TALLINN UNIVERSITY OF TECHNOLOGY
School of Business and Governance

Artemi Beljakov

EDUCATION-JOB MISMATCH – IMPLICATIONS FOR INDIVIDUAL EARNINGS IN ESTONIA

Master's thesis

Programme TAAM, specialisation Economic Analysis

Supervisor: Simona Ferraro, PhD

Tallinn 2023

I hereby declare that I have compiled the thesis independently
and all works, important standpoints and data by other authors
have been properly referenced and the same paper
has not been previously presented for grading.

The document length is 14747 words from the introduction to the end of the conclusion.

Artemi Beljakov 09.05.2023
(date)

TABLE OF CONTENTS

ABSTRACT

The thesis aims to estimate the impact of education-job mismatch in the Estonian labour market on the earning levels of individuals. To the knowledge of the author, previous studies have been done on assessing mismatch incidence and also impact on wages, but no study has been done in recent years, and previous results have started to become outdated. This study analyses the Estonian labour market for the period from 2009 to 2019 using data from the Estonian Labour Force Survey. Instrumental variable fixed effect estimation is used in the thesis to account for heterogeneity and endogeneity alongside cohort transformation of individual observations. The estimation results for different mismatch methods are consistent for both Duncan & Hoffman (year) and Verdugo & Verdugo (year) models, while the effect on wage varies between a 3.1 percent penalty to a 17.6 percent premium, depending on the mismatch type and model specification. Overall findings suggest wage premium for overeducated individuals against those adequately matched in the same occupation and the opposite for undereducated. Similarly, the wage premium for undereducated individuals compared to those adequately matched in the same level of education and the opposite for overeducated. The aim of the thesis is archived, but further research with different data sources is required to validate the results and investigate the effects of alternative concepts like skill mismatch.

Keywords: education-job mismatch, overeducation, undereducation, wage penalty, Estonian Labour Force Survey.

INTRODUCTION

In the past few decades, there has been a significant rise in the level of educational attainment among populations in advanced industrialised societies. To make the most out of the investment in human capital, it is necessary to ensure a good match between competencies gained during schooling and those required in the labour market (Quintini, 2011). There is, however, substantial variation between countries when it comes to education-job mismatch (Davia et al., 2017). When it comes to individual workers, research indicates that those who are overeducated tend to earn less than their peers with similar education whose jobs match their qualifications because the proportion of investment into years of schooling and education is being underutilised (McGuinness, 2006). As pointed out by Allen & Van Der Velden (2001), referencing other studies, overeducation is known to increase labour turnover in certain cases and is associated with lower productivity and job satisfaction. Overeducated workers may push down less-educated workers in the labour market or become unemployed (Davia et al., 2017).

Education-job mismatch in Estonia has been addressed by multiple studies using different survey data. European Commission (2016, p. 14-16), in their report, pointed out that there are almost 30% of the workforce in Estonia is categorised as overqualified, which sets Estonia in fourth place among all European countries. Similar results are shown by European Centre for the Development of Vocational Training or CEDEFOP (2015, p. 33–35), indicating relatively high overqualified levels among the Estonian workforce at around 26 percent or the 6^{th} highest within the 28 European countries. In the latest study using the Programme for the International Assessment of Adult Competencies (PIAAC), the authors Halapuu & Valk (2013) show that over one-third of workers in Estonia are overeducated and the probability of being undereducated increases among older people and those with higher education. They place Estonia in first place among countries with overeducation that participated in the PIAAC study. The contribution of this thesis would be an update of the results for the Estonian labour market using data from the Estonian Labour Force Survey from 2009 to 2019.

The aim of the thesis is to estimate the effect of educational mismatch on individual earnings in the Estonian labour market. The thesis intends to analyse methods used to identify the educational mismatch and potential reasons which might be causing the mismatch to have an effect on the income levels of the individual.

The main research questions assessed in the thesis are:

1. Do overeducated (undereducated) workers experience wage bonus (penalty) when compared to ones with the same employment but matching levels of education in Estonia?
2. Do overeducated (undereducated) workers experience wage penalty (bonus) when compared to ones with the same level of education but matching employment in Estonia?

In the thesis the following hypotheses are tested:

H_1: Overeducated (undereducated) face a wage advantage (disadvantage) compared to individuals with the same employment but matching levels of education;

H_2: Overeducated (undereducated) face a wage disadvantage (advantage) compared to individuals with the same level of education but matching employment.

To answer the research questions, the thesis is structured in the following way. The first chapter provides the theoretical background for the analysis and gives an overview of the main theories behind the formulation of wages. Next, the author gives an overview of the methods used in the empirical framework to define educational mismatches. The final sub-chapter gives insights into the mismatch incidences reported in the previous studies and a closer overview of the educational mismatch in Estonia.

In the second chapter, the author provides an outline of the methodology and data used in the research. The methodology section entails the rationale for the chosen methods – Duncan & Hoffman (1981) and Verdugo & Verdugo (1989) based on Mincer's (1974) wage equation. The data part presents the Estonian Labour Force Survey (ELFS) covering the period from 2009 to 2019, with data sampling procedures and constraints. Sub-chapter ends with descriptive statistics for chosen variables.

The third chapter presents the results of the empirical analysis of the effect of educational mismatches on the wages of individuals participating in the Estonian labour market. By using educational mismatches derived from job analysis and realised matches methods, the author shows the effect on wages, followed by the robustness analysis. Lastly, the author discusses the results based on the empirical and theoretical framework.

The author of the thesis would like to acknowledge his supervisor Simona Ferraro for her support and advice during the process of writing the thesis.

1. THEORETICAL FRAMEWORK

This chapter provides an overview of the theoretical framework behind education-job mismatch and definitions used to outline the mismatch in empirical studies. First, there is an overview of the main theories behind wage formation and how education-job mismatch could be described within these theories. The overview also includes examples of education-job mismatches and rationale, then ends with providing potential reasons for a mismatch on an individual level. The following sub-chapter describes all the main definitions of mismatch and how they are measured. The last sub-chapter provides an empirical overview of the implications and importance of education-job mismatch in different countries, including Estonia.

1.1. Theories behind over and under-education

Overeducation outlines an individual's excessive level of education compared to the required level for any particular job. Such excess in the level of education is potentially costly to the individual himself, as well as firms and economies. (McGuinness, 2006, p. 388) When it comes to conceptual theories explaining the overeducation phenomenon, there are three commonly referred to: the human capital theory, the job assignment theory, and the job competition theory.

The human capital theory by Gary Becker (1962) stipulates that firms fully utilise the capacity of their available workforce. The individual is thus paid by his/her marginal productivity, which is set by the level of the accumulated education and working experience, i.e. the human capital. Overeducation due to extra time spent in schooling could be seen as a substitute for the experience usually gained through work, which is more common among young individuals. Within this theoretical framework, overeducation is the result of low-quality human capital or its insufficiency. (Caroleo & Pastore, 2018, p. 1002)

Lester Thurows' Job competition theory, published in the book Unemployment and Inflation (Piore, 1979), describes the labour market from the perspective of training costs. The labour market is not a place for individuals to exchange their existing skills and competencies but instead, a training marketplace where training slots are allocated between individuals. This observation is based on Thurow examination of the United States (US) surveys showing that a greater part of the skills necessary in the labour market is attained via on-site job training and not through formal education. Allocation of training slots is based on the position within the queue, which is based on the number of training costs that should be invested into them. The fewer costs there are, the higher the individual is in the queue. Once individuals reach the top of the queue, their wage is defined exclusively by the characteristics of the job. McGuinness (2006, p. 392) points out that this opposes the idea that marginal productivity is defined by individuals' human capital but instead relies fully on the job. He concludes that within the job competition model, overeducation results from individuals protecting their places within the queue and trying to increase their placement in the queue by attaining an additional level of education.

A middle ground, taking points from both the human capital and job competition theories, is the assignment theory introduced by Sattinger (1993, p. 831–834). His theory also uses income or utility maximisation to explain the choice of a particular job by the individuals in question. Workers within specific sectors are not a result of random allocation but rather the result of their choices to maximise utility. The wage rates for individuals are an equilibrium outcome between individuals' own characteristics, job sector specifics, and available resources in the economy. Over- and undereducation in this model is explained as a result of the equilibrium between income and the factors mentioned earlier, where education-job match can be ignored in favour of higher income.

Other theories supporting assignment theory are job matching and turnover theory by Jovanovic (1979) and Sicherman & Galor (1990) career mobility. Turnover theory suggests overeducation as a result of misinformation in the labour market and, once realised by individuals, is addressed through repeated job searches. Similarly, career mobility describes overeducation as a willful choice to accelerate career progression and acquire necessary skills on-site. (Davia et al., 2017, p. 68)

Skill heterogeneity plays an important role in explaining educational mismatches, as highly skilled individuals could be occupied in positions with higher formal qualification requirements than they have and will appear as undereducated despite having competencies for the job. Contrary, individuals with a lack of skills and competencies will end up in positions requiring lower formal qualifications than originally owned and will appear as overeducated. (Quintini, 2011, p. 20)

Education-related mismatches due to over- and undereducation can also be explained with factors related to education aspects on an individual level, labour market, job-related specifics, and individual determinants (Somers et al., 2019, p. 583–587). Education-related aspects resulting in mismatch depend on the characteristics of study programs, as certain programs like liberal arts have high rates of mismatch, whereas health-related fields have the lowest. This can be explained by the fact that health-related fields equip graduates with occupation-specific competencies and skills. Those from liberal arts are more likely to end up searching for jobs outside their education field. (Wolbers, 2003, p. 255–257) Similar high mismatch rates have been found in the arts and humanities fields, while the lowest have been among health sciences and education (Boudarbat & Chernoff, 2012, p. 1925).

Job tenure results in the accumulation of company-specific skills that might make an individual less attractive to other employers, making it difficult for an employee with long tenure to find matching employment outside the company (Somers et al., 2019, p. 585). On the other hand, temporary job tenure does not provide enough incentive for employers to invest in employees, as one is expected to leave the company earlier when compared to full-time employees (Becker, 1962; Wolbers, 2003). Therefore, individuals with temporary contracts have a higher possibility of being mismatched (Boudarbat & Chernoff, 2012, p. 1925). When it comes to job tenure, there is typically a better match between education and employment in the public sector compared to the private sector on a sector-by-sector basis. The reasons could be attributed to education-related aspects, as the public sector typically consists of all healthcare and educational organisations. (Wolbers, 2003, p. 255–257)

For individual determinants, women may report themselves as being mismatched due to family-related reasons (Bender & Heywood, 2011, p. 264). Marital status and the existence of children have not been found to be related to overeducation. Working full-time, however, decreases the likelihood of underqualification while the existence of children increases it. These combined provide some support to the idea that women report themselves as mismatched, as they end up choosing less demanding jobs for which they are matching skill-wise but overeducated. (Quintini, 2011, p. 26) Men would report mismatch due to career-related decisions, including better wages or promotion possibilities. When looking at career paths for both women and men, a mismatch is more likely to occur in the later stages: individuals might take new jobs due to structural changes and certain occupations becoming obsolete or simply because there was no job available. (Bender

& Heywood, 2011, p. 261–264) Individual decisions could also be driven by overall economic conditions, as recessions might put pressure on the labour force to adjust their expectations and goals to a lower level, thus accepting jobs with lower educational requirements or from another field (Wolbers, 2003, p. 251).

1.2. Defining educational mismatches

Over or undereducation, further characterised as an educational mismatch, is a consequence of an individual level of schooling being over or under the level required by his/her particular job. This mismatch can be captured and classified using different methods, that in turn can be divided into “objective” and “subjective” types. The objective classification method can be further broken down into the job evaluation method and realised matches or the empirical method. (Groot & Maassen Van Den Brink, 2000; Leuven & Oosterbeek, 2011). Finally, the educational mismatch can be of two types: vertical and horizontal (Pecoraro, 2016, p. 537). All classification methods and mismatch types are discussed below.

The subjective type, namely the workers' self-assessment (SA), involves surveying individuals on education levels and whether they identify themselves as over(under-)educated or if their education-job level matches. An alternative approach is to ask about individuals’ perceived minimum educational requirements for their jobs. In the second case, the reported minimum requirements are compared with the actual education level of the surveyed to identify the educational mismatch type. (Groot & Maassen Van Den Brink, 2000, p. 150). There are some examples of survey questions from other studies that Hartog (2000, p. 132) brings up in his work, which have been used to directly or indirectly question the assessment of requirements for the job: “How much formal education is required to get a job like yours?” (Sicherman, 1991) or “What kind of education does a person need in order to perform your job?” (Alba-Ramirez, 1993).
The advantage of the subjective approach and SA is that the relevant information from a primary source is not subject to aggregation. However, the quality and reliability depend on the primary source and how accurate and unbiased response the surveyee can provide. Individuals might overemphasize the job requirements to give a better impression of their current position. They could also reflect the present hiring standards that are already subject to inflated schooling levels, whereas the job requirements have been unchanged. (Hartog, 2000, p. 200)

The job evaluation (JA) method involves using metrics like Standard Occupational Classification System (for the United Kingdom) from professional organisations to determine required education levels and comparing with the actual level of an individual in question (McGuinness, 2006, p. 396). Other metrics like the International Standard Classification of Occupations (ISCO) can also be used, but the job and education requirements set within these metrics should match in all countries using ISCO classification (McGuinness et al., 2018; Quintini, 2011). It is also possible to use the JA method based on the occupational dictionaries prepared by professional job analysts. Such an approach is deemed more accurate as it uses field expertise. However, it also has the following disadvantages: expensive to conduct, outdated if not updated regularly, and still involves a certain degree of subjectivity. (McGuinness et al., 2018; Verhaest & Omey, 2012)

A second option, or the realised matches (RM) method (also referred to as the empirical approach), is to calculate the mean (introduced by Verdugo & Verdugo, 1989, p. 633) or modal (introduced by Kiker et al., 1997) education level for all individuals within one occupation, and compare with actual levels. For that, duration of schooling or highest attained degree levels is used. The individual is then placed on a scale relative to the mean or modal level, and if more than one standard deviation is above or below the mean level, assigned as over- or undereducated. (McGuinness, 2006, p. 396)

Ease of use is the primary advantage of this method, as it can be applied to micro datasets like labour force surveys, given that educational attainment and occupation are available. Whereas actual skill requirements are not available due to aggregation showing average levels of all individuals, reflecting education levels necessary to get employment and not the ones to do it. Another constraint is the sample size, as the mode level of education would be derived for a broad group of occupations like health professionals. Such an approach will hide the education levels for specific occupations within the group, like nurses. (McGuinness et al., 2018, p. 988; Verhaest & Omey, 2010)

Mismatch types, as mentioned previously, are horizontal and vertical. When an individual has attained a degree within a particular field of education but is involved in an occupation from a different field, then such mismatch is referred to as horizontal mismatch (Robst, 2007; Somers et al., 2019). A vertical mismatch is between the required level of education to perform a job and the actual level of education (Pecoraro, 2016, 537).

