(November 14, 2013)

Mother tongue, host country income and return migration

Jan Saarela (University of Helsinki and Åbo Akademi University)

Kirk Scott (Lund University)

Abstract. Using a unique database constructed through the merging of administrative records from Sweden and Finland, this study is the first to provide a detailed examination of differential return-migration risks by people's mother tongue within a given nationality. We estimate hazard models of the propensity to return migrate among Finnish-speaking and Swedish-speaking Finns in Sweden, accounting for income and standard socio-demographic variables. In line with previous research, Swedish-speaking Finns are found to have notably lower return-migration risks than Finnish-speaking Finns. Our primary aim is to study whether this divergence relates to between-group differences in host country income. The motivation is that labour market outcomes of Swedish-speaking Finns are known to be in parity with native Swedes, which presumably is because they have the same mother tongue. The research hypothesis is not confirmed by our analyses, however, since host country income and the other background variables only explain a modest part of the language-group difference in return-migration risk.

--

Corresponding author: Jan Saarela, Åbo Akademi University, PO Box 311, 65101 Vasa, Finland, jan.saarela@abo.fi, +358-50-5962692.

1. Introduction

During the second half of the 20th century, migration attained previously unforeseen proportions and importance in Western Europe, as throughout the rest of the world (Zimmermann, 2005). Despite the importance of this population dynamic, the mechanisms behind the post-migration decision to remain or to return home have received much less attention than the general labour market success of these individuals. Using unique cross-country register data, this study will shed light on the previously unexamined aspect of origin-country heterogeneity, measured here as variation in mother-tongue among immigrants from Finland.

An important measure of immigrant integration is host country income (Dustmann, 1993; Borjas, 1999). Performing worse than expected might consequently make an immigrant more inclined to make a corrective move (Borjas and Bratsberg, 1996). Accordingly, labour migrants with low relative income tend to have elevated return migration rates (Edin et al., 2000), which is also the case for Finnish immigrants in Sweden (Saarela and Rooth, 2012). There is, nevertheless, substantial heterogeneity by native language.

The population registers in Finland, unlike those in Sweden, record each citizen's unique mother tongue. Information about the native language of an immigrant in Sweden is therefore available only via linkage to registers from Finland. Studies that have utilised such linked cross-boundary registers have found that Swedish-speaking immigrants from Finland have income and employment levels that are in parity with those of native Swedes, whereas the labour market performance of Finnish-speaking immigrants is clearly inferior (Rooth and Saarela, 2007). This language-group difference supports the common view that immigrants with good abilities to read, write and speak the language of the host country generally have better labour market outcomes than those with poorer abilities (Chiswick and Miller, 1995; Dustmann and Fabbri, 2003).

Swedish speakers in Finland have for many decades had higher emigration rates than Finnish speakers, whereas their return migration rates have been lower (Saarela and Finnäs, 2012). Among people who migrated from Finland in the 1980s, approximately half of the Swedish-speaking men had returned within eight years, compared with two-thirds of the Finnish-speaking men. The corresponding proportions for Swedish-speaking and Finnish-speaking women, respectively, were approximately 43 and 54 per cent. In the 1970s, over 80 per cent of all Finns who moved abroad went to Sweden, whereas in the 1980s and 1990s, the share was approximately 65 per cent and barely 40 per cent, respectively (Saarela and Finnäs, 2013). Since Sweden has been the primary destination country, it is evident that the Swedish-speaking immigrants from Finland have been much more likely to remain in the host country, at least for longer periods. Over one-fifth of the Finnish-born population in Sweden, which amounted to almost 200,000 persons at the turn of the millennium, has Swedish as their native language (Rooth and Saarela, 2007). Since this language group constitutes less than six percent of the total population in Finland, they are clearly overrepresented in the Finnish-born immigrant stock in Sweden.

Quantifying return migration dynamics by mother tongue, and particularly the interrelation with income, still requires detailed data, which have not been available before now. The present paper uses unique register-based information on Finnish immigrants in Sweden who can be observed also before they migrated from Finland. Since we observe each individual on an annual basis, our aim is to provide the first explicit assessment of return migration risks by native language of Finnish immigrants in Sweden, and investigate the potential role of host-country income on the language-group difference in return migration risks. We focus on labour migrants who made their first move to Sweden in the period 1988-2004.

