UNIVERSITY OF GOTHENBURG
SCHOOL OF BUSINESS, ECONOMICS AND LAW

Graduate School - Master's Thesis in Economics

Overeducation and its impact on wages in Vietnam

Anh Vu Tran

Abstract

The study uses the most recent Vietnam Household Living Standard Survey in 2014 and 2016 to describe the incidence of overeducation and estimate the wage effect of educational mismatch. Overeducation is more prevalent in male than in female groups. In addition, younger workers are more likely to be overeducated than older counterparts. Using the extended Mincer equation in which education is decomposed into over, required, and under years of schooling, the cross-sectional estimates are consistent with the literature. On the one hand, overeducated workers earn less, while undereducated workers earn more than their matched peers holding the same educational level. On the other hand, overeducated workers receive a higher wage, whereas undereducated co-workers gain a lower wage than their adequately educated colleagues. Unlike previous studies, the fixed effects model cannot be identified because of low within-individual variation. However, the panel data enables us to cross-check the years of education between two survey rounds. Although data inconsistencies exist, the estimation results are robust across different samples.

Supervisor: Nadine Ketel

22 May 2018

Acknowledgements

I would like to express my deepest gratitude to my supervisor, Nadine Ketel, for her continuous support and dedication while I was conducting my thesis. I have learnt a lot and improved myself beyond my limits from her useful feedback and careful guidance.

I would also like to thank Swedish Institute for the scholarship for my two-year master program at University of Gothenburg. Without their financial support, my study would not have been possible.

Finally, I would like to thank my dear family and their endless support throughout my study. Special thanks should also be given to my partner, Duc. Thank you for your support, love and care during my study and in the future.

Table of Contents

List of Tables and Figures

1. Introduction

Overeducation is a situation where the acquired educational level of a worker is higher than the job requirement. This issue was first investigated by Freeman (1976) who observed a significant decline in return to college in comparison to high school during the 1970s in the United States. Since then, the topic has been studied extensively because overeducation is a costly phenomenon for both society and individuals. If overeducated workers have smaller returns to education than correctly matched peers, misallocation will be associated with lower earnings. Earlier studies also find that overeducated people are less productive and less satisfied with the job than well-allocated workers (Tsang, Rumberger and Levin, 1991). Therefore, from a firm's perspective, hiring overeducated employees potentially reduce the overall output. For the society, because education is underutilized and investment in education yields smaller returns, the existence of overeducation is inefficient, especially when it is heavily subsidized by the government (McGuinness, 2006).

The incidence of overeducation and its impact on earnings have been well documented in developed countries including the US and many European countries such as Germany, Italy, Portugal, Spain, the Netherlands, and the UK thanks to the early development of higher education. In the last two decades, parallel to economic growth, educational attainment has improved in developing countries, as higher education becomes more affordable and accessible to a broader population. Therefore, if the labor supply increases faster than the job creation, overeducation will be a potential concern in these countries. However, so far evidence on the incidence and effects of overeducation in developing countries is scarce.

To contribute additional evidence to this scarce literature, this thesis attempts to measure the incidence of overeducation and its effect on wages in Vietnam using a subsample of wage earners from the most recent Vietnam Household Living Standard Survey 2016 (VHLSS). Since 50% of the 2016 sample were interviewed in the previous round, a balanced panel data from the VHLSS 2014 and 2016 is constructed to possibly control for individual heterogeneity.

We take Vietnam as the study object for several reasons. First, student enrolment at all levels across the country has increased considerably in recent years. For example, the primary and lower secondary gross enrolment rates were 0.99 and 0.97 respectively in 2013 (UNESCO, 2016). Higher education has also experienced rapid expansions. To illustrate, the number of higher education institutions (universities and colleges) in 2015 was 442 schools, a three-fold increase from 1999 (MOET, 2017). Likewise, total student enrolments in 2015 reached more than 2.2 million students, which is more than three times higher than that of 1997 (Sheridan, 2010; MOET, 2017). One possible explanation is that education and human resources development has been one of the top priority of the government of Vietnam. In fact, the government expenditure on education accounted for around 20% of total government expenditure during the 2010-2013 period. As a percentage of GDP, it is always among the largest in Asia (Vietnam (2013) 5.7%, Singapore (2013) 3%, Thailand (2012) 5%, Japan (2013) 4%, Rep. of Korea (2012) 4.5%) (UNESCO, 2016). The total expenditure per student by the government and households also follows an increasing trend for the period 2009-2013 (UNESCO, 2016). Second, a recent study using household data up to 2014 by Doan, Le and Tran (2017) shows that average return on education in Vietnam has reduced since 2009 possibly due to the oversupply of higher education. Another potential explanation for the decreasing return to education is the lack of quality higher education: skills of graduates are under the labor market

requirements (UNESCO, 2016; Dang and Glewwe, 2017). Finally, media attention has been paid to the fact that graduate students increasingly struggle to find a job and must accept underqualified jobs. All these three observations suggest the existence of overeducation in Vietnam; however, there have been few studies on this topic. Hence, addressing this knowledge gap will help both policymakers and individuals to reconsider and assess their level of investment in education.

The leading strand of the literature estimates the wage penalty of overeducation using the extended Mincer wage equation by Duncan and Hoffman (1981). In this specification, years of schooling are decomposed into surplus, required and deficit years of education. Most studies find consistent OLS estimates of wage effects of educational mismatch across different datasets and time. However, capturing the causal effect is challenging because of omitted variable bias and measurement error of human capital. To address the endogeneity concern, three major approaches are instrumental variables, fixed effect estimators, and propensity score matching, but they are far from reaching a thorough conclusion (Leuven and Oosterbeek, 2011).

