931,346	1.26%
4. Exclude individuals with *other* mother tongue or background	1,892,945	1.99%
5. Exclude individuals who cannot be observed at age 17 in a municipality in Finland	1,891,297	0.09%
6. Exclude individuals who move prior to 1987	1,888,438	0.15%
7. Exclude individuals if first recorded move is an immigration	1,886,457	0.10%
8. Exclude individuals if the second recorded move is an emigration	1,885,978	0.03%
9. Exclude individuals with missing information on both parents' characteristics	1,885,288	0.04%
10. Exclude individuals who die prior to turning 19 or die in the same month as they move	1,885,239	0.01%
11. Exclude those where we have no information on any end point	1,884,732	0.03%
12. Exclude those who move prior to turning 19	1,873,974	0.57%
Final population	**1,873,974**	
Migrations	86,667	
Return migrations	52,931	

Table A2. Hazard ratios from piecewise exponential models on the likelihood of migrating from Finland

	(1)		(2)		(3)	
Background						
Finnish uniform background (ref.)	1		1		1	
Swedish uniform background	4.456***	(0.046)	4.219***	(0.044)	2.777***	(0.038)
Swedish mixed background	3.107***	(0.052)	2.678***	(0.046)	2.106***	(0.038)
Finnish mixed background	1.491***	(0.034)	1.488***	(0.034)	1.287***	(0.030)
Parental characteristics						
Parental education						
Both parents have primary educ. (ref.)			1		1	
One parent has secondary educ.			1.169***	(0.017)	1.165***	(0.017)
Both parents have secondary educ.			1.327***	(0.019)	1.321***	(0.019)
One parent has tertiary educ.			1.983***	(0.026)	1.782***	(0.025)
Both parents have tertiary educ.			2.745***	(0.038)	2.302***	(0.034)
Parental employment status						
Neither parent is employed (ref.)			1		1	
One parent is employed			1.027	(0.023)	0.999	(0.023)
Both parents are employed			1.014	(0.023)	0.963	(0.022)
Parental income in quartiles						
Quartile 1 (bottom, ref.)			1		1	
Quartile 2			1.022*	(0.011)	1.002	(0.011)
Quartile 3			1.088***	(0.012)	1.019	(0.011)
Quartile 4 (top)			1.315***	(0.015)	1.185***	(0.015)
Not single parent (ref.)			1		1	
Single parent			1.052***	(0.012)	1.060***	(0.012)

Continued

Table A2. Continued

	(1)		(2)		(3)	
Individual characteristics						
Gender						
Male (ref.)					1	
Female					1.473***	(0.011)
Educational attainment						
Primary educ. (ref.)					1	
Secondary educ. with*out* matric. exam					0.583***	(0.007)
Secondary educ. with matric. exam					1.411***	(0.012)
Birth cohort						
1970-1974 (ref.)					1	
1975-1979					1.098***	(0.011)
1980-1984					0.950***	(0.010)
1985-1989					0.882***	(0.011)
1990-1994					0.953***	(0.012)
1995-2001					1.097***	(0.019)
Controls for region of residence at age 17					Yes	
Number of events	86,667		86,667		86,667	
Number of observations	1,873,974		1,873,974		1,873,974	
Number of person-years	23,233,366		23,233,366		23,233,366	
AIC	809,313		794,093		775,146	
BIC	809,597		794,527		775,984	

Note. Clustered standard errors at the individual level in parentheses. Constants are included in the model specifications. * p < 0.05; ** p < 0.01; *** p < 0.001 (two-tailed).

Table A3. Hazard ratios from competing risk models on the likelihood of migrating to Sweden, the other Nordic countries, the rest of the world or Estonia