2. Data and methods

The data used were constructed by integrating records on Finnish immigrants in Sweden from population registers in both Sweden and Finland. The information was provided by Statistics Sweden (permission number 8547689/181453) and Statistics Finland (permission number TK-52-215-11). The data for Finland refer to the period 1987-2008, and contain socioeconomic, demographic, and labour market variables for each individual, whereof most are measured at the end of each calendar year. The information from Sweden has a similar structure and refers to the period 1985-2005.

In the analyses, we study Finns who migrate to Sweden in the period 1988-2004, and who are under risk of return migration during the period 1988-2005. By doing so, we can observe all persons in the year before they migrated from Finland and have a follow-up period of 0-18 years. Since the registered moves in and out of each country refer to the exact date of the event, time spent abroad can be measured in days.

Our study population consists of immigrants in Sweden who were born in Finland and migrated directly from Finland (not via any third country). They amount to 35,271 persons. Approximately 15 per cent of all moves between the countries are repeat moves, meaning that the same individual emigrates more than once. Since subsequent migrations may induce bias, we choose to focus only on the first move of each person. To ensure the first migration, we exploit data in the Swedish records, which supply the date of any previous migration, even when it falls outside of our study window. All immigration consequently refers to each individual's first experience of the Swedish labour market.

People aged 18-65 years amount to 28,265 of all migrants, with the rest predominantly consisting of tied migrants of minor age. Statistics Finland has a policy of not providing detailed information on complete populations and has therefore randomly excluded 22.5 per cent of these persons. The remaining sample therefore consists of 21,903 individuals for whom we

have information from both Finland and Sweden. For an additional 782 persons there is missing information on central variables, and 21 persons have a mother tongue other than Finnish or Swedish. These are excluded, resulting in 21,575 persons aged 18-65 years for whom there is complete information from each country's registers.

To place focus on labour migrants, we have chosen to further restrict the data to persons who were aged 25-55 years and in the labour force at the time of migration from Finland. Since most migration occurs around 20 years of age, we end up with 8,842 study persons, of whom 3,611 are Finnish-speaking men, 1,212 are Swedish-speaking men, 3,058 are Finnish-speaking women, and 961 are Swedish-speaking women.

Our primary aim is to estimate the difference in the return migration risk between Swedish-speaking and Finnish-speaking Finns, and particularly whether it relates to language-group differences in host country income. We estimate hazard models, where the risk of return migration is piecewise constant for each calendar year spent in Sweden. Duration is consequently time spent in Sweden, the event of interest is return migration, and the key variables of interest are each migrant's unique mother tongue (Finnish or Swedish) and income in Sweden (inflation adjusted earnings per time unit in the country, plus a separate category for zero earnings). Persons who die or move to a third country from Sweden are treated as right-censored observations at the time of the event. Separate analyses are consistently undertaken for men and women.

We account for a number of confounders. The control variables used, which are described in Table 1 in the next section, are each person's age at migration, year of migration, educational level and field, region of residence in the home country, marital status, number of children under seven years of age, number of children under 18 years of age, region of residence in Sweden, and time spent in Sweden. There is also information about income in the home country before migration. Including this variable has a marginal effect on the estimates of interest. Due

to potential endogeneity problems, however, we excluded the variable from the models whose results are reported in the next section.

All variables except the first three mentioned are time-varying. The reason that education is treated as a time-consistent variable is that there are notable deficiencies on immigrants' education in the Swedish registers, and particularly for those who recently have immigrated (Statistics Sweden, 2011). These shortcomings are also evident in these data. Since many Finns return migrate after only a few years, immigrant education as measured from the Swedish records cannot consequently be used. This is, nevertheless, a marginal problem, since all persons studied here are at least 25 years of age.

3. Results

The return migration risk is very high during the first four years after immigration, and particularly during the second year (Figure 1). The profile is similar for both Swedish speakers and Finnish speakers, but the former lie at a notably lower level. These discrepancies result in considerable variation across the groups in the proportion of return migrants (Figure 2). After six years, as much as 61 percent of all Finnish-speaking immigrants in Sweden had returned to their home country, compared with 48 percent of all Swedish-speaking men, 42 percent of all Finnish-speaking women, and only 30 percent of all Swedish-speaking women.