Finally, there is a return to education specification, derived initially by Mincer (1958). The idea is that returns to schooling result in increases in individual earnings due to an increase in educational level. As described by Harmon et al. (2003, p. 116–118), Mincer specifies returns to schooling as an increase in earnings equivalent to the interest rate, which can be expected with every additional year of schooling.

Many studies estimate Mincer's specification to compare educational choices and compare mismatched individuals to adequately matched ones, as the difference in ratios highlights the potential penalty from making a specific education-related decision.

1.3. Empirical studies

In the overview of the Spanish labour market, where a study by Alba-Ramirez (1993) shows that recently graduated young workers are more likely to be overeducated. Furthermore, both over- and undereducated workers have higher turnover rates than adequately educated workers, with average occupancy duration lower by 16.6 to 5.5 percent, respectively. Using the RM method with mean values, Alba-Ramirez also finds out that the overeducation phenomenon tends to disappear with workers' age. Around 30 percent of workers of age 35 and over are overeducated, and the percentage further decreases to only 4 percent for workers older than 59 years. Finally, overeducation has been found to be connected with higher mobility rates, as changing occupations tends to result in an educational mismatch.

Similar results are in the study by Groot & Maassen Van Den Brink (2000, p. 153), where overeducation is seen as part of career mobility or entering the labour market. Using meta-analysis of multiple studies, they show that younger workers are more likely to be overeducated, as they start with less demanding jobs to earn experience and move to positions with matching educational requirements. Overall, the incidence of overeducation in the European Union (EU) and the US labour market appears to be about 26 percent and has not seen any significant changes during the observed period from the 1970s to the 1990s.

When looking at wage penalties due to mismatch, a study by Allen & van der Velden (2001, p. 443–445) of EU 11 countries and Japan show an 8 percent decrease each year due to overeducation, and when also accounting for skill mismatch, education mismatch seems to be more significant. In another study using PIAAC data in Spain, overeducated workers have a 3

percent premium in wage compared to those adequately educated within the same occupation, but about 17 percent penalty when compared to similarly educated and matched occupation (Nieto & Ramos, 2017, p. 229–232). They also point out that skill heterogeneity is important in determining wages but does not completely explain the effect of educational mismatch on wages.
Caroleo & Pastore (2018, p. 1026-1029) show that wage penalty related to educational mismatch is more significant than previously found in similar studies, especially for overeducation. AlmaLaurea data for Italy suggests a penalty between 21-25 percent, indicating low human capital capability for mismatched individuals, as the penalty for overskilling is also high (16 to 21 percent).

In the comprehensive work by McGuinness et al. (2018), 98 papers on the topic of overeducation are analysed, with 73 of them reporting an incidence of overeducation. Among 39 countries with an incidence of overeducation, the ones with consistently high rates are Italy, Spain, Greece, and Ireland. On the other hand, low rates are seen in the Czech Republic, Switzerland, Norway, and Finland. The prevalence of overeducation using subjective, empirical, and JA methodologies remains similar across the European countries, but as McGuinness et al. (2018) point out to the European Commission (2016), overeducation has significant variance depending on the measuring method used. One of the examples is Spain, where based on the JA method, the country has the highest incidence of overeducation in the EU. With empirical measures, it is third to last, emphasizing the challenges with estimating and interpreting the overeducation indicators.

Not all studies, however, show significant wage penalties arising from education mismatch. Work by Quintini (2011, p. 51), a working paper on employment and migration, looks at issues of overqualification and underskilled workers. Several data sources are used, including the European Survey of Working Conditions and the International Social Survey Programme (both for the 2005 period), and the European Community Household Panel. Model outcomes show a 20 percent wage penalty for overqualified workers compared to similarly educated but job-matched counterparts. When including controls for unobserved individual heterogeneity, Quintini's (2011, p. 33) model shows a more minor penalty of 3 percent for overqualified workers. Given also minor over-skilling penalties at about 1 percent, it could suggest that the level of attained education is a primary determinant of workers' wages.

Similar results are shown in a study by Pecoraro (2016, p. 542–547) for years between 1999 and 2004, using the Swiss Household Panel survey. He utilises RM and indirect SA methods to define overeducation and finds that returns to overeducation are associated with about a 2 percent wage penalty when compared to returns to actual (adequate) schooling. Within that penalty, a horizontal mismatch is deemed more serious and related to a higher wage penalty. It is important to note that in the study by Pecoraro, the share of adequately educated was between 62 and 64 percent, and the share of overeducated was about 20 percent but decreased significantly by 2004.

In general, the incidence of education-job mismatch is quite common but varies from country to country and based on the measurement method used. Also, all studies indicate wage penalties of different magnitude due to overeducation, starting as low as 3 percent and all the way to 25 percent.

1.4. Overview of the Estonian labour market

A study by Galasi (2008, p. 4–7) has included Estonia among other EU & European Economic Area countries while using European Social Survey data from 2004 to 2006. Using the subjective method, they report that Estonian workforce education matching is as follows: properly educated 8.2 percent, overeducated forming the central part with 78.9 percent, followed by 12.9 percent for undereducated. Estimated returns to education for proper education is close to 13 percent, while overeducation yields 4.4 percent, and undereducation results in a 2.2 percent loss. The required years of education are taken from the United Nations Educational, Scientific and Cultural Organization (UNESCO) statistical database, which is typical of other studies.

Looking at the Estonian case from the perspective of the ELFS, the study by Lamo & Messina (2010) analyses the period from 2003 to 2006. It focuses on overeducation and omits undereducation since only 2.5 percent of ELFS respondents classify themselves as undereducated. The empirical analyses in their study do not show any statistical significance between undereducated and well-matched workers. This leaves 12.6 percent of overeducated workers in the sample, which tends to be concentrated in the private sector, particularly in manufacturing, with an average job tenure of 4.5 years compared to 7.3 years for well-matched individuals. As for the outcomes, Lamo & Messina (2010) find out that the average wage penalty due to overeducation mismatch ranges between 24 to 27 percent for females and 18 to 24 percent for males.

Furthermore, when decomposing to age groups, their study indicates a smaller penalty for younger workers aged 16-29, between 4 and 9 percent for females and 8 to 13 percent for males but increases gradually for higher-aged groups. The highest age group, workers between 50 to 64, have the highest wage penalties from overeducation, around 33 to 35 percent for males and 29 to 30 percent for females. They conclude that, on average, the overeducation wage penalty in Estonia is much higher than in other EU countries, where similar studies were held.

Estonia participated in the PIAAC study in 2012 and has the results from an analysis of over 7500 responses. The main findings are that over one-third involved in the labour market are overeducated, which puts Estonia on the highest ranking among countries that participated in the PIAAC study. Education is also more important compared to skills in Estonia, where regardless of their proficiency in information processing, individuals with tertiary education have a distinct advantage when it comes to getting higher salaries. (Halapuu & Valk, 2013)

The estimates for over-qualification among tertiary graduates aged between 15-64 are available in the report by European Commission (2016, p. 14-16). The report is based on Eurostat's Labor Force Survey from 2013 and shows the following: using objective (or JA) indicators, Estonia has close to 30 percent of an over-qualified workforce. This puts Estonia in fourth place among countries with the highest overeducation. Using empirical indicators, the change is not as significant as in the case of Spain described above, with over 30 percent (closer to 35 percent) of the Estonian high-skilled workforce being subject to overeducation, and this time puts Estonia in 8 place out of 25. Regarding potential factors driving the mismatch, the report indicates Estonia as one of the countries where the customer service clerk position is predominantly occupied by high-skilled individuals with tertiary degree levels. In other countries like Germany or Sweden, these are more often occupied by individuals with upped secondary qualifications. (European Commission, 2016)

A similar report, focusing on EU-28 countries, carried out by the CEDEFOP (2015, p. 33–35) also indicates relatively high overqualified levels among the Estonian workforce, around 26 percent or the 6th highest within the sample (EU-28 average about 17 percent). The data is from a 2014 survey, and the overqualification levels were measured using the subjective (SA) method. When it comes to high levels of education, 15 percent of individuals with tertiary education are overqualified, which is now the fourth highest among the sample. Further, the report shows that over 70 percent of adult employees see their skills becoming obsolete in the next five years (European Centre for the Development of Vocational Training, 2015, p. 76-77).

2. METHODOLOGY AND DATA

This chapter of the thesis describes the methodology and the data used to estimate the effect of educational mismatch on the earning levels of individuals in the Estonian labour market. The first sub-chapter discusses the methodology, while the subsequent chapter introduces the data used in the empirical analysis and its limitations.

2.1. Methodology

This chapter will explain the methodology used in the empirical analysis to estimate the effect of educational mismatch on individual earnings. The methodology follows the previous studies conducted in the field and, first of all, the pioneering work by Mincer (1974) wage equation. The wage model looks at formal education as a proxy for individual human capital and is defined as follows in Equation (1):

$$\log(W_i) = \alpha + \beta_1 S_i^a + \beta_2 x_i + \beta_3 x_i^2 + u_i \quad (1)$$

where $\log(W_i)$ is the logarithm for an hourly real wage of worker i, βS_i^a is years of formal education, x_i is an individual experience, x_i^2 is square of individual experience and u_i is an error term.

A modified version of this equation has been used in two fundamental works by Duncan & Hoffman (1981) and Verdugo & Verdugo (1989). The Mincerain wage Eq. (1) is further specified into Over-Required-Under (ORU) education function by Duncan & Hoffman (1981), where the years of formal education are split into three variables: *(i)* years of education necessary to do the job, *(ii)* years of overeducation and *(iii)* years of undereducation. To calculate the years of over and undereducation, the following approach is used:
$S^o = S^a - S^r$ for the overeducated worker and 0 otherwise,
$S^u = S^r - S^a$ for undereducated worker and 0 otherwise.

With the additional specification, the ORU function takes the following form (Equation 2):

$$\log(W_i) = \alpha + \beta_1 S_i^r + \beta_2 S_i^o + \beta_3 S_i^u + \beta_4 x_i + \beta_5 x_i^2 + u_i \quad (2)$$

where $\log(W_i)$ is the logarithm for the hourly real wage of worker i, $\beta_1 S_i^r$ is the required education measured in years, $\beta_2 S_i^o$ is years of overeducation and $\beta_3 S_i^u$ years of undereducation, and the other covariates are the same as from Mincerian wage (Eq. 1).

The decomposition of years of education allows estimating the coefficients for over- and undereducation compared to adequately matched workers in the same job. A common finding in the literature shows that $\beta_1 > \beta_2 > |\beta_3|$ (Nieto & Ramos, 2017).

Verdugo & Verdugo (1989), in their study, are also concerned about the overeducation effect on earnings and introduce dummy variables to compare both over- and undereducation, to analyse whether there are diminishing returns on earning to an additional year of schooling. The model is specified in Equation (3):

$$\log(W_i) = \alpha + \beta_1 S_i^a + \beta_2 OE_i + \beta_3 UE_i + \beta_4 X_i + u_i \qquad (3)$$

where $\log(W_i)$ is the logarithm for the hourly real wage of worker i, $\beta_1 S_i^a$ is years of attained schooling, $\beta_2 OE_i$ and $\beta_3 UE_i$ are dummy variables for over- and undereducation respectively, X_i represents individual specific characteristics like years of working experience, region of residence, sector of employment and occupation categories, marital status. They also included the number of hours worked and weeks unemployed in 1979. Since in Eq. 3, the years of attained schooling are used, when compared to Eq. 2 with years of required schooling, the model interpretation is slightly different: Verdugo & Verdugo (1989) model shows the wage effect of over- and undereducation compared to workers with the same level of education.

The years of attained schooling are derived from the International Standard Classification of Education for 1997 (ISCED-97) and 2011 (ISCED-11) years by UNESCO Institute for Statistics (1999, 2012), and Estonian-centric acquired level of education reported in ELFS. These show the highest attained education level by an individual at the moment of participating in the survey. Because of that, the years of attained education are not actual years of schooling, which are not questioned in the ELFS. As Kemelbayeva (2020) notes in her work, having a similar issue estimating years of schooling would be rather an estimation of returns to credentials. She refers to Harmon et al. (2000) that this is still a reasonable approximation. The main indicators used to derive the attained years of schooling are those by UNESCO, but ISCED-97 has less detailed levels for those having first-stage tertiary education. To overcome it and correctly map ISCED-11 levels, the Estonian-centric variable is used, which reports level 5 first-stage tertiary education at a more granular level.

There are some issues related to ORU specification that has been pointed out by Verhaest & Omey (2012), which are related to heterogeneity and measurement error. When it comes to the specification of Eq. 3, McGuinness (2006), in his literature overview, points out the work by Cohn & Khan (1995), who have replicated the analysis carried out by Verdugo & Verdugo. They found that overeducated workers earn less than individuals with the same level of schooling but not overeducated. However, that does not imply negative returns to years of schooling. Using alternative definitions of required, over- and undereducation, the return to overeducation is rather positive.

The heterogeneity comes from hidden levels of ability that might be different from one individual to another, with the same level of education. Harmon et al. (2003) and McGuinness (2003) highlight that overeducation is only a sign of low ability level and the effect of earning greater income relative to people without academic degrees, the so-called sheepskin effect, should be studied in the context of both education mismatch and skill levels. To address this issue, studies include different proxies for skills levels like worker self-assessment on skill levels and their utilisation at work (Wen & Maani, 2022), worker skill proficiency score and self-assessment in a study by Nieto & Ramos (2017), or workers cognitive skills by Romero et al. (2017) and cognitive & technical skills by Wu & Wang (2018).

These types of skill proxies are not available in the ELFS, so alternative measures like work tenure (time an employee has worked in their current employment) and whether an individual is working full-time or part-time have been included. The first can serve as a potential estimate of an individual's ability levels acquired from current employment, while the second indicates the intensity of the learning by working. This, however, does not fully solve the potential omitted variable bias, so an additional step is to generate cohorts based on year of birth and gender so each individual is placed in the corresponding cohort. This provides multiple observations through the years of the survey instead of one observation per individual. Then, the means for numerical variables within each cohort are calculated and regressed using the fixed effects approach to demean the observations and remove the individual fixed effect.

As pointed out in multiple studies by Verhaest & Omey (2006, 2010, 2012), measurement error results in biased outcomes when overeducation and its effect are underestimated. Different measures of education mismatch all have shortcomings, and studies have addressed those by using instrumental variables. One measure of education mismatch is instrumented using an alternative measure of education mismatch. Wen & Maani (2022) use JA with RM methods (mean & mode) using one measure at a time while the remainder as instrumental variables. Iriondo & Pérez-Amaral (2016) derive mean and mode values from the JA method and use them as instruments, while Verhaest & Omey (2012) use two direct and indirect measures from the SA method to instrument for JA measure. As seen from a variety of methods used by different studies, there is no universally applicable one, and it depends on the underlying data and use case.