	Sweden		other Nordics		rest of the world		Estonia	
Background								
Finnish uniform background (ref.)	1		1		1		1	
Swedish uniform background	7.081***	(0.153)	3.690***	(0.126)	1.044	(0.028)	0.622**	(0.102)
Swedish mixed background	4.586***	(0.125)	2.874***	(0.126)	1.226***	(0.036)	1.046	(0.161)
Finnish mixed background	1.979***	(0.080)	1.470***	(0.089)	1.052	(0.034)	1.134	(0.170)
Parental characteristics								
Parental education								
Both parents have primary educ. (ref.)	1		1		1		1	
One parent has secondary educ.	1.033	(0.026)	1.177***	(0.044)	1.247***	(0.025)	1.196	(0.114)
Both parents have secondary educ.	1.140***	(0.030)	1.337***	(0.052)	1.440***	(0.029)	1.251*	(0.124)
One parent has tertiary educ.	1.399***	(0.034)	1.794***	(0.065)	2.043***	(0.039)	1.665***	(0.156)
Both parents have tertiary educ.	1.813***	(0.047)	2.413***	(0.093)	2.619***	(0.053)	2.009***	(0.198)
Parental employment status								
Neither parent is employed (ref.)	1		1		1		1	
One parent is employed	0.856***	(0.035)	1.114	(0.069)	1.035	(0.033)	1.190	(0.179)
Both parents are employed	0.834***	(0.034)	1.054	(0.065)	0.989	(0.031)	1.196	(0.180)
Parental income in quartiles								
Quartile 1 (bottom, ref.)	1		1		1		1	
Quartile 2	1.025	(0.020)	0.969	(0.027)	1.006	(0.015)	0.917	(0.067)
Quartile 3	1.023	(0.021)	0.935*	(0.027)	1.052***	(0.016)	0.849*	(0.066)
Quartile 4 (top)	1.151***	(0.026)	1.010	(0.032)	1.257***	(0.021)	1.181*	(0.096)
Not single parent (ref.)	1		1		1		1	
Single parent	1.053*	(0.022)	0.944	(0.030)	1.091***	(0.016)	1.200**	(0.083)

Continued

Table A3. Continued

	Sweden		other Nordics		rest of the world		Estonia	
Individual characteristics								
Gender								
Male (ref.)	1		1		1		1	
Female	1.467***	(0.019)	1.615***	(0.030)	1.504***	(0.014)	0.517***	(0.025)
Educational attainment								
Primary educ. (ref.)	1		1		1		1	
Secondary educ. with*out* matric. exam	0.617***	(0.013)	0.760***	(0.024)	0.519***	(0.009)	0.472***	(0.037)
Secondary educ. with matric. exam	1.277***	(0.020)	1.460***	(0.034)	1.485***	(0.017)	1.093	(0.059)
Birth cohort								
1970-1974 (ref.)	1		1		1		1	
1975-1979	1.121***	(0.022)	1.074**	(0.029)	1.085***	(0.014)	1.042	(0.078)
1980-1984	0.945**	(0.020)	0.903***	(0.026)	0.948***	(0.014)	1.234**	(0.094)
1985-1989	0.848***	(0.019)	0.781***	(0.024)	0.903***	(0.014)	1.769***	(0.142)
1990-1994	0.921***	(0.021)	0.729***	(0.025)	1.013	(0.018)	3.136***	(0.259)
1995-2001	1.022	(0.029)	0.620***	(0.031)	1.403***	(0.037)	3.991***	(0.421)
Controls for region of residence at age 17	Yes		Yes		Yes		Yes	
Number of events	24,822		12,102		47,836		1,907	
Number of observations	1,873,974		1,873,974		1,873,974		1,873,974	
Number of person-years	23,233,366		23,233,366		23,233,366		23,233,366	
AIC	271,275		157,321		467,649		31,185	
BIC	272,113		158,159		468,487		32,022	

Note. Clustered standard errors at the individual level in parentheses. Constants are included in the model specifications. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ (two-tailed).

Table A4. Hazard ratios from models estimating the likelihood of migrating from Finland among individuals with a mixed background, interacting the share of Swedish speakers in the municipality with ethno-linguistic background