(Figure 1 here)

(Figure 2 here)

The difference in the relative risk of return migration risk between the two groups is also highest during the first years subsequent to immigration. In men, the unadjusted risk ratio is somewhat over 0.50 the first two to five calendar years, whereas a similar risk ratio can be observed for women during the first two to sex calendar years (Table 1). In terms of an average over time, the unadjusted risk ratio is 0.58 for men and 0.57 for women (Table 2).

(Table 1 here)

(Table 2 here)

Finnish speakers and Swedish speakers differ on background characteristics (Table 3). The former are older, more of them are childless, not married, earlier immigrated (after 1987), and particularly the Finnish-speaking men are relatively low educated. Differences in the region of origin are mainly because Swedish speakers in Finland live concentrated along the coastline, whereas Finnish speakers are more dispersed around the country. People from Lappi are although overrepresented among Finnish-speaking migrants, and those from Pohjanmaa and the Åland Islands among Swedish speakers. When settling in Sweden, more Swedish speakers than Finnish speakers choose the Stockholm region, whereas a notably higher proportion of the Finnish speakers live in Norrbotten. There are also notable income differences between the two groups, no matter which alternative measure of income is applied (Table 4). In men, the proportion of people with zero registered earnings is approximately 2.5 times higher among Finnish speakers than among Swedish speakers, and in women about two times higher, somewhat depending on the measure applied. Swedish speakers, and particularly Swedish-speaking men, are also heavily overrepresented in the upper income deciles.

(Table 3 here)

(Table 4 here)

The lower return migration risk of Swedish speakers cannot be attributed to their generally more favourable background or income characteristics, however. The risk ratios change only marginally when we account for all variables except income (Tables 1 and 2). Including income moderates the adjusted risk ratio somewhat closer to parity, but even with the most successful measure on this account, it is on average 0.65 for each sex. Thus, even after adjusting for income differences, Swedish speakers have approximately 35 percent lower return migration risk than Finnish speakers.

Finally, we show that, for both men and women, people with zero earnings are by far most inclined to return migrate (Table 5). There also seems to be a U-shaped relationship between income and the return migration risk. Within the lower income deciles, the risk of return migration tends to fall with increased income, whereas it increases in the upper deciles and is relatively high among the top decile earners, corresponding to previous research using the same data (Saarela and Rooth, 2012). A potential explanation to the relatively high return migration risk at the top of the income distribution might be that, if some people migrate as part of an optimal maximisation plan, those who have reached their income target might be more inclined to move back home (cf. Dustmann, 2003). Men in the fifth decile, for instance, have 83 percent lower return migration risks than people without earnings, while those in the tenth decile have only 23 percent lower return migration risks. In spite of the large income differences between Swedish speakers and Finnish speakers, the interrelation between zero earnings and the return migration risk is basically similar in each group, at least among men, and there is a U-shaped relationship between income and the return migration risk in each group.

(Table 5 here)

4. Discussion and conclusion

(To be continued)

References

Borjas, G. J. (1999). The economic analysis of immigration. In *Handbook of Labor Economics*, Volume 3A, Ashenfelter, O. and Card, D. (eds.). Elsevier, North Holland, NY, 1697-1760.

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

Chiswick, B. R. & Miller, P. W. (1995). The endogeneity between language and earnings: International analyses. *Journal of Labor Economics*, **13**, 246-288.

Dustmann, C. (1993). Earnings adjustment of temporary migrants. *Journal of Population Economics*, **6**, 153-168.

Dustmann, C. (2003). Return migration and the optimal migration duration. *European Economic Review*, **47**, 353-367.

Dustmann, C. & Fabbri, F. (2003). Language proficiency and labour market performance of immigrants in the UK. *Economic Journal*, **113**, 695-717.

Edin, P.-A., LaLonde, R. & Åslund, O. (2000). Emigration of immigrants and measures of immigrant assimilation: Evidence from Sweden. *Swedish Economic Policy Review*, **7**, 163-204.

Rooth, D.-O. & Saarela, J. (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**, 207-221.