The education mismatch assessment method used in this thesis is based on previous studies and utilises the JA and RM methods. The design of ELFS using the International Standard Classification of Occupations (ISCO-08) is accurate and easily applicable due to respondents' occupation codes collected using ISCO-08 standards. Skill requirement levels for each major ISCO-08 group are shown in the table below:

Table 1. Mapping of ISCO-08 major groups to skill levels

ISCO-08 Major Group	Skill level
Managers	3 + 4
Professionals	4
Technicians and associate professionals	3
Clerical support workers	2
Services and sales workers Skilled agricultural, forestry and fishery workers Craft and related trades workers Plant and machine operators, and assemblers Elementary occupations	1
Armed forces occupations	1 + 2 + 4

Source: International Standard Classification of Occupations (2012); composed by the author

Occupations from armed forces are not present among respondents of the ELFS, thus being out of scope for this thesis. However, due to the broad range of accepted skill levels, armed forces are often not considered altogether (Nieto & Ramos, 2017). Using occupations from Table 1, which correspond to a one-digit ISCO-08 code, the skill levels are then matched with formal education levels by terms of ISCED-97 and ISCED-11 using the mapping from Table 2:

Table 2. Mapping of ISCO-08 skill levels, ISCED-97 and ISCED-11 codes and years of education

ISCO-08 skill level	ISCED-97 groups	ISCED-11 groups	Years of education
4	6 Second stage tertiary	8 Doctoral	20
	5a First stage tertiary (medium)	7 Master's	17
	5a First stage tertiary (medium)	6 Bachelor's	15
3	5b First stage tertiary (short & medium)	5 Short-cycle tertiary	14
2	4 Post-secondary	4 Post-secondary	13
	3 Upper secondary	3 Upper secondary	12
	2 Lower secondary	2 Lower secondary	9
1	1 Primary	1 Primary	9

Source: (International Labour Office, 2012; Statistics Estonia, 2023a; UNESCO Institute for Statistics, 1999, 2012); composed by the author

When multiple skill levels correspond to the same occupation, or if multiple educational levels are within the same skill level, then each individual within that category is assigned a corresponding individual value. This means that managers at skill levels 3 and 4 are matched, while only managers below are considered undereducated. Similarly, those in skill level 2 have required years of education based on their specific ISCED group.

International Labour Office (ILO) describes skill levels as the scope and complexity of tasks performed in any specific job and considers the nature of performed work, level of attained education and informal on-site training with work experience (International Labour Office, 2012). Description of skill levels lists the minimum level of education required and, therefore, can be used as a proxy for assessment of education-job mismatch.

The main approach for the JA method uses the methodology by Eurostat (2011) and identifies individuals as overeducated when individuals are involved in low to medium-level occupations (ISCO-08 occupation levels 4 to 9) with tertiary education (levels 5 and 6 for ISCED-97 classification). The same approach is used in the work by Ramos et al. (2012), but the focus is only on overeducated individuals. To continue the logic, those in high-level occupations (levels 1 to 3 in ISCO-08 classification) but without tertiary education (1 to 4 in ISCED-97 classification) are assigned as undereducated. Finally, those individuals not belonging to over or undereducated groups are assigned as adequately educated since their occupation and education levels match. To get the intensive form for Eurostat methodology and years of over and undereducation, the average of attained education (by years, from Table 2) is calculated by a two-digit ISCO-08 occupation code and then used with ORU specification.

An alternative approach for the JA method is based on a similar step from Wen & Maani (2022), where the years of required schooling to do a specific job are assigned using a work classification framework. Individuals are assigned to ISCO-08 skill groups based on their occupation, as shown in Table 1. They belong to the over- or undereducated group if their attained education is at least one standard deviation above or below the mean within their occupation according to ISCO-08 requirements. Those not assigned to any of the two groups are adequately educated within their occupation. A similar method has also been adopted by Iriondo & Pérez-Amaral (2016) and Nieto & Ramos (2017).

The years of attained schooling described above are calculated using mean and mode methods over individuals grouped by occupational two-digit ISCO-08 classified code for the RM method. Everyone above or under the mode value is assigned as over(under)educated, while the rest are defined as adequately educated or matching. For the mean measure, two approaches are taken. The individual is compared to the mean, and if he/she is one standard (Mean-one) deviation above or under, he/she is identified as over- or undereducated, respectively. The same is repeated but with one-half (Mean-half) of a standard deviation to get the second measure for the mean method. The idea is to get two measures that are more and less strict at capturing individuals not belonging to adequately educated groups, and a similar approach has been used in the study by Wen & Maani (2022).

The selection bias is another point to consider, given that not all individuals decide to participate in the labour market. Some of the studies (Wu & Wang, 2018) or (Wen & Maani, 2022) completely ignore this problem, while others (Galasi, 2008; Harmon et al., 2003; Nieto & Ramos, 2017) account for it and include the probability of being employed in the model. Their findings indicate the importance of including this variable as it is statistically significant.

The author of this thesis uses Heckman's (1979) two-step selection model by using individual, human capital and family-related variables such as gender, years of attained education, experience, whether Estonian or not, number of children and regional dummies adopted from Nieto & Ramos (2017). Under this specification, an hourly wage is a function of education and experience, while the likelihood of observing wage (the likelihood of working) is a function of gender, whether being an immigrant, number of children under 18 years of age, marital status and region. The wage is also implicitly part of the likelihood function to be working via the inclusion of education and experience, which determines the wage itself. The Inverse Mills Ratio from the Heckman model is then included in the main models as the probability of being employed.

The fixed effects regression with instrumental variables (FEIV) technique is used to estimate both Eq. 2 and Eq. 3 as in Iriondo & Pérez-Amaral (2016), Nieto & Ramos (2017) and Wen & Maani (2022). The FEIV is chosen based on the literature review but is also validated with Hausman tests. The dependent variable of hourly real net wage is calculated, where all taxes and other withholding are eliminated to represent more comparative amounts. The first model to be estimated is as follows in Equation (4):

$$\log(W_c) = \alpha + \beta_1 S_c^r + \beta_2 S_c^o + \beta_3 S_c^u + \beta_4 X_c + \lambda_c + u_c \qquad (4)$$

where the dependent variable is $\log(W_c)$ is the logarithm for a real hourly wage, $\beta_1 S_c^r$ is the required education measured in years, $\beta_2 S_c^o$ is years of overeducation and $\beta_3 S_c^u$ years of undereducation. $\beta_4 X_c$ includes variables having a child (dummy), the logarithm of experience, marital status (dummy), tenure (years), working part-time (dummy), nationality (whether being Estonian or not, dummy) as well as dummy variables of the region (north, south, west, east, central), Nomenclature of Economic Activities or NACE (total groups of five dummies) and year dummy variables for ELFS survey. The λ_c is the lambda coefficient or probability of being employed derived from the Heckman model, while u_c is an error term. Subscript c denotes cohort, and each numeric variable has been calculated as cohort average from the individual level.

For the second model, the specification is indicated in Equation (5):

$$\log(W_c) = \alpha + \beta_1 S_c^a + \beta_2 S_c^o + \beta_3 S_c^u + \beta_4 X_c + \lambda_c + u_c \qquad (5)$$

where similar to Eq. 3, the dependent variable is the logarithm of real hourly wage $\log(W_i)$, and the only different variable is $\beta_1 S_c^a$ as years of attained schooling. The other variables $\beta_2 S_c^o$, $\beta_3 S_c^u$, $\beta_4 X_c$, λ_c, u_c as well as subscript c are the same as in Eq. 4. In the original Eq. 3, the over- and undereducation are used with dummy variables. However, later studies like the ones by Hartog (2000) and Iriondo & Pérez-Amaral (2016) argue that using years instead of dummies is a better approach, also referencing the work of Cohn & Khan (1995), which argued the misinterpretation of dummy coefficients.

Both Eq. 4 and Eq. 5 are estimated three times with three different methods for educational mismatch and two instruments. The first estimates are done with the JA (Eurostat) method as a key explanatory variable and mode as the instrumental variable. Second estimates use Mean-one as a key explanatory variable and mode value as the instrumental variable. Finally, as part of the alternative assessment and robustness check, the specification of JA (ISCO-08) is instrumented using the mean-half variable. This gives the opportunity to compare multiple outputs based on different educational mismatch methods. The reason behind those estimates is that previous studies show different outcomes depending on the method used, as there is no universally applicable and accurate method for each case.

In line with studies by Iriondo & Pérez-Amaral (2016) and Wen & Maani (2022), the derived years of required education (in Eq. 4) are instrumented with alternative measures of required education, following the same pairs described above. This is done to account for potential bias when determining the required years of education. For Eq. 5, years of attained education are used and thus are not instrumented by any alternative measure since there is no other way to measure the highest degree obtained at the moment of a survey.

2.2. Data

Data used in the empirical analysis in this thesis are from the Estonian Labour Force Survey (ELFS) carried out by Statistics Estonia. It is a household and individual level survey carried out with a quarterly frequency starting from the year 1995. The survey is designed as rotational panel data. Each household or individual is interviewed in two consecutive quarters and then once again in the following year, during the same quarters, for a total of four times (Statistics Estonia, 2023). As a result, the households or individuals are not observed through the full timespan of a survey, resulting in unbalanced panel data. The statistical population in ELFS are individuals aged between 15 and 74 years who are permanent residents in Estonia. The information available covers employment status and sector, occupation, income, and socio-demographic aspects such as age, gender, education level, marital status, and more.

In order to estimate the mismatch effect, only the data for individuals is used. Duplicate values resulting from the rotational panel data design are removed. The sample is limited to employed individuals without missing wage information. Outliers in the wage category are assessed, and one individual with the highest wage among all respondents but with the lowest education and only 6 years of working experience compared to 21 years average among high earners (monthly net wage above 6,000 EUR) was excluded, while other high earners remain in the sample selection. The sample is also restricted to those who have attained the minimum required educational level, which according to the Basic Schools and Upper Secondary Schools Act (2010), is a basic education: 9 years of schooling. The control for the retirement age is also concluded, and those individuals working after the respective retirement age are removed. The years of survey used are 2009-2019, marking the period between the Great Recession and COVID-19 pandemic, where effects on wages from educational mismatch should not be affected by any major shocks or discrepancies. The ELFS methodology changed in 2009, making the study of previous years challenging.

The length of actual work experience is recorded under the ELFS, so the same approach is used as described in the study by Jalakas (2020), who also analysed the ELFS study. The length of actual work experience is calculated as the difference between the survey year and the starting date of the respondent's first job.

The wage data collected by ELFS includes net and gross wages. Since the studied period from ELFS data starts from 2009, while Kroons were still in place in Estonia, the wages until 2010 have been converted to Euros. The wages are then further normalised using the consumer price index, with the year 2015 being in the middle of the whole period taken as the baseline. Because ELFS collects responses on the number of hours worked during the typical working week, it is possible to further specify it to hourly wages. This is also in line with the majority of the studies looking at the effects of education mismatch and wages.

A detailed explanation of applying the methods to assess education mismatch in ELFS sample is described in the methodology section, while the table below shows the overall incidence of it on the cohort level. Individual level mismatch incidence can be found in Appendix 1. JA method based on Eurostat methodology shows the lowest incidence of education-job mismatch, while ISCO-08 specification with mode and mean-half indicates over 50 percent of mismatched individuals.

Table 1. Incidence of educational mismatch by different evaluation methods on a cohort level

Education mismatch type / Method	JA (Eurostat)	JA (ISCO-08)	Mode	Mean-one	Mean-half
Matched	71%	42%	42%	62%	40%
Undereducated	17%	28%	32%	21%	32%
Overeducated	12%	30%	26%	17%	28%
Total (Percentage)			100%		
Total (Amount)			24,556		

Source: ELFS 2009-2019, calculated by the author

Overall instruments are assigned based on examples from previous studies and also in accordance with the strength of pairwise correlation. The Strenght of pairwise correlation between measures of overeducation measured in years according to ORU specification is similar to the correlation between undereducation. Therefore, Table 2 below will focus only on them. Correlation for undereducation measures is available under Appendix 2. The JA (ISCO-08) and Mean-half formulate one group showing a higher incidence of educational mismatch compared to Mean-one and mode. Since the correlation is stronger between these pairs, they are used as dependent-instrumented variable combinations. For mismatch, according to the Eurostat method, mean years of attained education as a required level for the job is used and therefore is instrumented using the mode variable. In this way, the mean of attained education can be instrumented by mode.

Table 2. Correlation between dependent variable and years of overeducation by different methods

Variable	log wage	S^o (Eurostat)	S^o (ISCO-08)	S^o (Mode)	S^o (Mean-one)	S^o (Mean-half)
log wage	1	–	–	–	–	–
S^o (Eurostat)	-0.04	1	–	–	–	–
S^o (ISCO-08)	-0.03	0.71	1	–	–	–
S^o (Mode)	0.09	0.69	0.69	1	–	–
S^o (Mean-one)	0.15	0.75	0.74	0.79	1	–
S^o (Mean-half)	0.19	0.71	0.76	0.82	0.91	1

Source: ELFS 2009-2019, calculated by the author

A small correlation between measures of education mismatch and the dependent variable of wage still exists, while instrument validity relies on much stronger correlation between measures of education mismatch themselves. The same instruments and similar pairs are used in other studies, while instrument validity is also assessed under first-stage regression and reported in the results section.

For the first hypothesis, comparing mismatched individuals with the ones matched and from the same employment, Eq. 4 with Duncan & Hoffman model is estimated. The same variables as from Eq. 5, but the required years of schooling for the job are interpreted as difference from the individual with a matching level of education.

Studies analysing the overeducation effect on wages consistently underline the existence of wage penalties for those individuals being overeducated when compared to individuals with the same level of education but matched employment (McGuinness et al., 2018). This is in line with the second hypothesis, and to test this, Eq. 5 by Verdugo & Verdugo is estimated with the variables described in the methodology section. The opposite is tested for individuals being undereducated, who have a wage premium compared to ones with the same level of education but matching employment (McGuinness et al., 2018; Verhaest & Omey, 2012).

Other theories that are not of primary interest to this thesis but are usually discussed briefly in other studies are also assessed. For example, the human capital theory states that wages would always equal the workers' marginal product and reflect human capital accumulated through years of schooling or on-the-job training. Therefore, by including the variables related to human capital like experience, tenure and years of education, the effect from overeducation should not be negative unless there are some omitted variables. Similarly, it would follow for undereducated, implying a lack of formal education resulting in lower wages if all else held constant. This will be tested with Verdugo & Verdugo model under Eq. 5 to see if accumulated years of over- and undereducation contribute to higher (lower) wages the same way as attained education. The human capital theory entails that $S^a = -S^o = S^u$, or that attained education has the same effect on wage as years of undereducation or negative of years of overeducation (Iriondo & Pérez-Amaral, 2016).

