LUND UNIVERSITY
School of Economics and Management

Master programme in Economic Demography

Pre- and post-migration labour market mismatch in Sweden 1970-1990

Martin Önnerfors
martin@onnerfors.se

Abstract

Labour market outcomes for immigrants in general is a well researched field, but the mechanisms behind labour market mismatches among immigrants post-migration is still in need of empirical research. Using a unique and newly compiled dataset on Swedish immigrants, also containing pre-migration occupational and educational information data, this study aims makes use of maximum likelihood models to measure the influence of an individual's pre-migration history on the post-migration outcomes. The results show significant associations between pre- and post-migration mismatch status, and a persistence in the mismatched state over time. Possible explanations include individual ability, discrimination, imperfect transferability of human capital, and changes in labour demand. The method used is not causal, and the associations shown should be researched further using a causal approach.

Keywords: educational mismatch, labour market integration, immigration, Sweden

EKHM52
Master thesis, second year (15 credits ECTS)
June 2016
Supervisor: Kirk Scott
Examiner: Kerstin Enflo
Word count: 17.670

Contents

List of Tables

List of Figures

1 Introduction

The success or failure of immigrants on their host country labour markets is an increasingly discussed topic today. Crises in, and inequalities between, different parts of the world are driving migration flows to increasingly high levels. In light of the large global disparities in prosperity and influence, some researchers consider even the high levels of recent years to be modest, compared to what might theoretically be the case (Massey et al. 1999, p.7).

The modern Swedish history of immigration and labour market integration has gone through periods of labour migration, refugee migration, as well as economic boom and bust. Along with the economic crisis of the 1990s came a structural change of the economy that put further demand on host country-specific skills for immigrants (Rosholm, Scott, and Husted 2006). There are, in other words, both supply and demand forces at work in shaping labour market outcomes for immigrants in Sweden. In order to succeed in labour market integration, policy needs to be aptly designed to handle these issues.

Labour market mismatches, meaning that an individual's acquired education does not match with the required education for his/her occupation, is an issue that can be costly to society. When it comes to the unfavourable form of mismatch (overeducation), the state is paying via education costs and loss of efficiency, and the overeducated individual by lesser labour market outcomes and emotional cost (Piracha, Tani, and Vadean 2012). Generally, labour market mismatches are more common among immigrants (Leuven, Oosterbeek, et al. 2011), and this is also the case for the Swedish labour market (Joona, Gupta, and Wadensjö 2014).

A common problem in researching labour market outcomes for immigrants in general, and mismatches in particular, is that most often, pre-migration data is not available. Information on an individual's occupational and educational history is important when trying to find explanations for observed outcomes, but for most studies, only post-migration data is available. This study uses a unique data source on Swedish immigrants, providing information on pre-migration education and occupation, and allows for taking this important piece of the explanation into account when studying labour market mismatches in Sweden.

Using this newly complied data source of immigrants arriving to Sweden between 1970 and 1990, the objective of this thesis is to know more about what role an individual's pre-migration history might have in explaining post-migration labour market mismatches. The study adds to the field both by contributing knowledge on the Swedish situation using unique pre-migration information, but also to the general field, by producing results that can be added to a field in need of empirical research.

1.1 Research question

This thesis aims to answer the following research question: does an immigrant's pre-migration labour market mismatch influence the possibility of being mismatched in the Swedish labour market? If an association is found, what is the direction and magnitude?

1.2 Outline

This thesis is organised as follows: first (section 2), a theoretical framework and previous research on the possible aspects of immigrant labour market mismatches are presented. Sec-

tion 3 will go through the data used, and in section 4, methods and econometric models are discussed. Descriptive results, and later the results from the models, are presented in section 5, and synthesised with the hypotheses in section 6. The final discussion of the findings is found in section 7.

2 Theory and previous research

The study of labour market outcomes for immigrants is a field of research that incorporates many sub-fields. Even without introducing the complexity of an individual's trajectory before arriving in a new country, there are still many different theories concerning what predicts labour market outcomes. With the added information of an individual's pre-migration history, the mechanisms become even more complicated. Since many perspectives are needed to cover the possible explanations for the labour market mismatches of immigrants, this theory section will draw from three fields of research: firstly (section 2.1), theory and results from general studies on labour market mismatches, and secondly (section 2.2) theory and results on general labour market outcomes of immigrants. The third field (section 2.3) presents studies on the combined subject of labour market mismatches among immigrants. Finally (section 2.5), the presented research is used to construct hypotheses and a priori expectations.

2.1 Labour market mismatches

A labour market mismatch can arise from a number of sources. An income mismatch exists when a person's income deviates from what can be expected for the current occupation. A horizontal educational mismatch exists when the field of a person's education deviates from the field of the current occupation. The most commonly researched, and also the one that is the focus of this paper, is the vertical educational mismatch. This type of mismatch is defined as a discrepancy between the level of education acquired by a person, and the required level of education for this person's current occupation. In general, the state of overeducation is the most researched. Overeducation can be argued to be of greater policy interest, in the sense that it is associated with costs for both state (providing the education) and individual (monetary and emotional cost). This section will give an overview of the existing theoretical perspectives on vertical mismatches, and common results from previous research in the field.

2.1.1 Human capital theory

The definition, and even the possible existence of a labour market mismatch, depends on one's theoretical perspective. An influential work by Gary Becker (1964, *Human capital*) defines the wage outcome of a worker as representative of the worker's marginal productivity (Sala et al. 2011, p.1027). In this framework, even if a worker has more education than what could be considered as required, the wage reflects the true value of the worker's productivity (without market imperfections). By this definition, no mismatches can exist. According to McGuinness (2006, p.387), the interest for labour market mismatches, and educational mismatches in particular, was brought to real interest by Richard Freeman's study of the US labour market in 1976. In this study, it was concluded that the declining returns to education among American college graduates was due to excess supply from general overeducation. Even though this study did not consider overeducation at the individual level, but rather as a market collapse,

it drew the attention of researchers (Sala et al. 2011, p.1026).

The seminal work by Mincer (1974) introduces the human capital earnings function, which has become widely used within economics. Using this function, an educational mismatch can exist within the Human Capital framework: according to Mincer's model, surplus schooling exists as a compensation for a lack of work-specific human capital McGuinness (2006, p.390). In this sense, educational mismatch can be viewed as a problem of omitted variables. In general, Human Capital theory can be argued to view educational mismatches mostly as a short-term problem of insufficient human capital, if the concept is at all recognized.

2.1.2 Job competition theory

Job competition theory was introduced by Thurow (1975), and it puts the demand side of the labour market in command of an individual's outcomes. According to this theory, prospective workers are put in a queue and ordered by how much on-the-job training would be required to make them productive. A highly educated worker is always selected before a lower-educated worker, which means that if the supply of workers is higher than the demand, overeducated workers will be selected. Since the employer sets the wage according to the job characteristics, not the individual characteristics, the overeducated worker then ends up with a wage penalty (Tsai 2010, p.607).

2.1.3 Job assignment theory

Where Human Capital theory focuses on the individual's characteristics, and job competition theory on the occupation's characteristics, job assignment theory does a bit of both. The theory was introduced by Sattinger (1993), and it empathises the interplay between the worker's choice of a job/sector and how wages are set. In job assignment theory, a wage outcome is influenced both by a worker seeking to maximise utility, and an employer seeking a certain kind of worker (McGuinness 2006, p.398). An educational mismatch according to this theory can arise from both the individual and the job, thus leaving the field open for explanations including individual ability as well as the characteristics of the occupation.

2.1.4 Career mobility theory

Complementing the different theoretical perspectives on mismatch, Sicherman and Galor (1990) presented a general theory on mobility in the labour market. Introducing the "probability of promotion" for a worker within the current company (ibid., p.177), the theory makes it possible to regard overeducation as a first career step. In other words, being overeducated in a short-term perspective in the beginning of your career is considered rational. The authors also theorise that overeducation could stem from a worker compensating his/her lack of job-specific human capital. Since both formal and informal human capital are needed to be qualified for an occupation, a worker with excess formal education need not necessarily be overqualified, considering the demands for informal on-the-job skills (Robst 1995, p.539). In a test of the career mobility theory using longitudinal data, Büchel and Mertens (2004, p.803) argue that their results "cast serious doubts" on the ability of career mobility theory to explain the presence of overeducation. Considering the data used in this thesis, the career mobility theory can not be fully utilised, since all individuals have had a job prior to

migration. It could, theoretically, be argued that the "starting over" that faces immigrants post-migration is comparable to a native's first job, but it would be harder to find a credible strategy to identify this using the data at hand. This theory will, therefore, not be considered in the hypotheses.

2.1.5 Signalling and screening

Within the signalling theory, the value of a persons's education does not necessarily lie in the education itself, but in the signal that the completed education sends to a potential employer: a signal of higher productivity. So, education increases earnings not because it increases a person's productivity, but because it signals that a person is "..cut out for 'smart' work." (Borjas 2005, p.241). According to the signalling theory, the potential employee invests both direct and opportunity costs to acquire the right signalling towards an employer, but this is dependent on the potential wage benefit exceeding the cost (Weiss 1995). As the signalling theory takes the employee's perspective, the screening theory takes the employer's perspective on the same concept, assuming that the employer uses acquired education level to screen and filter among applicants. The education level can then be seen as a proxy for (often unobserved) positive individual characteristics, such as productivity (Arrow 1973).

2.1.6 Previous results

The empirical research concerning outcomes from labour market mismatches came into focus again (after a spell of lower interest during the 1970s) with an article by Duncan and Hoffman (1981). This article introduced the ORU model (Overeducation - Required education - Undereducation) that is a modified version of the human capital earnings function (Mincer 1974), and the ORU model is now commonly seen in empirical research on mismatches. The results from Duncan and Hoffman (1981) showed that the returns to one extra year of overeducation are positive, although only half of the premium that comes from adding another year of required education (that is, education pays off better while employed on in an occupation with the same required education as the acquired). In general, many studies seem to come to this conclusion: overeducation is associated with a wage penalty (compared to the correctly matched workers of the same education), and this penalty is costly in many perspectives (Leuven, Oosterbeek, et al. 2011, p.290). The cost is carried both by individual and state, to different degrees, depending on the educational system. In later years, however, many researchers have started to question these findings. Two issues are often brought up as problematic: the self-selection problem, and the measurement-error problem, and these will be discussed below.

2.1.7 Self-selection and ability

Much of the research that finds wage penalties from overeducation are based on the OLS model, which often suffers from identification issues when used in this field. The classical problem of non-random assignment to treatment means that it is impossible to know if it's variation in the independent variable or variation in the error term that is affecting the outcomes variable (Angrist and Pischke 2008, p.12). The widely used ORU model introduced by Duncan and Hoffman (1981), when used in combination with OLS, is no exception to that rule (Leuven, Oosterbeek, et al. 2011, p.304). Up until quite recently, these issues have been

more or less overlooked (Leuven, Oosterbeek, et al. 2011). The use of standard OLS to model labour market outcomes from mismatches has been heavily criticised by some researchers (Tsai 2010, Korpi and Tåhlin 2009, Pecoraro 2014), and methods such as IV or fixed effects have been employed to try and get rid of the apparent omitted variable bias.

It can be argued that who ends up in an overeducated state and who is correctly matched is not a random process, but is influenced by the individual's characteristics (such as the commonly mentioned "innate ability"). If this is the case, the labour market outcome of a worker can not be argued to only come from the mismatched state itself, but might as well come from variation in ϵ. Using individual fixed effects to account for this individual ability, Tsai (2010) aims to find outcomes that have not been influenced by self-selection. Her results suggest that the proposed wage penalty from overeducation disappears when individual effects are being controlled for. This is interpreted by the author as an indication of the real reason behind the wage penalty being selection into overeducation by lower-ability workers (ibid., p.611).

Pecoraro (2014) mentions the same critique regarding omitted ability variables. He addresses the issue using a fixed effects approach, but also tries to control for ability bias in an OLS setting using a proxy variable (ibid., p.311). The proxy variable consists of the difference between the expected and the realised wage for an occupation. Using this variable needs the assumption that the set wage measures a worker's individual productivity/ability that is unrelated to the acquired education. This method was first used by Chevalier (2003), and both Chevalier and Pecoraro reach the conclusion that there is indeed negative selection on ability into overeducation.

2.1.8 The measurement-error problem

The second issue that is often mentioned concerning the identification of mismatch outcomes is measurement errors in the variable of interest. The fact that a mismatch variable consists of a difference between two schooling level variables (acquired and required education) makes potential measurement errors an even bigger problem than if schooling were used as a single variable (Leuven, Oosterbeek, et al. 2011, p.306). Also, using a fixed-effects framework to account for unobserved factors may inflate the error even further, since the fixed-effects frameworks are known to be unforgiving towards measurement errors (Angrist and Pischke 2008, p.168).

The case of these measurement errors has been discussed and approached in different ways. Tsai (2010, p.613) is aware of the fact that her fixed-effects results could be argued to be due to measurement-error bias. Both a numerical approach and survey data are used to test the sensitivity of the results, and Tsai finds that the results hold. These tests are, however, questioned by Leuven, Oosterbeek, et al. (2011, p.309), who argue that they are not shown to be consistent. Verhaest and Omey (2012, p.77) also discuss the measurement problem at length, and note that the bias in overeducation outcome studies usually is directed downwards. This would mean that the wage penalties usually found are understated.

A common strategy to combat measurement-error bias in mismatch models is to instrument different types of mismatch variables on each other. Robst (1994) uses this technique and

finds that the wage penalty is even higher than first estimated. Verhaest and Omey (2012, p.86) also use the technique, combined with a fixed-effects model, and find that measurement error is a substantial source of downward bias. A problem with the IV approach as corrector of measurement error is, as pointed out by Leuven, Oosterbeek, et al. (2011, p.308), that it only applies when the measurement error is classical (that is, when the measurement error is uncorrelated to the true mismatch value). In many cases, it can be argued that the measurement error is non-classical, which means that IV correction will not eliminate the bias.

2.2 Labour market outcomes for immigrants

Separate from the impact of labour market mismatch in general, labour market outcomes for immigrants is an area of research by its own right. Early works of Barry Chiswick found results indicating that immigrant wage outcomes are significantly lower than natives, but that they catch up and eventually surpass native earnings (Chiswick 1978). The initial dip in earnings is attributed to a lack of country-specific human capital, which is said to be acquired with growing labour market experience. The approach and conclusions by Chiswick were criticised by Borjas (1985), who argued that Chiswick's results are a product of his cross-sectional approach, and that the results more likely come from differences between cohorts. Borjas finds, instead, that the "quality" of the cohorts are different, with the earlier cohort being of higher quality. This leads to a situation that might look like an improvement with more years spent in the host country, when it is actually the earlier cohort having a higher quality than the later.

In later years, this field of research has widened considerably, and many explanations for immigrant outcomes have come into focus. In this section, theories and results concerning the potential mechanisms behind immigrant labour market outcomes will be presented and discussed.

2.2.1 Selection into migration

Self-selection is an important issue in most parts of empirical microeconomics and this is also certainly true for the study of migrants. A model to analyse self-selection into migration was developed by Borjas (1987), based on the notion that people who migrate can not be expected to be chosen at random. The model presented by Borjas includes a framework to analyse the mechanisms behind positive and negative selection on skills (and other factors). The outcomes from this selection process depend on a number of factors: the transferability of skills, the income distribution in the source country, and the returns to education in both source and host country (Rooth and Saarela 2007, p.91). The combination of these sources can create a selection that is very specific: for example, a host country can attract migrants that are negatively selected on observable skills (low education), but positively selected on unobservable skills (high ability). Return migration, which also has a large effect on the composition of migrants in the host country, can be argued to be driven by the same forces that drive the selection into first migration (ibid.).

Concerning the current state of selection into migration in the OECD countries, Belot and Hatton (2012) conclude that positive selection on education has more to do with low physical and cultural distance than the usually mentioned wage incentives. Also, having a colonial

legacy in a source country lowers the poverty constraints that usually block poorer people from migrating (Belot and Hatton 2012, p.1125). In the case of Sweden, Rooth and Saarela (2007) find that Finnish immigrants are negatively selected on education, since returns to education for highly educated Finnish workers are higher in Finland than in Sweden.

2.2.2 Discrimination

An important factor affecting the labour market outcomes of immigrants is ethnic discrimination in the labour market. The observed ethnic income gap in Sweden means a -15/-22% wage penalty for southern EU/non-EU second generation immigrants, compared to natives (Nordin and Rooth 2009, p.488). Together with Denmark and Belgium, Sweden has the highest ethnic employment gap in the OECD countries (OECD 2015, p.69). In theory, these gaps in labour market outcomes can arise from discrimination, but also from unobserved individual heterogeneity, and the nature of the discrimination issue makes it especially difficult to show empirically that it exists.