Saarela, J. & Finnäs, F. (2011). Återvandring och upprepad emigration i ljuset av vår befolkningsstatistik (Return and repeat migration in the light of our population statistics). In *Hälsa och välfärd i ett föränderligt samhälle - festskrift till Gunborg Jakobsson*, Nygård, M. & Finnäs, F. (eds.), Åbo Akademis förlag, Turku, 41-51.

Saarela, J. & Finnäs, F. (2013). The international family migration of Swedish-speaking Finns: The role of spousal education. *Journal of Ethnic and Migration Studies*, **39**, 391-408.

Saarela, J. & Rooth, D.-O. (2012). Uncertainty and international return migration: Some evidence from linked register data. *Applied Economics Letters*, **19**, 1893-1897.

Statistics Sweden (2011). *Återutvandring efter tid i Sverige* (Return migration by time spent in Sweden). Background material about demography, children and family 2011:1. Statistics Sweden, Forecast Institute, Stockholm.

Zimmermann, K. (ed.) (2005). *European migration: What do we know?*. Oxford University Press, New York.

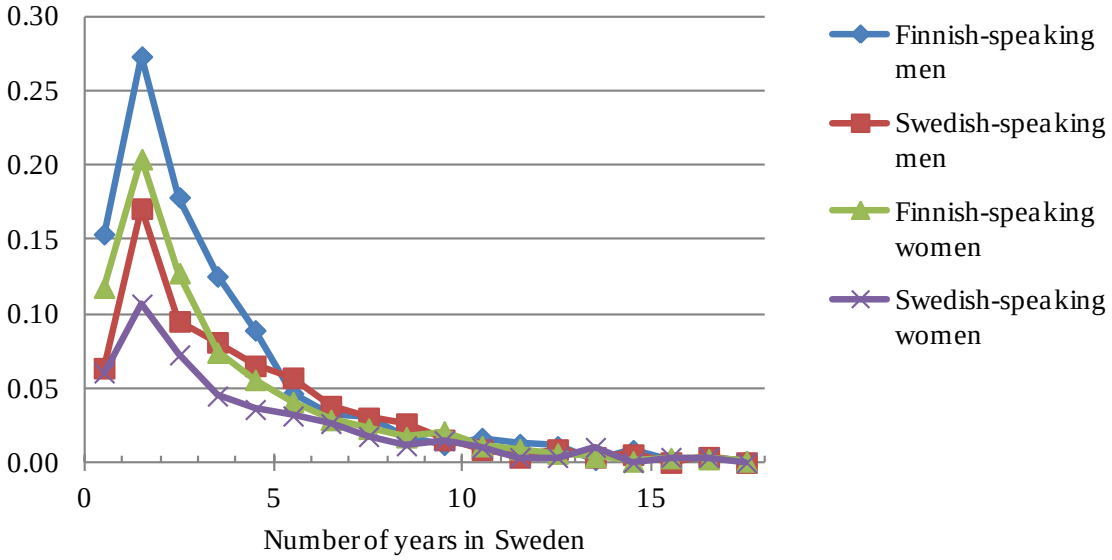

Figure 1. One-year return migration risk by time abroad, sex and mother tongue

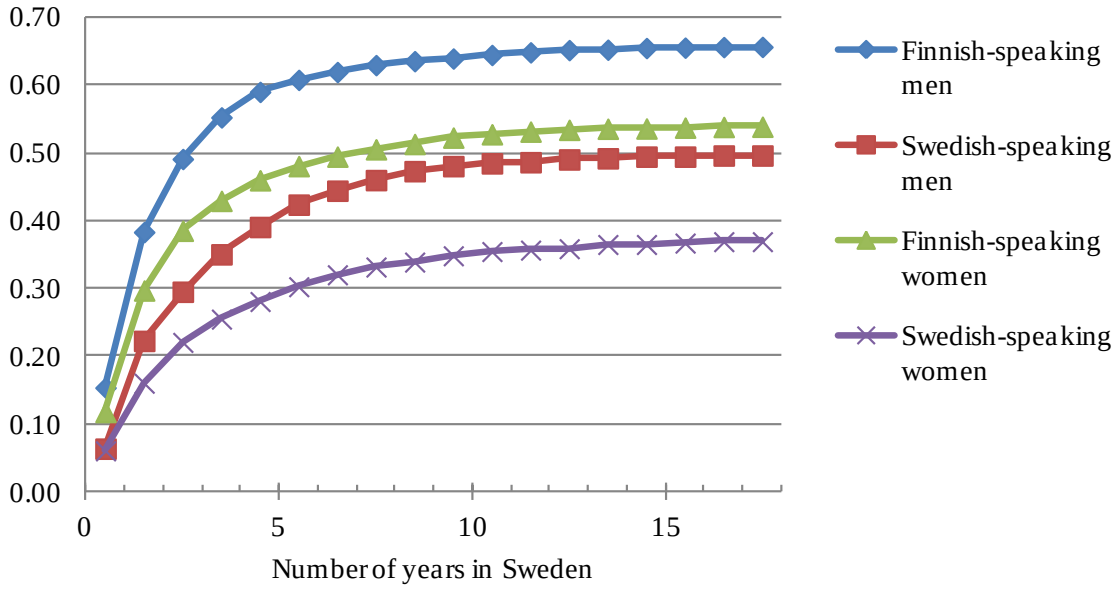

Figure 2. Proportion return migrants by time abroad, sex and mother tongue

Table 1. Return migration risk of Swedish speakers as compared with that of Finnish speakers (with 95% confidence interval) by sex and number of calendar years in the host country

	Men			Women		
	No controls	All except income	+ Income	No controls	All except income	+ Income
Time abroad						
1	0.79 (0.46-1.54)	0.82 (0.42-1.60)	0.86 (0.44-1.69)	1.36 (0.67-2.76)	1.41 (0.69-2.88)	1.48 (0.72-3.02)