Contrary to human capital theory, the Thurow model stipulates that wages will depend entirely on the level of education required for the job, which implies that coefficients for over- and undereducation are both equal to zero (Iriondo & Pérez-Amaral, 2016; McGuinness, 2006). This implies including job-related characteristics as variables and accurately identifying education mismatch. In this case, both models can be used to test this theory and, in line with other studies, should reject it (McGuinness, 2006, p. 413).

The career mobility and turnover theories by Sicherman & Galor (1990) and Jovanovic (1979) would require repeated observations of one individual to assess the temporary effect and whether it becomes stable over a longer period of time. Due to ELFS design, these theories could not be tested within this framework, as newly added observations could continuously contribute to the short-term effect of career mobility and turnover. Similarly, Sattinger (1993) assignment theory requires a wider approach with the inclusion of industry and economy-specific variables and is usually only overall discussed in the mismatch studies and literature (Davia et al., 2017).

To finalise an overview of the data, the summary statistics are presented in Table 3. The final sample size after all data cleaning includes 24,556 observations at the cohort level. The same overview but at an individual level is available under Appendix 3. Standard deviations are not shown for variables where they give little information to improve the table's readability.

Table 3. Summary statistics for variables included in the main analysis

Variable	Obs	Mean	Std. Dev.	Min	Max
log rhwage	24556	2,05	0,53	0,07	4,60
S^o (Eurostat)	24556	0,27	0,83	0,00	6,01
S^u (Eurostat)	24556	0,29	0,77	0,00	6,60
S^o (ISCO-08)	24556	0,75	1,31	0,00	8,00
S^u (ISCO-08)	24556	0,60	1,12	0,00	7,12
S^o (Mean-one)	24556	0,41	1,01	0,00	6,59
S^u (Mean-one)	24556	0,50	1,06	0,00	6,60
S^o (Mean-half)	24556	0,59	1,04	0,00	6,59
S^u (Mean-half)	24556	0,63	1,08	0,00	6,60
S^o (Mode)	24556	0,61	1,17	0,00	8,00
S^u (Mode)	24556	0,79	1,34	0,00	8,00
S^a	24556	12,98	2,13	9,00	20,00
S^r (Mean)	24556	13,01	1,27	10,99	15,84
S^r (ISCO-08)	24556	12,76	1,98	9,00	16,16
S^r (Mode)	24556	13,16	1,90	12,00	17,00
Estonian	24556	0,71	–	0,00	1,00
married	24556	0,68	–	0,00	1,00
has child	24556	0,39	–	0,00	1,00
part time	24556	0,13	–	0,00	1,00
log experience	24556	2,78	1,06	-2,48	4,00
tenure	24556	7,39	7,75	0,00	46,92
NACE	24556	2,57	–	1,00	5,00
region	24556	2,71	–	1,00	5,00
year	24556	2014,34	–	2009,00	2019,00
lambda	24556	0,30	0,07	0,11	0,60
cohort	24556	55,51	25,01	1,00	114,00

Source: ELFS 2009-2019, calculated by the author

3. RESULTS AND DISCUSSION

In this section, the primary outcomes are presented and discussed. First, the results of the Heckman two-step selection model are presented, followed by the formulation of cohorts and estimating the fixed effects model. The results for the Duncan & Hoffman (Eq.4) and Verdugo & Verdugo (Eq.5) models are analysed. Robustness checks to the main results are also included and evaluated. The chapter concludes with a discussion of the results with references to the empirical and theoretical framework.

3.1. Main results

First, the author employs the Heckman two-step selection model to evaluate the presence of selection bias in the employment for a full selection at an individual level. The result of the Heckman estimation shows a statistically significant Inverse Mills Ratio ($p - value \leq 0.001$), which indicates the selection bias problem. The Inverse Mills Ratio is estimated for each individual from the Heckman model and is used in the models going forward. Results for selection model are presented in Table 3.

Table 3. Heckman two-step selection model

Variables	Function of wage	Likelihood of observing wage	Lambda
Female	–	0.197***	–
Immigrant	–	-0.024	–
Number of children	–	0.035***	–
Marital status	–	-0.014	–
Region	–	-0.085***	–
S[a]	0.081***	-0.032***	–
Experience	0.013***	-0.007***	–
Experience^2	-0.001***	0.001***	–
Lambda	–	–	0.324***
Constant	0.880***	1.456***	–
Observations	40,744	40,744	40,744

* p ≤ 0.05, ** p ≤ 0.01, *** p ≤ 0.001.
Source: ELFS 2009-2019; author’s calculations

Heckman model results also show statistical significance ($p - value$ ≤0.001) of all included variables in both first and second step estimation except those of immigrant and marital status. They remain in the model due to their importance outlined in the literature despite not showing any statistical significance.

Following the empirical studies, the basic Mincer wage equation is estimated. Since cohorts are derived using gender and year of birth, the fixed effects model reports results without including a gender variable. The results of the Mincer wage equation are illustrated in Table 4.

Table 4. Mincer wage equation

Variables	Pooled OLS	Fixed Effects	Random Effects
S[a]	0.103***	0.115***	0.106***
Female	-0.366***	–	-0.349***
Experience	0.013***	0.059***	0.023***
$Experience^2$	-0.001***	-0.001***	-0.001***
Lambda	-1.070***	-0.785***	-0.991***
Constant	1.226***	-0.269***	0.907***
Observations	24,556	24,556	24,556
Number of Cohorts	–	114	114
R^2	0.212	0.304	0.140

* p ≤ 0.05, ** p ≤ 0.01, *** p ≤ 0.001
Source: ELFS 2009-2019; author's calculations

Variable for attained years of education shows an expected positive coefficient and varies slightly between the three models, while higher variance is seen for years of experience. The fixed effects model has the highest value for goodness of fit, while overall, all models indicate the statistical significance of coefficients for included variables. As the results are in line with previous studies, the Mincer equation results give a good indication to continue with the regression analysis.

Several diagnostic tests are concluded to assess the presence of potential issues like multicollinearity, heteroscedasticity and functional form of the model. While testing for multicollinearity and going through Variance Inflation Factor (VIF), variables for experience and experience squared had to be removed and replaced by their logarithmic form. After that, no signs of multicollinearity are present in the models, and VIF test results can be seen in Appendix 4. The Breush-Pagan test was used for the heteroscedasticity assessment and shows the presence of heteroscedasticity for all three model specifications. Following the test results and good practices of working with variables grouped at cohort levels, robust standard errors clustered at the cohort level are used. For functional form testing of the model, Ramsey Regression Equation Specification Error Test (RESET) was used and showed that all of the models are misspecified. Trying different specifications with interaction effects, quadratic and cubic forms still failed to reject the null hypothesis of the RESET test. Therefore, the analysis continued by adding all important variables outlined in the studies specified in the methodology section with consideration that all relevant variables are included in the model specification. It has to be noted that due to the inability to reject the null hypothesis of the RESET test, values may be biased.

Finally, plots for residuals versus fitted values and kernel density estimators were used to assess the adequacy of the regression model. The residuals are distributed evenly across the zero line and have constant variance. Plots for both Eq. 4 and 5 can be seen in Appendix 6. Neither significant outliers are seen under plots for kernel density estimator, also indicating normal distribution, and can be checked under Appendix 7.

To choose between the appropriate modelling approach and whether to use pooled Ordinary Least Squares (pooled OLS), fixed effects or random effects method, the Hausman test is used. The literature and previous studies almost exclusively use fixed effects estimation as the main method due to heterogeneity and endogeneity reasons (Leuven & Oosterbeek, 2011). This is also suggested in this study as the Hausman test shows that a model with fixed effects estimation is better than the model with random effects, indicating that individual-specific effects are correlated with the explanatory variables, and the fixed effects model can control for that. Similarly, the F-test that all of the u_i fixed effects are zero suggests that the fixed effects model is better than pooled OLS.

Therefore, the analysis continues with estimating fixed effects models, and instrumental variables are assessed based on the outputs of first-stage regressions. First, Eq. 4 is tested for all three measures of JA and RM mismatch. Reported statistics include coefficients that measure the instruments' effect on the endogenous variable as a within-group effect, accompanied by statistical significance. Then, there is a t-statistic as an indicator of the strength and significance of the relationship between the instrument and the endogenous variable. Finally, the F-statistic of the first stage within regression is shown to test for joint significance of the instruments for each endogenous variable and following the rule of thumb, it has to be higher than 10. This will indicate that instruments are strong predictors of endogenous variables. Results are reported in Table 5 below.

Table 5. First-stage within regression for instrumental variables in Equation 4

Model specification	Variable	Instrument	Coefficient	t	F(28, 113)	Prob > F
JA (Eurostat)	S^o	S^o (mode)	0.499***	33.87	108.29	0
JA (Eurostat)	S^u	S^u (mode)	0.296***	37.20	110.21	0
JA (Eurostat)	S^r	S^r (mode)	0.551***	120.92	4655.51	0
JA (ISCO-08)	S^o	S^o (mean-half)	0.941***	106.92	1667.71	0
JA (ISCO-08)	S^u	S^u (mean-half)	0.880***	129.08	1370.92	0
JA (ISCO-08)	S^r	S^r (mean-half)	1.536***	136.34	2234.77	0
RM (Mean-one)	S^o	S^o (mode)	0.707***	67.37	306.41	0
RM (Mean-one)	S^u	S^u (mode)	0.788***	94.35	1776.83	0
RM (Mean-one)	S^r	S^r (mode)	0.551***	120.92	4655.51	0

* $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$
Source: ELFS 2009-2019; author's calculations

For Eq. 4, the strongest instruments are for JA (ISCO-08) model specification, while the weakest are for JA (Eurostat), which can be explained by the fact that using Eurostat methodology to capture education-job mismatch has given the lowest incidence, while other methods have shown higher rates of over- and undereducated. Despite that, all of the instruments pass the requirements and can be assessed as valid.

Then, similar first-stage regressions are estimated to assess the instrument validity in Eq. 5, and results are reported in Table 6 below.

Table 6. First-stage within regression for instrumental variables in Equation 5

Model specification	Variable	Instrument	Coefficient	t	F(28, 113)	Prob > F
JA (Eurostat)	S^o	S^o (mode)	0.490***	37.27	108.29	0
JA (Eurostat)	S^u	S^u (mode)	0.311***	41.65	110.21	0
JA (ISCO-08)	S^o	S^o (mean-half)	1.261***	84.07	1609.53	0
JA (ISCO-08)	S^u	S^u (mean-half)	1.065***	130.08	1190.31	0
RM (Mean-one)	S^o	S^o (mode)	0.571***	62.38	306.04	0
RM (Mean-one)	S^u	S^u (mode)	0.578***	120.61	1776.83	0

* $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$
Source: ELFS 2009-2019; author's calculations

Similar to instruments from Eq. 4, the reported coefficients indicate valid instruments for Eq. 5. Again, due to the specifics of Eurostat methodology, the instruments indicate the weakest strength between the endogenous variables when compared to other mismatch detection methodologies.

Having the necessary tests complete and instrument validity checked, the study moves to the main part of the mismatch analysis and runs fixed effects regressions with instrumental variables (FEIV) for Eq. 4 and 5. First, the JA (Eurostat) and RM (Mean-one) methods are evaluated. The results of the main variables for the methods mentioned above are reported in Table 7 and are used to assess the initial outcomes of the models and whether they can be used for hypothesis testing.

Table 7. Main variable results of FEIV models for Equations 4 & 5

Variables	Eq. 4 JA (Eurostat) IV - Mode	Eq. 4 RM (Mean-One) IV – Mode	Variables	Eq. 5 JA (Eurostat) IV - Mode	Eq. 5 RM (Mean-One) IV - Mode
S^o	0.027***	0.023***	S^o	-0.107***	-0.131***
S^u	-0.121***	-0.044***	S^u	0.176***	0.106***
S^r	0.174***	0.149***	S^a	0.093***	0.134***
Obs	24,556	24,556	Obs	24,556	24,556
Cohorts	114	114	Cohorts	114	114
R^2	0.460	0.481	R^2	0.468	0.470

* $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$
R^2 refers to R-squared overall
Source: ELFS 2009-2019; author's calculations

Looking at the results for Eq. 4, the coefficients for mismatch are significant at the level of $p \leq 0.001$. The coefficients for overeducation and required education are quite similar, while the coefficient for undereducation varies a lot between JA and RM methods, highlighting the difference between the mismatch reporting methods. The signs for mismatch coefficients align with findings from previous studies, and ratios for the goodness of fit are at 0.46 and 0.48, meaning that almost half of the variation can be explained by selected variations.

Overall results of the main variables for Eq. 5 are similar to Eq. 4, where coefficients for mismatch are significant at the same level, and the signs also follow the logic in previous studies. The difference between mismatch measurement methods is more pronounced, and a 0.47 ratio of R-squared overall indicates sufficient goodness of fit. Results for both Eq. 4 and 5 provide a good indication of the models' adequacy and indicate that they can be used for hypothesis testing. With that, the full model output for Eq. 4 is provided in Table 8, followed by an analysis of all included variables.

Table 8. Full results of the FEIV model for Equation 4

Groups	Variables	JA (Eurostat) IV - Mode	RM (Mean-One) IV - Mode
Years	S^o	0.027***	0.023***
	S^u	-0.121***	-0.044***
	S^r	0.174***	0.149***
	Log(exp)	0.041***	0.039***
	Tenure	0.004***	0.004***
Dummy	Estonian	0.144***	0.138***
	Marital	0.034***	0.029***
	Has child	0.040***	0.042***
	Part time	0.029**	0.028**
NACE	G-J trade	-0.066***	-0.072***
	K-N finance	-0.117***	-0.111***
	O-S service	-0.179***	-0.189***
	A,F,T,U agr & constr	-0.009	-0.011
Region	Centre	-0.177***	-0.171***
	East	-0.197***	-0.203***
	West	-0.241***	-0.243***
	South	-0.279***	-0.285***
Survey year	2010	-0.008	-0.006
	2011	0.079***	0.080***
	2012	0.151***	0.154***
	2013	0.246***	0.248***
	2014	0.312***	0.313***
	2015	0.334***	0.331***
	2016	0.435***	0.441***
	2017	0.529***	0.532***
	2018	0.681***	0.678***
	2019	0.767***	0.768***
Mills Ratio	Lambda	0.789***	0.862***
Model	Constant	-0.861***	-0.556***
	Observations	24,556	24,556
	Number of cohorts	114	114
	R^2	0.460	0.481

* $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$
The reference groups are as follows: Estonian = not Estonian; marital = not married; has child = no child; part-time = full-time work; NACE = B-E Industry; region = North; survey year = 2009.
Source: ELFS 2009-2019; author’s calculations

For the first hypothesis, the model under Eq. 4 is used. In the Eurostat approach, the coefficient suggests about a 2.7 percent wage premium for each additional year of overeducation when compared to individuals in the same occupation but whose education-job is matching. In the Mean-one method, a wage premium is estimated at 2.3 percent for each year of overeducation, which is quite similar to the Eurostat approach. Higher differences are seen in returns to a year of undereducation, where the Eurostat method indicates a three times higher wage penalty at 12.1 percent compared to 4.4 percent in Mean-one. The results for education-job mismatch cannot reject the first hypothesis for both over- and undereducated individuals.