	All destinations		Sweden		other Nordics		rest of the world		Estonia	
Background										
Finnish mixed background (ref.)	1		1		1		1		1	
Swedish mixed background	1.638***	(0.053)	2.248***	(0.124)	1.850***	(0.153)	1.262***	(0.059)	1.131	(0.255)
Share Swedish speakers in municipality										
Below 50 % (ref.)	1		1		1		1		1	
50 % or above	1.317***	(0.093)	1.819***	(0.190)	1.683**	(0.291)	0.931	(0.113)	1.492	(0.858)
Interaction										
Swedish mixed background × 50% or above	0.861*	(0.066)	0.750**	(0.082)	0.767	(0.140)	0.779	(0.109)	0.317	(0.249)
Parental characteristics										
Parental education										
Both parents have primary educ. (ref.)	1		1		1		1		1	
One parent has secondary educ.	1.277***	(0.076)	1.084	(0.096)	2.081***	(0.364)	1.274**	(0.119)	1.868	(0.837)
Both parents have secondary educ.	1.567***	(0.098)	1.354***	(0.123)	2.566***	(0.467)	1.535***	(0.154)	1.692	(0.853)
One parent has tertiary educ.	1.888***	(0.104)	1.532***	(0.124)	2.918***	(0.492)	2.045***	(0.176)	2.016	(0.927)
Both parents have tertiary educ.	2.483***	(0.144)	2.039***	(0.173)	4.333***	(0.748)	2.598***	(0.237)	1.347	(0.669)
Parental employment status										
Neither parent is employed (ref.)	1		1		1		1		1	
One parent is employed	0.859	(0.082)	0.845	(0.123)	1.435	(0.459)	0.798	(0.114)	0.386	(0.255)
Both parents are employed	0.874	(0.083)	0.923	(0.133)	1.519	(0.483)	0.740*	(0.105)	0.424	(0.269)
Parental income in quartiles										
Quartile 1 (bottom, ref.)	1		1		1		1		1	
Quartile 2	0.935	(0.043)	0.912	(0.064)	1.009	(0.117)	0.917	(0.067)	1.340	(0.538)
Quartile 3	0.932	(0.043)	1.028	(0.070)	0.919	(0.107)	0.837*	(0.062)	0.805	(0.341)
Quartile 4 (top)	1.138**	(0.054)	1.134	(0.080)	1.180	(0.139)	1.085	(0.083)	2.004	(0.837)
Not single parent (ref.)	1		1		1		1		1	
Single parent	1.039	(0.042)	1.069	(0.067)	1.120	(0.114)	0.965	(0.062)	1.256	(0.362)

Continued

Table A4. Continued

	All destinations		Sweden		other Nordics		rest of the world		Estonia	
Individual characteristics										
Gender										
Male (ref.)	1		1		1		1		1	
Female	1.506***	(0.041)	1.511***	(0.063)	1.561***	(0.107)	1.543***	(0.068)	0.533**	(0.116)
Educational attainment										
Primary educ. (ref.)	1		1		1		1		1	
Secondary educ. with*out* matric. exam	0.678***	(0.035)	0.683***	(0.052)	0.914	(0.114)	0.595***	(0.053)	0.585	(0.251)
Secondary educ. with matric. exam	1.458***	(0.048)	1.439***	(0.073)	1.590***	(0.136)	1.417***	(0.073)	1.928*	(0.512)
Birth cohort										
1970-1974 (ref.)	1		1		1		1		1	
1975-1979	1.146***	(0.049)	1.233**	(0.086)	1.108	(0.122)	1.093	(0.067)	0.990	(0.298)
1980-1984	1.075	(0.048)	1.200**	(0.085)	1.151	(0.129)	0.957	(0.064)	0.754	(0.265)
1985-1989	1.030	(0.048)	1.134	0.084)	0.992	(0.117)	0.957	(0.069)	1.069	(0.371)
1990-1994	1.174***	(0.058)	1.329***	(0.099)	1.028	(0.126)	1.078	(0.088)	1.778	(0.637)
1995-2001	1.306***	(0.080)	1.480***	(0.126)	1.014	(0.159)	1.292*	(0.156)	1.854	(0.907)
Controls for region of residence at age 17	Yes		Yes		Yes		Yes		Yes	
Number of events	5,778		2,528		908		2,249		93	
Number of observations	67,979		67,979		67,979		67,979		67,979	
Number of person-years	809,641		809,641		809,641		809,641		809,641	
AIC	45,933		24,295		10,457		20,858		1,502	
BIC	46,582		24,945		11,106		21,508		2,140	

Note. Clustered standard errors at the individual level in parentheses. Constants are included in the model specifications. * p < 0.05; ** p < 0.01; *** p < 0.001 (two-tailed).

Table A5. Hazard ratios from piecewise exponential models on the likelihood of return migrating to Finland

	(1)		(2)		(3)	
Background						
Finnish uniform background (ref.)	1		1		1	
Swedish uniform background	0.776***	(0.010)	0.778***	(0.010)	0.752***	(0.014)
Swedish mixed background	0.844***	(0.018)	0.849***	(0.019)	0.840***	(0.019)
Finnish mixed background	0.921**	(0.027)	0.920**	(0.027)	0.940*	(0.028)
Parental characteristics						
Parental education						
Both parents have primary educ. (ref.)			1		1	
One parent has secondary educ.			1.007	(0.018)	0.988	(0.017)
Both parents have secondary educ.			1.010	(0.018)	0.971	(0.017)
One parent has tertiary educ.			0.982	(0.016)	0.962*	(0.016)
Both parents have tertiary educ.			0.969	(0.017)	0.958*	(0.017)
Parental employment status						
Neither parent is employed (ref.)			1		1	
One parent is employed			0.941*	(0.028)	0.957	(0.028)
Both parents are employed			0.903***	(0.026)	0.955	(0.028)
Parental income in quartiles						
Quartile 1 (bottom, ref.)			1		1	
Quartile 2			0.993	(0.013)	0.990	(0.013)
Quartile 3			0.970*	(0.013)	0.967*	(0.013)
Quartile 4 (top)			0.993	(0.015)	0.961**	(0.014)
Not single parent (ref.)			1		1	
Single parent			1.018	(0.014)	1.023	(0.014)