Trying to find evidence of labour market discrimination in Sweden, Carlsson and Rooth (2007) construct an experiment using fake job applications with ethnic/native-sounding names being the variation of interest. They find large differences in callback rates, where Middle Eastern-sounding names names receive as much as 50% less callbacks. The authors discuss the potential mechanisms behind the results, but cannot safely say which type of discrimination is behind. On the same topic, Nordin and Rooth (2009) exploit military enlistment intelligence tests as a proxy for individual ability to identify discrimination. Their results suggest that ethnic discrimination in Sweden affects employment possibilities, but not wages (p.504). This is attributed to either labour market discrimination and/or unobserved variables (p.496).

2.2.3 Labour demand

The demand side of the labour market for immigrants affects both labour market outcomes and migrant flows. The paper by Borjas (1985) emphasises changes in the supply side (cohort quality) but does not rule out a "fall in demand for immigrant labour" (p.485). In later years, the interplay between macro-economic fluctuations and immigrant outcomes have been the focus of several studies. Dustmann, Glitz, and Vogel (2010) research the labour market responses of immigrants during economic downturns in the UK and Germany. These two countries are shown to have large differences in immigrant population composition, both regarding levels of education and countries of origin (p.4). These differences might in turn produce differences in outcomes. Dustmann et al. find, however, similar responses to economic crisis in both countries, as lower-educated immigrants experience a heavier employment penalty than highly educated immigrants. The main immigrant outcome from economic downturn (in both countries) is higher unemployment, not lower wages (p.14).

A similar study on the Scandinavian labour market was presented by Rosholm, Scott, and Husted (2006), where different immigrant cohorts in Denmark and Sweden are followed on the labour market. Similar to the results of Dustmann et al., labour market outcomes in the two countries show a common pattern, despite the differences in unemployment patterns during the study years (p.335). The results show that immigrants in both countries experienced

declining opportunities in employment during the study period (1985-1995). The authors present the theory that this decline is due to a change in the labour market structure (the demand side), shifting to a market that is more demanding in terms of informal human capital and country-specific skills. The change can be described as the switch from "Fordism" to "Toyotaism" (Helgertz 2010). Similar results are found by Bevelander and Nielsen (2001), who use a decompositional method to find sources of variation in the Swedish immigrant-native employment gap. They find that the deteriorating employment conditions for Swedish immigrants between 1970-1990 arise from changes in unobserved variables rather than what can be observed using regular socio-economic variables (p.463). These unobserved variables are suggested to be labour market discrimination, or the same structural changes pointed out by Rosholm, Scott, and Husted (2006).

2.2.4 Transferability of human capital

Education and experience that has been acquired abroad does not always hold its value in a new country. Transferability of human capital for migrants can be an issue, as quality of schools, educations and certificates might differ between countries. Also, cultural and linguistic distances between source and host country can have an effect on this process in both directions.

Friedberg (1996) makes a distinction between formal human capital (years of education) acquired abroad and domestically in her article on Israeli immigrants. It is found that human capital is imperfectly transferred between countries, but that the rate of portability varies with certain groups of countries. Immigrants to Israel from "Europe and the Western Hemisphere" have higher rates of return to their pre-migration human capital than do migrants from Africa and Asia (p.246). Friedberg theorises that this might be due to differences in school quality, or discrimination. It is also found that the returns to labour market experience acquired abroad is insignificant in general. Chiswick and Miller (2002) study foreign-born men from non-English speaking countries in the US, and find that English language proficiency has a significant positive effect on wage outcomes. The concept of linguistic distance (meaning how far away an immigrant's mother tongue is from the host country language, linguistically) is used as a measure of "skill transferability". Also, it is found that living in a "linguistically concentrated area" has a negative effect on wage outcomes (ibid., p.49). In the case of Sweden, Helgertz (2013) researches the impact of linguistic distance of immigrants on their labour market outcomes (in the years 1970-1990). The results show that an increasing linguistic distance is associated with negative labour market outcomes: proficiency in a language from the Germanic language family gives an advantage in the likelihood of acquiring a job (ibid., p. 462). There does not, however, seem to be any large differences between the non-Germanic languages. Judging from the results from ibid., language skills are important in order to succeed in the Swedish labour market during the study period (which is also the study period in this thesis).

2.3 Mismatches among immigrants

The combined topic of labour market mismatches among immigrants is generally underresearched, but on the rise (Piracha, Tani, and Vadean 2012). The field exists in the cross-

section between the mismatch literature (section 2.1) and the literature on immigrant labour market outcomes (section 2.2), and carries with it the complexities from these two fields. Depending on one's choice of outcome variable, different theories and results have been presented. Regarding incidence, it is generally reported that immigrants have a higher degree of educational mismatch than natives (Chiswick and Miller 2008, Leuven, Oosterbeek, et al. 2011, Piracha, Tani, and Vadean 2012). Possible explanations as to why immigrants experience labour market mismatches are individual ability (section 2.1.7), selection into migration (section 2.2.1), first-job tenure (section 2.1.4), discrimination (section 2.2.2), signalling (section 2.1.5) and language/transferability (section 2.2.4).

There are two studies on the subject that are of special importance in relation to this thesis: the study on pre- and post-migration mismatch mechanisms in Australia by Piracha, Tani, and Vadean (ibid.), and the study of mismatch outcomes and state dependence of Swedish immigrants by Joona, Gupta, and Wadensjö (2014). The first study utilises data collected by Australian authorities, in a similar manner to the SLI database used in this thesis (described further in section 3.2). The authors use signalling as their hypothesis for why immigrants could experience both pre- and post-migration mismatch. The signalling proposed is not an education signal, but a signal from the most recent employment pre-migration (Piracha, Tani, and Vadean 2012, p.2). A notable difference between Sweden and Australia is the Australian migration policy implemented in 1995, described as the "skill stream" of migrants. Through a point system, a large share (around 50% in 1999) of the migrants accepted into Australia are graded according to the benefit they could bring to the country, which puts emphasis on occupational experience and language skills (Miller 1999, p.193). Since "business skills" and references are important in the point scheme, the migrants accepted will need to have this formally in order, which makes the signalling theory more likely. It is, on the contrary, likely that some groups of migrants (such as refugees) will have a harder time producing the necessary work credentials in order for signalling to take place. The results from this study show that the existence of a pre-migration mismatch is the strongest predictor of a post-migration mismatch, after controlling for a number of demographic and occupational variables (these results are compared to the findings of this thesis in section 6). The authors attribute the results to "ability signals" from the pre-migration mismatch status (Piracha, Tani, and Vadean 2012, p.19).

The second study by Joona, Gupta, and Wadensjö (2014) focuses on the post-migration mismatch outcomes of immigrants in Sweden. By using a rich source of register data, individuals are followed over time in order to know more about the possible differences in state dependence of mismatches. The previous research presented by the authors suggests that there is indeed a strong state dependence within overeducation, but that the potential heterogeneous effects between natives and immigrants are underresearched (ibid., p.4). With regards to incidence of mismatches among Swedish immigrants, the authors find a higher incidence of overeducation among immigrants compared to natives (p.10). State dependence is modelled using a dynamic random effects model with Mundlak correction, and the results show a very high degree of state dependence in overeducation in general, for both natives and immigrants. Immigrants, and especially non-Western immigrants, have even higher rates than natives, which is attributed in part to imperfect transferability of human capital (p.20). In contrast to this thesis, Joona, Gupta, and Wadensjö (ibid.) do not have access to pre-migration mismatch

data, so there can be no direct comparison of results, but the study is still a good indication of the Swedish labour market experience for immigrants.

2.4 Swedish immigration history

Prior to the Second World War, immigration to Sweden was mainly made up by North American return migrants. After the Second World War, Sweden had an intact infrastructure and industry, and simultaneously, demand grew for raw materials to re-build Europe. Growing demand increased the pressure on the industry, and made the Swedish government recommend an increased inflow of labour migrants of around 10.000 per year (Helgertz 2010, p.3). The first cohorts to arrive in the 1940s and 1950s were labour migrants, but later, inflows of political asylum seekers began to arrive: political instability in Greece, Poland, Chile and Yugoslavia (amongst others) led to refugees becoming the dominant part of the migrant stock. For Poland, the repression of the Jaruzelski regime made immigration to Sweden peak in 1982, and for Chile, the years just after the 1973 military coup marked the height of immigration to Sweden (Klinthäll 2007, p.584). In recent years, apart from work and study migrants, waves of immigration have come consisting of refugees from crises in Iraq, Afghanistan and Syria, with following waves of tied movers (SCB 2016a).

2.5 Hypotheses

Following the presented theory and results, a number of a priori expectations can be stated. As a clarification, it should be mentioned that the topic and research question of this thesis models a mechanism that is not the most commonly researched: the relation and persistence of a mismatch over time, with the intermediate disturbance of an international migration. Much of the existent theory and research on mismatches focuses on a secondary labour market outcome, such as wage, and not the existence and persistence of the mismatch per se. This means that the validity of comparison of these theories and results with the research question in this thesis can be discussed. In order to make use of the theories and results, it is necessary to make the comparison: higher wage and higher employment can be considered "better" labour market outcomes, which $\approx$ less likelihood of overeducation. The five proposed hypotheses are:

- *H1 - signalling*: the effect of a "pure" (randomly assigned) pre-migration mismatch can be theorised to transfer from pre- to post-migration via signalling, which would manifest through a higher likelihood of a post-migration mismatch from a pre-migration mismatch.
- *H2 - individual ability*: if a pre-migration mismatch can be theoretically attributed to individual ability, the presence of a pre-migration mismatch should make a post-migration mismatch more likely. Since ability can be considered time-invariant, this effect should also be constant over time.
- *H3 - discrimination*: if labour market discrimination can be argued to contribute negatively also to matching, a general shift downwards (i.e. undereducated $\rightarrow$ required education or overeducated, required education $\rightarrow$ overeducated) would be an outcome from arriving in Sweden for some groups of immigrants.

- *H4 - labour demand*: if lower labour demand for immigrants also affects mismatch status, a higher likelihood for downwards shift will be visible in cohort effects for years of boom or bust.

- *H5 - transferability*: considering the heterogeneity between countries of origin, immigrants from different countries are expected to have different likelihoods of mismatch based on different levels of transferability and linguistic/cultural distance. Also, the likelihood of overeducation is expected to decline with increasing years since migration, as country-specific human capital is accumulated.

These hypotheses will be commented on using the results from this thesis in section 6.

3 Data

There are two principal data sources behind this thesis. For data on the Swedish side, official Swedish data sources such as the tax registry and 5-year censuses have been used to construct a panel dataset for each individual. This dataset includes income, civil status and internal migrations, all of which are data and events occurring in Sweden (post-migration).

For the pre-migration information, the Swedish Longitudinal Immigrant database (SLI) has been used. This database consists of a sample of immigrants to Sweden starting in 1968, and was collected continuously until 2005 (Helgertz 2010). From the SLI, a random sample of 17.074 individuals is the base sample for this thesis. The SLI is a unique source of pre-migration information, and contains information on origin, education, occupation and language. For this particular research question, the information on pre-migration occupation is of special importance: this variable allows for the construction of a pre-migration mismatch status, which is rarely seen in this field of research. As mentioned in section 2.3, the data source used in Piracha, Vadean, et al. (2013) and Piracha, Tani, and Vadean (2012) is to the author's knowledge the only comparable source containing this information. This section will describe how variables of interest have been constructed, and how the final sample was selected.

3.1 Sample

As stated previously, the SLI sample contains 17.074 individuals to begin with. After removing individuals without a job title, without data on acquired education and whose job titles could not be matched against the official list of Swedish occupational titles (SSYK), 8.848 individuals remain. Also, individuals with previous occupations in the military sector (ISCO code 0, table 3) were excluded, since the military occupational field does not operate in the same way as the rest of the labour market. 17 individuals with inconsistent information on birth year (more than one unique year present) were also removed. Before joining the SLI data to any additional data, the sample size is 8769 individuals (the removal progress can be followed in table 1).

There are some additional aspects of the sample that need to be handled when modelling. Already mentioned are the military occupations, but there are also other special groups: occupations within performing culture and sports are very hard to assign a believable required education, since one can reach higher levels within these fields both with and without higher

Reason for removal	*Count*	*Left in sample*
Original sample		17074
No data on occupation (pre-migration)	3886	13188
No data on education (pre-migration)	2055	11133
Not matched against SSYK	2285	8848
Military sector	62	8786
Conflicting birth years	17	8769

Table 1: Sample size

education. Individuals who were self-employed prior to migration might not be representative, since they can be considered to have a higher likelihood of starting their own business also in Sweden, which would put them outside the regular labour market. Furthermore, individuals outside the age range of 16 to 54 at arrival in Sweden can not be considered representative: the younger individuals might receive education in Sweden, and the older individuals have a higher likelihood of entering directly into retirement (Helgertz 2010, p.40). These categories (individuals within performing arts or sports, self-employed, and outside the age range) will be excluded from the main model results, but included in as a part of sensitivity tests (section 5.5).

3.1.1 Censuses

Income data from the tax register was joined to the SLI without losing any individuals from the sample. Information on occupations and civil status, however, are only present in the 5-year censuses. Censuses are available, for this thesis, for the years 1970, 1975, 1980, 1985 and 1990. These censuses are the data points available for following migrants in Sweden, and as migrants arrive continuously during the study period, many different combinations of presence in data are possible. Table 2 contains the number of individuals present on number of censuses. This number of individuals corresponds to the final sample used for modelling, and thus excludes the categories mentioned in section 3.1, and also individuals with missing data.

Number of censuses present	*1*	*2*	*3*	*4*	*5*
Individuals	3875	2038	1216	594	181
% of final sample	100,00%	52,53%	31,34%	16,93%	4,66%

Table 2: Number of individuals per census

As seen in the table, 3875 individuals are present in at least one census (this one being their first), and only 181 individuals are present for all five censuses. "Present" is defined as having a mismatch status, which implies being employed at the time of census. Note that this can be in any of the five censuses - when modelling, all data is evaluated at time of census and the individuals are pooled from different immigration cohorts. No individual appears twice within one census, but individuals are followed over time in as many censuses as possible.

3.2 Pre-migration data

The SLI pre-migration data contains title of occupation for every individual before migrating to Sweden. This information has been collected by the Swedish immigration authority during meeting concerning permits (Helgertz 2010). The occupational title information was hand-typed and contained numerous spelling errors. In order to clean and synthesize the present occupations, the hand-typed data were matched against the SSYK (SCB 2016c). During the matching, a Levenshtein string distance score was calculated between the two sources, and this was used to facilitate the subsequent manual work of finding a correct occupational title for each pre-migration title. In total, around 3500 unique strings were matched.

3.2.1 Required education

In order to calculate a mismatch, information on required education for an individual's current occupation is needed. There are several methods that can be used to obtain this (presented in Leuven, Oosterbeek, et al. 2011):

- *self-assessment*: individuals are asked how much education they believe is needed to perform (or be recruited to) their current job. The advantage of this approach is that it is gives a direct insight in the requirements, but workers might also be misinformed about or exaggerate how much education is needed.

- *job analysis*: this approach makes use of the existing international (and national) occupational classifications (for example, ISCO), in which a required education for each occupation has been decided upon. Using this technique might give more consistent results, but might also remove valuable heterogeneity.

- *realised matches*: making use of the already employed individuals within an occupation, it is possible to calculate a representative education from these. Both the mean and mode education within an occupation have been used by different researchers. Using a numerical variable such as years of education, the definition of a mismatch has often been defined as +/- 1 σ from the mean, which can be argued to be an arbitrary definition. Also, this method has been criticised for only reflecting the supply/demand forces and not a true general required education (ibid., p.293).

Since the SSYK (which has been matched against the SLI occupation titles) also contains a required education for each occupation (originating from the ISCO-08 classification), the job analysis approach would be a first suggestion for the current data. In the case of the SLI, however, which contains immigrants from many different countries, the job analysis method was deemed inappropriate. At closer inspection of the data, it was found that specific countries and occupations contained heterogeneity that would have created bias if disregarded: for example, more than 90% of the nurses from Iran in the SLI sample had a secondary education, while the SSYK/ISCO classification dictates that being a nurse requires a post-secondary education. If job analysis were used here, almost all Iranian nurses would turn out as mismatched, even though it can easily be argued that this is probably not the case. The self-assessment method was impossible to use, since data had already been collected, so the realised matches method was chosen to compute the required education.