Returns to years of required education are higher than mismatch coefficients, which is also in line with the previous studies. In the Eurostat method, the returns for each additional year of required education indicate 17.4 percent wage premium and slightly lower 14.9 percent are reported in the Mean-one. The remaining variables in the years' category – the logarithm of experience and tenure, are either matching or very close and show an expected positive effect on wage, while both are statistically significant.

All dummy variables except working part-time or full-time are statistically significant at the level of $p \leq 0.001$, while the part-time work is significant at the 1 percent level ($p \leq 0.01$). The part-time coefficient is also the only one surprising at first glance, showing a positive effect on wages compared to those working full-time. However, as the dependent variable is wage in hourly form, there are potential reasonings behind that and it will be discussed among other results in the following section.

In the NACE group, the reference category is the mining, manufacturing, electricity and water supply (B-E) industry. All other industries suggest a negative effect on wages. A full list of NACE codes is available in Appendix 8. The A,F,T,U agriculture and construction industry remains statistically insignificant for both mismatch methods, while other industries are significant at the 0.1 percent level ($p \leq 0.001$). Region-wise, all coefficients are again statistically significant at 0.1 percent level compared to the baseline northern region, where also the capital of Estonia is located, and indicate a negative effect on wages. The years of the survey show the highest positive coefficients among all groups, while the year 2010 is not statistically significant for both methods. Finally, lambda is statistically significant and indicates the importance of accounting for selection bias using the Heckman model. With that, the analysis moves to assess the results of Eq. 5, reported in Table 9.

Table 9. Full results of the FEIV model for Equation 5

Groups	Variables	JA (Eurostat) IV - Mode	RM (Mean-One) IV - Mode
Years	S^o	-0.107***	-0.131***
	S^u	0.176***	0.106***
	S^a	0.093***	0.134***
	Log(exp)	0.036***	0.038***
	Tenure	0.004***	0.005***
Dummy	Estonian	0.134***	0.153***
	Marital	0.026***	0.036***
	Has child	0.045***	0.044***
	Part time	0.029**	0.030**
NACE	G-J trade	-0.061***	-0.051***
	K-N finance	-0.083***	-0.091***
	O-S service	-0.179***	-0.158***
	A,F,T,U agr & constr	-0.026***	-0.026***
Region	Centre	-0.169***	-0.186***
	East	-0.215***	-0.213***
	West	-0.251***	-0.257***
	South	-0.299***	-0.298***
Survey year	2010	-0.005	-0.008
	2011	0.085***	0.083***
	2012	0.158***	0.154***
	2013	0.250***	0.246***
	2014	0.307***	0.312***
	2015	0.318***	0.328***
	2016	0.436***	0.431***
	2017	0.523***	0.526***
	2018	0.663***	0.673***
	2019	0.760***	0.768***
Mills Ratio	Lambda	0.971***	0.883***
Model	Constant	0.117***	-0.400***
	Observations	24,556	24,556
	Number of cohorts	114	114
	R^2	0.468	0.470

* $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$
The reference groups are as follows: Estonian = not Estonian; marital = not married; has child = no child; part-time = full-time work; NACE = B-E Industry; region = North; survey year = 2009.
Source: ELFS 2009-2019; author's calculations

The model from Eq. 5 is used to test the second hypothesis. Coefficients for years of overeducation are both negative, implying negative returns for each year of overeducation when compared to an individual who holds the same level of education and is working in a job that matches their educational qualifications. Investment in each year of overeducation is associated with a 10.7 to 13.1 percent wage penalty.

In contrast, the coefficient for years of undereducation indicates positive returns when compared to the same individual as before. The wage premium for each year of undereducation is estimated at 17.6 to 10.6 percent in Eurostat and Mean-one methods, respectively. Investment in one year of attained education yields positive returns between 9.3 to 13.4 percent. These outcomes also cannot reject the second hypothesis.

Similar to the outcome of Eq. 4, the remaining groups for dummy, NACE, region, survey years and mills ratio variables are statistically significant at the same 0.1 percent level ($p \leq 0.001$), with a similar exception in the survey year 2010 being not statistically significant. Because outcomes follow the same logic as in Eq. 4, they are not discussed separately in this part. The only main exception in group variables is related to A,F,T,U agriculture and construction in the NACE category. This time the coefficient is with the same statistical significance as other variables at 0.1 percent level while also remaining negative. In conclusion, overeducated individuals face wage advantage when compared to matched individuals within the same occupation, followed by wage disadvantage compared to matched individuals within the same education level. Contrary to that, undereducated individuals experience wage disadvantages when compared to matched individuals within the same occupation but get wage premiums compared to matched individuals within the same education level. Returns to years of education, whether attained or required, remain higher compared to mismatch returns but with an exception in the Eurostat method for Eq. 5.

3.2. Robustness check

To assess the robustness of the results presented in the previous sub-chapter, an additional approach adopted from Wen & Maani (2022) is estimated, where JA is based on ISCO-08 guidelines and provides more granular requirements among four skills groups compared to two groups in levels of education and occupation within JA (Eurostat) method. The results of the robustness check are presented in Table 10.

Table 10. Full results of FEIV models for Equation 4 & 5 using JA (ISCO-08) method

Groups	Variables	Eq. 4 JA (ISCO-08) IV - Mean-half	Variables	Eq. 5 JA (ISCO-08) IV - Mean-half
Years	S^o	0.018***	S^o	-0.086***
	S^u	-0.031***	S^u	0.075***
	S^r	0.097***	S^a	0.106***
	Log(exp)	0.033***		0.035***
	Tenure	0.004***		0.003***
Dummy	Estonian	0.136***		0.133***
	Marital	0.025***		0.025***
	Has child	0.047***		0.043***
	Part time	0.051***		0.055***
NACE	G-J trade	-0.054***		-0.051***
	K-N finance	-0.062***		-0.059***
	O-S service	-0.151***		-0.150***
	A,F,T,U agr&constr	-0.018**		-0.021**
Region	Centre	-0.185***		-0.179***
	East	-0.226***		-0.214***
	West	-0.281***		-0.258***
	South	-0.333***		-0.300***
Survey year	2010	-0.003		-0.005
	2011	0.088***		0.085***
	2012	0.167***		0.162***
	2013	0.261***		0.254***
	2014	0.327***		0.313***
	2015	0.346***		0.332***
	2016	0.457***		0.442***
	2017	0.551***		0.537***
	2018	0.696***		0.680***
	2019	0.791***		0.776***
Mills Ratio	Lambda	1.213***		0.961***
Model	Constant	0.042		-0.015
	Observations	24,556		24,556
	Cohorts	114		114
	R^2	0.503		0.498

* $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$

The reference groups are as follows: Estonian = not Estonian; marital = not married; has child = no child; part time = full-time work; NACE = B-E Industry; region = North; survey year = 2009.
Source: ELFS 2009-2019; author's calculations

The outcome of the alternative assessment indicates similar positive and negative results on wages, while coefficients for both variables remain statistically significant at the 0.1 percent level ($p \leq 0.001$).

In Eq. 4, the mismatch coefficients are less pronounced compared to JA (Eurostat) and RM (Mean-one) methods, as values are closer to zero. The returns to years of overeducation suggest that about 1.8 percent wage premium when compared to individuals in the same occupation but whose education and job levels match. Compared to the same individuals, undereducation can be associated with a 3.1 percent wage penalty. The returns to years of required education are higher than both mismatch coefficients and show almost a 10 percent wage increase for each subsequent year.

Variables from groups show similar behaviour, while the NACE variable of A,F,T,U agriculture & construction becomes statistically significant at the 1 percent level ($p \leq 0.01$) when compared to Eq. 4, in the main results section. Model goodness of fit improves slightly, and the ratio is at 0.5, indicating that half of the variation in wages can be explained by selected variables.

For the results in Eq. 5, there is a similar conclusion, as main mismatch indicators and other group variables follow a similar pattern compared to Eq. 5 from the main results section. The coefficient for years of overeducation indicates negative returns. Therefore, an overeducated individual suffers from an 8.6 percent wage penalty compared to an individual who holds the same level of education but matching employment. An investment in each year of undereducation gives a 7.5 percent wage premium in a similar scenario. For years of attained education, the returns show a 10.6 percent wage premium.

In conclusion, for robustness analysis, the JA (ISCO-08) approach shows similar variable behaviour compared to the main analysis, while the magnitude of positive and negative effects on wages is slightly lower among mismatch indicators.

3.3. Discussion

The results of the education-job mismatch models (section 3.1.) are in line with the previous studies and literature. Two main models with JA (Eurostat) and RM (Mean-one) show statistically significant results for both Eq. 4 made using Duncan & Hoffman (1981) specification and Eq. 5 based on an initial model by Verdugo & Verdugo (1989) with changes adopted from more recent literature by Iriondo & Pérez-Amaral (2016), Verhaest & Omey (2012) and Wen & Maani (2022).

When looking at the general incidence of education-job mismatch, the percentages vary quite a bit, with 77 percent matched according to JA (Eurostat) approach and down to 49 percent utilising JA (ISCO-08) method. On the one hand, the previous studies mentioned in Chapter 1.4., including one by Halapuu & Valk (2013) on the Estonian labour market, reported a rather high incidence for either overeducation or both over- and undereducation, and the results from JA (ISCO-08) tend to mirror these percentages. On the other hand, the study, also working with the ELFS data by Lamo & Messina (2010), shows high wage penalties for overeducation, which is more in line with findings for JA (Eurostat) and RM (Mean-one) approach. From studies in chapter 1.4., the ones by CEDEFOP (2015) and European Commission (2015) are the closest by their relevance in terms of years studied, while this thesis is the only one to analyse the period, including the latest data until 2019. The difference in periods can be attributed to the less severe education-job mismatch reported in this thesis compared to previous studies, as the overall situation in Estonia could have improved.

The models with required years of education (Eq. 4) are consistent with different mismatch measure methods. Overeducated individuals receive higher wages compared to matched individuals within the same occupation, which could indicate a premium for more advanced theoretical and practical knowledge achieved from a higher level of education. Potentially lost working experience due to attainment of education does not apply since models control for individual differences in experience and tenure. Due to that, it could be suggested that an undereducated individual has less knowledge due to a lack of schooling for the currently attained position, and that is reflected in the earnings of the individual. These findings are to be expected based on the literature overview in Chapter 1.1., while differences in wage can be attributed to Sattinger (1993) assignment model, where lower education can be related to lower productivity and vice versa.

In models with years of attained education (Eq. 5), the opposite effect is observed, which is to be expected following the methodology based on previous studies and described in Chapter 2.1. When looking at the perspective of the same education levels, an overeducated individual earns less than matched individuals in the same occupation. Potential reasoning it is described in the literature by

(McGuinness, 2006), as well as his later works, also mentioned in this thesis. Those with an education higher than required by the job performed are more likely to underutilise their competencies, hence having lower productivity and getting lower wages. Contrary, individuals with an education level lower than required possess other specific characteristics that make it possible to work in more demanding occupations. Therefore, when compared to matched individuals with the same education, undereducated individuals get a wage premium from working in higher occupations and having better competencies.

The magnitude of the effect on individual earnings using different mismatch definition methods is generally more pronounced for models in main results utilising the JA (Eurostat) and RM (Mean-one) approach. While the coefficients for mismatch using ISCO-08 guidelines are more subdued. Potential reasoning might be in the overall incidence of education-job mismatch captured by these methods. Where the Eurostat method has a relatively low mismatch incidence of 23 percent (Table 1.), those classified as mismatched are more likely to be truly over- and undereducated because of the structure that this method imposes. Those with higher education can only be overeducated and have to work in the occupation without any tertiary requirements. The same is true for individuals with lower education, where only undereducation is possible. This leaves edge cases where implications are more pronounced.

A similar logic can be applied to the model outcomes of the RM (Mean-one) method. While the overall incidence is still relatively small at 32 percent, the individual has to be significantly different from the mean to be classified as mismatched. This is in line with results from Wen & Maani (2022) and their outcomes of the mean and one standard deviation approach. Somewhat similar can be seen in a study by Lamo & Messina (2010) in chapter 1.4., where relatively low levels of overeducation result in high wage penalties. However, the magnitude of the wage penalty there is over 20 percent which is not the case for this thesis.

When comparing model outcomes from the main results with JA (ISCO-08) method, the overall mismatch incidence is almost 20 percent higher than RM (Mean-one) and 28 percent higher than the JA (Eurostat). More borderline individuals are included with the more aggressive assignment of mismatched individuals. This can contribute to less pronounced wage effects from the mismatch, as the individual who is over or undereducated by only a small margin is less likely to face wage difference compared to an adequately matched individual.

Overall, both JA (Eurostat) and JA (ISCO-08) should be seen as mismatch measures based on the work classification framework, thus indicating potential mismatch from the perspective of general evaluation for multiple countries and regions. This provides comparative grounds when comparing to other studies using the same methods but does not necessarily imply a good fit nor accurate estimation for Estonia in particular. If the Eurostat and ILO assessments are more biased towards larger countries and regions, then estimates for small countries like Estonia could suffer as well. That could be the reason for the extreme results in Galasi (2008) study also using UNESCO classification framework, as these results are nowhere to be seen in this study, but there is also a significant difference in covered periods.

Moreover, statistical measures are used both as alternative measures and as instrumental variables since measures based on RM tend to indicate what is actually happening in the labour market without making any assumptions. Then again, if the Estonian labour market or some specific occupations and NACE groups suffer from certain effects outside the scope that can be captured by ELFS data, like long-term lack of professionals; poorly developed industries; private or government-owned monopolies fully controlling one or multiple sectors – then RM methods would be biased towards them and show what is happening within certain industry or occupation without correctly capturing the actual effect of mismatch.

The often-used subjective SA measure by studies such as work by Alba-Ramirez (1993) or (Robst, 2007) and also many more recent studies like Kemelbayeva (2020) and Verhaest et al. (2017) is also used to assess the incidence of mismatch and its implications on wages. However, this method is unavailable due to a lack of data in the ELFS. There are also no questions or any form of assessment related to the skill levels of individuals.

The latest studies analysing the effect of job mismatch also focus heavily on the concept of skill mismatch. The vast literature overview by McGuinness et al. (2018, p. 997) points to numerous studies also assessing the effect of over and underskilling, and the average overskilling wage penalty among 38 estimates is found to be 7.5 percent. Similarly, more recent researches also collecting survey data dedicate separate parts of their work towards assessing skill mismatch and its joint effect with education mismatch (Cedefop, 2015).