Continued

Table A5. Continued

	(1)		(2)		(3)	
Individual characteristics						
Gender						
Male (ref.)					1	
Female					0.775***	(0.007)
Educational attainment						
Primary educ. (ref.)					1	
Secondary educ. with*out* matric. exam					1.053***	(0.016)
Secondary educ. with matric. exam					0.912***	(0.009)
Birth cohort						
1970-1974 (ref.)					1	
1975-1979					1.063***	(0.013)
1980-1984					1.106***	(0.014)
1985-1989					1.192***	(0.017)
1990-1994					1.348***	(0.022)
1995-2001					1.372***	(0.034)
Controls for region of residence at age 17					Yes	
Number of events	52,931		52,931		52,931	
Number of observations	86,667		86,667		86,667	
Number of person-years	434,848		434,848		434,848	
AIC	254,556		254,512		252,982	
BIC	254,765		254,830		253,597	

Note. Clustered standard errors at the individual level in parentheses. Constants are included in the model specifications. * p < 0.05; ** p < 0.01; *** p < 0.001 (two-tailed).

Table A6. Hazard ratios from competing risk models on the likelihood of migrating to Sweden, the other Nordic countries, the rest of the world or Estonia

	Sweden		other Nordics		rest of the world		Estonia	
Background								
Finnish uniform background (ref.)	1		1		1		1	
Swedish uniform background	0.554***	(0.016)	0.799***	(0.032)	0.783***	(0.029)	1.108	(0.219)
Swedish mixed background	0.653***	(0.023)	0.769***	(0.041)	0.875***	(0.036)	0.720	(0.133)
Finnish mixed background	0.802***	(0.042)	0.948	(0.067)	0.938	(0.041)	1.120	(0.172)
Parental characteristics								
Parental education								
Both parents have primary educ. (ref.)	1		1		1		1	
One parent has secondary educ.	1.018	(0.031)	0.983	(0.038)	1.001	(0.026)	0.972	(0.110)
Both parents have secondary educ.	1.008	(0.032)	0.968	(0.039)	0.982	(0.026)	0.933	(0.112)
One parent has tertiary educ.	1.049	(0.031)	0.987	(0.037)	0.953*	(0.023)	0.859	(0.094)
Both parents have tertiary educ.	1.018	(0.032)	1.049	(0.041)	0.930**	(0.024)	1.063	(0.121)
Parental employment status								
Neither parent is employed (ref.)	1		1		1		1	
One parent is employed	0.852**	(0.043)	1.053	(0.071)	1.015	(0.044)	1.187	(0.262)
Both parents are employed	0.834***	(0.042)	1.034	(0.069)	1.031	(0.045)	1.319	(0.289)
Parental income in quartiles								
Quartile 1 (bottom, ref.)	1		1		1		1	
Quartile 2	0.990	(0.024)	0.988	(0.029)	0.995	(0.019)	0.993	(0.085)
Quartile 3	0.990	(0.024)	0.966	(0.030)	0.979	(0.019)	0.881	(0.078)
Quartile 4 (top)	0.969	(0.027)	0.916*	(0.032)	1.014	(0.022)	0.777**	(0.071)
Not single parent (ref.)	1		1		1		1	
Single parent	1.066*	(0.027)	1.009	(0.034)	1.011	(0.020)	1.092	(0.089)