Code	*Group*
1	Managers
2	Professionals
3	Technicians and associate professionals
4	Clerical support workers
5	Service and sales workers
6	Skilled agricultural, forestry and fishery workers
7	Craft and related trades workers
8	Plant and machine operators, and assemblers
9	Elementary occupations
0	Armed forces occupations

Table 3: ISCO-08 major groups

3.2.2 Creating the required education-variable

In order to best capture the nature of an occupation within the country context that it exists, the matched and cleaned occupation titles from SLI were grouped by title and country, and from these groups, the mode education was selected. The chosen cut-off group size was 10 individuals. If an occupation had less than 10 individuals within a country, it was grouped on title and country-group. And, if title and country-group was not enough, the remaining required educations were chosen using country groups and the 10 ISCO "major groups" (listed in table 3). These ISCO groups exist with the SLI data as a part of the connected SSYK classification, and consist of a broad classification of occupations, both horizontally (sector) and vertically (level of management).

The share of individuals classified with each method is listed in table 4. The distribution these individuals, as shown per the calculated required education, is fairly stable across the sample. A deviation can be seen in the "Countrygroup+ISCO"-method group, which has a larger share of post-secondary individuals. This can be explained by the fact that there exists a larger number of unique occupation titles in the higher spectrum of the education scale (in SSYK, 51% of the titles are at post-secondary level, and only 15% at primary level). This means that there are more titles to group by in the higher end, and less chance of passing the 10-individual cut-off.

Classification method	*Primary*	*Secondary*	*Post-secondary*
Country	56,35%	20,61%	23,04%
Countrygroup	57,77%	18,20%	24,03%
Countrygroup+ISCO	46,44%	8,07%	45,49%

Table 4: Percentage of individuals by classification method

3.2.3 Acquired education

The SLI information on an individual's acquired education prior to migration is available in several variables: years of education (numeric), vocational education (free-text) and education category (free-text). The "years of education" variable turned out to be unusable, since data

was missing for 43% of the individuals. The "education category" variable, however, is present for 81% of the individuals and contains the three education groups "Primary", "Secondary" and "Post-secondary". This variable is used as the acquired education level for individuals prior to migration. The variable "vocational education" contains free-text information that was found to be too incoherent for use in this analysis. For example, 78% of the nurses in the sample have information in this variable, but there are 37 different strings containing the information just for this occupation. Since the cleaning of this data would be manual for the most part, there was not enough time in this project to make use of this information.

3.3 Constructing a mismatch variable

To construct a vertical educational mismatch variable, two components are needed: acquired education level of an individual, and the required education level of the current occupation of the same individual.

The variable is computed according to this scheme (AE = acquired education, RE = required education):

$$AE > RE = Overeducation$$
$$AE = RE = Matched$$
$$AE < RE = Undereducation$$

For the pre-migration data, the distribution of educational mismatches can be seen in Table 5. The distribution of the pre-migration mismatches shows an amount of mismatches that is similar both over and under the "required education" level. These figures differ somewhat from the corresponding distribution presented by Piracha, Tani, and Vadean (2012), where only 8% of the migrants where overeducated, and 24% where undereducated before migrating. These figures, however, reflect the underlying Australian immigrant policy, which is selective towards highly educated immigrants (ibid., p.8). A higher degree of undereducation is therefore to be expected in Australia compared to Sweden, where the immigration policy does not favour highly educated migrants (Klinthäll 2007).

Pre-migration mismatch status	*Individuals*	*% individuals*
Overeducated	630	16,26%
Required Education	2775	71,61%
Undereducated	470	12,13%
Sum	3875	100,00%

Table 5: Pre-migration mismatches

The mismatch variable should also be broken down on other variables: table 6 contains the distribution of pre-migration mismatches by acquired education (also pre-migration). Firstly, it can be noted that there are no overeducated individuals with primary education, and no undereducated individuals with post-secondary education. This is by definition of the variable. It is noteworthy that only 12,7% of the correctly matched individuals have a secondary education, which indicates that this educational category is more prone to mismatches. This can be seen better in the column-wise percentages in Table 7.

Mismatch status	*Primary*	*Secondary*	*Post-secondary*	Sum	*Primary*	*Secondary*	*Post-secondary*	Sum
Overeducated		385	245	630		61,11%	38,89%	100,00%
Required Education	1629	353	793	2775	58,70%	12,72%	28,58%	100,00%
Undereducated	259	211		470	55,11%	44,89%		100,00%
Sum	1888	949	1038	3875				

Table 6: Pre-migration mismatches by acquired education

From this table, it is obvious that the secondary education level has a higher share of mismatches than the primary and post-secondary levels. Given, the secondary class can be mismatched in both directions, but it is still noteworthy that the majority of individuals in this education class is mismatched. This issue is discussed further in section 3.5.

Mismatch status	*Primary*	*Secondary*	*Post-secondary*
Overeducated		40,69%	23,68%
Required Education	86,30%	37,12%	76,32%
Undereducated	13,70%	22,19%	
	100,00%	100,00%	100,00%

Table 7: Pre-migration mismatches by acquired education - column-wise percentages

For a small number of individuals, two occupations were reported. In these cases, there will only be a deviation in case the two occupations has different required educations (which was the case for 147 individuals). For these cases, the highest required education was chosen to compute the mismatch status (but models were also run using the lowest, see section 5.5).

3.4 Post-migration data

The post-migration data from the Swedish registries is generally straightforward to use. Most of the data used is constructed as long panels, where demographic events, such as migrations and civil status changes, are recorded with an exact time-stamp. As data was processed and selected, a number of decisions had to be made: firstly, the data was down-sampled to a yearly basis. In the case of individuals changing their citizenship, civil status, or residence several times during a year, only the last event within the variable was used. Secondly, repeat migrations were considered: if an individual immigrates to Sweden, emigrates, and then immigrates again, the "years since immigration" variable will only count the years spent in Sweden. Generally, peculiar edge cases that were found (such as multiple birth/death dates) were dropped, but these were very few in total (<50).

3.4.1 Required education

The required education for an occupation in Sweden is provided by SCB as part of the census, via the socioeconomic classification variable SEI (SCB 2016b). This variable, however, was not delivered with the 1970 and 1975 censuses. To mediate this problem, the SEI scores from the 1980/1985/1990 censuses were used to calculate a mode value (using the "realised matches"-method as described in section 3.2.2). This calculation, compared to the calculation on pre-migration data, did not suffer from lack of observations, and no lower threshold was needed (the three censuses together contain >750.000 observations). The method might create

a bias, arising from the potential difference in educational attainment within occupations over time in Sweden (but this can be considered to be low-risk, since the periods are close in time).

3.4.2 Country of origin

Distribution of country of origin for all individuals in the sample can be seen in Table 8. Countries that had a very low number of individuals (<10) have been excluded from the sample (examples of these countries are Eritrea, Russia, Bosnia and North Vietnam). Germany has been chosen as the base category for modelling, since it is well represented in the sample and culturally close to Sweden.

Country of origin	*Individuals*	
Chile	438	11,30%
Germany	658	16,98%
Greece	542	13,99%
Iran	165	4,26%
Poland	687	17,73%
Turkey	366	9,45%
USA	376	9,70%
Yugoslavia	643	16,59%
Sum	3875	100,00%

Table 8: Pre-migration mismatches by country of origin

3.4.3 Metropolitan place of residence

All Swedish municipalities have an urban status category of metropolitan/non-metropolitan, which is delivered by SCB. The metropolitan municipalities are the ones in and around the three largest cities in Sweden (Stockholm, Göteborg and Malmö), and the distribution of individuals in the sample on this status can be seen in Table 9. Since around 17% of the total Swedish population lives in these three cities (SCB 2015), it is clear that living in a metropolitan area is overrepresented in the sample. There does not seem to be, however, any skewed representation within the mismatch categories.

Pre-migration mismatch status	*Non-metropolitan*	*Metropolitan*	Sum	*Non-metropolitan*	*Metropolitan*	
Overeducated	214	416	630	33,97%	66,03%	100,00%
Required Education	1085	1690	2775	39,10%	60,90%	100,00%
Undereducated	176	294	470	37,45%	62,55%	100,00%
Sum	1475	2400	3875	38,06%	61,94%	

Table 9: Pre-migration mismatches by residence status

3.5 Data issues and bias

Concerning the data used, there are several issues that might compromise the validity and reliability of the results:

- *precision of census data*: an important restriction on the data used in this thesis is the interval between censuses. These censuses are only performed every five years, which means that individuals have time to switch jobs several times between censuses. According to the figures of Joona, Gupta, and Wadensjö (2014), around 14% of Swedes switch jobs each year, but with no significant differences in frequency between natives and immigrants. Theoretically, this could bias the results in both directions, and it is only the jobs both that come and go between two censuses that pose the real issue, as these may be changing the mismatch status. There is unfortunately little to be done about this issue other than keep it in mind when interpreting the results.

- *overrepresentation in secondary education*: As shown in section 3.3, there seems to be an overrepresentation of educational mismatches in the "secondary education" category. It is, by definition, possible to be mismatched in both directions in this educational category, but it is also possible that there might be a larger inherent likelihood of being mismatched being in this category - due to unmeasurable labour market variables. As noted in section 5.5, different cut-off values for the mismatch variable generation have been tried but this did hardly influence the mismatch distributions. The possible bias arising from this overrepresentation is thus hard to demonstrate, or to predict in direction.

- *attrition*: the panel data sample suffers from attrition of several kinds: mortality/old age retirement, outmigration and unemployment. Out of the three, mortality/retirement attrition is the least problematic, since it is likely to be random enough for the purposes of this thesis. Concerning outmigration, however, there is good reason to suspect that this is not random (as discussed in section 2.2.1), and it is reasonable to expect that individuals experiencing failure in the labour market are more likely to emigrate. This would bias the likelihood of overeducation downwards, in the theoretical case that these individuals still were on the labour market. The problem of only considering employed individuals (under which the unemployment attrition falls) is discussed closer in section 5.6.

- *measurement error*: as discussed in section 2.1.8, mismatch variables are especially sensitive to measurement errors, since both errors in the acquired and required education variables can influence the result. As for the required education variable (pre-migration), sensitivity tests using different cut-offs have been tried, without finding any differences. The pre-migration acquired education and occupation information is self-reported, and it is possible that these suffers from social desirability bias. This type of bias is mostly mentioned in cases of self-reported data concerning behaviour and political views (Kaminska and Foulsham 2013), and is therefore considered low-risk in the case of this thesis. The post-migration data on occupation and education used in this thesis is Swedish registry, which usually is considered to be of high quality.

- *post-migration education*: when calculating mismatch status, both pre- and post-migration, the acquired education level pre-migration is used. However, it is possible that individuals educate themselves further in Sweden, but data on this has not been available in this thesis. If a large number of individuals were to re-educate themselves in Sweden and thereby change education category, this could bias the results upwards, and overestimate the likelihood of being overeducated.

- *vocational data*: as mentioned in section 3.2.3, free-text information on an individual's completed vocational education was not included because of time constraints. Some rudimentary inspection of the data shows that the vocational title in most cases seem to correspond to the acquired education level entered (i.e. doctors having post-secondary education) - but more time would be needed to confirm that this holds for all of the sample. Depending on how common it is to report non-corresponding combinations (i.e. primary education + doctor), this issue can lead to an overestimation of the likelihood of being overeducated.

4 Method

The research question states that it is the probability of experiencing a labour market mismatch that will be the outcome variable of the analysis. This phenomenon is identified with the categorical mismatch variable (described in section 3.3), where the post-migration mismatch status will be used as dependent variable, and the pre-migration mismatch status being the independent variable of interest.

To answer the research question, two models are proposed: Model 1 uses an ordered logistic regression setup, and Model 2 uses a multinomial regression setup. Both models have an associative approach rather than causal, and the endogeneity problems connected to these models are discussed in section 4.4. In this section, the theoretical foundation of the methods (logistic regression and the ML estimator) will first be presented, and then the specifications of the two models.

4.1 Logistic regression

When estimating using probabilities using a linear model, the probability of an event $y_i|x_i$ occurring can go below 0 or above 1. In a Logit model, this does not occur, since a distributional assumption is made about the probability distribution of y_i: common choices are the standard normal distribution (probit) or the standard logistic distribution (logit). In this paper, the logit model will be used as a starting point, but sensitivity tests using the probit model will also be made (section 5.5).

Defining the function of interest as $w = x_i'\beta$, the logistic distribution function is given by:

$$F(w) = \frac{e^w}{1 + e^w}$$

Defining the probability of $y_i = 1$ as $p_i = P(y_i = 1|x_i)$, the log odds ratio of an event can then be defined as:

$$log\frac{p_i}{1 - p_i} = x_i'\beta$$

4.1.1 Ordered logit models

Using an ordered model for categorical data makes sense if the categories have an inherent order that can be exploited. The idea can be explained using the existence of a latent variable

y^* that controls the transition between the categories. In a simplified case, this latent variable could for example be individual ability, which can be argued to influence labour market mismatches. The latent variable (which is unobserved) can be defined as $y_i^* = x_i'\beta + u_i$, without an intercept (Cameron and Trivedi 2010, p.519). As the value of y_i^* increases, the probability of y_i taking a certain value changes (Verbeek 2008, p.210). For example, if ability could be measured on a scale from 0 to 12, the threshold points between overeducated and matched / matched and undereducated could lie at 4 and 8. This would give the logic:

$$y_i^* < 4 \rightarrow y_i = overeducated$$
$$4 <= y_i^* < 8 \rightarrow y_i = matched$$
$$y_i^* >= 8 \rightarrow y_i = undereducation$$

As the outcome of the ordered logit is interpreted, it can be referred to coefficients as being results of an increase in the underlying latent variable. In this case, that would mean a hypothetical ability increase (hypothetical, since ability is often seen as constant) that would result in a change in the mismatch status.

The ordered logit model uses the assumption of proportional odds between the categories, which means that the probability distributions for each of the outcome variable categories are assumed to be identical (Long and Freese 2006, p.151). This assumption is tested in section 5.2.3.

4.1.2 Multinomial logit models

In contrast to the ordered logit model, a multinomial model does not assume that there is a monotonic latent variable behind the choice of category, instead, the categories are treated as independent and mutually exclusive (Verbeek 2008, p.229). When calculating the probabilities, coefficients are usually interpreted in relation to a base category in the outcome variable, but base probabilities can be predicted for all categories using marginal effects methods. A frequently mentioned issue with multinomial models is the Independence of Irrelevant Alternatives (IIA), which is a situation that can arise when two or more alternatives in the dependent variable have the same practical implication. A common example of this is the choice between travelling by train, blue bus, or red bus - choosing one of the latter alternatives would imply a high utility also for the other (ibid., p.230). For the variable used in this thesis, however, this will not be an issue: the dependent variable is not a choice variable, and the categories are mutually exclusive by definition of the variable. Using a multinomial model will provide a different angle on the data, by not locking the outcome variable to an ordinal form. If there is variation that does not fall into the pre-defined ordinal pattern of rising ability/labour market success, it should be captured by a multinomial model.

4.2 Maximum Likelihood estimator

A non-linear limited dependent variable model is often modelled using a Maximum Likelihood (ML) estimator (ibid., p.211). The choice between Linear Probability Models, LPM (using OLS estimator), and ML estimators in limited dependent variable models depends on context and the underlying data-generating process, but it is also an issue of some debate. Using

OLS gives access to many of its wanted characteristics (for example, easier interpretation of marginal effects), but it has also been argued that it gives inconsistent estimates in limited dependent variable models (Horrace and Oaxaca 2006). In this thesis, an ML estimator will be used.

The ML estimator gets its name from the fact that once a probability distribution for the outcome is assumed, this function is maximised to give the most likely value of β (in a regression setting). Instead of estimating a value for y, an ML estimator estimates the likelihood that y takes a certain value. In general, the ML principle builds on the notion that a random variable y has a probability density function that depends on a set of unknown parameters θ, which gives the function $f(y|\theta)$. If n observations from this process can be argued to be independent and identically distributed (IID), the joint density function can be written as:

$$f(y_1|\theta) = \prod_{i=1}^{n} f(y_i|\theta)$$

This is then the product of the individual densities (Greene 2003, p.509), and is called the likelihood function. It is often written as $L(\theta|y)$, with θ first, which is to indicate that we are interested in the unknown function behind the values of y that we can observe. This function is unobservable and a part of the data-generating process, but using our measured values of y, we can estimate it. Usually, the logged version of this likelihood function (the log-likelihood) is used, since it is then easier to maximise (Verbeek 2008, p.180). In a regression setting, where we also have independent variables (in a matrix X), the log-likelihood function can be written as:

$$logL(\theta|y, X) = \sum_{i=1}^{n} logf(y_i|x_i, \theta)$$

Since the log-likelihood function allows for summing probabilities.