In this study, we first describe the rate of overeducation using the objective measures in which either the average or the mode educational level of each occupation is proxied for job requirements. We find that the incidents of overeducation and undereducation vary with the measures of required education. In fact, the percentages of overeducated workers in 2016 are 14.5% and 33.4% using the mean and mode indexes respectively. This large gap is typical in the literature and not unique for the case of Vietnam, explained by the difference in the definition of two measures. Furthermore, men on average have a higher rate of overeducation than women. Also, the incidence of overeducation differs across age groups. For the mean index, the share of overeducated workers is the highest for the 25-year-old group at about 25% before decreasing to around 10% for the over 40-year-old groups. This pattern supports the flourishing of higher education in the last 20 years.

Next, we estimate the effect of education mismatch on wages by OLS using the Duncan and Hoffman (1981) specification. We confirm a positive association between education and earnings when controlling for age, gender, marital status, ethnicity, urban, six geographical regions and job characteristics. We also find the return to over, required and undereducation to be in line with the literature. On the one hand, overeducated workers earn more, while undereducated workers earn less than their adequately educated co-workers. On the other hand, overeducated workers have a lower wage, whereas undereducated workers have a higher wage than their matched peers with the same educational level. In addition, we observe different returns to overeducation for different groups categorized by individuals' characteristics such as age cohorts, urban/rural areas and contract status, while we do not find the difference in return to overeducation of female and male groups.

We cannot utilize the panel structure of the data because of the low within-individual variation of years of schooling and years of required education. In other words, the fixed effect or first-difference model cannot be identified. Alternatively, we use the 2014 data to check the robustness of the estimation results since inconsistencies in years of education are observed between two survey rounds. We find that the estimators are consistent across different samples. Therefore, although we cannot derive a causal relationship, we can conclude that there is evidence of wage penalty for educational mismatch in Vietnam.

The thesis is structured as followed. The next section reviews existing studies. Section 3 summarizes three dominant theories explaining the existence of overeducation and motivates three hypotheses. Section 4 describes the data and the variables of interest. Section 5 presents the empirical strategy and some attempts to measure causal effects. Section 6 discusses the estimation results and robustness check. Section 7 provides the conclusions of the study.

2. Literature review

The topic of overeducation and return to overeducation has been widely studied due to its significant costs to individual workers, employers and society, together with its importance for policy (McGuinness, 2006; Pouliakas, 2012). We hereby give a brief review of the existing empirical evidence in regard to the measure of overeducation, countries of study, and identification strategies to form the basis for our choices of methodological approaches.

Overeducation is determined based on how large a person's educational attainment is above the required qualification of the job. Particularly, there are three measures of overeducation, including one subjective and two objective methods (Hartog, 2000). The subjective measure can be obtained through a worker self-report approach. Under this approach, individual workers report the minimum requirement of their job, which is then compared with their acquired education. The method can also involve asking the respondents directly whether they are overeducated. This approach was first used by Duncan and Hoffman (1981), and later by Büchel and van Ham (2003), McGuinness and Pouliakas (2017) and many others, in which the specific survey questions about overeducation are available.

Alternatively, overeducation can be captured objectively by comparing workers' educational level with the evaluation of professional job analysts based on the job title. Another objective measure is the realized matches which also assumes jobs with the same titles have the same requirements (Kler, 2005). Accordingly, an over (under)educated worker is the person whose years of schooling is higher (lower) than the job requirement. The required educational level is defined as the most common qualification of employers within a job such as the mean or the mode of the distribution (Hartog, 2000). According to the pioneer, Verdugo and Verdugo (1989), a person is overeducated (undereducated) if her educational level is one standard deviation higher (lower) than the *average* qualification of her occupation, while she is adequately educated if her educational level belongs to the one standard deviation range. While the mean of realized matches is beneficial when the educational requirement for a job is unknown, the symmetry pattern and the choice of one standard deviation are arbitrary and not theoretically motivated (Leuven and Oosterbeek, 2011). As an alternative, Kiker, Santos and de Oliveira (1997) consider the *mode* educational level as the job requirement which is less sensitive to outliers and can address the symmetry issue. However, both measures are unable to capture the variation of job requirement across jobs within an occupation. Instead, they only exploit the difference between professions.

Groot and Maassen Van Den Brink (2000) summarize 25 studies and find that the standard deviation approach produces a considerably lower estimate of the incidence of overeducation than other methods. Also, the rates of overeducation and undereducation differ significantly among three measures. In an attempt to explain that, McGuinness (2006) states that the mean of realized matches method imposes a distance of one standard deviation, while the others do not have any restrictions. However, these

differences do not lead to a significant difference in the estimate of the return to education (Groot and Maassen Van Den Brink, 2000; Rubb, 2003). From the practical perspective, the choice of measurement mainly depends on data availability (McGuinness, 2006). In this study, we can only analyze with the mean and mode of realized matches measures of overeducation.