Continued

Table A6. Continued

	Sweden		other Nordics		rest of the world		Estonia	
Individual characteristics								
Gender								
Male (ref.)	1		1		1		1	
Female	0.785***	(0.013)	0.881***	(0.018)	0.717***	(0.009)	1.148*	(0.068)
Educational attainment								
Primary educ. (ref.)	1		1		1		1	
Secondary educ. with*out* matric. exam	0.923**	(0.024)	1.111***	(0.036)	1.108***	(0.026)	0.854	(0.085)
Secondary educ. with matric. exam	0.837***	(0.015)	1.049*	(0.025)	0.954***	(0.014)	0.826**	(0.051)
Birth cohort								
1970-1974 (ref.)	1		1		1		1	
1975-1979	1.048*	(0.024)	1.083**	(0.030)	1.052**	(0.018)	1.159	(0.107)
1980-1984	1.056*	(0.026)	1.102***	(0.033)	1.109***	(0.020)	1.405***	(0.136)
1985-1989	1.055*	(0.028)	1.145***	(0.038)	1.241***	(0.026)	1.603***	(0.155)
1990-1994	1.056	(0.030)	1.205***	(0.046)	1.515***	(0.036)	2.007***	(0.187)
1995-2001	0.907*	(0.037)	1.129	(0.071)	2.009***	(0.074)	1.798***	(0.247)
Controls for region of residence at age 17	Yes		Yes		Yes		Yes	
Number of events	15,942		9,330		26,388		1,271	
Number of observations	24,822		12,102		47,836		1,907	
Number of person-years	125,721		42,937		259,452		6,738	
AIC	73,908		37,655		131,224		5,361	
BIC	74,453		38,140		131,810		5,743	

Note. Clustered standard errors at the individual level in parentheses. Constants are included in the model specifications. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ (two-tailed).

Table A7. Hazard ratios from models estimating the likelihood of migrating among individuals with a mixed background, interacting the share of Swedish speakers in the municipality with ethno-linguistic background

	All destinations		Sweden		other Nordics		rest of the world		Estonia	
Background										
Finnish mixed background (ref.)	1		1		1		1		1	
Swedish mixed background	0.907*	(0.037)	0.875	(0.062)	0.770**	(0.071)	0.910	(0.059)	0.379*	(0.169)
Share Swedish speakers in municipality										
Below 50 % (ref.)	1		1		1		1		1	
50 % or above	0.747**	(0.075)	0.745*	(0.106)	0.616*	(0.144)	0.685	(0.135)	0.158*	(0.140)
Interaction										
Swedish mixed background × 50% or above	1.136	(0.119)	0.996	(0.147)	1.531	(0.368)	1.399	(0.300)	0.001***	(0.001)
Parental characteristics										
Parental education										
Both parents have primary educ. (ref.)	1		1		1		1		1	
One parent has secondary educ.	1.077	(0.083)	1.036	(0.116)	1.227	(0.278)	1.088	(0.143)	2.652	(2.071)
Both parents have secondary educ.	1.034	(0.084)	0.953	(0.114)	1.364	(0.300)	1.097	(0.158)	0.822	(0.596)
One parent has tertiary educ.	0.993	(0.071)	0.980	(0.103)	1.341	(0.281)	0.989	(0.119)	0.725	(0.439)
Both parents have tertiary educ.	0.987	(0.073)	0.917	(0.100)	1.289	(0.275)	0.988	(0.121)	1.620	(1.363)
Parental employment status										
Neither parent is employed (ref.)	1		1		1		1		1	
One parent is employed	1.127	(0.149)	0.907	(0.166)	1.490	(0.474)	1.226	(0.306)	1.761	(1.717)
Both parents are employed	1.280	(0.168)	1.024	(0.185)	1.511	(0.476)	1.455	(0.362)	2.086	(1.892)
Parental income in quartiles										
Quartile 1 (bottom, ref.)	1		1		1		1		1	
Quartile 2	0.926	(0.055)	0.883	(0.078)	0.985	(0.134)	0.944	(0.100)	2.543	(1.395)
Quartile 3	0.931	(0.054)	0.892	(0.078)	0.877	(0.119)	0.966	(0.097)	1.909	(1.268)
Quartile 4 (top)	0.888*	(0.053)	0.828*	(0.076)	0.874	(0.120)	0.940	(0.095)	1.307	(0.684)
Not single parent (ref.)	1		1		1		1		1	
Single parent	1.047	(0.055)	1.070	(0.086)	0.919	(0.110)	1.007	(0.091)	1.144	(0.458)