2	0.51 (0.44-0.60)	0.51 (0.43-0.61)	0.56 (0.47-0.67)	0.47 (0.39-0.58)	0.50 (0.40-0.62)	0.52 (0.41-0.65)
3	0.47 (0.38-0.58)	0.45 (0.36-0.56)	0.51 (0.41-0.63)	0.49 (0.38-0.63)	0.50 (0.38-0.65)	0.51 (0.39-0.67)
4	0.57 (0.44-0.73)	0.53 (0.41-0.68)	0.62 (0.48-0.81)	0.61 (0.44-0.83)	0.60 (0.43-0.83)	0.62 (0.45-0.86)
5	0.57 (0.42-0.78)	0.53 (0.38-0.73)	0.62 (0.45-0.86)	0.65 (0.43-0.99)	0.62 (0.40-0.95)	0.64 (0.42-0.98)
6	1.04 (0.75-1.45)	0.96 (0.68-1.35)	1.13 (0.80-1.59)	0.56 (0.34-0.92)	0.53 (0.32-0.88)	0.55 (0.33-0.92)
7-11	1.15 (0.84-1.57)	1.04 (0.75-1.43)	1.22 (0.88-1.68)	0.81 (0.56-1.18)	0.72 (0.49-1.05)	0.75 (0.51-1.10)
12-18	0.66 (0.36-1.18)	0.58 (0.32-1.06)	0.70 (0.38-1.27)	1.25 (0.73-2.16)	1.09 (0.63-1.89)	1.12 (0.64-1.94)

The income measure used here is type 1, as referred to in the footnotes of Table 4.

Estimates in each column stem from the same model. We have just switched the reference group to be able to statistically test the difference between Swedish speakers and Finnish speakers for each year.

Table 2. Adjusted risk of return migration (with 95% confidence interval for Swedish speakers as compared with Finnish speakers in alternative models when applying different measures of income

	Men	Women
Control variables included		
No	0.58 (0.53-0.63)	0.57 (0.51-0.64)
All except income	0.57 (0.51-0.64)	0.58 (0.50-0.67)
+ income type 1	0.65 (0.58-0.74)	0.65 (0.56-0.74)
+ income type 2	0.65 (0.57-0.73)	0.64 (0.56-0.74)
+ income type 3	0.59 (0.52-0.67)	0.63 (0.55-0.73)
+ income type 4	0.59 (0.52-0.66)	0.64 (0.55-0.73)

The different types of income measures are described in the footnotes of Table 4.