In the literature survey and meta-analysis, Hartog (2000), Groot and Maassen Van Den Brink (2000), McGuinness (2006), and Leuven and Oosterbeek (2011) conclude that the empirical studies yield consistent estimates of the wage effect of education. Overeducated workers have a wage penalty in comparison to their perfectly matched peers who have the same educational level. The converse result is true for undereducated labors. In addition, in one job, overeducated workers earn the highest wage, followed by adequately educated and undereducated workers. In terms of data structure, most studies use cross-sectional data, while a few utilize household panel data. There are two main model specifications: the over-required-undereducation (ORU) developed by Duncan and Hoffman (1981) and the dummy variable proposed by Verdugo and Verdugo (1989). The main feature of the ORU specification is that individual's educational attainment is decomposed into three parts corresponding to the amount of education required by the job, surplus and deficit amount of schooling. Duncan and Hoffman (1981) use the 1976 Panel Study of Income Dynamics and the decomposed Mincer wage equation to reveal that the return to an additional year of schooling beyond the job requirement is positive, but its magnitude is significantly lower than the return to a year of required education. Verdugo and Verdugo (1989) use a slightly different approach, including two dummy variables on over- and undereducation, and controlling for actual educational attainment in the wage education. They find that overeducated workers on average earn less than other groups. However, this approach is criticized because they do not control for required education, but interpret the result as if they did (Leuven and Oosterbeek, 2011). We, hence, use the ORU specification for our analysis to utilize the decomposed specification.

Leuven and Oosterbeek (2011) emphasize that all the previous studies had not adequately explained the wage effect of education. Besides the measurement error of human capital, endogeneity is the primary concern that prevents researchers from capturing the causal effect. The number of studies that attempt to address this issue is limited. They can be classified into three methods: instrumental variables (Korpi and Tåhlin, 2009), fixed effect model (Bauer, 2002; Korpi and Tåhlin, 2009; Lindley and McIntosh, 2009; Tsai, 2010) and propensity score matching (McGuinness, 2006). Regarding the instrumental variable approach, the specification requires at least three instrumental variables for three endogenous variables, which is very challenging especially for the required educational level (Leuven and Oosterbeek, 2011). Worse still, the choice of variables is further limited to the survey content. For example, Korpi and Tåhlin (2009), one of the few studies in this strand, use four variables on childhood conditions as instruments for educational level. Although the J-test shows they are all exogenous, we argue that childhood conditions such as sibling size, economic conditions can affect wages via other channels beside education. In addition, the instruments are weak; hence, they could not derive a solid conclusion. Regarding the fixed effect model, there is little variation that can be exploited because people are less likely to change their educational level after entering the job market (Leuven and Oosterbeek, 2011). The analysis hence mostly relies on the job change of individuals, but the strict exogeneity assumption usually fails. Bauer (2002) uses the German Socioeconomic Panel for the 1984-1998 period to estimate the wage effect of overeducation. While the pooled OLS estimation result is similar to earlier studies, the panel estimation shows that the difference in

return to over, required and undereducation disappears when unobserved individual heterogeneity is controlled, and the mode index of required education is used. They also emphasize that the significantly lower within-group variation in years of schooling for the mean index leads to unreasonable fixed effect estimators. Korpi and Tåhlin (2009) use the same method for the Swedish panel data and confirm that time-invariant individual heterogeneity factors can explain the large part of wage effect of education. However, they mention but do not explicitly discuss the within and between variations of years of schooling. While they all find that the wage penalty disappears when addressing the endogeneity issue, none of the approaches seems very convincing.

The majority of studies focus on developed countries where higher education has a long history of development. In particular, the research interest began in the US (Freeman, 1976) and gradually shifted to European countries such as Germany, Italy, the Netherlands, Spain, and the UK. In recent years, educational attainment in all levels has been rapidly improving in developing countries. Notably, higher education sector becomes more accessible and affordable, moving from an elite to a broader population system. Therefore, overeducation may become a potential issue in those countries where there has been little evidence on this topic due to data unavailability. Several first studies are Mexico (Quinn and Rubb, 2006), India, Thailand, and Philippines (Mehta *et al.*, 2011), and Thailand (Paweenawat and Vechbanyongratana, 2015), and Armenia, Yunnan (China), Georgia, Laos, Sri Lanka, and Vietnam (Chua and Chun, 2016). In general, their findings are consistent with those in developed countries. For example, Quinn and Rubb (2006) use the cross-sectional data of Mexico during the 1987-1997 period with three different measures of overeducation (based on the mean, mode and OLS estimation) and the ORU specification to capture the return to education. The dataset consists of 4945 men working in more than 100 occupational categories. They confirm that overeducated workers have a wage penalty in comparison to their perfectly-matched peers. More recently, Chua and Chun (2016) use a novel survey in urban areas of several developing countries which is rich in labor skill information. Their survey data contains specific questions on the job requirement. Therefore, they are able to use the subjective measure of education and the ORU specification. Controlling for human capital (tenure, cognitive and non-cognitive skills and self-reported deficiency), job characteristics (industries, sectors), and demographic characteristics, they find a similar result to the literature. Specifically, the coefficients on under and overeducation are significantly lower than that on required education. For a subsample of Vietnam, in the full model, they conclude that the return on an additional year of required, over and undereducation are 6.7, 3.1 and -3.7%. Using a different dataset and measure of overeducation from the first study in Vietnam, our research would contribute to the literature additional evidence from a developing country.