4.3 Average marginal effects

Marginal effects (ME) are a way to make the magnitude of coefficients more tangible, and since the coefficients from non-linear estimators are harder to interpret than those of linear models, marginal effects are often needed to say something about magnitude. A marginal effect can be defined as the change in conditional mean of y as an independent variable changes by one unit (which can also be described as a partial derivative). In a linear model, this gives

$$E[y|x] = x'\beta \rightarrow \frac{dE[y|x]}{dx} = \beta$$

Which means that the coefficients can be directly interpreted as ME:s (Cameron and Trivedi 2010, p.122). In non-linear models, however, the interpretation is not as straightforward, which is why ME:s are popular together with these models. One of the most common ME:s is the Marginal Effect at Means (MEM), which computes the effect of change in one x while holding the other independent variables at their means (ibid., p.347). This method can, however, produce results that are hard to interpret: for example, if a dichotomous variable such as gender is included, MEM might hold this constant at its mean, which can give the "effect of a change in another x given that you're 17% female". The Average Marginal Effect

(AME), on the contrary, makes use of all data in a different way when calculating the effects. Formally, the calculation of the AME of the ith variable can be described as:

$$AME_i = \beta_i \frac{1}{n} \sum_{k=1}^{n} f(\beta x^k)$$

In the continuous variable case, where βx^k is the linear combination of parameters (model output) for the kth observation (Bartus et al. 2005). In words, the calculation of AMEs can be explained as a counterfactual calculation of each individual. For example, if the independent variable of interest is gender: for each individual, the probability of outcome is calculated using the model parameters, first as if the person was male, and then female. The difference in probability between these two is then the marginal effect of gender for this individual. This calculation is done for all individuals, and the average of these summed MEs is the AME (Williams et al. 2012). It could always be argued that this counterfactual calculation does not make sense, since variables are taken out of context. For the model used in this thesis, I would argue that AMEs are the most logical choice of ME method, but MEM will also be calculated for comparison.

4.4 Endogeneity

One of the major problems within social sciences regarding inference is the fact that non-experimental data almost always suffers from omitted variable bias. Doing a simple comparison between groups of individuals often requires assumptions that are hard to verify and hard to argue convincingly. The potential outcomes framework allows for formalising the different aspects of a supposedly causal chain (Angrist and Pischke 2008, p.27). In the current case, two individuals can be given as example: individual A was not overeducated pre-migration, but individual B was. The simple comparison between these two would be to look at their labour market outcomes post-migration and then calculate the effect of overeducation as $Y_{Bi} - Y_{Ai}$. This would give the labour market outcome as treated with overeducation, minus the labour market outcome as not treated with overeducation, and the difference would be the effect of pre-migration overeducation. The problem with this comparison is that we cannot be sure that these two individuals are comparable, and that the effect of being overeducated would have been the same for individual A, had he/she been overeducated. Individuals are different in many ways that are not included in control variables, or even measurable, so using A as a counterfactual for B may be incorrect. This can be written formally as:

$$(Y_{Ai}|D_i = 1) \neq (Y_{Bi}|D_i = 0)$$

Where D_i denotes treatment. It can be read as: the potential outcome as not treated for the individual who was treated (which is unobservable, since this individual was in fact treated) is not the same as the potential outcome as not treated for the individual who was not treated (which is observable). In practise, this means that this type of model can not give causal estimates, since the two individuals can not be each other's counterfactuals. This is often referred to as self-selection, but it is in essence the same as the problem of not having all available control variables (omitted variable bias) (ibid., p.12).

In the case of labour market mismatches and migrants, there are several sources of self-selection. Concerning mismatch, a person with high ability can be argued to self-select into undereducation. A person can also self-select into migration based on ability or other omitted variables. It is also likely that individuals are selected into employment, which means that the results for a sample of employed individuals might lack in external validity. The solution for the self-selection problem, in lack of a randomizing experimental situation (which is often not an option in social sciences), is some kind of quasi-experimental design. The options concerning this thesis are discussed in section 4.7, and the consequences of the current design are discussed in section 5.6.

4.5 Model 1

The first model is an ordered logit model using the following specification:

$$y_c = \beta_1 x_{iPM} + X'_{ic}\beta + u_{ic}$$

Where y_c is the likelihood of an upwards post-migration mismatch status transition at census c, x_{iPM} is the independent variable of interest - which is the pre-migration mismatch status for individual i, X_{ic} is a vector of control variables for individual i at census c, and u_{ic} is the error term. The model can be seen as a modified ORU model, where over/undereducation dummies replace the number of years spent in a mismatched state, that are part of the original ORU model by Duncan and Hoffman (1981).

In an alternative specification, lagged census mismatch status will be included,

$$y_c = \beta_1 x_{iPM} + \beta_2 x_{iMMc-1} + X'_{ic}\beta + u_{ic}$$

Where x_{iMMc-1} is the mismatch status of the previous census (there can be up until four of these). The potential problem of included a lagged variable in this setting is commented on in section 4.8.

The vector $X'_{ic}\beta$ contains the following variables:

- *pre-migration mismatch status*: this is the variable of interest, and it contains the three categories OE, RE and UE.
- *cohort*: the immigration cohort variable consists of the categories "70 and earlier","71-75","76-80","81-85" and "86-90". It is relevant to include as there is reason to expect differences in labour market experiences due to when an immigrant arrives in Sweden.
- *gender*: differences in labour market mismatches between genders has earlier been shown, which makes this variable important to include (Leuven, Oosterbeek, et al. 2011, p. 298).
- *age and age squared*: as with gender, age is expected to have an impact on the likelihood of mismatch and is therefore included in the model (ibid., p. 298).
- *pre-migration required education*: including an education variable in this model would be problematic if the acquired education was used, since it would be colinear with the mismatch variable. The pre-migration required education, however, is not - but it

might capture valuable variation of possible heterogeneous effects on the likelihood of mismatch.

- *country of origin*: as has been shown in previous research, there are many reasons to expect different labour market outcomes for immigrants from different countries (regarding cultural and linguistic distance, for example), which is why this variable is included.
- *years since migration*: years since migration is usually an important variable in migration studies, since it can capture the time effect of living in the host country. Theoretically, an individual can be expected to perform better in the labour market with rising years since migration, conditional on the individual being able to enter the labour market.
- *metropolitan*: the dynamics of labour supply and demand can be expected to be different depending on the size of the city an immigrant lives in (shown in the case of Sweden by Åslund and Rooth 2007), which is the reason for including this variable.
- *Visa category*: the labour market outcomes of refugees, compared to immigrants on a work visa, can easily be argued to turn out differently, which is why this variable is important to include.
- *civil status*: previous research has shown that a person's civil status can affect the subsequent labour market outcomes (Loughran and Zissimopoulos 2009) - so it is important to include also in a mismatch model.

4.6 Model 2

The second model is a multinomial logit model using the following specification:

$$Pr(y = MM_m) = F(\alpha_m + \beta_1 x_{iPM} + X'_{ic}\beta + u_{ic})$$

Where $Pr(y = MM_m)$ is the probability that the post-migration mismatch status will take the status m (where m can take the three values OE, RE, UE), F is the likelihood function, x_{iPM} is the independent variable of interest - which is the pre-migration mismatch status for individual i, X_{ic} is a vector of control variables for individual i at census c, and u_{ic} is the error term. By definition in a probability model with a limited dependent variable, the resulting probability will always be in relation to the base category chosen. Generally in these models, RE (required education = correctly matched) is chosen as the base category, since the focus of this thesis is the probability of mismatch and not correct match (although some marginal effects on the probability of being matched will also be presented in section 5.4, for reference). The included control variables in Model 2 will be the same as in Model 1.

4.7 Other possible specifications

To try and address the endogeneity described in section 4.4, a fair amount of time was spent exploring possible causal methods. A fixed-effects setup would be a first choice, but by definition, it does not give estimates for time-invariant variables (Angrist and Pischke 2008). Since the variable of interest itself is time-invariant, this would beat the purpose. An alternative to the fixed-effects setup was therefore considered, using a random effects setup combined

with a Mundlak correction. Since the random effects setup by itself relies on the same zero conditional mean assumption as the regular OLS, it does not by itself solve the endogeneity problem. The method proposed by Mundlak (1978) makes use of the independent variables to correct for individual heterogeneity: all time-variant variables are averaged over time at the panel level and used to correct the estimation. If the assumption holds that these variables are correlated to the unobserved heterogeneity (the method is also sometimes called "correlated random effects"), this correction renders the same result as a fixed-effects model, but also provides coefficients for time-invariant variables (Joona, Gupta, and Wadensjö 2014, p.9). After some testing it was decided that the available independent variables in this thesis were not sufficient to build a credible Mundlak correction on the individual level. For future studies, however, this could be a viable alternative in order to increase the credibility of the results.

4.8 Including lagged variables

In separate variations of both models 1 and 2, the lagged census mismatch status is included in the list of independent variables. It is fairly easily argued that the labour market mismatch status of an individual in the second census can be correlated with the same status in the first census (and correspondingly for third, fourth and fifth censuses) - this is the reason for trying to include these lagged results. Including these will, however, create a theoretical situation that is similar to including a lagged version of the dependent variable (an unknown proportion of the explaining variance on the current census status can be argued to come from the pre-migration status, but also from earlier censuses, even if the correlation can not be argued to be =1).

Including lagged dependent variables in a model with panel data is a practise that has been criticised by, among others, Angrist and Pischke (2008, p.244) and Keele and Kelly (2006): since the residual at time *t-1* can be easily argued to be correlated with the residual at time t , the model will suffer from autocorrelation in the residuals, and render incorrect standard errors. Keeping these dynamics in mind, model variations including the lagged census mismatch status will be run and their marginal effects interpreted (section 5.4).

4.9 Persistence vs. state dependence

Another method to view the mechanism behind mismatches over time is to employ a time-series perspective, using hazard models to determine if there is dependence in the mismatched state. The models in this thesis are not models of state dependence, but rather a lighter form of it, which is persistence. This distinction was pointed out by Mavromaras and McGuinness (2012), and the difference is said to be that a state dependence shows the direct causal effect of a previous mismatch on the subsequent mismatch status (controlling for factors that caused the mismatch to begin with). Persistence, on the other hand, can be interpreted as the duration of time an individual stays in a mismatched state (Joona, Gupta, and Wadensjö 2014, p.16). Going further, it is also possible to model the combinations of mismatch sequences (and not just the entering/leaving of a single state), using sequence analysis. This is an analysis technique that is common in sociology and well suited for analysis of labour market outcomes (Abbott and Tsay 2000). Both hazard models and sequence analysis would be interesting to use on the data used in this thesis, but is out of scope because of time limitations. This thesis

will only consider simple persistence in labour market mismatches.

5 Results

This section will present results from the data and models described above. Firstly, descriptive results will be shown, before presenting the results from the two models.

5.1 Descriptive results

The variable means and distributions by census can be seen in Table 10 (page 30). As described in section 3.1.1, individuals are modelled per their census progression, and thus, it is only possible to be present in five censuses if you immigrated in 1970 or earlier. Judging from the distributions, the most common immigrant (at first census) is 38 years of age, arrives in 1971-1975, has primary education, is matched pre-migration, lives in a Swedish metropolitan region, is a tied mover, is born in Poland, is married, and male. As censuses progress, some details can be noted: the earlier cohorts have a much higher share of worker migrants than the later cohorts, something that is also mentioned in section 2.4. The distributions within variables such as gender, civil status, education are generally stable over the censuses, whereas age and years since migration naturally follow the progression upwards.

5.1.1 Transition

An important descriptive part of this research question is the transition from pre- to post-migration mismatch status, which is essentially what the question is all about. The collected transitions, based on first/second/third/fourth/fifth census for each individual, can be seen in Table 11. The table has percentages row-wise and can be read as follows: in their first census, 450 individuals were reported as being overeducated in their current occupation. In total, 630 individuals were overeducated pre-migration, which equals 71,43% staying in the state of overeducation pre- and post. Looking at the second census, this figure is down to 64,48%. In general, the diagonal of the percentage matrix represents the share of individual remaining in their pre-migration mismatch status. It is noteworthy that the first census shows a clear downward shift in status for the UE/RE categories (only 24,47% of all UE stay UE, 26,23% of RE become OE). Going forward in time, this seems to recuperate somewhat: on the third census, 36,42% of all UE are now UE again, and only 17,34% of the RE are OE).

It can also be informative to look at the progression in Sweden, without considering the pre-migration status. Table 12 shows the transition between the first and third census in Sweden, and it suggests that the persistence seen in Table 11 is strong even without considering pre-migration mismatches. These host-country mismatch persistence percentages are very similar to the corresponding figures presented by Piracha, Tani, and Vadean (2012, p.10) for the Australian situation.

The transition of different categories can be more easily seen in Table 13. In this table, the diagonals from Table 11 are by rows, so the percentage of "persisters" progresses rightward. The share of RE staying RE remains fairly stable, the share of UE rises, and the share of OE drops and stabilises around 52%. The large initial change from first to second census could very well reflect the notion that the initial years after immigration are especially difficult when it comes to immigrant's labour markets.

	First	*Second*	*Third*	*Fourth*	*Fifth*
Age	37.83	42.16	45.41	48.50	50.88
Cohort: 70 and earlier	0.21	0.28	0.37	0.51	1.00
Cohort: 71-75	0.25	0.32	0.40	0.49	
Cohort: 76-80	0.23	0.26	0.23		
Cohort: 81-85	0.17	0.14			
Cohort: 86-90	0.14				
Pre-migration education: Primary	0.49	0.54	0.56	0.59	0.59
Pre-migration education: Secondary	0.24	0.21	0.20	0.19	0.23
Pre-migration education: Post-secondary	0.27	0.24	0.23	0.22	0.19
Pre-migration matching: Overeducated	0.16	0.14	0.12	0.10	0.12
Pre-migration matching: Required Education	0.72	0.74	0.75	0.78	0.76
Pre-migration matching: Undereducated	0.12	0.12	0.12	0.12	0.12
Years since migration	4.85	9.53	14.11	18.76	21.81
Non-metropolitan	0.38	0.38	0.38	0.38	0.43
Metropolitan	0.62	0.62	0.62	0.62	0.57
Visa category: Refugee	0.16	0.12	0.08	0.02	0.03
Visa category: Tied Mover	0.52	0.49	0.45	0.39	0.15
Visa category: Worker	0.32	0.39	0.46	0.59	0.82
Country of birth: Chile	0.11	0.11	0.08	0.04	0.02
Country of birth: Germany	0.17	0.18	0.20	0.21	0.23
Country of birth: Greece	0.14	0.14	0.14	0.17	0.18
Country of birth: Iran	0.04	0.03	0.02	0.02	0.02
Country of birth: Poland	0.18	0.19	0.19	0.16	0.13
Country of birth: Turkey	0.09	0.09	0.09	0.09	0.08
Country of birth: USA	0.10	0.08	0.07	0.08	0.04
Country of birth: Yugoslavia	0.17	0.19	0.20	0.23	0.30
Civil status: Divorced	0.11	0.15	0.17	0.18	0.14
Civil status: Married	0.72	0.72	0.71	0.74	0.75
Civil status: Unmarried	0.15	0.10	0.08	0.06	0.08
Civil status: Widow/widower	0.01	0.03	0.03	0.02	0.02
Gender: female	0.42	0.42	0.40	0.34	0.30
Gender: male	0.58	0.58	0.60	0.66	0.70
Individuals	3875	2038	1216	594	181