Table 3. Summary of variables distribution by sex and mother tongue (%)

	Men		Women	
	Finnish-speakers	Swedish-speakers	Finnish-speakers	Swedish-speakers
Age at migration				
25-29	34.7	43.1	33.7	40.9
30-39	39.3	34.0	38.1	34.8
40-55	26.0	22.9	28.2	24.3
Immigration year				
1988-1991	44.5	35.4	41.0	32.8
1992-1995	22.9	26.9	25.6	26.6
1996-1999	13.6	19.7	13.6	20.6
2000-2004	19.0	17.9	19.8	20.0
Education				
Primary	33.8	24.9	26.0	22.9
Secondary	39.3	32.8	31.9	28.2
Tertiary	26.9	42.4	42.1	49.0
Region in home country				
Uusimaa	29.1	33.5	37.2	35.0
Varsinais-Suomi	8.2	8.4	7.7	8.6
Pirkanmaa	8.6	1.1	8.0	0.6
Pohjanmaa	3.4	37.1	3.4	34.4
Lappi	13.9	0.2	13.1	0.2
Åland Islands	0.6	14.3	0.9	15.4
Any other	36.2	5.3	29.5	5.7
Marital status				
Unmarried	44.3	46.3	40.9	46.4
Married	38.0	42.3	37.8	36.1
Previously married	17.7	11.3	21.3	17.5
Children under 7 years				
0	78.6	74.4	76.0	74.3
1	13.2	16.1	16.1	15.9
2+	8.1	9.5	7.9	9.8
Children under 18 years				
0	69.9	65.1	60.6	61.8
1	13.2	15.4	19.1	16.8
2+	16.9	19.5	20.3	21.5
Region in host country				
Stockholm	43.4	54.5	45.8	54.0
Norrbotten	13.9	1.7	12.9	1.5
Any other	42.7	43.8	41.3	44.5
Time in host country				
1-2 years	37.3	30.1	31.9	27.6
3-6 years	34.3	37.9	34.5	36.8
7+ years	28.4	32.0	33.6	35.6
Total number of persons	3,611	1,212	3,058	961
Total risk time in years	16,043	7,083	16,564	6,248

Distribution refers to percentages of total risk time.

Time in the host country gives the contribution of each calendar year to total risk time (not the inverse of the proportion of return migrants in Figure 2).

In the analyses, all continuous variables are used at the count-data level, education is used as the level and field of education with 41 categories, region in the home country has 20 categories, marital status consists of the categories not married, married, and previously married, and region in the host country has 24 categories.

Table 4. Income distribution by sex and mother tongue for alternative measures (%)

	Men								Women							
	Type 1		Type 2		Type 3		Type 4		Type 1		Type 2		Type 3		Type 4	
	Fi.-sp.	Sw.-sp.	Fi.-sp.	Sw.-sp.	Fi.-sp.	Sw.-sp.	Fi.-sp.	Sw.-sp.	Fi.-sp.	Sw.-sp.	Fi.-sp.	Sw.-sp.	Fi.-sp.	Sw.-sp.	Fi.-sp.	Sw.-sp.
Income decile																
No earnings	27.1	11.4	27.1	11.4	10.7	3.6	10.7	3.6	19.3	9.8	19.3	9.8	6.6	2.5	6.6	2.5
First	8.7	5.5	7.5	4.7	11.0	4.6	10.7	4.6	9.1	6.3	10.3	7.1	11.0	5.2	11.3	5.4
Second	8.8	5.3	7.4	4.4	11.3	3.9	11.4	3.8	8.8	6.9	10.1	8.2	10.7	5.8	10.8	5.6
Third	8.4	6.2	6.2	4.6	10.5	6.1	8.6	4.4	8.4	8.3	10.3	10.8	9.9	8.5	11.9	9.4
Fourth	7.8	7.7	6.2	4.8	10.2	6.4	7.3	4.2	8.3	8.1	10.4	10.5	10.5	6.7	13.0	10.7
Fifth	7.3	8.8	6.5	6.3	9.2	8.9	7.8	7.2	8.2	8.8	9.1	11.4	8.9	10.7	10.2	13.1
Sixth	6.9	9.9	7.3	8.8	8.6	10.5	8.2	7.4	7.7	10.0	7.4	10.8	8.9	11.0	9.0	15.2
Seventh	6.5	10.7	7.2	10.3	7.3	13.4	8.1	11.9	7.4	10.8	7.0	10.5	7.9	13.5	8.5	11.9
Eighth	6.3	11.2	7.5	12.2	7.3	13.8	8.3	14.0	7.3	11.2	6.6	8.7	8.6	11.7	7.7	11.2
Ninth	5.7	12.8	7.5	14.6	6.5	15.5	8.0	16.7	7.7	10.2	5.9	7.7	8.1	13.0	6.9	10.5
Tenth	6.6	10.5	9.6	17.8	7.5	13.2	10.8	22.1	7.9	9.6	3.6	4.4	8.8	11.4	4.1	4.4