Table B-1 in the Appendix summarizes 16 studies using the realized matches measure and Duncan and Hoffman specification. First, for the mean index, the incidence of overeducation is from 7 to 18%, except for the markedly high rate of the Australian graduates (Kler, 2005). Meanwhile, for the mode index, the proportion of overeducated workers is significantly higher, varying from 25 to over 30%. It should also be noted that the occupational classification differs among studies. The least detailed occupational code is one-digit code, while the most specific classification is three-digit code. Second, regarding the model specification, the set of control variables is quite similar among all studies. They find that the return to an additional year of overeducation is positive, ranging from 3 to 9%. Meanwhile, the return to one more year of undereducation is negative, fluctuating from -8 to -3%. In addition, their absolute values are smaller than

the return on years of required education which is between 6 and 18%. In contrast, some studies by Groot (1993, 1996, and 1997) find negative and positive coefficients on years of surplus and deficit schooling respectively. They explain that overeducated workers have less experience and lower productivity; hence, they receive a lower wage than correctly matched coworkers. Based on that summary, we expect the share of overeducated workers in Vietnam differs between two measures of overeducation and falls into the range of previous studies. For the estimation of return to surplus, required and deficit years of education, we expect there exists wage penalty for incorrectly-matched workers; however, the sign of these effects cannot be predicted based on that summary.

3. Theoretical framework and hypothesis

Although the number of studies on overeducation has rapidly increased, there is no consensus theory of overeducation. Instead, most studies describe the incidence of overeducation in relation to existing labor market literature (McGuinness, 2006). Three dominant theories that explain the presence of overeducation are (i) Human capital theory, (ii) Job competition model, and (iii) Assignment theory.

Human capital theory

The human capital theory was developed by Becker (1964) using the labor supply approach. In this model, workers earn their marginal product which is determined by their accumulated human capital such as formal education, job training, and experience.

In a competitive market, both firms and individual workers react to the change in labor supply. Therefore, the existence of overeducation is associated with a higher number of skilled workers available at a relatively lower wage. On the one hand, firms' production technology is so flexible that they are able to substitute the previous low skilled to higher skilled labors, and hence fully utilize labor forces. On the other hand, workers will have a lower return to education and subsequently adjust their investment in human capital. As a result, overeducation is temporary and should not exist in equilibrium. In other words, the rate of return to education is independent of the overeducation status of a worker. From this theory, we can derive the first hypothesis.

Hypothesis 1: There is no difference in return to education among overeducated, adequately educated and undereducated workers.

Job competition theory

A concern with the human capital theory is that the rigidity of production process may prevent firms from quickly responding to the change in labor input (Duncan and Hoffman, 1981). Different firms probably have various levels of rigidity. Thus, workers' earnings should also be affected by the characteristics of their jobs.

To address this issue, Thurow (1975) proposed the job competition model with a demand-side approach. In this model, workers are in a queue system for each job and jobs are in another wage-based queue system. Thus, workers have to compete to maintain their place to get a higher chance of getting a job. The ranking is determined by their educational level as firms realize a substitution between formal schooling and job training, and a negative correlation between formal education and training cost. The existence of overeducation is associated with an increase in the average educational attainment of the labor force. Again, more high-skilled labors are supplied at relatively lower wages. With a fixed labor

demand, lower skilled workers will be replaced by the abundant high-skilled group. While receiving a lower return to education, high-skilled workers continue to invest in education to maintain their position in the job market. Thus, overeducation persists and workers' skills are underutilized. Based on this theory, we can motivate the second hypothesis.

Hypothesis 2: There is no difference in return to education between overeducated and undereducated individuals because earnings are determined by required schooling.

Assignment theory

Combining the two above theories, Sattinger (1993) developed the assignment theory considering both demand and supply sides of the labor market. In the model, different workers have different job-specific skill sets. Therefore, the productivity and earnings differ among workers even though they have acquired the same level of education. This implies that earnings depend on both human capital and the matching between workers and jobs. Thus, overeducation exists when workers underutilize their skills due to inefficient job allocation. Like the job competition theory, overeducation persists if the job assignment does not change.

Hypothesis 3: There are differences in return to over, adequate and under education

In summary, although all three theories explain the incidence of overeducation based on the supply and demand of labor market, they use different mechanisms through which human capital affects workers' wages. Human capital theory states that employees' actual level of education determines their earnings, while job competition theory concludes that only job requirement matters. Meanwhile, assignment theory introduces the heterogenous skill assumption: workers having the same level of education may possess different skill sets. Therefore, workers' productivity depends on both their human capital and their job characteristics. In other words, both actual and required education levels affect workers' earnings. All these hypotheses will be tested using the national household data from Vietnam.

4. Data

The data for this study is obtained from the last two rounds of the Vietnam Household Living Standard Survey (VHLSS) conducted by the General Statistics Office of Vietnam (GSO) in 2014 and 2016. The VHLSS has a sample size of around 46000 households in every round, representative at the national, regional and provincial levels (GSO, 2016). The survey covers information about demographic characteristics, education, health, employment of all household members, as well as total income and expenditure of the households.

4.1 Panel data construction and Sample restriction

The sampling design of the VHLSS enables us to construct individual-level panel data. Particularly, half of the households in the previous round remain in the subsequent sample, while the other half is randomly and newly selected from the master frame. Individual data can be matched between two survey years using individual-level identification code. Since there exists mismatch in household and individual codes, matched observations are further checked with individual information such as gender, years of birth and name.