Continued

Table A7. Continued

	All destinations		Sweden		other Nordics		rest of the world		Estonia	
Individual characteristics										
Gender										
Male (ref.)	1		1		1		1		1	
Female	0.774***	(0.027)	0.754***	(0.040)	0.912	(0.074)	0.713***	(0.042)	1.754	(0.593)
Educational attainment										
Primary educ. (ref.)	1		1		1		1		1	
Secondary educ. with*out* matric. exam	0.949	(0.064)	0.851	(0.083)	0.815	(0.127)	1.100	(0.137)	2.453	(2.138)
Secondary educ. with matric. exam	0.916*	(0.037)	0.829**	(0.052)	1.112	(0.104)	0.941	(0.066)	2.252	(1.274)
Birth cohort										
1970-1974 (ref.)	1		1		1		1		1	
1975-1979	1.076	(0.055)	1.060	(0.088)	0.954	(0.114)	1.135	(0.092)	0.639	(0.280)
1980-1984	1.086	(0.061)	1.001	(0.088)	1.152	(0.135)	1.042	(0.099)	1.615	(0.973)
1985-1989	1.145*	(0.067)	1.027	(0.094)	0.927	(0.123)	1.265*	(0.128)	1.859	(0.944)
1990-1994	1.327***	(0.081)	0.956	(0.092)	1.166	(0.156)	1.945***	(0.209)	1.993	(1.025)
1995-2001	1.285**	(0.111)	1.057	(0.125)	0.929	(0.208)	2.159***	(0.409)	0.300	(0.416)
Controls for region of residence at age 17	Yes		Yes		Yes		Yes		Yes	
Number of events	3,312		1,469		633		1,148		62	
Number of observations	5,778		2,528		908		2,249		93	
Number of person-years	30,066		13,337		3,544		12,804		381	
AIC	16,448		7,198		2,826		5,985		296	
BIC	16,914		7,610		3,166		6,403		473	

Note. Clustered standard errors at the individual level in parentheses. Constants are included in the model specifications. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ (two-tailed).