Since instrumental variables are derived using the same methods for defining mismatches, the heavy reliance is on the accuracy of the ELFS data and the lack of errors or bias when collecting results from respondents. The outcomes should be tested with alternative sources like PIAAC, CEDEFOP or European labour force survey to get a more representative picture of mismatch incidence and its effect on wages in Estonia. However, some of these do not cover the same periods as ELFS. Therefore the results of this thesis should be taken with caution and more as a general interpretation and not as a direct assessment of mismatch incidence and wage impact severity.

Results for additional explanatory variables from groups mentioned in Chapter 3.1. are in line with expected outcomes and consistent across different model specifications and mismatch assessment methods. Being Estonian contributes to a larger salary due to potentially better knowledge of the local language and other aspects like culture and industry. The marital status and having children both contribute to small hourly wage bonuses, as these are more likely to occur in the later stages of adulthood. One less obvious outcome for the dummy variable is that working part-time yields slightly higher returns than working full-time, yet a potential explanation could be sufficient or large income, where those earning more might be able to afford to work part-time. For the NACE group, the highest wages are in the B-E Industry category, as it includes more hard-working and intensive industries like mining & quarrying and manufacturing, while overall having fewer wages reported in low-income areas. The survey years capture the overall trend for wage growth from 2009 to 2019, while the year 2010 is quite similar to baseline 2009, thus remaining statistically insignificant in all models. Finally, the inclusion of the Inverse Mills Ratio from Heckman's (1979) selection model or Lambda coefficient in the model specification shows statistical significance at 0.1 percent level ($p \leq 0.001$) and highlights the importance of accounting for selection bias into employment.

Other models mentioned in the theoretical framework are the human capital model by Becker (1962) and Lester Thurows' Job competition model from a book by Piore (1979). In order for the human capital theory to hold, as described in Chapter 2.2., the coefficients of education mismatch should equal the coefficient of attained years of education. Results for Eq. 4 indicate that the theory can be rejected at 0.1 percent confidence level ($p \leq 0.001$). As for the competition model, the coefficient for overeducation should equal the coefficient for undereducation, which is not something indicated by the results of both Eq. 4 and 5. In line with other studies like Iriondo & Pérez-Amaral (2016), the model does not hold in this thesis.

CONCLUSION

The aim of the thesis was to estimate the effect of educational mismatch on individual earnings in the Estonian labour market. The study tries to understand the incidence of educational mismatch in Estonia and analyse the potential effect on individual earnings. In order to attain the objective, the thesis searched for the answers to the following questions: Do overeducated (undereducated) workers experience wage bonus (penalty) when compared to ones with the same employment but matching levels of education in Estonia? Do overeducated (undereducated) workers experience wage penalty (bonus) when compared to ones with the same level of education but matching employment in Estonia?

To answer the first research question, the author employed Duncan & Hoffman (year) model with separate measures of job assessment and realised matches methods to estimate the mismatch incidence and the effect of individual income. The empirical analysis was carried out using a fixed effect model with instrumental variables and individuals grouped at cohort levels to address the endogeneity and unobserved heterogeneity, which are often present in social studies. The results indicate an overeducation wage bonus when compared to individuals with the same employment but matching levels of education and the opposite for undereducation in Estonia.

In order to answer the second research question and the mismatch effect in the same level of education, the model from Verdugo & Verdugo (year) was used. In the empirical part, the model specification was adjusted to reflect the intensive form of education-job mismatch and show results per additional year attained, similar to Duncan & Hoffman (year) model. The same measures were used to estimate the mismatch incidence and effect of wages with instrumented fixed effect model. The outcomes indicate a wage penalty for overeducation when compared to individuals with the same education but matching employment, and the opposite can be suggested for undereducated individuals in Estonia.

The general results reveal a wage premium for returns to years of overeducation under Duncan & Hoffman model specification, ranging from 1.8 to 2.7 percent. Under the same model, the returns to years of undereducation show a 3.1 to 12.1 percent wage penalty. In the modified Verdugo & Verdugo model, the results suggest a wage premium from 7.5 to 17.6 percent for each additional year of undereducation. In comparison, an 8.6 to 13.1 percent wage penalty is associated with returns to years of overeducation. In both models, the returns to either attained or required education show the highest increase in wage from 9.7 to 17.4 percent with one exception for Verdugo & Verdugo model and job evaluation approach under the Eurostat method.

The advantage of ELFS is a rotationary panel data design that allows for more broad sample coverage and potentially decreases sample selection bias of fixed panel data. There are also many control variables present in the survey that is used in other studies on education-job mismatch and the effects on wages. However, certain variables are still missing, such as indirect subjective assessment of individual education and job match, as well as work experience, while included individuals are observed only over a short period of time. Finally, there is no indication of the skills levels included in ELFS. The latest studies show the importance of distinguishing between skill and educational levels when assessing mismatch and its implication. Therefore, it is important to take into account that the values may be subject to bias when interpreting them.

From the individual perspective, the results could be seen as an indication of the importance of matching the level of education and occupation. However, as returns to both required and attained education outweigh penalties from the mismatch, it should be safe to assume within a reasonable period that investing in education, regardless of mismatch, will still pay off if an individual gets matching employment. As some of the study results indicate over 50 percent mismatch incidence, the public authorities should also monitor the situation with mismatch and implement better education system or career guidance if the percentage gets too high, as short-term effects of mismatch still have a negative impact on the individual outcome. Finally, high mismatch incidence from realised matches method indicates the potential for better job requirements in terms of education level. Companies should review the job requirements for each position and ensure that they are appropriate for the education needed to perform the job effectively.

To the knowledge of the author, there have not been any recent studies in Estonia focusing on the impact of education-job mismatch on individual wages, and the contribution of this thesis is to update the findings of the previous studies for a period until 2019. Further studies could validate the results by using different sources like PIIAC or the European labour force survey while also focusing on evaluating the skill mismatch concept and the joint significance with education mismatch.

KOKKUVÕTE

HARIDUSTASEMETE JA TÖÖKOHTADE MITTEVASTAVUS - MÕJU INDIVIDUAALSETE SISSETULEKUTELE EESTIS

Artemi Beljakov

Käesoleva magistritöö eesmärk on hinnata haridustasemete ja töökohtade mittevastavuse mõju individuaalsetele sissetulekutele Eesti tööturul ning analüüsida haridustasemete ja töökohtade mittevastavuse tuvastamise meetodeid ja võimalikke põhjuseid, mis võivad mõjutada mittevastavuse mõju indiviidi sissetulekutasemele. Eesmärgi täitmiseks otsitakse vastuseid järgmistele püstitatud küsimustele:

1. Kas üleharitud (alaharitud) töötajad saavad Eestis kõrgema (madalama) palka võrreldes sama töökohataseme inimestega kuid samaväärse haridustasemega?
2. Kas üleharitud (alaharitud) töötajad saavad Eestis kõrgema (madalama) palka võrreldes sama haridustaseme inimestega kuid samaväärse töökohatasemega?

Lähtudes varasemast kirjandusest, testiti magistritöös järgmisi hüpoteese:

H_1: Üleharitud (alaharitud) saavad kõrgema (madalama) palka võrreldes nendega, kes on sama töökohatasemega kuid samaväärse haridustasemega;

H_2: Üleharitud (alaharitud) saavad madalama (kõrgema) palka võrreldes nendega, kes on sama haridustasemega kuid samaväärse töökohatasemega;

Uurimisküsimuse vastuste leidmiseks on töö struktureeritud järgmiselt. Esimene peatükk annab ülevaade peamistest palga teooriatest ja kuidas nende alusel selgitada hariduse ja töökoha mittevastavuse tekkimist. Kõige levinum on inimkapitali ja töökohtadele konkureerimise teooriad. Inimkapitali teooria järeldab, et töötaja tootlikkus ja seejärel palgatase sõltuvad inimkapitali mahust (Becker, 1962). Töökogemuse puudumise korral võib liigne haridus olla nähtav nagu asendaja. Selle teooria raames on üleharidus tingitud madala kvaliteedi inimkapitaliga või selle ebapiisavusega (Caroleo & Pastore, 2018, p. 1002). Tööhoha konkureerimise teooria, mis avaldati (Piore, 1979) raamatus, kirjeldab tööturgu koolituskulude vaatenurgast. Teooria eeldab, et vajalikud oskused töö tegemiseks on omandatud vaid kohapeal ja palgatase sõtlub täielikult töökoha taseme pealt. Seetõttu käsitletakse igat üksikisikut kui koolituskulusid, ja mida madalam kulu, seda rohkem tõenäosus töökoha saavutamiseks. Täiendava haridustasemega katsuvad üksikisikut enda koolituse kulutusi tööandja vaatenurgast vähendada. Eelneva kahe teooria vahel on Sattingeri (1993, p. 831–834) määramise teooria, kus palk sõltub individuaalsetest, ettevõtte- ja majanduslikest teguritest. Määramis teooria puhul võib üle- ja alaharidus tekkida, kui haridusest sõltumatu tegurid annavad tasakaalus kõrgema palka. Lisaks on välja toodud karjääri mobiilsuse teooria, oskuste heterogeensus, õppeprogrammi tüüp ja muid individuaalsed tegurid koos nende mõjuga töö ja hariduse mittevastavusele.

Seejärel autor tutvustab mõõtmismeetodite ülevaade, mis kasutatakse hariduse ja töökoha mittevastavuse tuvastamiseks. Peamised hariduse ja töökoha nõuete vastavuse hindamise meetodid on subjektiivsed (otsene ja kaudne), objektiivsed (eksperthinnang või töönõuete analüüs) ja empiirilised (keskmine ja mood omandatud hariduse kohta). Tavaliselt nendest parem on eksperthinnang, kuid selle läbiviimine on kallis ja täpsus ikka sõltub eksperdi kvalifikatsioonist (McGuinness et al., 2018). Ülejäänud meetodite kasutamine sõltub andmete struktuurist ja kättesaadavusest ning universaalselt sobivat meetodit ei ole olemas. Lühidalt on selgitatud vertikaalsed ja horisontaalsed hariduse ja töö mittevastavused, ning alampeatükk lõpeb kirjeldusega, kuidas Minceri võrrandi kasutatakse mittevastavuse uuringutes.

Viimased kaks alapeatüki annavad ülevaate varasemate uurimustele üle- ja alaharitud juhtumite kohta ning nende püsivuse ja mõju peale, seejärel keskendes uuringutele Eesti tööjõu turule. Üldiselt on hariduse-töö sobitamatus üsna tavaline ja varieerub riigiti ja kasutatava mõõtmismeetodi järgi, samal ajal kui teatatud palga karistused ulatuvad 3 protsendist kuni 25 protsendini. Euroopa uuringud, sealhulgas sisaldavaid Eesti andmed, näitavad, et Eesti kuulub mittevastavuse juhtumite arvu poolest kõrgemate riikide hulka.

Teine peatükk keskendub uurimistöö metoodika ja andmete tutvustusele. Metoodika osas põhjendab autor valitud lähenemist – Duncan & Hoffman (1981) ja Verdugo & Verdugo (1989) meetodit tuginedes Minceri (1974) palgavõrrandile. Mõlemas mudelis kasutatakse üle- ja alahariduse aastate arvu, et näidata tuulemusi intensiivvormi kujul. Hinnatud Heckmani mudel kinnitab selektsiooni probleemi ning edaspidi analüüsis kasutakse Millsi pöördmäär. Mittevastavuse tuvastamiseks kasutab autor objektiivse ja empiirilise meetodi. Objektiivne meetod sisaldab töönõuete analüüsi Eurostat ja Ametite Rahvusvahelise klassifitseerimissüsteemi

(International Standard Classification of Occupations, ISCO) metoodika alusel. Empiirilise meetodi puhul on valitud keskmine koos ühe ja pool-ühe standardhälbe ulatusega ja mood hinnangud. Fikseeritud efektide mudelit koos instrumentidega kasutatakse endogeensuse ja heterogeensusega tegelemiseks. Selle mudeli kasutamiseks grupeeritakse individuaalsed vaatlused kohordi tasemel sugu ja sünniaasta järgi ning peale seda arvutakse kohordi-kesksed väärtused numbriliste näitajate jaoks.

Andmeplokk esitab Eesti tööjõu-uuringu (ETU) andmed aastatest 2009. – 2019. koos valimi loomise protsessiga ja seotud andmepiirangutega. Valitud periood peaks näitama tavalist olukorra tööturul ilma suurte katkestusteta nagu 2008 aasta maailma kriis või COVID-19 pandeemia. Lisaks valitud perioodi tõttu oli autor enne 2011. aastat olevad palgandmed teisendanud eurodeks ja seejärel normaliseerinud tarbijahinnaindeksi järgi, võttes 2015. baasaastaks. Valim on piiratud tööjõuga enne pensioniaja saabumist. Kasutades hariduse ja töö mittevastavuse tuvastamise meetodid on esitatud mittevastavuse sagedus – kõige madalam mittevastavuse osakaal on Eurostati metoodika puhul (29 protsenti) ja kõige kõrgem keskmise koos pool-ühe standardhälbega (60 protsenti). Seejärel esitakse mittevastavuse korrelatsiooni tulemused ja instrumentide valikute põhjendus. Teine peatükk lõpeb valitud muutujate kirjeldava statistikaga.

Kolmandas peatüki alustab autor peamiste tulemuste esitamisega. Duncan & Hoffmani mudeli tulemus näitab palgaboonus iga ülehariduse veedetud aasta kohta vahemikus 1.8 kuni 2.7 protsenti võrreldes töötajaga kes on sama töökohatasemega kuid samaväärse haridustasemega. Selle sama mudeli all on alahariduse iga veedetud aasta kohta palk 3.1 kuni 12.1 protsendi võrra väiksem. Muudetud Verdugo & Verdugo mudelis näitavad tulemused palgalisa 7.5 – 17.6 protsenti iga täiendava alaharise aasta eest võrreldes nendega, kes on sama haridustasemega kuid samaväärse töökohatasemega. Samuti on iga ülehariduse aasta kohta palk 9.7 kuni 17.4 protsenti vähem. Robustsuse analüüsi tulemused on kooskõlas põhimudeli tulemustega, näidates suuremat mittevastavuse sagedust, kuid madalamat palgakaristust. Mõlemad hüpoteesid on kinnitatud nii üleharitud ja alaharitud töötataje jaoks Eestis.

Selle magistritöö panus on värskendatud tulemused Eesti jaoks. Kuid töötaja oskuste tase ja subjektiivse hinnastamismeetodi puuduliku andmete tõttu peab tulemustesse suhtuda ettevaatusega, kuna jääb tõenäosus et hinnangud on kallutatud. Järgnevad uuringud võiksid valideerida tulemusi ja lahendada puuduva muutuja ettepanek, kasutades erinevaid andmeallikaid nagu PIAAC või Euroopa tööjõuuuring.