From the original sample of the VHLSS, we include respondents in their working age (female: 15-55, male: 15-60). We further restrict the sample to people who are currently working and having a wage job as the primary occupation in the last 12 months. The employment-to-population ratio is around 83% in both

years, and 40-44% of them are wage earners in their main job in 2014 and 2016 respectively. The composition of this sample is slightly different from the Labor Force Survey (LFS) which collects national statistics on employment. Specifically, for the LFS, the percentage of paid workers are 35.6% and 41.2% for 2014 and 2016, while the employment-to-population ratio is around 75-77% (GSO, 2017). As expected, the labor market in a developing country is characterized by the small share of wage employment. This observation raises the concern of sample selection bias because we cannot observe earnings of the self-employed group. According to the human capital, job competition and job assignment theory, the probability of employment depends on both human capital and job characteristics, implying that unobserved factors in the wage regression also affect whether a worker gets a job. For example, if we assume that employed workers have higher ability than non-employed counterparts, selection bias will lead to a downward bias of the return to education. Previous studies show that correcting for selection bias leads to a higher wage penalty for overeducated workers (Caroleo and Pastore, 2017). Therefore, without addressing the selection bias, we can only derive the conclusion for the wage earners instead of the entire population.

Occupations are classified into 45 occupations according to the two-digit level of the Vietnam Occupational Standard Classification 2009. Following the standard in the literature, we exclude military officers and three other occupations of less than ten observations for each year[1]. As a result, we have a sample of 36829 and 38596 observations for 2014 and 2016 respectively, and a balanced panel of 10533 observations.

Table 1: Sample restriction

Selection criteria	2014		2016	
	No of observations	%	No of observations	%
Original sample (Age ≥ 6)	176205		175241	
Working age (Male 15-60, Female 15-55)	110748	100.0	107125	100.0
Currently working	91694	82.8	88879	83.0
Wage earners in the main job	37117	33.5	38883	36.3
Non-military job with more than 10 obs.	36829	33.3	38596	36.0
Final sample	**36829**		**38596**	
Balanced panel			**10533**	

4.2 Measure of overeducation

Data availability does not allow us to use the job analysis and worker self-assessment measure of overeducation. Thus, required educational level is measured by either the mean or the mode of realized matches. For the first definition, a worker is overeducated (undereducated) if her educational level is one standard deviation higher (lower) than the average educational level of all workers in her occupation. For the second definition, a person qualifying higher (lower) than the mode educational attainment of each occupational group is regarded as overeducated (undereducated).

This survey contained information on the highest completed grade (0-12) and the highest educational qualifications which are general schooling (primary, lower and upper secondary, college, graduate and

[1] The 2014 data excludes occupations code 12 – Leaders working in the National Assembly and Office of the President, 13 – Leaders working in the Government Office, and 14 – Leaders working in the People's Courts and people's procuracy. The 2016 data excludes occupations code 13, 14, and 17 – Leaders working in humanitarian organizations; organizations for other particular benefits.

post-graduate) and vocational training (short-term, long-term, professional secondary, and vocational college). The number of acquired schooling years is calculated as the highest completed grade plus additional years of higher education and vocational training. The conversion table is specified in Table A-1 in the Appendix. For example, a person who completed grade 12 and hold a university degree (four years) and a short-term vocational training certificate (one year) will have total schooling years of 17. The summary of years of education is shown in the third row of Table 3. Accordingly, the average years of education are 10.28 years in 2016, which is marginally higher than that in 2014.

The sample size of previous studies using the same measure of overeducation (A summary in Table B-1 in the Appendix) ranges from 1 to 30 thousand observations for cross-sectional data. In addition, most of them categorize occupation at the two-digit level. We have a relatively large sample of more than 38000 observations and 42 two-digit coded occupations. The number of observations for each occupation ranges from 60 to 4466 observations with an average of 920 for the 2016 sample. Therefore, the concern of unreliable estimates of required education due to the small sample size is minor. As can be seen in Table A-3 in the Appendix, the required years of schooling are almost unchanged across two survey rounds. For the average years of schooling, the gap between two years for each occupational category ranges from 0 to 0.5 year. Similarly, the mode years of education remain unchanged for all occupational categories except for the low skilled labors in agriculture, forestry and fisheries. Another critique towards the realized-matches measure is the ignorance of within-occupation variation. Hence, instead of using nine broader occupational categories, we use the most detailed code available of 45 two-level professional codes.

Using the mean index, the average years of required education is 10.2 years, while the surplus or deficit years of schooling is around one year. For the mode index, the required years of schooling is slightly higher of 10.7 years. Meanwhile, the years of overeducation and undereducation are 0.8 and 1.4 years respectively (Data description in the lower panel of Table 3). The incidence of overeducation is presented in Table 2. The proportion of overeducated workers varies with the measures of overeducation. For the 2016 sample, the percentages of overeducated respondents are 14.5% and 33.0% for the mean and mode definition of required education respectively. Likewise, these figures for the 2014 sample are 14.5% and 27.5% respectively.

The large gap between two measures, which is in line with previous research in Portugal (Kiker, Santos and de Oliveira, 1997), and Mexico (Quinn and Rubb, 2006), can be explained by the difference in nature of the definition of required education. The incidence of overeducation and undereducation fall into the range of existing studies. Compared to developed countries, the share of overeducated workers appears to be smaller. In addition, the rate of overeducation measured by the mean index is considered as the lower bound of the true incidence since it produces the lowest estimates among all measures (Groot and Maassen Van Den Brink, 2000; Leuven and Oosterbeek, 2011).