Table 10: Variable means, per census

	First census							
Pre-migration	*OE*	*RE*	*UE*	*Sum*	*OE*	*RE*	*UE*	
OE	450	169	11	630	71,43%	26,83%	1,75%	100,00%
RE	728	1672	375	2775	26,23%	60,25%	13,51%	100,00%
UE	101	254	115	470	21,49%	54,04%	24,47%	100,00%
Sum	1279	2095	501	3875	33,01%	54,06%	12,93%	100,00%
	Second census							
Pre-migration	*OE*	*RE*	*UE*	*Sum*	*OE*	*RE*	*UE*	
OE	187	100	3	290	64,48%	34,48%	1,03%	100,00%
RE	284	951	270	1505	18,87%	63,19%	17,94%	100,00%
UE	24	137	82	243	9,88%	56,38%	33,74%	100,00%
Sum	495	1188	355	2038	24,29%	58,29%	17,42%	100,00%
	Third census							
Pre-migration	*OE*	*RE*	*UE*	*Sum*	*OE*	*RE*	*UE*	
OE	93	53	2	148	62,84%	35,81%	1,35%	100,00%
RE	159	557	201	917	17,34%	60,74%	21,92%	100,00%
UE	19	77	55	151	12,58%	50,99%	36,42%	100,00%
Sum	271	687	258	1216	22,29%	56,50%	21,22%	100,00%
	Fourth census							
Pre-migration	*OE*	*RE*	*UE*	*Sum*	*OE*	*RE*	*UE*	
OE	34	24	2	60	56,67%	40,00%	3,33%	100,00%
RE	57	274	132	463	12,31%	59,18%	28,51%	100,00%
UE	8	44	19	71	11,27%	61,97%	26,76%	100,00%
Sum	99	342	153	594	16,67%	57,58%	25,76%	100,00%
	Fifth census							
Pre-migration	*OE*	*RE*	*UE*	*Sum*	*OE*	*RE*	*UE*	
OE	11	10		21	52,38%	47,62%	0,00%	100,00%
RE	18	84	36	138	13,04%	60,87%	26,09%	100,00%
UE	2	13	7	22	9,09%	59,09%	31,82%	100,00%
Sum	31	107	43	181	17,13%	59,12%	23,76%	100,00%

Table 11: Mismatch status transitions

	Mismatch, third census							
Mismatch, first census	*OE*	*RE*	*UE*	*Sum*	*OE*	*RE*	*UE*	
OE	240	101	3	344	69,77%	29,36%	0,87%	100,00%
RE	31	547	125	703	4,41%	77,81%	17,78%	100,00%
UE		39	130	169	0,00%	23,08%	76,92%	100,00%
Sum	271	687	258	1216	22,29%	56,50%	21,22%	100,00%

Table 12: Transition between first and third census in Sweden

	Census				
Pre-migration	*First*	*Second*	*Third*	*Fourth*	*Fifth*
OE	71,43%	64,48%	62,84%	56,67%	52,38%
RE	60,25%	63,19%	60,74%	59,18%	60,87%
UE	24,47%	33,74%	36,42%	26,76%	31,82%

Table 13: Average share of "persisters" per census, by pre-migration mismatch status

5.2 Model 1: ordered logistic regression

The ordered logit, as presented in section 4.1.1, focuses on levels rather than letting the individual categories of the dependent variable being detached from each other. The interpretation of coefficients is then, accordingly, different from in a multinomial setting. The dependent variable categories have been ordered according to the underlying hypothetical ability distribution: 1 is overeducated, and 3 is undereducated. The coefficients for the independent variables are to be interpreted as the odds of being in a "higher" level, meaning higher in labour market performance or ability. By design, this upwards level shift could be both rising from overeducation to required education, or from required education to undereducation, with the distance assumed equal.

The results from the ordered logit model can be seen on page 34. The independent variable coefficients will be commented on below.

- *pre-migration undereducation*: the coefficients from moving to an undereducated state (relative to a matched state) have positive signs for the likelihood of an upwards level shift, and this effect is present and significant throughout all five census progressions.

- *pre-migration overeducation*: the coefficients from moving to an overeducated state (relative to a matched state) have negative signs for the likelihood of an upwards level shift. This effect is significant and present for all five census progressions.

- *cohort*: the coefficient signs from arrival cohorts indicate that arriving in later than the reference cohort "70 and earlier" yields a positive sign on the likelihood of upwards shift (with the exception of "71-75" for first census, which is not significant).

- *gender*: for all five census progressions, being a female is associated with a negative likelihood of upwards shift, all coefficients are significant (except for the fifth census).

- *age*: the coefficients for linear age are mostly insignificant, but the third and fourth census show a negative sign that is weakly significant. The coefficients for squared age are small (as expected for a squared-variable coefficient) but mostly show no effect.

- *pre-migration required education*: the coefficient for being in a pre-migration occupation that requires secondary or post-secondary education is associated with a higher likelihood of being overeducated (in relation to primary education). This is not surprising, since the mismatch variable in itself depends on the initial education level (the reason for including this variable is explained in section 4.5).

- *country of origin*: in relation to the base category (Germany), all included countries have a negative coefficient on the likelihood of upwards shift, in the first census. These coefficients largely keep their signs and significance up until the fourth census progression.

- *years since migration*: years since migration has a positive sign, which means that each added year in Sweden increases the likelihood of an upwards shift. These coefficients are significant until the fourth census.

- *metropolitan*: residing in a metropolitan area at the time of census (compared to a non-metropolitan area) gives a negative coefficient for the likelihood of an upwards shift, but this is only weakly significant in the first census and not in the later censuses.
- *Visa category*: arriving as a tied mover or refugee (in relation to work migrants, which is the base category) has a negative coefficient in the first census - which means a lower likelihood of upwards shift in comparison. This coefficient is stable and significant over the five censuses for tied movers, but not for refugees.
- *civil status*: the coefficients for being married at time of census, compared to the base category (unmarried) are negative but not significant.
- *cuts*: the coefficients for the level cut points in the dependent variable are not interesting by themselves, but only if they were to significantly overlap each other (which would mean that the levels were not separate enough to use in an ordered model). A performed t-test shows that the levels in all five models are significantly different from each other.

	First census	Second census	Third census	Fourth census	Fifth census
Pre-migration UE	2.58***	3.33***	3.11***	2.25***	2.98***
	(0.14)	(0.22)	(0.28)	(0.38)	(0.78)
Pre-migration RE	Ref	Ref	Ref	Ref	Ref
Pre-migration OE	-3.94***	-4.21***	-4.24***	-4.45***	-4.41***
	(0.14)	(0.21)	(0.28)	(0.40)	(0.77)
Cohort: 70 and earlier	Ref	Ref	Ref	Ref	Ref
Cohort: 71-75	0.15	0.47***	0.17	0.35	
	(0.11)	(0.14)	(0.15)	(0.19)	
Cohort: 76-80	0.66***	0.60***	0.27		
	(0.12)	(0.15)	(0.19)		
Cohort: 81-85	0.76***	0.70***			
	(0.13)	(0.19)			
Cohort: 86-90	0.50***				
	(0.15)				
Gender: male	Ref	Ref	Ref	Ref	Ref
Gender: female	-0.60***	-0.80***	-0.64***	-0.85***	-0.85
	(0.08)	(0.12)	(0.15)	(0.21)	(0.44)
Age of Individual	0.03	0.01	-0.18*	-0.30*	-0.58
	(0.03)	(0.05)	(0.08)	(0.13)	(0.32)
Age squared	-0.00	-0.00	0.00	0.00*	0.01
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Pre-migration RE: Primary	Ref	Ref	Ref	Ref	Ref
Pre-migration RE: Secondary	-3.06***	-3.08***	-3.02***	-3.18***	-3.37***
	(0.14)	(0.21)	(0.28)	(0.40)	(0.82)
Pre-migration RE: Post	-4.11***	-4.12***	-4.01***	-3.68***	-4.28***
	(0.14)	(0.21)	(0.26)	(0.35)	(0.73)
Chile	-1.97***	-1.73***	-1.92***	-1.59**	-2.71
	(0.18)	(0.25)	(0.33)	(0.49)	(1.46)
Germany	Ref	Ref	Ref	Ref	Ref
Greece	-1.58***	-1.63***	-2.14***	-2.30***	-2.37***
	(0.14)	(0.19)	(0.24)	(0.31)	(0.61)
Iran	-1.10***	-0.44	-0.97*	0.51	-0.20
	(0.22)	(0.34)	(0.49)	(0.78)	(1.31)
Poland	-1.29***	-1.14***	-1.42***	-1.01**	-0.37
	(0.13)	(0.17)	(0.22)	(0.31)	(0.65)
Turkey	-1.31***	-1.08***	-1.56***	-1.66***	-1.21
	(0.15)	(0.20)	(0.26)	(0.37)	(0.68)
USA	-0.80***	-0.94***	-1.30***	-1.05*	-1.43
	(0.15)	(0.22)	(0.29)	(0.42)	(1.01)
Yugoslavia	-1.31***	-1.26***	-1.20***	-0.96***	-1.13*
	(0.13)	(0.17)	(0.21)	(0.27)	(0.51)
Years since Migration	0.08***	0.11***	0.05*	0.04	0.02
	(0.01)	(0.02)	(0.02)	(0.03)	(0.08)
Non-metropolitan area	Ref	Ref	Ref	Ref	Ref
Metropolitan area	-0.16*	-0.20	0.01	-0.06	-0.49
	(0.08)	(0.11)	(0.13)	(0.19)	(0.37)
Visa category: Worker	Ref	Ref	Ref	Ref	Ref
Visa category: Tied mover	-0.59***	-0.45***	-0.34*	-0.48*	-0.32
	(0.10)	(0.13)	(0.16)	(0.21)	(0.50)
Visa category: Refugee	-0.54***	-0.29	0.25	0.47	-1.18
	(0.16)	(0.22)	(0.31)	(0.63)	(1.16)
Civil status: Unmarried	Ref	Ref	Ref	Ref	Ref
Civil status: Married	-0.12	-0.27	-0.14	0.58	0.81
	(0.11)	(0.17)	(0.24)	(0.39)	(0.66)
Civil status: Widow/widower	0.06	0.10	-0.30	0.08	1.28
	(0.32)	(0.36)	(0.44)	(0.68)	(1.33)
Civil status: Divorced	-0.17	-0.12	-0.04	0.55	1.27
	(0.15)	(0.20)	(0.28)	(0.44)	(0.80)
cut1	-3.93***	-4.66***	-9.54***	-12.59***	-19.11*
cut2	0.44	0.12	-4.94**	-7.90*	-14.24
Observations	3875	2038	1216	594	177
Pseudo R^2	0.322	0.329	0.325	0.310	0.318

5.2.1 Interactions and heterogeneous effects

As a part of modelling and interpreting results, a number of different interactions between independent variables were tested in order to see if there were heterogeneous effects. Among these are the interactions of pre-migration mismatch status with country of origin, Visa category, gender and immigration cohort (coefficients can be seen in the Appendix). The results of the interaction regressions show no interaction coefficients that deviate from the base effects in an interesting way. Many of the interaction coefficients are also insignificant. This can be interpreted as there being too much heterogeneity within these categories to find any distinct mechanisms on such a detailed level - and that a larger sample size would help. Or, of course, it can be interpreted as a lack of heterogeneous effects in this area, but it seems theoretically plausible that there would be some. Since there exists, to my knowledge, no previous studies with interactions using similar method and data, there is also little to compare with.

5.2.2 Goodness-of-fit

The provided goodness-of-fit measure (pseudo R-squared) calculates the difference between a bare-bones model and the current model, to try and measure the added value of the model. There are many different ways of calculating a goodness-of-fit measure for non-linear models, and there seems to be no consensus as to which is better. Though not as easily interpreted as a regular R-squared from a linear model, the STATA output can still be interpreted as higher = better. Tests of different model specifications have been made (section 5.5), and generally, the pseudo R-squared increases as the included variables are added stepwise to the model.

5.2.3 Testing the proportional odds assumption

The proportional odds assumption for ordered logit models can be tested using the Brant test, which is a part of the *oparallell* package in STATA (M. L. Buis 2013). Results from the first-census model shows that the assumption of the odds not being proportional can not be rejected - i.e., the model fails the test. For models with relatively small sample sizes, this is not uncommon, and is not by itself a reason not to use the ordinal model setup (M. Buis, Williams, et al. 2013). It does, however, make the use of another setup (like the multinomial) logical, since it will view the data from another angle where ordinality is not assumed.

5.3 Model 2: multinomial logistic regression

The results from the multinomial regression model can be seen on pages 37 and 38. In this table, not all independent variable coefficients are reported[1].

Firstly, it should be noted that the outcome variable has the three familiar categories, OE/RE/UE. In this regression, RE is the reference outcome, and the coefficients in the other outcomes are related to this. Starting with the outcome of being overeducated in Sweden at the first census (page 37), we see that the coefficient for "Pre-migration UE" is -1,69. This is to be read as: the relative log odds of being overeducated vs. being matched in Sweden will decrease by 1,69 if moving from being matched pre-migration to being undereducated pre-migration (all other variables held constant). Similarly, the odds of being overeducated in Sweden will

[1]The most relevant independent variables from the multinomial model are included in the table, a complete table is readily available from the author by request.

increase by 4,77 if moving from pre-migration match to pre-migration overeducation. The outcomes for pre-migration match status on second/third/fourth and fifth census are reported in columns progressing rightward. Note that in this model, previous census match status is not included as a control variable - this is discussed in section 5.5). The independent variable coefficients will be commented on below.

	(1) First census b/se	(2) Second census b/se	(3) Third census b/se	(4) Fourth census b/se	(5) Fifth census b/se
Outcome: Overeducated					
Pre-migration UE	-1.69***	-2.04***	-1.49***	-1.27**	-2.61*
	(0.15)	(0.26)	(0.30)	(0.46)	(1.19)
Pre-migration RE	0.00	0.00	0.00	0.00	0.00
	(.)	(.)	(.)	(.)	(.)
Pre-migration OE	4.77***	4.79***	4.87***	5.45***	5.47***
	(0.23)	(0.34)	(0.51)	(0.82)	(1.56)
Cohort: 70 and earlier	0.00	0.00	0.00	0.00	0.00
	(.)	(.)	(.)	(.)	(.)
Cohort: 71-75	-0.06	-0.62**	0.03	-0.22	
	(0.16)	(0.21)	(0.24)	(0.31)	
Cohort: 76-80	-0.58***	-0.67**	0.12		
	(0.17)	(0.22)	(0.26)		
Cohort: 81-85	-0.43*	-0.98***			
	(0.17)	(0.26)			
Cohort: 86-90	-0.14				
	(0.19)				
Gender: male	0.00	0.00	0.00	0.00	0.00
	(.)	(.)	(.)	(.)	(.)
Gender: female	0.33**	0.43**	0.07	0.47	-0.72
	(0.11)	(0.16)	(0.22)	(0.33)	(0.83)
Age of Individual	-0.12**	-0.07	0.28*	0.36	0.76
	(0.04)	(0.07)	(0.13)	(0.24)	(0.73)
Age squared	0.00**	0.00	-0.00*	-0.00	-0.01
	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)
	(0.23)	(0.35)	(0.51)	(0.81)	(1.53)
Chile	1.35***	1.68***	1.77***	1.71*	3.83
	(0.22)	(0.32)	(0.44)	(0.72)	(1.96)
Germany	0.00	0.00	0.00	0.00	0.00
	(.)	(.)	(.)	(.)	(.)
Greece	0.63**	1.10***	1.07*	1.69**	1.91
	(0.22)	(0.32)	(0.44)	(0.56)	(1.14)
Iran	0.71**	-0.03	0.20	-0.72	1.04
	(0.25)	(0.43)	(0.57)	(0.95)	(1.69)
Poland	0.71***	1.19***	0.90**	0.63	-0.27
	(0.16)	(0.24)	(0.29)	(0.43)	(0.91)
Turkey	0.37	0.28	0.52	1.20	-1.16
	(0.24)	(0.41)	(0.49)	(0.72)	(2.06)
USA	0.11	0.58*	0.47	0.68	1.20
	(0.17)	(0.26)	(0.34)	(0.48)	(1.21)
Yugoslavia	0.35	1.00**	0.75	-0.80	-0.24
	(0.20)	(0.32)	(0.40)	(0.66)	(0.98)
Years since Migration	-0.05**	-0.10***	-0.05	0.00	-0.08
	(0.02)	(0.03)	(0.04)	(0.06)	(0.16)
Visa category: Worker	0.00	0.00	0.00	0.00	0.00
	(.)	(.)	(.)	(.)	(.)
Visa category: Tied mover	0.65***	0.72***	0.49*	0.40	-0.45
	(0.13)	(0.19)	(0.23)	(0.34)	(0.88)
Visa category: Refugee	0.60**	0.36	-0.47	-0.34	1.81
	(0.20)	(0.29)	(0.40)	(0.78)	(1.52)
_cons	-2.63**	-3.67*	-11.99***	-13.92*	-19.31
	(0.83)	(1.56)	(3.03)	(6.00)	(19.67)