LIST OF REFERENCES

Alba-Ramirez, A. (1993). Mismatch in the Spanish labor market: overeducation?. *Journal of Human Resources*, 259-278. https://doi.org/10.2307/146203

Allen, J., & Van der Velden, R. (2001). Educational mismatches versus skill mismatches: effects on wages, job satisfaction, and on-the-job search. *Oxford economic papers*, 53(3), 434-452. https://doi.org/10.1093/oep/53.3.434

Basic Schools and Upper Secondary Schools Act. RT I 2010, 41, 240.

Becker, G. S. (1962). Investment in human capital: A theoretical analysis. *Journal of political economy*, 70(5, Part 2), 9-49. https://doi.org/10.1086/258724

Bender, K. A., & Heywood, J. S. (2011). Educational mismatch and the careers of scientists. *Education Economics*, 19(3), 253-274. https://doi.org/10.1080/09645292.2011.577555

Boudarbat, B., & Chernoff, V. (2012). Education–job match among recent Canadian university graduates. *Applied Economics Letters*, 19(18), 1923-1926. https://doi.org/10.1080/13504851.2012.676730

Caroleo, F. E., & Pastore, F. (2018). Overeducation at a glance. Determinants and wage effects of the educational mismatch based on AlmaLaurea data. *Social Indicators Research*, 137, 999-1032. https://doi.org/10.1007/s11205-017-1641-1

Cohn, E., & Khan, S. P. (1995). The wage effects of overschooling revisited. *Labour Economics*, 2(1), 67-76. https://doi.org/10.1016/0927-5371(95)80008-L

Davia, M. A., McGuinness, S., & O'Connell, P. J. (2017). Determinants of regional differences in rates of overeducation in Europe. *Social science research*, 63, 67-80. https://doi.org/10.1016/j.ssresearch.2016.09.009

Duncan, G. J., & Hoffman, S. D. (1981). The incidence and wage effects of overeducation. *Economics of education review*, 1(1), 75-86. https://doi.org/10.1016/0272-7757(81)90028-5

Cedefop. (2015). Skills, qualifications and jobs in the EU: The making of a perfect match? Evidence from Cedefop's European skills and jobs survey. *CEDEFOP reference series 103, Publications Office of the European Union.*

European Commission, Directorate-General for Employment, Social Affairs and Inclusion, Eurostat, (2016). *Analytical web note: measuring skills mismatch. 7/2015, Publications Office*. https://doi.org/10.2767/620480

European Commission. (2010). List of NACE codes. Retrieved from: https://ec.europa.eu/competition/mergers/cases/index/nace_all.html

Eurostat. (2011). Migrants in Europe. A statistical portrait of the first and second generation. *Publications Office of the European Union.* Retrieved from: https://ec.europa.eu/eurostat/documents/3217494/5727749/KS-31-10-539-EN.PDF.pdf/bcf27a60-7016-4fec-98c5-e8488491ebbd?t=1414775922000

Galasi, P. (2008). *The effect of educational mismatch on wages for 25 countries* (No. BWP-2008/8). Budapest Working Papers on the Labour Market.

Groot, W., & Van Den Brink, H. M. (2000). Overeducation in the labor market: a meta-analysis. *Economics of education review*, 19(2), 149-158. https://doi.org/10.1016/S0272-7757(99)00057-6

Halapuu, V., & Valk, A. (2013). Adult skills in Estonia and the world: Primary results of the PIAAC study. *Ministry of Education and Research*. Retrieved from: https://www.hm.ee/media/1813/download

Harmon, C., Oosterbeek, H., & Walker, I. (2003). The returns to education: Microeconomics. *Journal of economic surveys*, 17(2), 115-156. https://doi.org/10.1111/1467-6419.00191

Harmon, C., Oosterbeek, H., & Walker, I. (2000). *The returns to education: a review of evidence, issues and deficiencies in the literature* (No. 5). Centre for the Economics of Education, London School of Economics and Political Science.

Hartog, J. (2000). Over-education and earnings: where are we, where should we go?. *Economics of education review*, 19(2), 131-147. https://doi.org/10.1016/S0272-7757(99)00050-3

Heckman, J. J. (1979). Sample selection bias as a specification error. Econometrica: *Journal of the econometric society*, 153-161. https://doi.org/10.2307/1912352

International Labour Office. (2012). *International Standard Classification of Occupations: ISCO-08* (1st ed.). Geneva. Retrieved from: https://www.ilo.org/wcmsp5/groups/public/@dgreports/@dcomm/@publ/documents/publication/wcms_172572.pdf

Iriondo, I., & Pérez-Amaral, T. (2016). The effect of educational mismatch on wages in Europe. *Journal of Policy Modeling*, 38(2), 304-323. https://doi.org/10.1016/j.jpolmod.2015.12.008

Jalakas, K. (2020). *The impact of parenthood on the divergence of women's and men's wages in Estonia.*

Jovanovic, B. (1979). Job Matching and the Theory of Turnover. *Journal of Political Economy*, 87(5, Part 1), 972–990. https://doi.org/10.1086/260808

Kemelbayeva, S. (2020). Returns to schooling in Kazakhstan: an update using a pseudo-panel approach. *Eurasian Economic Review*, 10, 437-487. https://doi.org/10.1007/s40822-020-00148-z

Kiker, B. F., Santos, M. C., & De Oliveira, M. M. (1997). Overeducation and undereducation: evidence for Portugal. *Economics of Education Review*, *16*(2), 111-125. https://doi.org/10.1016/S0272-7757(96)00040-4

Lamo, A., & Messina, J. (2010). Formal education, mismatch and wages after transition: Assessing the impact of unobserved heterogeneity using matching estimators. *Economics of Education Review*, *29*(6), 1086–1099. https://doi.org/10.1016/j.econedurev.2010.06.002

Leuven, E., & Oosterbeek, H. (2011). Overeducation and mismatch in the labor market. *Handbook of the Economics of Education*, 4, 283-326. https://doi.org/10.1016/B978-0-444-53444-6.00003-1

McGuinness, S. (2003). Graduate overeducation as a sheepskin effect: evidence from Northern Ireland. *Applied Economics*, *35*(5), 597–608. https://doi.org/10.1080/00036840220000029284

McGuinness, S. (2006). Overeducation in the labour market. *Journal of economic surveys*, 20(3), 387-418. https://doi.org/10.1111/j.0950-0804.2006.00284.x

McGuinness, S., Pouliakas, K., & Redmond, P. (2018). Skills mismatch: concepts, measurement and policy approaches. *Journal of Economic Surveys*, *32*(4), 985–1015. https://doi.org/10.1111/joes.12254

Mincer, J. (1958). Investment in human capital and personal income distribution. *Journal of Political Economy*, *66*(4), 281–302. https://doi.org/10.1086/258055

Mincer, J. (1974). Schooling, Experience, and Earnings. Human Behavior & Social Institutions No. 2.

Nieto, S., & Ramos, R. (2017). Overeducation, skills and wage penalty: Evidence for Spain using PIAAC data. *Social Indicators Research*, 134, 219-236. https://doi.org/10.1007/s11205-016-1423-1

Pecoraro, M. (2016). The incidence and wage effects of overeducation using the vertical and horizontal mismatch in skills: Evidence from Switzerland. *International Journal of Manpower*, *37*(3), 536–555. https://doi.org/10.1108/IJM-10-2014-0207

Piore, M. (Ed.). (1979). *Unemployment and Inflation: Institutionalist and Structuralist Views* (1st ed.). Routledge. https://doi.org/10.4324/9781315084374

Quintini, G. (2011). *Right for the Job: Over-Qualified or Under-Skilled?* https://doi.org/10.1787/5kg59fcz3tkd-en

Ramos, R., Surinach, J., & Artís, M. (2012). Regional economic growth and human capital: the role of over-education. *Regional Studies*, 46(10), 1389-1400. https://doi.org/10.1080/00343404.2012.675140

Robst, J. (2007). Education and job match: The relatedness of college major and work. *Economics of Education Review*, 26(4), 397-407. https://doi.org/10.1016/j.econedurev.2006.08.003

Romero, L. M., Huertas, I. P. M., & Jiménez, M. D. M. S. (2017). Wage effects of cognitive skills and educational mismatch in Europe. *Journal of Policy Modeling*, 39(5), 909-927. https://doi.org/10.1016/j.jpolmod.2017.08.001

Sattinger, M. (1993). Assignment models of the distribution of earnings. Journal of economic literature, 31(2), 831-880.

Sicherman, N. (1991). “Overeducation” in the labor market. *Journal of Labor Economics, 9*(2), 101–122. https://doi.org/10.1086/298261

Sicherman, N., & Galor, O. (1990). A theory of career mobility. *Journal of Political Economy, 98*(1), 169–192. https://doi.org/10.1086/261674

Somers, M. A., Cabus, S. J., Groot, W., & van den Brink, H. M. (2019). Horizontal mismatch between employment and field of education: evidence from a systematic literature review. *Journal of Economic Surveys*, *33*(2), 567–603. https://doi.org/10.1111/joes.12271

Statistics Estonia. (2021). *Handbook of educational statistics*. Retrieved from: https://www.hm.ee/sites/default/files/documents/2022-10/haridusstatistika_kasiraamat_2021.pdf

Statistics Estonia. (2023). Estonian Labour Force Survey. Methodology https://www.stat.ee/en/find-statistics/methodology-and-quality/esms-metadata/40013

UNESCO Institute for Statistics. (1999). *International standard classification of education: ISCED 1997.* Retrieved from*: http://uis.unesco.org/sites/default/files/documents/international-standard-classification-of-education-1997-en_0.pdf*

UNESCO Institute for Statistics. (2012). *International standard classification of education: ISCED 2011*. UNESCO Institute for Statistics. Retrieved from: https://uis.unesco.org/sites/default/files/documents/international-standard-classification-of-education-isced-2011-en.pdf

Verdugo, R. R., & Verdugo, N. T. (1989). The impact of surplus schooling on earnings: Some additional findings. Journal of human resources, 629-643. https://doi.org/10.2307/145998

Verhaest, D., & Omey, E. (2010). The determinants of overeducation: different measures, different outcomes? *International Journal of Manpower*, *31*(6), 608–625. https://doi.org/10.1108/01437721011073337

Verhaest, D., & Omey, E. (2012). Overeducation, undereducation and earnings: further evidence on the importance of ability and measurement error bias. *Journal of Labor Research*, *33*(1), 76–90. https://doi.org/10.1007/s12122-011-9125-6

Verhaest, D., Sellami, S., & van der Velden, R. (2017). Differences in horizontal and vertical mismatches across countries and fields of study. *International Labour Review*, *156*(1), 1–23. https://doi.org/10.1111/j.1564-913X.2015.00031.x

Wen, L., & Maani, S. A. (2022). Earnings penalty of educational mismatch: a comparison of alternative methods of assessing over-education. *New Zealand Economic Papers*, *56*(2), 169–194. https://doi.org/10.1080/00779954.2022.2034175

Wolbers, M. H. (2003). Job mismatches and their labour-market effects among school-leavers in Europe. *European sociological review*, 19(3), 249-266. https://doi.org/10.1093/esr/19.3.249

Wu, N., & Wang, Q. (2018). Wage penalty of overeducation: New micro-evidence from China. *China Economic Review*, *50*, 206–217. https://doi.org/10.1016/j.chieco.2018.04.006

APPENDICES

Appendix 1. Incidence of educational mismatch by different evaluation methods on individual level

Education mismatch type / Method	JA (Eurostat)	JA (ISCO-08)	Mode	Mean (S.D.)	Mean (half S.D.)
Matched	77%	49%	49%	68%	46%
Undereducated	13%	23%	26%	17%	26%
Overeducated	10%	28%	25%	15%	28%
Total (Percentage)	100%				
Total (Amount)	33,710				

Source: ELFS 2009-2019, calculated by the author

Appendix 2. Correlation matrix for measures of undereducation

Variable	log rhwage	S^o (Eurostat)	S^o (ISCO-08)	S^o (Mean-one)	S^o (Mean-half)	S^o (Mode)	S^u (Eurostat)	S^u (ISCO-08)	S^u (Mean-one)	S^u (Mean-half)	S^u (Mode)	S^r (Mean)	S^r (ISCO-08)	S^r (Mode)
log rhwage	1,00	–	–	–	–	–	–	–	–	–	–	–	–	–
S^o (Eurostat)	-0.04	1,00	–	–	–	–	–	–	–	–	–	–	–	–
S^o (ISCO-08)	-0.03	0.71	1,00	–	–	–	–	–	–	–	–	–	–	–
S^o (Mean-one)	0.09	0.69	0.69	1,00	–	–	–	–	–	–	–	–	–	–
S^o (Mean-half)	0.15	0.75	0.74	0.79	1,00	–	–	–	–	–	–	–	–	–
S^o (Mode)	0.19	0.71	0.76	0.82	0.91	1,00	–	–	–	–	–	–	–	–
S^u (Eurostat)	0.09	-0.11	-0.19	-0.16	-0.13	-0.18	1,00	–	–	–	–	–	–	–
S^u (ISCO-08)	0.02	-0.16	-0.28	-0.24	-0.19	-0.27	0.77	1,00	–	–	–	–	–	–
S^u (Mean-one)	0.01	-0.17	-0.3	-0.27	-0.21	-0.29	0.54	0.77	1,00	–	–	–	–	–
S^u (Mean-half)	-0.07	-0.14	-0.24	-0.22	-0.17	-0.24	0.5	0.8	0.85	1,00	–	–	–	–
S^u (Mode)	-0.06	-0.17	-0.3	-0.27	-0.21	-0.29	0.6	0.89	0.82	0.93	1,00	–	–	–
S^r (Mean)	0.4	-0.19	-0.34	-0.02	0.02	0.1	0.37	0.26	0.25	-0.01	0.06	1,00	–	–
S^r (ISCO-08)	0.35	-0.17	-0.17	-0.02	0.03	0.12	0.36	0.2	0.28	0,00	0.06	0.92	1,00	–
S^r (Mode)	0.32	-0.17	-0.16	-0.21	0,00	0.08	0.27	0.17	0.42	0.07	0.06	0.8	0.87	1,00