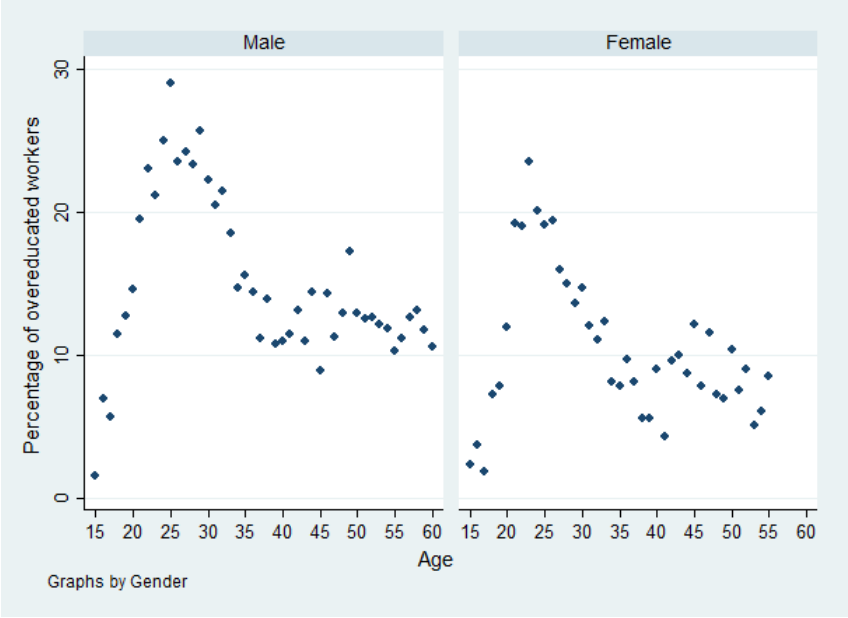
Table 2: The distribution of overeducation by two definitions of required education and years

	Mean index		Mode index	
	2014	2016	2014	2016
Undereducated	14.3	15.2	38.5	32.7
Adequately educated	71.2	70.3	34.0	34.0
Overeducated	14.5	14.5	27.5	33.4

Figure 1 shows the relationship between the proportion of overeducated workers and age, similar for both men and women. In general, men experience a higher rate of overeducation than women. For workers under 22 years old, the proportion of overeducation is quite low because most of them are still at school and have low skilled jobs. As soon as they enter the labor market, we observe a higher rate of overeducation. This proportion reaches a peak at the age of 25 and decreases sharply afterward before remaining stable at around 10% from the age of 40. This pattern is similar when required education is measured by the mode although the actual value is different

This observation is in line with the rapid development of higher education in Vietnam. Since 2000, the number of higher education institutions have increased markedly, followed by the number of student enrolment. For instance, the number of higher education institutions in 2015 was 442 schools, which is three and four times larger than those figures in 1999 and 1987. In the same vein, total student enrolments in 2015 reached more than 2.2 million students, a three-fold increase from 1997 (Sheridan, 2010; MOET, 2017). We hence expect the incidence of overeducation is higher among young cohorts, especially for those under 35 years old who experience the booming of higher education sector. Despite that significant growth, the gross enrolment rate in tertiary remains low at 16% in comparison with other countries in the region (Sheridan, 2010); hence, graduates may not be oversupplied. However, it is widely concerned that the quality of higher education cannot meet the requirements of the society (UNESCO, 2016). Specifically, graduates lack relevant working skills and need on-the-job training (Sheridan, 2010). If firms realize the importance of on-the-job training and consider extra years of education as a reduction of training cost, they will recruit people with higher education level. To compete for jobs, graduates will get more formal years of education. Thus, we expect a high proportion of overeducated workers in the labor market, which is the prediction of the job competition theory (Thurow, 1975). This observation can also be explained by the career mobility theory (Sicherman and Galor, 1990). Accordingly, younger cohort tends to accept lower skilled job since it has a higher chance of promotion in the future.

Figure 1: The incidence of overeducation by age and gender (mean index)



4.3 Outcome variable and Control variables

In the survey, respondents are asked to report up to three jobs. We use the hourly wage of the primary job as the outcome variable which is calculated by the total salary (including bonus, allowance), the number of working days in the last 12 months and the average number of working hours in one day. The respondents also report their wage in the last 30 days which is highly correlated with their total salary. As can be seen in Table 3, the average hourly wage is 25.85 thousand VND (1.21 USD) in 2014, which is slightly less than that in 2016, 28.73 thousand VND (1.27 USD).

The description and descriptive statistics of other variables are presented in Table 3. As can be seen, the mean and standard deviation of all variables are quite similar between two survey rounds. Regarding the demographic characteristics, around 40% of the sample are female workers. The average age is 35 years old. Furthermore, most of the respondents are married (72%) and from the majority ethnic group (89-91%). Regarding the job characteristics, half of the workers have a labor contract. The majority of them are working in the private sectors, while around 23% and 10% of the respondents are in the public and foreign sector respectively.