	(1) First census b/se	(2) Second census b/se	(3) Third census b/se	(4) Fourth census b/se	(5) Fifth census b/se
Outcome: Undereducated					
Pre-migration UE	2.08***	2.85***	3.23***	2.45***	3.77**
	(0.24)	(0.36)	(0.50)	(0.59)	(1.43)
Pre-migration RE	0.00	0.00	0.00	0.00	0.00
	(.)	(.)	(.)	(.)	(.)
Pre-migration OE	-2.16***	-3.20***	-3.41***	-2.83***	-17.20
	(0.33)	(0.60)	(0.75)	(0.78)	(1152.17)
Cohort: 70 and earlier	0.00	0.00	0.00	0.00	0.00
	(.)	(.)	(.)	(.)	(.)
Cohort: 71-75	0.10	0.34	0.25	0.35	
	(0.16)	(0.18)	(0.21)	(0.24)	
Cohort: 76-80	0.52**	0.59**	0.65*		
	(0.18)	(0.21)	(0.27)		
Cohort: 81-85	0.95***	0.37			
	(0.21)	(0.28)			
Cohort: 86-90	0.93***				
	(0.23)				
Gender: male	0.00	0.00	0.00	0.00	0.00
	(.)	(.)	(.)	(.)	(.)
Gender: female	-0.78***	-1.07***	-0.99***	-1.12***	-1.94*
	(0.13)	(0.17)	(0.21)	(0.29)	(0.76)
Age of Individual	-0.05	-0.02	-0.08	-0.19	-0.78*
	(0.04)	(0.07)	(0.11)	(0.18)	(0.39)
Age squared	0.00	-0.00	0.00	0.00	0.01
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
	(0.25)	(0.37)	(0.51)	(0.60)	(1.40)
Chile	-1.87***	-1.36***	-1.05*	-1.17	-0.96
	(0.30)	(0.39)	(0.49)	(0.65)	(1.58)
Germany	0.00	0.00	0.00	0.00	0.00
	(.)	(.)	(.)	(.)	(.)
Greece	-2.11***	-1.78***	-2.57***	-2.67***	-3.35***
	(0.20)	(0.24)	(0.33)	(0.40)	(0.90)
Iran	-0.63	-0.95	-0.95	0.71	1.45
	(0.34)	(0.60)	(0.95)	(1.01)	(2.53)
Poland	-1.38***	-0.51*	-1.39***	-1.01*	-2.05
	(0.20)	(0.24)	(0.33)	(0.47)	(1.33)
Turkey	-1.65***	-1.30***	-1.81***	-1.76***	-2.09*
	(0.20)	(0.24)	(0.31)	(0.43)	(0.86)
USA	-2.16***	-1.71**	-3.18**	-0.56	-16.88
	(0.42)	(0.57)	(1.07)	(0.76)	(1533.93)
Yugoslavia	-1.64***	-1.22***	-1.19***	-1.31***	-2.58***
	(0.17)	(0.20)	(0.24)	(0.32)	(0.77)
Years since Migration	0.09***	0.10***	0.05	0.05	-0.01
	(0.02)	(0.02)	(0.03)	(0.04)	(0.12)
Visa category: Worker	0.00	0.00	0.00	0.00	0.00
	(.)	(.)	(.)	(.)	(.)
Visa category: Tied mover	-0.47**	-0.25	-0.21	-0.51	-0.63
	(0.14)	(0.17)	(0.21)	(0.27)	(0.74)
Visa category: Refugee	-0.49	-0.29	-0.27	-0.30	-12.14
	(0.25)	(0.35)	(0.52)	(1.56)	(2259.27)
_cons	1.11	0.36	2.51	5.94	22.33*
	(0.82)	(1.33)	(2.40)	(4.20)	(10.21)
Observations	3875	2038	1216	594	177
Pseudo R^2	0.362	0.352	0.362	0.345	0.419

- *pre-migration undereducation*: the coefficients from moving to an undereducated state (relative to a matched state) have negative signs on post-migration overeducation (page 38), and positive signs on post-migration undereducation. These results are stable and significant for both outcome categories and throughout the censuses.

- *pre-migration overeducation*: the coefficients from moving to an overeducated state (relative to a matched state) have positive signs on post-migration overeducation, and negative signs on post-migration undereducation. These results are also according to theoretical a priori expectations, and are significant and stable for all censuses (with the exception of the fifth census, UE outcome from OE - very few individuals succeed in making this status shift).

- *cohort*: the coefficient signs from arrival cohorts indicate that arriving in "76-80", "81-85" or "86-90" (relative to the "70 and earlier" cohort) have negative signs for likelihood of post-migration overeducation, and corresponding positive effects on post-migration undereducation. The coefficients are significant in the first census, but become less significant as the sample size falls with progressing number of censuses.

- *gender*: for undereducation, the coefficients show a negative association for women, indicating that women in general are less likely to experience post-migration undereducation compared to men. These coefficients are significant and stable through all census progressions. The coefficients on being overeducated post-migration from gender show a positive association between being a woman and being overeducated post-migration. These coefficients are significant only for the first two censuses.

- *age*: the outcome of being overeducated post-migration has a negative association with linear age, but is only significant for some census progressions. For squared age, the coefficient is small (as expected for a squared-variable coefficient) and has signs in both directions, which makes it hard to interpret in a meaningful way. The outcome as undereducated post-migration has a negative linear age coefficient for all censuses (but they are mostly not significant), and an in-existent coefficient for squared age.

- *country of origin*: in relation to the base category (Germany), Greece, Poland, Chile and Iran have positive coefficients on the likelihood of being overeducated post-migration. The coefficients for Greece, Chile and Poland are significant for first, second and third census, while Iran is only significant in the first census. For the outcome of being undereducated post-migration, almost all countries have negative coefficients compared to Germany, many of which are significant and persistent over several censuses.

- *years since migration*: years since migration has negative coefficients for overeducation outcome, and positive coefficients for undereducation outcome, both of which are significant in the first and second census.

- *Visa category*: arriving as a tied mover or refugee (in relation to work migrants, which is the base category) has a positive coefficient on the outcome as overeducated post-migration. These coefficients are significant in first census (for refugees) and first and second census (for tied movers). The coefficients for outcome as undereducated post-migration are correspondingly negative, but only sporadically significant.

It can be noted that the results from the multinomial model seem to confirm the findings from the ordinal model. While this is not surprising since the same data source is used, it is still an important indication that the results hold for different model specifications. The pseudo R-squared from the multinomial models are around the same as for the ordinal models ($\approx 0,3$).

5.4 Marginal effects

The interpretation of the magnitude of log-odds coefficients from logit regressions can be difficult: using the described Average Marginal Effects (AME) method, some more readily interpretable effects can be produced. The AME:s are computed using the method described in 4.3, and are presented by the predicted category of the outcome variable. Marginal effects are reported for both model 1 (ordered logit) and model 2 (multinomial logit). The effects are calculated for the independent variable of interest (pre-migration mismatch status).

5.4.1 Model 1: ordered logistic regressions

The AME:s for post-migration overeducation based on pre-migration match status from the ordered model can be seen in Table 14.

	(1)	(2)	(3)	(4)	(5)
	First census	Second census	Third census	Fourth census	Fifth census
Undereducated	-0.233***	-0.196***	-0.185***	-0.120***	-0.137***
	(0.00844)	(0.00875)	(0.0113)	(0.0147)	(0.0261)
Required Education	0	0	0	0	0
	(.)	(.)	(.)	(.)	(.)
Overeducated	0.515***	0.538***	0.533***	0.559***	0.522***
	(0.0136)	(0.0211)	(0.0290)	(0.0433)	(0.0793)
Observations	3875	2038	1216	594	177
Pseudo R^2	0.322	0.329	0.325	0.310	0.318

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 14: Average Marginal Effects: outcome as overeducated from pre-migration mismatch status

The interpretation of these marginal effects (ME:s) follow the same logic as the coefficients, but can be interpreted as magnitude in percentages. The first column in Table 14 can be read as: the likelihood of being overeducated in Sweden will decrease by 23,3% if moving from being matched pre-migration to being undereducated pre-migration (measured at first census in Sweden, all other variables held constant). The corresponding effect for being overeducated pre-migration gives an increased likelihood of post-migration overeducation of 51,5%. As the censuses progress, the effect of pre-migration mismatch status on outcome as overeducated does not change much, but the standard errors grow as the sample size decreases.

The marginal effects for outcome as matched (required education) post-migration can be seen in Table 15. The ME:s in this table show that, firstly, being overeducated pre-migration

(in relation to being matched) is associated with a 37,3% decrease in the likelihood of being matched post-migration in the first census. This effect decreases as the censuses progress, but the standard errors grow as well. The effect from being undereducated pre-migration (in relation to being matched) is smaller (a decrease of -10,5%) but increases as censuses progress.

	(1) First census	(2) Second census	(3) Third census	(4) Fourth census	(5) Fifth census
Undereducated	-0.105***	-0.248***	-0.227***	-0.197***	-0.260***
	(0.0100)	(0.0183)	(0.0226)	(0.0353)	(0.0628)
Required Education	0	0	0	0	0
	(.)	(.)	(.)	(.)	(.)
Overeducated	-0.373***	-0.364***	-0.322***	-0.302***	-0.270***
	(0.0136)	(0.0212)	(0.0287)	(0.0426)	(0.0776)
Observations	3875	2038	1216	594	177
Pseudo R^2	0.322	0.329	0.325	0.310	0.318

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 15: Average Marginal Effects: outcome as matched (required education) from pre-migration mismatch status

The marginal effect of being mismatched (relative to being matched) pre-migration on a post-migration match is generally negative for both over- and undereducation, and persistent over censuses. The magnitudes are larger for overeducation than undereducation.

For the outcome of being undereducated post-migration, the marginal effects can be seen in Table 16. These ME:s show that it is 33,7% more likely to be undereducated post-migration if pre-migration status is undereducated (in relation to being matched, ceteris paribus), with the corresponding likelihood for pre-migration overeducation being a 14,2% decrease. Variations of these AMEs, but in the form of MEMs, can be seen in the Appendix.

As discussed in section 4.8, it is questionable how far in the future it can be argued that the pre-migration mismatch status has an effect. Depending on how the mechanism behind the mismatch is seen, earlier censuses in Sweden might be taking over / confounding the effect of the pre-migration mismatch as time progresses. The marginal effects from pre-migration mismatch on post-migration outcomes, including controls for mismatch status in previous Swedish censuses, can be seen in Tables 17, 18 and 19. The models contain: second census outcome, controlling for first census outcome - third census outcome, controlling for first and second census outcome, and fourth census outcome, controlling for first, second and third census outcome (the fifth-census model sample size, together with the lagged variables, made it unreliable - a non-concave error - and it is thus not included in the output).

General for these marginal effects, compared to the models without controlling for lagged census results, is that the effect in most cases is of lower magnitude but the same signs. For example, the marginal effect of being overeducated (compared to matched) pre-migration on the outcome of overeducation in second census is 51,5% in the uncontrolled model, and 24,6% in the controlled model. Keeping the potential confounding mechanism in mind, it could be argued that the mismatch status of the first census captures some of the effect from

	(1) First census	(2) Second census	(3) Third census	(4) Fourth census	(5) Fifth census
Undereducated	0.337***	0.443***	0.411***	0.317***	0.397***
	(0.0165)	(0.0213)	(0.0268)	(0.0445)	(0.0739)
Required Education	0	0	0	0	0
	(.)	(.)	(.)	(.)	(.)
Overeducated	-0.142***	-0.174***	-0.211***	-0.256***	-0.252***
	(0.00556)	(0.00808)	(0.0107)	(0.0152)	(0.0306)
Observations	3875	2038	1216	594	177
Pseudo R^2	0.322	0.329	0.325	0.310	0.318

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 16: Average Marginal Effects: outcome as undereducated from pre-migration mismatch status

	(1) Second census	(2) Third census	(3) Fourth census
Undereducated	-0.139***	-0.0784***	-0.0222
	(0.0106)	(0.0170)	(0.0215)
Required Education	0	0	0
	(.)	(.)	(.)
Overeducated	0.246***	0.122***	0.0667*
	(0.0276)	(0.0312)	(0.0294)
Observations	2038	1012	470
Pseudo R^2	0.489	0.581	0.723

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 17: Average Marginal Effects: outcome as overeducated from pre-migration mismatch status (with lagged census controls)

	(1) Second census	(2) Third census	(3) Fourth census
Undereducated	-0.101***	-0.0654**	-0.0194
	(0.0187)	(0.0220)	(0.0197)
Required Education	0	0	0
	(.)	(.)	(.)
Overeducated	-0.121***	-0.00438	0.0644*
	(0.0229)	(0.0181)	(0.0281)
Observations	2038	1012	470
Pseudo R^2	0.489	0.581	0.723

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 18: Average Marginal Effects: outcome as matched (required education) from pre-migration mismatch status (with lagged census controls)

	(1) Second census	(2) Third census	(3) Fourth census
Undereducated	0.240***	0.144***	0.0416
	(0.0279)	(0.0374)	(0.0404)
Required Education	0	0	0
	(.)	(.)	(.)
Overeducated	-0.125***	-0.118***	-0.131**
	(0.00767)	(0.0184)	(0.0424)
Observations	2038	1012	470
Pseudo R^2	0.489	0.581	0.723

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 19: Average Marginal Effects: outcome as undereducated from pre-migration mismatch status (with lagged census controls)

the pre-migration mismatch. It is also important to remember that the potential residual autocorrelation in these models may render the standard errors incorrect. The complexity surrounding different combinations of state progressions have been discussed in section 4.9, and further empirical tests of these progressions are out of scope for this thesis.

The coefficients for age shown in model 1 were not especially large or significant, but it might nevertheless be interesting to view the marginal effects of an individual's age at arrival in Sweden at the probability of being mismatched (in this case, overeducated). These predictive margins are shown in Figure 5.4.1. Firstly, it can be noted that there is a large difference in the base effect from pre-migration, which is seen in the distance between the three lines. Also, it is apparent that the likelihood of being overeducated is larger if an individual arrives at a younger age, regardless of his/her pre-migration mismatch status. When interpreting this graph, though, it is important to keep the confidence intervals in mind: especially at the ends of the age scale, they grow larger, so the predictions should be interpreted with caution. Should the predicted pattern prove to be right, is shows an interesting pattern that also falls in line with previous research - which is a lower likelihood of overeducation with rising age (Leuven, Oosterbeek, et al. 2011, p.298).

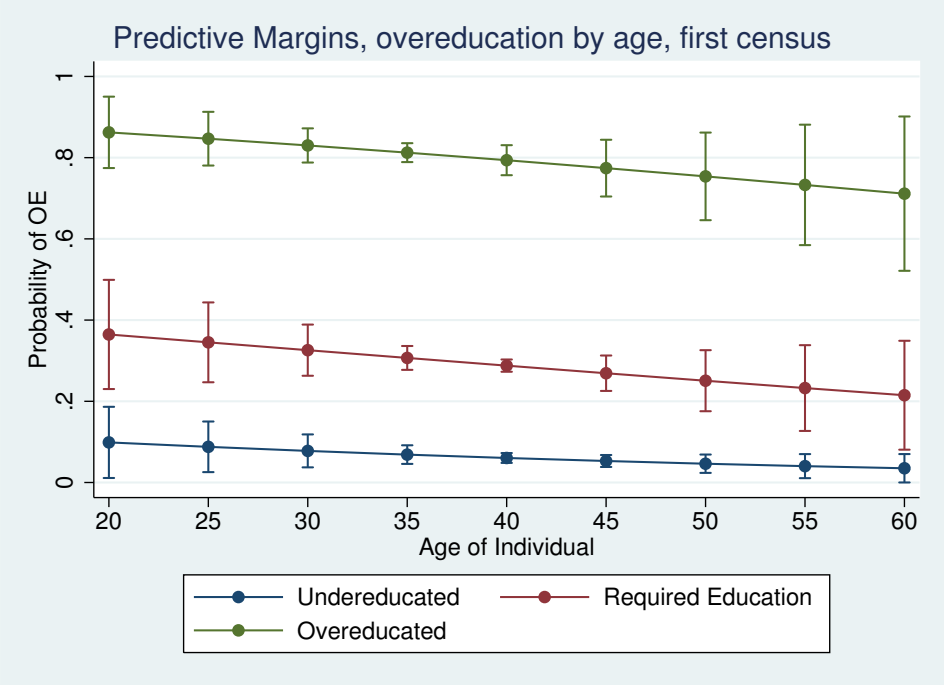

Figure 1: Predictive margins - OE by age, first census

5.4.2 Model 2: multinomial logistic regressions

Corresponding marginal effects from the multinomial model can be found in the Appendix. Generally and with few exceptions, these effects are the same as in the ordered logit model.