All measures of income are constructed by dividing annual income with time in the host country, which is number of days/365.
Type 1 refers to deciles based on the distribution of income for each year in the data for men and women, respectively.
Type 2 refers to deciles based on the distribution of income for each year in the data for both sexes.
Type 3 refers to deciles based on the distribution of accumulated income in the host country for men and women, respectively.
Type 4 refers to deciles based on the distribution of accumulated income in the host country for both sexes.

Table 5. Risk ratio of return migration (with 95% confidence intervals) by income decile

			Interaction between income and mother tongue			
			Men		Women	
	Men	Women	Finnish-speak.	Swedish-speak.	Finnish-speak.	Swedish-speak.
Income decile						
No earnings	1	1	1	1.00 (0.84-1.19)	1	0.84 (0.67-1.04)
First	0.30 (0.26-0.35)	0.30 (0.25-0.35)	0.31 (0.26-0.37)	0.24 (0.17-0.34)	0.27 (0.23-0.33)	0.22 (0.16-0.31)
Second	0.23 (0.19-0.27)	0.19 (0.15-0.23)	0.24 (0.20-0.29)	0.16 (0.11-0.25)	0.21 (0.17-0.25)	0.16 (0.12-0.23)
Third	0.25 (0.21-0.30)	0.24 (0.20-0.28)	0.24 (0.20-0.29)	0.24 (0.17-0.33)	0.15 (0.13-0.19)	0.06 (0.04-0.09)
Fourth	0.18 (0.15-0.22)	0.14 (0.12-0.18)	0.20 (0.16-0.25)	0.10 (0.07-0.16)	0.18 (0.14-0.22)	0.08 (0.05-0.12)
Fifth	0.17 (0.14-0.20)	0.12 (0.09-0.15)	0.20 (0.16-0.25)	0.08 (0.05-0.12)	0.16 (0.12-0.21)	0.10 (0.06-0.15)
Sixth	0.20 (0.16-0.24)	0.16 (0.13-0.21)	0.25 (0.20-0.31)	0.08 (0.05-0.13)	0.17 (0.13-0.23)	0.09 (0.06-0.15)
Seventh	0.19 (0.16-0.24)	0.15 (0.12-0.20)	0.22 (0.18-0.29)	0.11 (0.08-0.16)	0.25 (0.20-0.33)	0.17 (0.12-0.26)
Eighth	0.27 (0.23-0.33)	0.17 (0.14-0.22)	0.32 (0.26-0.40)	0.16 (0.12-0.21)	0.32 (0.25-0.41)	0.17 (0.11-0.28)
Ninth	0.38 (0.32-0.45)	0.25 (0.20-0.30	0.52 (0.44-0.63)	0.15 (0.11-0.20)	0.61 (0.48-0.77)	0.29 (0.18-0.47)
Tenth	0.73 (0.63-0.84)	0.67 (0.57-0.79)	0.77 (0.65-0.91)	0.54 (0.43-0.69)	2.16 (1.66-2.79)	0.91 (0.53-1.55)

The income measure used here is type 1, as referred to in the footnotes of Table 4. The estimates in the first two columns for each sex consequently stem from the same specification as referred to by the last row and first two columns for each sex in Table 2.

The results from models with interaction mean that we have substituted the main effects of income decile and mother tongue with a variable that captures the joint effects of these two variables. This procedure does improve the statistical fit of each model at any reasonable level of statistical significance.