Table 3: Variable definition and descriptive statistics

Variable description	2014 (N=36829)		2016 (N=38598)	
	Mean	Std. dev.	Mean	Std. dev.
Hourly wage (thousand VND)	25.85	63.78	28.73	33.06
Logarithm of hourly wage	3.04	0.59	3.18	0.56
Acquired years of schooling	10.21	4.45	10.28	4.43
Female dummy (female=1)	0.40	0.49	0.41	0.49
Age of the respondents	35.45	10.76	35.83	10.73
Marital status (married=1)	0.72	0.45	0.72	0.45
Ethnicity (majority=1)	0.91	0.29	0.89	0.31
Urban area (urban=1)	0.38	0.49	0.39	0.49
Having a labor contract (contract=1)	0.49	0.50	0.50	0.50
Working in public sector (state=1)	0.26	0.44	0.23	0.42
Working in foreign sector (foreign=1)	0.09	0.28	0.10	0.30
Mean index				
Surplus years of schooling	1.09	1.66	1.11	1.66
Required years of schooling	10.21	3.33	10.28	3.29
Deficit years of schooling	1.09	1.90	1.11	1.91
Mode index				
Surplus years of schooling	0.83	1.60	1.23	2.24
Required years of schooling	10.73	3.40	10.24	4.33
Deficit years of schooling	1.35	2.20	1.20	2.17

5. Empirical strategy

5.1 Baseline model

In this study, we follow previous studies and propose the ORU specification. The model is built upon the standard Mincer wage equation to capture the overall return to schooling (Eq. 1). The wage for an individual i is a function of years of schooling, years of experience and its squared term. When the zero-conditional mean assumption holds, the coefficient on schooling captures the causal effect of education on earnings.

$$lnWage_i = \alpha_0 + \alpha_1 Education_i + \alpha_2 Experience_i + \alpha_3 Experience_i^2 + \varepsilon_i \text{ (Eq. 1)}$$

The ORU specification is an extension of Mincer equation by Duncan and Hoffman (1981), in which the actual years of schooling is decomposed into three parts, required education for a job (RE), over (OE) and undereducation (UE) as in Eq. 2. Accordingly, an exactly matched worker would have the education level equaling to the required education (AE = RE), while an overeducated and undereducated worker would have the educational level defined as AE = RE + OE (UE = 0), and AE = RE – UE (OE = 0) respectively. The definition of required education could be either *the mean* or *the mode* of the distribution of each occupation.

$$\text{AE} = \text{RE} + \text{OE} - \text{UE} \text{ (Eq. 2)}$$

The above defined educational variables are then substituted in Eq. 1 to obtain ORU specification as in Eq. 3.

$$lnWage_i = \alpha_0 + \alpha_1 RE_i + \alpha_2 OE_i + \alpha_3 UE_i + \beta X_i + \varepsilon_i \text{ (Eq. 3)}$$

X is a vector of individual characteristics with a corresponding coefficient vector β. The first term $\alpha_1$refers to the return to the one more year of the job requirement. The second α_2 and third term α_3 indicate the return to additional year of surplus and deficit years of schooling for overeducated and undereducated educated workers respectively.

The ORU specification allows us to test three hypotheses specified in Section 3. The first hypothesis, motivated by the human capital theory, states that wage is determined by actual educational attainment of workers, which is equivalent to test $H1: \alpha_1 = \alpha_2 = -\alpha_3$. In contrast, the second hypothesis informed by the job competition theory indicates that only required education affects workers' wage. For this specification, it is similar to test $H2: \alpha_2 = \alpha_3 = 0$. Finally, the third hypothesis that supports the assignment theory implies that wages depend on both human capital and job-specific factors. Therefore, we can test $H3: \alpha_1 \neq \alpha_2 \neq \alpha_3$. If we cannot accept either the first or the second hypothesis, we can conclude that there is evidence supporting the third hypothesis.

Empirical studies have provided consistent estimated results (Rubb, 2003): the coefficients on over and required education is positive, while the coefficient on undereducation is negative. In addition, the absolute values of α_1 and α_3 are smaller than α_2. This finding can be interpreted that over- (undereducated) workers have a higher (lower) wage than correctly matched workers in the same occupation, but a lower (higher) wage than exactly matched workers having the same level of education. Thus, empirical results do not support the job competition and human capital theory.

The control variables in previous empirical studies can be categorized into three groups: human capital, job and social-demographic characteristics of respondents and informed by relevant theories. According to human capital theory, human capital contains both formal and informal education. The former is measured by years of schooling, while the latter could be captured by ability and skill test, experience, tenure, and on-the-job and other training. Job characteristics are other important control variables as motivated by the job competition theory, which include contract type, working hours, firm size, firm ownership (state, private, and foreign ownership) and industry. In addition, social-demographic characteristics of respondents consist of age, gender, marital status, race, dependency ratio, and living locations.

We follow the standard specification and consider the data availability to choose the set of control variables. Specifically, wage is a function of gender, age, marital status, ethnicity, urban, six regional dummies and job characteristics (contract, firm ownership, and 21 industries). It should be noted that compared to the LFS, the VHLSS does not cover many employment information such as job mobility, job searching and screening, and job training. However, it has the advantage of measuring educational level by years of schooling rather than qualification. Hence, analysis using the VHLSS facilitates the interpretation.

A limitation of OLS estimator is that it is inconsistent and biased if we do not include sufficient controls in the wage regression. Therefore, estimated coefficients on years of schooling cannot be interpreted as causal due to omitted variable bias. For an omitted factor that is positively correlated with earnings, the rate of return to overeducation may be overestimated if the unobserved factor is positively associated with the surplus year of education. In contrast, the coefficient on years of overeducation may be underestimated if there is a negative association between overeducation and omitted variables. One may argue that workers tend to take extra years of education to compensate for the lack of other human capital components (Sicherman, 1991). A typical example is ability, which is both positively correlated with workers' earnings and the number of years of overeducation. Thus, the omission of variables in the wage regression could lead to an upward bias of the return to overeducation. This prediction is confirmed by earlier empirical studies: controlling for individual heterogeneity leads to a remarkable reduction in the wage effect of overeducation.