Source: ELFS 2009-2019, calculated by the author

Appendix 3. Summary statistics for selected variables at individual level

Variable	Obs	Mean	Std. Dev.	Min	Max
log rhwage	33,710	2.08	0.55	0.06	4.60
age	33,710	41.96	12.03	18.00	63.00
birth year	33,710	1971.95	12.38	1945.00	2001.00
gender	33,710	0.55	0.49	0.00	1.00
S^o (Eurostat)	33,710	0.26	0.87	0.00	6.01
S^u (Eurostat)	33,710	0.29	0.83	0.00	6.60
S^o (ISCO-08)	33,710	0.74	1.38	0.00	8.00
S^u (ISCO-08)	33,710	0.61	1.20	0.00	7.12
S^o (Mean-one)	33,710	0.42	1.09	0.00	6.75
S^u (Mean-one)	33,710	0.48	1.11	0.00	6.60
S^o (Mean-half)	33,710	0.61	1.12	0.00	6.75
S^u (Mean-half)	33,710	0.62	1.14	0.00	6.60
S^o (Mode)	33,710	0.61	1.27	0.00	8.00
S^u (Mode)	33,710	0.79	1.44	0.00	8.00
S^a	33,710	13.08	2.28	9.00	20.00
S^r (Mean)	33,710	13.08	1.35	10.99	15.84
S^r (ISCO-08)	33,710	12.88	2.10	9.00	16.16
S^r (Mode)	33,710	13.26	2.08	12.00	17.00
Estonian	33,710	0.74	0.44	0.00	1.00
experience	33,710	22.32	12.64	0.00	54.58
married	33,710	0.71	0.45	0.00	1.00
has child	33,710	0.41	0.49	0.00	1.00
tenure	33,710	7.59	8.28	0.00	46.92
part-time	33,710	0.10	0.30	0.00	1.00
NACE	33,710	2.55	1.23	1.00	5.00
region	33,710	2.67	1.67	1.00	5.00
year	33,710	2014.44	3.26	2009.00	2019.00
lambda	33,710	0.30	0.07	0.11	0.60

Source: ELFS 2009-2019, prepared by the author

Appendix 4. Multicollinearity tests – VIF

	Eq. 4 - Duncan & Hoffman			Eq. 5 - Verdugo & Verdugo		
Variable	JA (Eurostat)	JA (ISCO-08)	RM (Mean-One)	JA (Eurostat)	JA (ISCO-08)	RM (Mean-One)
S^o	1.09	1.35	1.14	1.19	1.32	1.71
S^u	1.21	1.22	1.12	1.09	1.38	1.81
S^r	1.87	1.85	1.55	–	–	–
S^a	–	–	–	1.83	2.28	3.42
tenure	1.31	1.32	1.31	1.30	1.32	1.31
log experience	1.48	1.53	1.52	1.53	1.54	1.53
Estonian	1.42	1.43	1.42	1.41	1.43	1.40
married	1.20	1.20	1.20	1.20	1.20	1.20
has child	1.12	1.12	1.12	1.12	1.12	1.12
part time	1.09	1.09	1.09	1.09	1.09	1.09
NACE						
G-J trade	1.62	1.61	1.63	1.61	1.61	1.61
K-N finance	2.04	1.88	2.02	1.87	1.86	1.94
O-S service	1.23	1.22	1.23	1.22	1.21	1.22
A, F, T, U agr & constr	1.39	1.39	1.39	1.39	1.39	1.39
Region						
Centre	1.34	1.36	1.36	1.36	1.36	1.35
East	1.38	1.40	1.40	1.40	1.40	1.40
West	1.70	1.79	1.77	1.79	1.79	1.78
South	2.67	2.88	2.81	2.88	2.90	2.86
Year						
year 2010	1.51	1.51	1.51	1.51	1.51	1.51
year 2011	1.53	1.53	1.53	1.53	1.53	1.53
year 2012	1.52	1.53	1.52	1.52	1.53	1.52
year 2013	1.50	1.50	1.50	1.50	1.50	1.50
year 2014	1.85	1.85	1.85	1.85	1.85	1.86
year 2015	1.57	1.57	1.57	1.57	1.57	1.58
year 2016	1.81	1.81	1.82	1.81	1.81	1.82
year 2017	1.65	1.65	1.66	1.65	1.65	1.66
year 2018	1.80	1.80	1.81	1.80	1.81	1.81
year 2019	1.81	1.82	1.82	1.82	1.82	1.82
lambda	2.44	2.74	2.66	2.76	2.77	2.75
Mean VIF	1.58	1.61	1.58	1.59	1.63	1.70

Source: ELFS 2009-2019, calculated by the author in Stata16

Appendix 5. Breusch-Pagan test for heteroskedasticity

Equation 4. JA (eurostat)

```
Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of ln_rhwage

         chi2(1)      =   151.07
         Prob > chi2  =   0.0000
```

Equation 4. JA (ISCO-08)

```
Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of ln_rhwage

         chi2(1)      =   147.80
         Prob > chi2  =   0.0000
```

Equation 4. RM (Mean-one)

```
Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of ln_rhwage

         chi2(1)      =   147.79
         Prob > chi2  =   0.0000
```

Equation 5. JA (eurostat)

```
Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of ln_rhwage

         chi2(1)      =    92.84
         Prob > chi2  =   0.0000
```

Equation 5. JA (ISCO-08)

```
Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of ln_rhwage

         chi2(1)      =   145.83
         Prob > chi2  =   0.0000
```

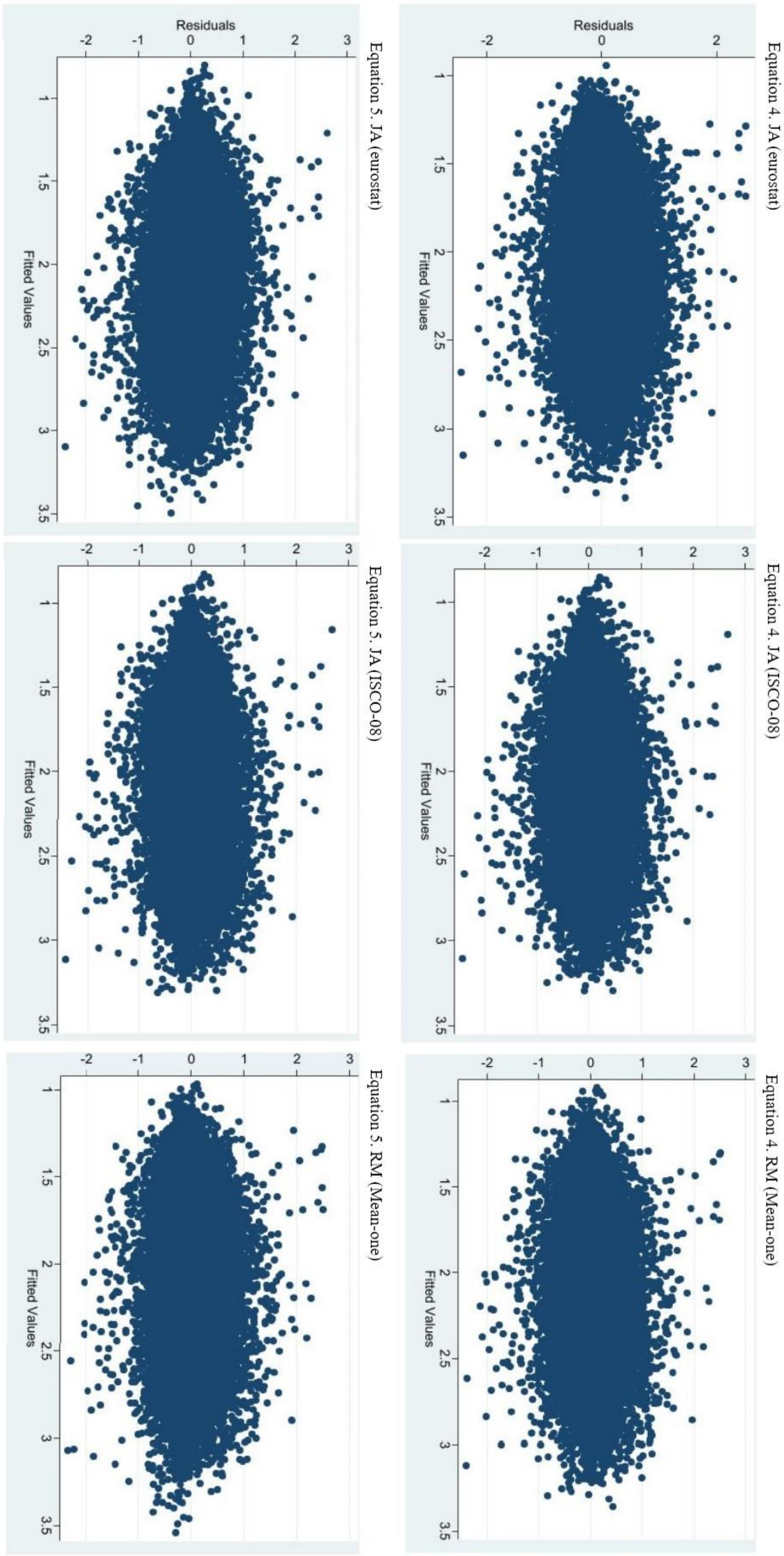
Equation 5. RM (Mean-one)

```
Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of ln_rhwage

         chi2(1)      =   141.29
         Prob > chi2  =   0.0000
```

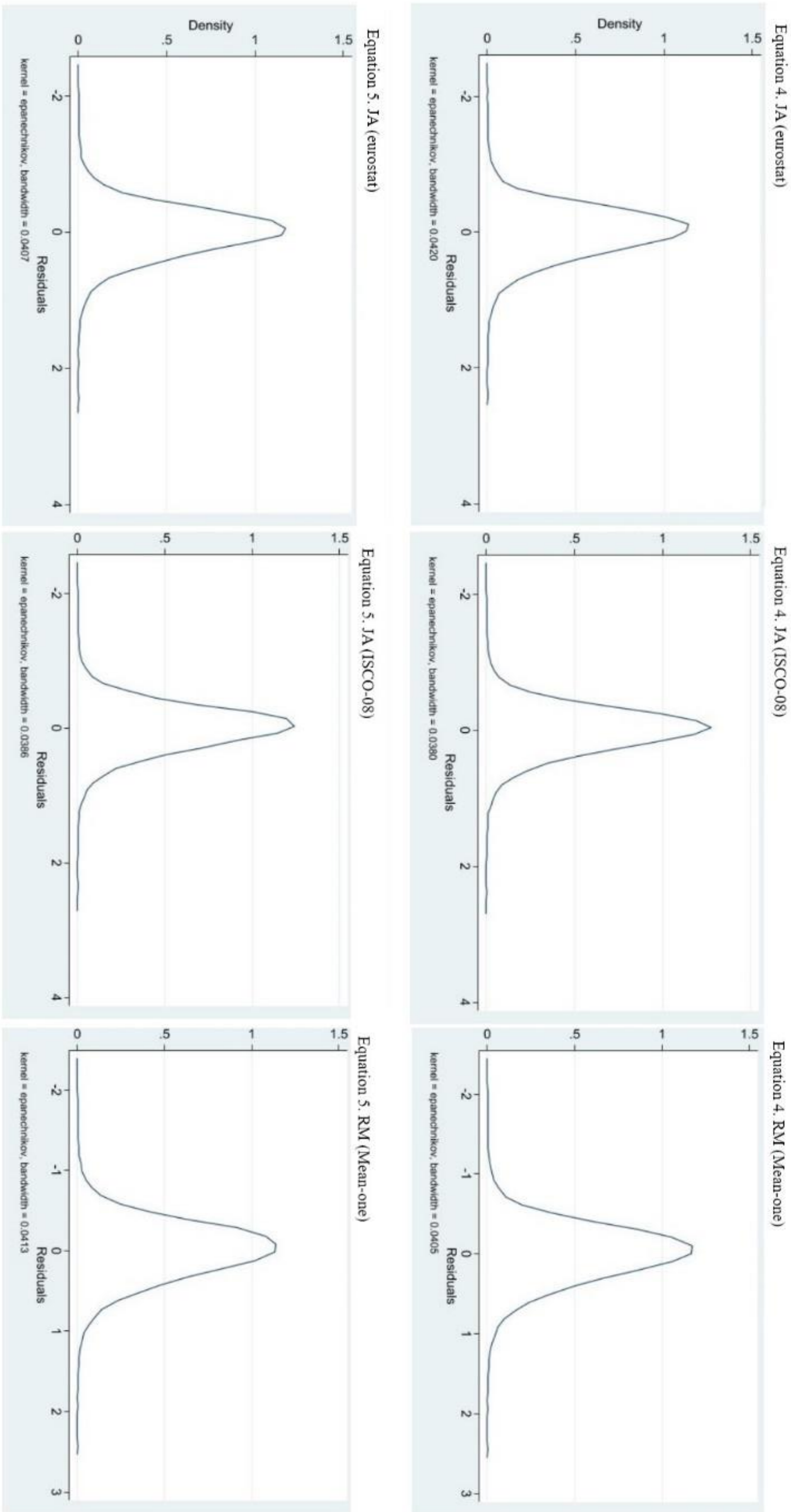
Source: ELFS 2009-2019, calculated by the author in Stata16

Appendix 6. Plot for residuals versus fitted values



Source: ELFS 2009-2019, composed by the author

Appendix 7. Kernel density estimation



Source: ELFS 2009-2019, composed by the author

Appendix 8. NACE 1-digit level codes and names

NACE Level-1 code	Description
A	Agriculture, forestry, and fishing
B	Mining and quarrying
C	Manufacturing
D	Electricity, gas, steam, and air conditioning supply
E	Water supply; sewerage, waste management and remediation activities
F	Construction
G	Wholesale and retail trade; repair of motor vehicles and motorcycles
H	Transportation and storage
I	Accommodation and food service activities
J	Information and communication
K	Financial and insurance activities
L	Real estate activities
M	Professional, scientific, and technical activities
N	Administrative and support service activities
O	Public administration and defense; compulsory social security
P	Education
Q	Human health and social work activities
R	Arts, entertainment, and recreation
S	Other service activities
T	Activities of households as employers; undifferentiated goods- and services-producing activities of households for own use
U	Activities of extraterritorial organizations and bodies

Source: (European Commission, 2010), composed by the author

Appendix 9. Non-exclusive license

A non-exclusive license for reproduction and publication of a graduation thesis[1]

I, Artemi Beljakov,

1. Grant Tallinn University of Technology free license (non-exclusive license) for my thesis

"Education-job mismatch – Implications for individuals earnings in Estonia",

supervised by Simona Ferraro,

1.1 to be reproduced for the purposes of preservation and electronic publication of the graduation thesis, incl. to be entered in the digital collection of the library of Tallinn University of Technology until expiry of the term of copyright;

1.2 to be published via the web of Tallinn University of Technology, incl. to be entered in the digital collection of the library of Tallinn University of Technology until expiry of the term of copyright.

2. I am aware that the author also retains the rights specified in clause 1 of the non-exclusive license.

3. I confirm that granting the non-exclusive license does not infringe other persons' intellectual property rights, the rights arising from the Personal Data Protection Act or rights arising from other legislation.

09.05.2023 (date)

[1] *The non-exclusive licence is not valid during the validity of access restriction indicated in the student's application for restriction on access to the graduation thesis that has been signed by the school's dean, except in case of the university's right to reproduce the thesis for preservation purposes only. If a graduation thesis is based on the joint creative activity of two or more persons and the co-author(s) has/have not granted, by the set deadline, the student defending his/her graduation thesis consent to reproduce and publish the graduation thesis in compliance with clauses 1.1 and 1.2 of the non-exclusive licence, the non-exclusive license shall not be valid for the period*