5.5 Sensitivity analysis

The sensitivity of the results has been tested in a number of ways:

The chi-square tests that all included variables equal zero (included in the STATA model outputs) are significant on the 1% level for all models run. While this is not a general sign of approval, it is a first indication that the models have some predictive value. Going further, separate likelihood ratio (LR) tests have been conducted to test if some combinations of independent variables are jointly redundant (redundance of an individual variable is available by the p-values). Models combining the independent variable of interest with all other covariates have been tested, and their joint redundance rejected at the 1% level. The question of heteroskedasticity in limited dependent variable models has been debated amongst economists, and although tests do exist, it is a bit of a contradictory situation: the outcome from the model is a probability which is only as good as what is included in the model - the predictions can therefore considered to be "locked" by the data by design. STATA can produce robust standard errors for limited dependent variable models, but these are of the same type (White/sandwich estimator) as for linear models (STATA 2016, p.12). The argument against using these would be that a heteroskedastic maximum likelihood estimator would produce inconsistent estimates to begin with, and it is on the standard errors of these supposedly inconsistent parameters that the robustness correction is made (Giles 2013). The White standard errors compared to the regular ones for the first-census ordered model can be seen in the Appendix, and the difference between errors these is very small. Lastly, the following sensitivity checks have been run:

- for the individuals that had two pre-migration occupations with different required educations, models were run using both the highest and lowest education level. The results using the lowest education level show no significant differences in outcomes compared to using the highest.
- instead of using the logistic distribution, models were also run using the normal distribution (probit). The probit results do not deviate in any notable way from the logit results, which is also often to be expected (Gujarati 2009, p.571).
- the groups excluded from the sample (mentioned in section 3.1) were included as sensitivity tests but did not change the results of either model 1 or 2 in any way that would change the interpretation of the results.
- the number of individuals chosen as a cut-off for realised matches when computing the required education using the "realised matches" could in theory bias the results. The number chosen (section 3.2.2) was 10 individuals, but running the computation with 5, 15 and 20 as a cut-off does not make a significant difference in the result.

5.6 Limitations and validity

Apart from the potential issues with the data used (mentioned in section 3.5), there are a number of issues that might compromise the validity of the results in this thesis.

As has been mentioned in section 4.4, the results from the two models in this thesis can

only show associations and not causal mechanisms, because of the individual heterogeneity/omitted variable bias that comes from the differences in potential outcomes. This can be also be seen as a compromise of the internal validity. This issue is discussed and incorporated in the interpretation of results in this thesis.

The sample of immigrants (from the complete SLI database) is a random sample and should not compromise the external validity by its own right. A concern regarding the sample, and to some extent the validity, is the fact that only employed individuals are included. Firstly, it should be mentioned that handling this issue would require a larger theoretical base, since the mechanism of mismatch $\rightarrow$ mismatch might be different from the mechanism of mismatch $\rightarrow$ unemployed. It would, however, most likely be relevant also for this thesis - the issue is mentioned and handled both by Piracha, Tani, and Vadean (2012) and Helgertz (2010). The problem with only observing employed individuals is that they may constitute a non-random sample of the population, which will bias the results and lower the external validity (Piracha, Tani, and Vadean 2012, p.7). The solution applied by both mentioned authors is a two-step estimator, that first estimates the probability of being employed, which is later used as an instrument in the main model. Doing the same correction in this thesis, however, was not possible because of data and time constraints.

6 Conclusion

Having presented results from two models, this section will put the results in perspective and discuss the possible explanations for these results. A starting point will be the hypotheses presented in section 2.5. Firstly, it is important to distinguish between what is being measured and what is not being measured. Given what has been presented of previous research in the field, it is unlikely that what is being picked up in coefficients are the "pure" effects of the included independent variables, and this is most certainly the case with the current variable of interest (the pre-migration mismatch status). The theoretical and empirical explanations presented in the hypotheses are not measured explicitly, and their presence will therefore have to be derived and discussed rather than read from a model output.

Before discussing the hypotheses, the primary research question (section 1.1) will be answered: The results indicate an association between the pre-migration mismatch status and the post-migration mismatch status. For the first recorded census in Sweden, a pre-migration overeducation (compared to a pre-migration match) is associated with a 51,5% increase in the likelihood of a post-migration overeducation. The corresponding likelihood for pre-migration undereducation with the outcome of post-migration undereducation is a 33,7% increase in likelihood. These marginal effects remain relatively stable also at second, third, fourth and fifth census. The results are stable in the presence of different control variables, and after sensitivity tests. The likelihood association on overeducation found in this thesis matches well with the results found by Piracha, Tani, and Vadean (ibid., p.19), where a 45% increased likelihood is reported. For undereducation, however, the corresponding figure is an increased likelihood of 61%, which is almost double of what is found in this thesis. A possible and theoretically appealing explanation to this difference is the Australian immigration policy (described in section 2.3), which in essence means a positive selection on ability for

non-humanitarian refugees (which would bias parts of the immigrant stock upwards). It is, however, not possible to verify this explanation.

The hypotheses presented in section 2.5 will now be revised using the results from this thesis:

- *H1 - signalling*: the signalling of "real" productivity manifested by a previous mismatch (described in section 2.1.5), is the preferred theoretical explanation used by Piracha, Tani, and Vadean (2012), and the authors also state that this hypothesis is confirmed by their results. The reason given for this conclusion is the found dependence between the pre- and post-migration mismatch status (p.19), but no further indication is given as to why this connection is comes from signalling and not, for example, discrimination. The results from this thesis show the same pattern of associated persistence between pre- and post-migration, but I do not find reason to attribute this separately to signalling. The mechanism of an applicant implicitly demonstrating his/her ability by presenting acquired and required education (for previous job) at application is of course relevant also in this case. It is probably not always applicable, as certain groups of immigrants (for example, refugees) might have a harder time producing the right paperwork, references and credentials to be able to show previous employment. Disregarding this, the results from this thesis can not be used to prove or disprove the existence of signalling.

- *H2 - individual ability*: the mechanism of unmeasurable ability leading to self-selection into mismatch can be viewed both as a theoretical explanation and an omitted variable problem (when using a model that does not account for this). As proposed in this hypothesis, ability would lead to a persistence of the mismatched state, and would not be affected by recuperating effects from gains in country-specific human capital. Also, the simple fact that an individual has a pre-migration mismatch that persists also post-migration is a hint that unmeasurable variables (such as ability or motivation) are at least part of the story (together with post-migration explanations such as discrimination and transferability) as pointed out by Piracha, Tani, and Vadean (ibid., p.5). The results from this thesis show a pattern that would fit well into the ability story, since a persistence is found both in first census in Sweden, but also for later censuses. This does not enable us to conclude that ability is the primary driver behind the exhibited mismatch pattern. It does, however, seem reasonable to assume that ability is part of the explanation behind the results seen in this thesis. A way to know more about how ability plays a role in the mismatch mechanism would be to use a causal method (as used by Tsai 2010) to control for individual heterogeneity, and investigate if the found associations remain.

- *H3 - discrimination*: the theoretical function of discrimination in the case of mismatches is twofold: either, a person is discriminated against by not getting a job, or by getting a job below the acquired education level. The available previous research on the subject (section 2.2.2) suggests that discrimination in Sweden affects the possibility of employment rather than other labour market outcomes (for example, wage discrimination). To what extent is is likely to results in a downwards status transition is hard to answer, but this effect is theoretically plausible, although complicated. A direct effect of discrimination on mismatch could be that a person applies for a job but is offered another one, of lower status. An indirect effect that sets different norms for natives and immigrants

(thereby making immigrants apply for lower-status jobs) is more plausible, but difficult to demonstrate. Using the results presented in this thesis, discrimination as an explanation for the persistence of mismatches can neither be confirmed not denied. However, some observations can be made: mismatch coefficients show differences between countries of origin, where likelihoods of upwards mobility are higher for immigrants that are culturally and linguistically closer to Sweden (such as Germany and the US). This could be driven by discrimination, but it might also be driven by differences in transferability of human capital or a signalling effect.

- *H4 - labour demand*: A decrease in demand for immigrant labour could, in theory, increase the likelihood of being overeducated, since applicants would be forced to apply for jobs below their acquired level of education. The fall in demand could stem from macroeeconomic changes (Dustmann, Glitz, and Vogel 2010) and/or institutional changes in the economy (Rosholm, Scott, and Husted 2006). The results from this thesis show small but significant coefficients indicating that cohorts later than the pre-1970 cohort had a better situation, which would be in line with Sweden experiencing an economic boom during the 1980s. Despite this, I would hesitate to conclude that this is unambiguous support for the demand theory. The results do by no means rule out the presence of demand effects, but a more detailed model would have to be used to show them.

- *H5 - transferability*: As mentioned in H3, the results in this thesis show differences in likelihoods of upwards mobility between countries of origin, where immigrants that are culturally and linguistically closer to Sweden have higher likelihoods. This can be interpreted as an indication of returns to lower transferability costs, but it is also important to point out that this could also be due to discrimination or signalling. The results do not indicate, however, any significant recuperation of likelihoods, which could otherwise be expected if the post-migration mismatch was due to lack in country-specific human capital. A similar result is found by Joona, Gupta, and Wadensjö (2014, p.20), where persistence in overeducation remains even after controlling for region of origin. In this sense, the results point more towards an ability explanation.

Regarding coefficients from other included covariates, some comments can be made: the marginal effects from age (as shown in the ME figure 5.4.1) correspond well with what could be expected from previous research. The age ME:s found in this thesis, however, have large standard errors and should therefore be interpreted with some caution. Also the coefficients on gender, with females experiencing a higher likelihood of overeducation, are in line with previous results (Leuven, Oosterbeek, et al. 2011, p.298). It can also be noted that generally, the results from the ordinal and multinomial model both show the same pattern.

7 Discussion

In this thesis, a number of angles on the pre- and post-migration mismatch phenomenon have been presented. Using a maximum likelihood method, persistence in the mismatched state pre- and post-migration, and over time, has been shown. Because of the non-causal, non-specific econometric model used, no causal effects can be shown, but rather associations. A

number of hypotheses based on theory and previous research was presented, and discussed in the light of the findings. Also as a result of the non-causal method, no hypothesis could be rejected, as more detailed models would be needed to achieve this. This does not mean that the findings of are of no value. Since the dataset that was constructed as a part of this thesis is unique, the results from this thesis give an important indication of the primary mechanisms concerning mismatches among Swedish immigrants.

Concentrating on this primary mechanism, meaning the possible explanations as to why Swedish immigrants might experience labour market mismatches, seemed like a natural first choice considering that the data source was newly constructed. When looking at the previous literature, however, outcomes that are theoretically further down the causal chain (most often wage outcomes) are by far the most studied. It is apparent from the theoretical part of this thesis that the mechanisms behind only the incidence of labour market mismatches among immigrants are complex enough to disentangle - showing also the explanations for wage outcomes would naturally be a second step after the incidence.

As has been shown, there are a number of issues to keep in mind when interpreting the results. The most important of these are the fact that the models can only show associations, not causal relations, because of the non-causal method used. Other studies employing causal methods (such as Tsai 2010) have found that when controlling for individual heterogeneity, some of the wage penalty from overeducation disappears, suggesting that low-ability selection is part of the story. This might be an interesting starting hypothesis for a future study using the same data as in this thesis. In any case, the results from this thesis can be regarded as upper-bound estimates, containing uncontrolled variation from some of the theoretical sources shown.

It is also important to consider the applicability of the results from this thesis. The data used are specific of the time and sample, and influenced both by the Swedish society at the time and the characteristics of the immigrant cohorts included. The extent to which the results can be extrapolated to other countries and time periods can always be debated, but I would be surprised if completely different associations were found, should the model be run using current Swedish data. The results from Joona, Gupta, and Wadensjö (2014) suggest, even though they do not have access to pre-migration data, that a similar persistence pattern can be found in Sweden today.

The question of which explanation that is behind the results can not be given a single and secure answer. It is clear, though, that there is more than one factor involved in shaping the labour market outcomes for immigrants in Sweden. The explanations of individual ability, transferability, discrimination, signalling and labour demand are all plausible, but in different ways. While the discrimination and ability explanations can considered to be constant over time (assuming that the level of discrimination does not change), transferability and signalling can be argued to be more self-reinforcing processes. In the case of an individual being able to enter the labour market, the strength of transferability and signalling as negative effects will diminish over time, as country-specific human capital is accumulated. If, on the other hand, an individual does not succeed in entering the Swedish labour market, these explanations can not be separated from the time-invariant explanations mentioned. Since the sample in

this thesis includes only employed individuals, and a high degree of persistence is found, the story that is being told is at least likely to include both explanations of discrimination and individual ability.

The results found in this thesis, consisting of the persistence associations found, will contribute valuable information that broadens the discussion on immigrant labour market integration in Sweden. In order to say more about which explanations are more important, and thereby what policy implications the results might have, more research is needed. But it is clear from the results of this thesis that not one explanation, and thereby not one solution, will be enough when creating policies for successful labour market integration.

7.1 Future research

An important future contribution using the SLI database would be to use a more causal approach to solidify and broaden the associations found in this thesis. For example, the possible explanations behind overeducation mentioned in this thesis are interesting to hypothesise individually, in order to know more about where to direct policy. An example of this might be to use a method similar to Joona, Gupta, and Wadensjö (2014) and Tsai (2010), in order to know more about how much individual heterogeneity plays a role.

References

Abbott, Andrew and Angela Tsay (2000). "Sequence analysis and optimal matching methods in sociology review and prospect". In: *Sociological methods & research* 29.1, pp. 3–33.

Angrist, Joshua D and Jörn-Steffen Pischke (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Arrow, Kenneth J (1973). "Higher education as a filter". In: *Journal of public economics* 2.3, pp. 193–216.

Bartus, Tamás et al. (2005). "Estimation of marginal effects using margeff". In: *Stata journal* 5.3, pp. 309–329.

Belot, Michèle VK and Timothy J Hatton (2012). "Immigrant Selection in the OECD*". In: *The Scandinavian Journal of Economics* 114.4, pp. 1105–1128.

Bevelander, Pieter and Helena Skyt Nielsen (2001). "Declining Employment Assimilation of Immigrant Males in Sweden". In: *Journal of Population Economics*.

Borjas, George J (1985). "Assimilation, changes in cohort quality, and the earnings of immigrants". In: *Journal of labor Economics*, pp. 463–489.

— (1987). *Self-selection and the earnings of immigrants*.

— (2005). *Labor economics*. Vol. 6. McGraw-Hill New York:

Buis, Maarten L. (2013). *OPARALLEL: Stata module providing post-estimation command for testing the parallel regression assumption*. URL: `https://ideas.repec.org/c/boc/bocode/s457720.html`.

Buis, Maarten, Richard Williams, et al. (2013). "Using simulation to inspect the performance of a test, in particular tests of the parallel regressions assumption in ordered logit and probit models". In: *German Stata Users' Group Meetings 2013*. 06. Stata Users Group.

Büchel, Felix and Antje Mertens (2004). "Overeducation, undereducation, and the theory of career mobility". In: *Applied economics* 36.8, pp. 803–816.

Cameron, Adrian Colin and Pravin K Trivedi (2010). *Microeconometrics using stata*. Vol. 2. Stata Press College Station, TX.

Carlsson, Magnus and Dan-Olof Rooth (2007). "Evidence of ethnic discrimination in the Swedish labor market using experimental data". In: *Labour Economics* 14.4, pp. 716–729.

Chevalier, Arnaud (2003). "Measuring over-education". In: *Economica* 70.279, pp. 509–531.

Chiswick, Barry R (1978). "The effect of Americanization on the earnings of foreign-born men". In: *The journal of political economy*, pp. 897–921.

Chiswick, Barry R and Paul W Miller (2002). "Immigrant earnings: Language skills, linguistic concentrations and the business cycle". In: *Journal of Population Economics* 15.1, pp. 31–57.

— (2008). "Why is the payoff to schooling smaller for immigrants?" In: *Labour Economics* 15.6, pp. 1317–1340.

Duncan, Greg J and Saul D Hoffman (1981). "The incidence and wage effects of overeducation". In: *Economics of Education Review* 1.1, pp. 75–86.

Dustmann, Christian, Albrecht Glitz, and Thorsten Vogel (2010). "Employment, wages, and the economic cycle: Differences between immigrants and natives". In: *European Economic Review* 54.1, pp. 1–17.

Friedberg, Rachel M (1996). *You can't take it with you? Immigrant assimilation and the portability of human capital*. Tech. rep. National Bureau of Economic Research.

Giles, Dave (2013). *Robust Standard Errors for Nonlinear Models*. URL: `http://davegiles.blogspot.se/2013/05/robust-standard-errors-for-nonlinear.html`.