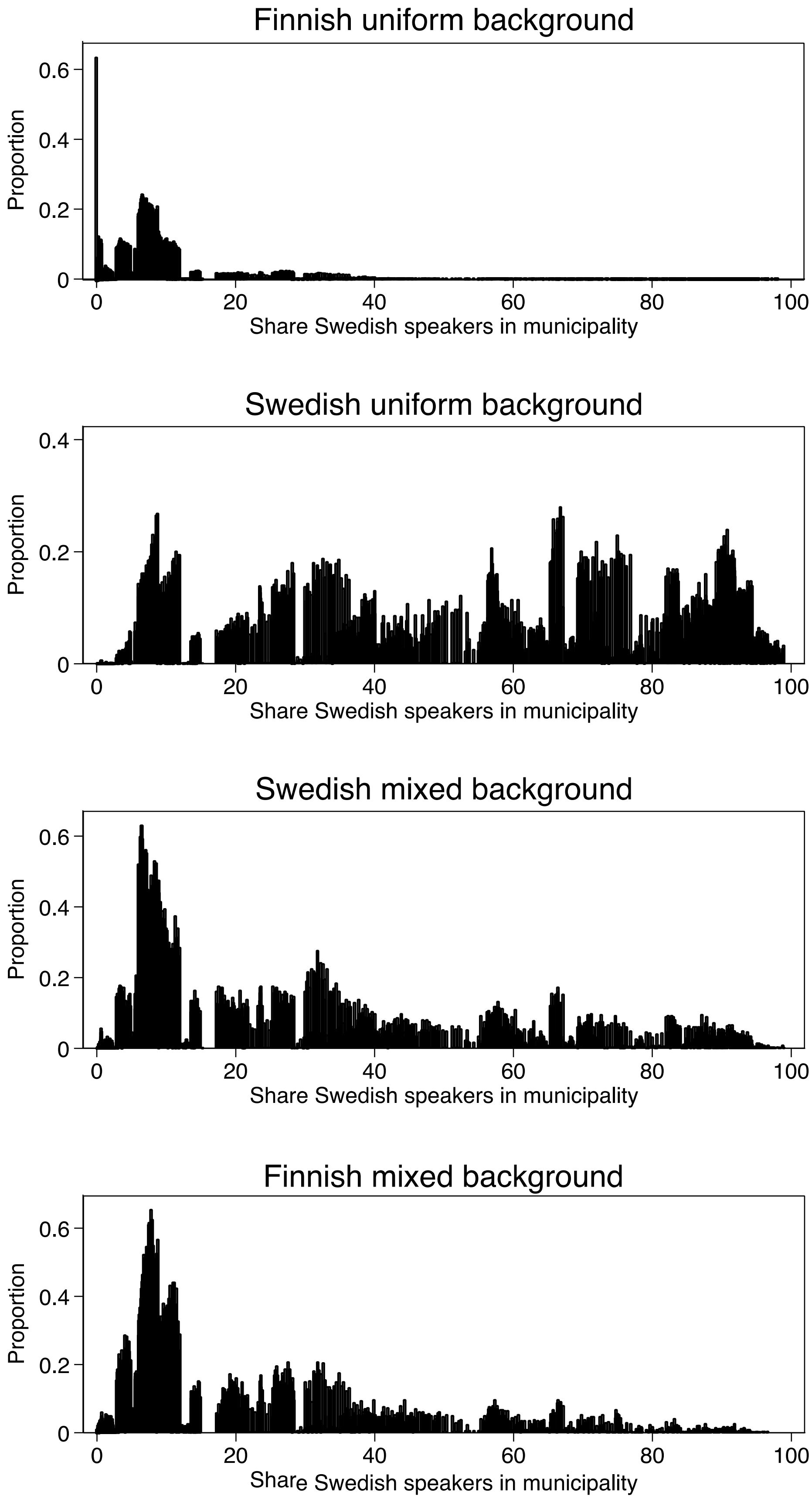

Figure A1. Density plots of the share of Swedish speakers in the municipality by ethno-linguistic background.

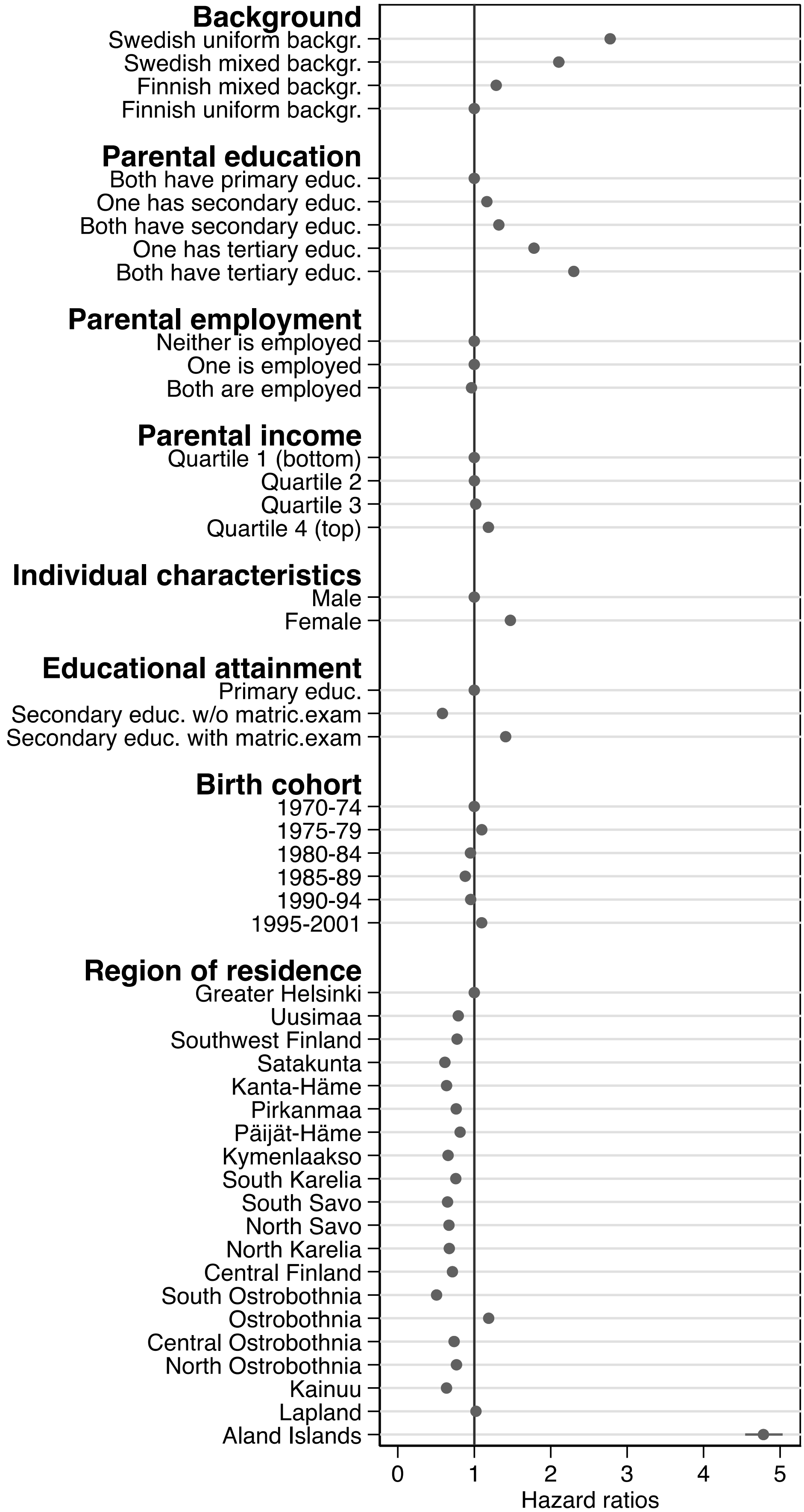

Figure A2. Forest plot of the likelihood of migrating by ethno-linguistic background, parental and individual socioeconomic characteristics.