Greene, William H (2003). *Econometric analysis*. Pearson Education India.

Gujarati, Damoder N (2009). *Basic econometrics*. Tata McGraw-Hill Education.

Helgertz, Jonas (2010). *Immigrant careers-why country of origin matters*. Vol. 53. Lund University.

— (2013). "Pre-to Post-Migration Occupational Mobility of First Generation Immigrants to Sweden from 1970–1990: Examining the Influence of Linguistic Distance". In: *Population Research and Policy Review* 32.3, pp. 437–467.

Horrace, William C and Ronald L Oaxaca (2006). "Results on the bias and inconsistency of ordinary least squares for the linear probability model". In: *Economics Letters* 90.3, pp. 321–327.

Joona, Pernilla Andersson, Nabanita Datta Gupta, and Eskil Wadensjö (2014). "Overeducation among immigrants in Sweden: incidence, wage effects and state dependence". In: *IZA Journal of Migration* 3.1, pp. 1–23.

Kaminska, Olena and Tom Foulsham (2013). *Understanding sources of social desirability bias in different modes: evidence from eye-tracking*. Tech. rep. ISER Working Paper Series.

Keele, Luke and Nathan J Kelly (2006). "Dynamic models for dynamic theories: The ins and outs of lagged dependent variables". In: *Political analysis* 14.2, pp. 186–205.

Klinthäll, Martin (2007). "Refugee return migration: return migration from Sweden to Chile, Iran and Poland 1973–1996". In: *Journal of Refugee Studies* 20.4, pp. 579–598.

Korpi, Tomas and Michael Tåhlin (2009). "Educational mismatch, wages, and wage growth: Overeducation in Sweden, 1974–2000". In: *Labour Economics* 16.2, pp. 183–193.

Leuven, Edwin, Hessel Oosterbeek, et al. (2011). "Overeducation and mismatch in the labor market". In: *Handbook of the Economics of Education* 4, pp. 283–326.

Long, J Scott and Jeremy Freese (2006). *Regression models for categorical dependent variables using Stata*. Stata press.

Loughran, David S and Julie M Zissimopoulos (2009). "Why wait? The effect of marriage and childbearing on the wages of men and women". In: *Journal of Human resources* 44.2, pp. 326–349.

Massey, Douglas S et al. (1999). *Worlds in Motion: Understanding International Migration at the End of the Millennium: Understanding International Migration at the End of the Millennium*. Clarendon Press.

Mavromaras, Kostas and Seamus McGuinness (2012). "Overskilling dynamics and education pathways". In: *Economics of Education Review* 31.5, pp. 619–628.

McGuinness, Seamus (2006). "Overeducation in the labour market". In: *Journal of economic surveys* 20.3, pp. 387–418.

Miller, Paul W (1999). "Immigration policy and immigrant quality: The Australian points system". In: *The American Economic Review* 89.2, pp. 192–197.

Mincer, Jacob A (1974). "Age and Experience Profiles of earnings". In: *Schooling, experience, and earnings*. NBER, pp. 64–82.

Mundlak, Yair (1978). "On the pooling of time series and cross section data". In: *Econometrica: journal of the Econometric Society*, pp. 69–85.

Nordin, Martin and Dan-Olof Rooth (2009). "The Ethnic Employment and Income Gap in Sweden: Is Skill or Labor Market Discrimination the Explanation?*". In: *The Scandinavian journal of economics* 111.3, pp. 487–510.

OECD (2015). *International Migration Outlook 2015*. Organisation for Economic Co-operation and Development.

Pecoraro, Marco (2014). "Is There Still a Wage Penalty for Being Overeducated But Well-matched in Skills? A Panel Data Analysis of a Swiss Graduate Cohort". In: *Labour* 28.3, pp. 309–337.

Piracha, Matloob, Massimiliano Tani, and Florin Vadean (2012). "Immigrant over-and under-education: The role of home country labour market experience". In: *IZA Journal of Migration* 1.1, pp. 1–21.

Piracha, Matloob, Florin Vadean, et al. (2013). "Migrant educational mismatch and the labor market". In: *The International Handbook on the Economics of Migration* 9, pp. 176–192.

Robst, John (1994). "Measurement error and the returns to excess schooling". In: *Applied Economics Letters* 1.9, pp. 142–144.

— (1995). "Career mobility, job match, and overeducation". In: *Eastern Economic Journal* 21.4, pp. 539–550.

Rooth, Dan-Olof and Jan Saarela (2007). "Selection in migration and return migration: Evidence from micro data". In: *Economics letters* 94.1, pp. 90–95.

Rosholm, Michael, Kirk Scott, and Leif Husted (2006). "The Times They Are A-Changin': Declining Immigrant Employment Opportunities in Scandinavia". In: *International Migration Review* 40.2, pp. 318–347.

Sala, Guillem et al. (2011). "Approaches to skills mismatch in the labour market: a literature review". In: *Papers: revista de sociologia* 96.4, pp. 1025–1045.

Sattinger, Michael (1993). "Assignment models of the distribution of earnings". In: *Journal of economic literature* 31.2, pp. 831–880.

SCB (2015). *Sveriges befolkning ökar – men inte i hela landet*. URL: `http://www.scb.se/sv_/hitta-statistik/artiklar/sveriges-befolkning-okar--men-inte-i-hela-landet/`.

SCB (2016a). *In- och utvandrare 19602015 och prognos 20162060*. URL: http://www.scb.se/sv_/Hitta-statistik/Statistik-efter-amne/Befolkning/Befolkningsframskrivningar/Befolkningsframskrivningar/14498/14505/Aktuell-befolkningsprognos/Sveriges-framtida-befolkning-20152060/91832/.

— (2016b). *Socioekonomisk indelning (SEI)*. URL: http://www.scb.se/sv_/Dokumentation/Klassifikationer-och-standarder/Socioekonomisk-indelning-SEI/.

— (2016c). *Standard för svensk yrkesklassificering (SSYK)*. URL: http://www.scb.se/sv_/Dokumentation/Klassifikationer-och-standarder/Standard-for-svensk-yrkesklassificering-SSYK/.

Sicherman, Nachum and Oded Galor (1990). "A theory of career mobility". In: *Journal of political economy*, pp. 169–192.

STATA (2016). *STATA manual, Robust variance estimates*. URL: http://www.stata.com/manuals13/p_robust.pdf#p_robustRemarksandexamplesMaximumlikelihoodestimators.

Thurow, Lester C (1975). *Generating Inequality: The Distributional mechanisms of the economy*. National Technical Information Service.

Tsai, Yuping (2010). "Returns to overeducation: a longitudinal analysis of the US labor market". In: *Economics of Education Review* 29.4, pp. 606–617.

Weiss, Andrew (1995). "Human capital vs. signalling explanations of wages". In: *The Journal of Economic Perspectives* 9.4, pp. 133–154.

Verbeek, Marno (2008). *A guide to modern econometrics*. John Wiley & Sons.

Verhaest, Dieter and Eddy Omey (2012). "Overeducation, undereducation and earnings: further evidence on the importance of ability and measurement error bias". In: *Journal of Labor Research* 33.1, pp. 76–90.

Williams, Richard et al. (2012). "Using the margins command to estimate and interpret adjusted predictions and marginal effects". In: *Stata Journal* 12.2, p. 308.

Åslund, Olof and Dan-Olof Rooth (2007). "Do when and where matter? initial labour market conditions and immigrant earnings*". In: *The Economic Journal* 117.518, pp. 422–448.

A Appendix

	(1) First census	(2) Second census	(3) Third census	(4) Fourth census	(5) Fifth census
Undereducated	-0.179***	-0.160***	-0.138***	-0.0887***	-0.133***
	(0.0108)	(0.0120)	(0.0168)	(0.0211)	(0.0323)
Required Education	0	0	0	0	0
	(.)	(.)	(.)	(.)	(.)
Overeducated	0.511***	0.529***	0.520***	0.534***	0.486
	(0.0144)	(0.0226)	(0.0327)	(0.0418)	(0.978)
Observations	3875	2038	1216	594	177
Pseudo R^2	0.362	0.352	0.362	0.345	0.419

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 20: Average Marginal Effects (multinomial model): outcome as overeducated from pre-migration mismatch status

	(1) First census	(2) Second census	(3) Third census	(4) Fourth census	(5) Fifth census
Undereducated	-0.127***	-0.269***	-0.295***	-0.255***	-0.295***
	(0.0330)	(0.0379)	(0.0410)	(0.0622)	(0.0846)
Required Education	0	0	0	0	0
	(.)	(.)	(.)	(.)	(.)
Overeducated	-0.385***	-0.362***	-0.320***	-0.298***	-0.239**
	(0.0154)	(0.0242)	(0.0344)	(0.0457)	(0.0860)
Observations	3875	2038	1216	594	177
Pseudo R^2	0.362	0.352	0.362	0.345	0.419

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 21: Average Marginal Effects (multinomial model): outcome as matched (required education) from pre-migration mismatch status

	(1) First census	(2) Second census	(3) Third census	(4) Fourth census	(5) Fifth census
Undereducated	0.306***	0.430***	0.433***	0.343***	0.428***
	(0.0318)	(0.0366)	(0.0387)	(0.0596)	(0.0795)
Required Education	0	0	0	0	0
	(.)	(.)	(.)	(.)	(.)
Overeducated	-0.126***	-0.167***	-0.200***	-0.235***	-0.247***
	(0.00673)	(0.00944)	(0.0126)	(0.0220)	(0.0279)
Observations	3875	2038	1216	594	177
Pseudo R^2	0.362	0.352	0.362	0.345	0.419

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 22: Average Marginal Effects (multinomial model): outcome as undereducated from pre-migration mismatch status

	(1) Outcome: OE	(2) Outcome: RE	(3) Outcome: UE
Undereducated	-0.157***	-0.230***	0.387***
	(0.00810)	(0.0292)	(0.0304)
Required Education	0	0	0
	(.)	(.)	(.)
Overeducated	0.742***	-0.686***	-0.0562***
	(0.0139)	(0.0151)	(0.00364)
Observations	3875	3875	3875
Pseudo R^2	0.322	0.322	0.322

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 23: Marginal Effects at Means (ordinal model): three outcomes from pre-migration mismatch status.

Note: MEM means for table 23 available on request from author.

Interaction models, first census	(1) Model 1 b/se	(2) Model 2 b/se	(3) Model 3 b/se	(4) Model 4 b/se
Pre-migration UE	1.90***	2.83***	2.43***	2.50***
	(0.25)	(0.25)	(0.18)	(0.30)
Pre-migration RE	0.00	0.00	0.00	0.00
	(.)	(.)	(.)	(.)
Pre-migration OE	-4.00***	-3.99***	-3.99***	-4.18***
	(0.26)	(0.22)	(0.16)	(0.27)
Chile	-0.70***	-0.69***	-0.67***	-0.70***
	(0.19)	(0.17)	(0.17)	(0.17)
Germany	1.12***	1.28***	1.29***	1.27***
	(0.15)	(0.13)	(0.13)	(0.13)
Greece	-0.49**	-0.30*	-0.28	-0.32*
	(0.17)	(0.15)	(0.15)	(0.15)
Iran	0.09	0.18	0.20	0.16
	(0.25)	(0.21)	(0.21)	(0.21)
Poland	0.00	0.00	0.00	0.00
	(.)	(.)	(.)	(.)
Turkey	-0.22	-0.01	-0.02	-0.03
	(0.18)	(0.16)	(0.15)	(0.16)
USA	0.58**	0.49**	0.50**	0.48**
	(0.18)	(0.15)	(0.15)	(0.15)
Yugoslavia	-0.18	-0.02	-0.01	-0.04
	(0.16)	(0.14)	(0.14)	(0.14)
Undereducated × Chile	0.44			
	(0.44)			
Undereducated × Germany	1.62***			
	(0.34)			
Undereducated × Greece	0.03			
	(0.65)			
Undereducated × Iran	0.64			
	(0.60)			
Undereducated × Poland	0.00			
	(.)			
Undereducated × Turkey	1.85***			
	(0.52)			
Undereducated × USA	-0.54			
	(0.45)			
Undereducated × Yugoslavia	1.04*			
	(0.42)			
Overeducated × Chile	-0.29			
	(0.40)			
Overeducated × Germany	-0.49			
	(0.33)			
Overeducated × Greece	0.40			
	(0.37)			
Overeducated × Iran	0.28			
	(0.57)			
Overeducated × Poland	0.00			
	(.)			
Overeducated × Turkey	-0.11			
	(0.52)			
Overeducated × USA	-0.14			
	(0.40)			
Overeducated × Yugoslavia	-0.03			
	(0.38)			

Interaction models, first census (cont.)	(1) Model 1 b/se	(2) Model 2 b/se	(3) Model 3 b/se	(4) Model 4 b/se
Undereducated × Worker		0.00		
		(.)		
Undereducated × Tied Mover		-0.39		
		(0.28)		
Undereducated × Refugee		-0.11		
		(0.41)		
Required Education × Worker		0.00		
		(.)		
Required Education × Tied Mover		0.00		
		(.)		
Required Education × Refugee		0.00		
		(.)		
Overeducated × Worker		0.00		
		(.)		
Overeducated × Tied Mover		-0.04		
		(0.25)		
Overeducated × Refugee		0.36		
		(0.33)		
Undereducated × male			0.00	
			(.)	
Undereducated × female			0.30	
			(0.24)	
Required Education × male			0.00	
			(.)	
Required Education × female			0.00	
			(.)	
Overeducated × male			0.00	
			(.)	
Overeducated × female			0.14	
			(0.23)	
Undereducated × 70 and earlier				0.00
				(.)
Undereducated × 71-75				-0.49
				(0.37)
Undereducated × 76-80				0.45
				(0.37)
Undereducated × 81-85				0.79
				(0.41)
Undereducated × 86-90				-0.24
				(0.43)
Overeducated × 71-75				0.34
				(0.35)
Overeducated × 76-80				0.22
				(0.34)
Overeducated × 81-85				0.28
				(0.34)
Overeducated × 86-90				0.27
				(0.35)
Observations	3875	3875	3875	3875
Pseudo R^2	0.330	0.323	0.323	0.325

Robust Standard Errors model, 1st census	(1) First census (without robust) b/se	(2) First census, robust standard errors b/se
Pre-migration UE	2.58***	2.58***
	(0.14)	(0.18)
Pre-migration RE	0.00	0.00
Pre-migration OE	-3.94***	-3.94***
	(0.14)	(0.15)
Cohort: 70 and earlier	0.00	0.00
Cohort: 71-75	0.15	0.15
	(0.11)	(0.11)
Cohort: 76-80	0.66***	0.66***
	(0.12)	(0.12)
Cohort: 81-85	0.76***	0.76***
	(0.13)	(0.14)
Cohort: 86-90	0.50***	0.50**
	(0.15)	(0.15)
Gender: male	0.00	0.00
Gender: female	-0.60***	-0.60***
	(0.08)	(0.09)
Age of Individual	0.03	0.03
	(0.03)	(0.03)
Age squared	-0.00	-0.00
	(0.00)	(0.00)
Pre-migration RE: Primary	0.00	0.00
Pre-migration RE: Secondary	-3.06***	-3.06***
	(0.14)	(0.15)
Pre-migration RE: Post	-4.11***	-4.11***
	(0.14)	(0.14)
Chile	-1.97***	-1.97***
	(0.18)	(0.19)
Germany	0.00	0.00
Greece	-1.58***	-1.58***
	(0.14)	(0.14)
Iran	-1.10***	-1.10***
	(0.22)	(0.25)
Poland	-1.29***	-1.29***
	(0.13)	(0.14)
Turkey	-1.31***	-1.31***
	(0.15)	(0.15)
USA	-0.80***	-0.80***
	(0.15)	(0.16)
Yugoslavia	-1.31***	-1.31***
	(0.13)	(0.13)
Years since Migration	0.08***	0.08***
	(0.01)	(0.01)
Non-metropolitan area	0.00	0.00
Metropolitan area	-0.16*	-0.16*
	(0.08)	(0.08)
Visa category: Worker	0.00	0.00
Visa category: Tied mover	-0.59***	-0.59***
	(0.10)	(0.10)
Visa category: Refugee	-0.54***	-0.54***
	(0.16)	(0.16)
Civil status: Unmarried	0.00	0.00
Civil status: Married	-0.12	-0.12
	(0.11)	(0.11)
Civil status: Widow/widower	0.06	0.06
	(0.32)	(0.30)
Civil status: Divorced	-0.17	-0.17
	(0.15)	(0.16)
Observations	3875	3875
Pseudo R^2	0.322	0.322