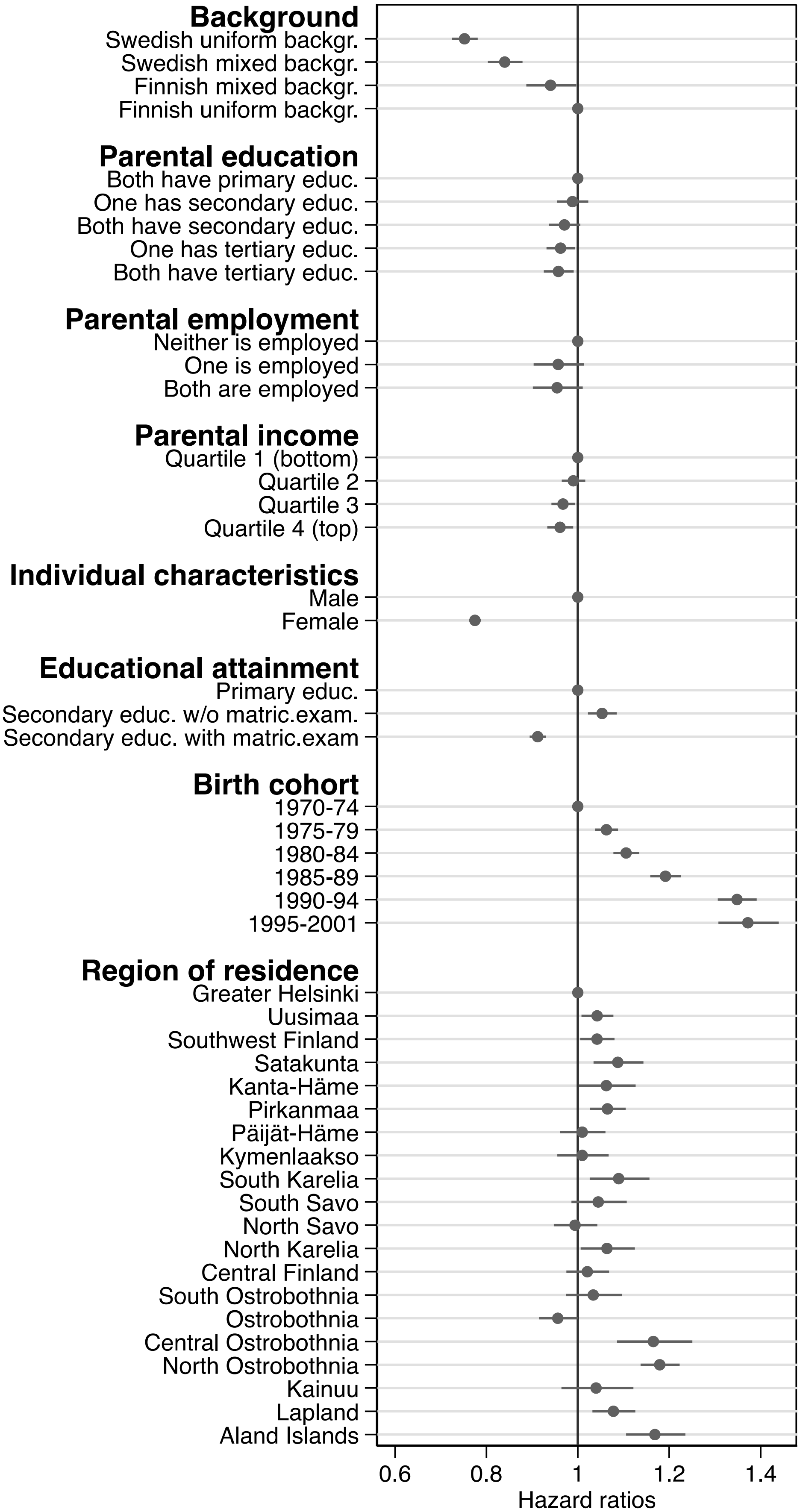

Figure A3. Forest plot of the likelihood of return migrating by ethno-linguistic background, parental and individual socioeconomic characteristics.

Stockholm Research Reports in Demography
Stockholm University,
106 91 Stockholm,
Sweden
www.su.se | info@su.se | ISSN 2002-617X

Stockholm University
Demography Unit