The literature suggests three possible approaches to address this limitation. First, instrument variable regression can capture the causal effect of educational mismatch on wages. However, as being discussed earlier, the choice of instruments is substantially limited due to the survey content. In addition, the test of validity of these variables remains a challenge. Thus, we cannot use this method in this study. Second, propensity score matching is another alternative. Yet, the identifying assumption is similar to that of OLS, no selection bias based on unobserved characteristics (Leuven and Oosterbeek, 2011). Therefore, this is not a reliable approach to solve omitted variable bias. Finally, fixed effect model can potentially control for time-invariant individual heterogeneity, which is discussed in detail in the next section.

5.2 An attempt to control for individual heterogeneity

In this section, we present one possible approach to control for individual heterogeneity: first-differenced estimator. Since the VHLSS has been conducted biannually with 50% sample rotation, this panel data structure may enable us to control for time-invariant and unobserved individual factors such as motivation and ability. The model is specified as in Eq. 4, in which i and t denote the person and survey year respectively; $year2016$ is the dummy for year 2016; a_i is the unobserved time-invariant factor. In this study, a_i could contain ability or motivation that is correlated with both educational attainment and wage.

$$lnWage_{it} = \alpha_0 + \delta_0 year2016 + \alpha_1 RE_{it} + \alpha_2 OE_{it} + \alpha_3 UE_{it} + \beta X_{it} + a_i + \varepsilon_{it} \quad \text{(Eq. 4)}$$

Because a_i is unobserved and constant over time, we can difference the data between two years to get the first-differenced equation (Eq. 5) which is then estimated by OLS.

$$\Delta lnWage_{it} = \delta_0 + \alpha_1 \Delta RE_{it} + \alpha_2 \Delta OE_{it} + \alpha_3 \Delta UE_{it} + \beta \Delta X_{it} + \Delta \varepsilon_{it} \quad \text{(Eq. 5)}$$

Although the first-differenced estimator has the advantage of controlling for time constant individual heterogeneity, some assumptions need to be satisfied to yield unbiased and consistent estimators. First,

the OLS estimator is imprecise if there is minor variation in the independent variable. We expect little variation in respondents' years of schooling since people are less likely to change their qualification after they enter the labor market. In addition, as the panel data covers only two years, the required years of education for an occupation may not change. Therefore, the estimation mainly relies on the variation in their job. However, this problem may still exist also because of the short span of the data set. Second, the first differencing estimator is inconsistent if the strict exogeneity assumption does not hold: $\Delta\varepsilon_{it}$ is correlated with the first differenced of the variable of interest. In this context, the assumption may not be satisfied since we expect there exist unobserved factors that affect the changing job decision and wage.

6. Empirical results

6.1 Baseline model – OLS estimations

Table 4 presents a consistent estimation result across different model specifications using the 2016 cross-sectional data. The 2014 cross-sectional data also yields a similar result (Table A-2 in the Appendix).

The first column confirms a positive relationship between education and earnings. One more year of schooling is associated with a 3.5% increase in hourly wage when socio-economic and job characteristics are held constant. The second column shows the evidence of wage penalty for overeducated and undereducated workers when required education is measured by the average years of schooling. The returns to a surplus and deficit year of education are 2%, and -2%, while the return to a year of required education is 7%. It implies that overeducated workers earn more than their colleagues, but they can earn higher if they can find a job requiring their actual educational level. Conversely, undereducated workers are paid a lower wage than their co-workers. These coefficients are slightly different if required education is measured by the mode as in the third column, which suggests that the estimation result is not sensitive to different measures of overeducation. These findings are consistent with most existing studies. For the pooled sample of six developing countries including Vietnam, Chua and Chun (2016) find the returns to over, required and undereducation to be 3.5%, 6.4% and -2.9% respectively.

The coefficients on other variables are statistically significant and have expected signs. On average, female workers earn about 18% less than male counterparts. In addition, the relationship between age and earnings is concave. Married workers receive 7% higher than other groups when other factors are held constant. The estimation result also reveals the positive ethnicity and urban gaps. Regarding job characteristics, as expected, there is a positive association between earnings and contract status. Besides, workers in foreign firm earn 16% higher than those in private and state enterprises on average.

6.2 Return to education mismatch by workers' characteristics

The return to education may vary among different groups of workers (Tsai, 2010). We first examine whether overeducation affects wage differently between men and women since female workers are paid significantly less than male counterparts on average. Table 5 shows the estimates of the return to education mismatch for male and female workers separately. The coefficients on over, required and under years of schooling in the two models are almost identical regardless of measures of required education. This result implies that there is no difference in returns to three educational components between men and women, which is in line with the meta-analysis by Groot sand Maassen Van Den Brink (2000). However, other studies such as Tsai (2010) find that women have higher return to surplus or deficit years of education than men for the US data.

Table 4: Return to over, required and under education (Eq. 3)

Log of hourly wages	(1) Base model	(2) Mean index	(3) Mode index
Years of schooling	0.035***		
	(0.001)		
Years of overeducation		0.019***	0.026***
		(0.002)	(0.002)
Years of required education		0.072***	0.057***
		(0.002)	(0.001)
Years of undereducation		-0.024***	-0.027***
		(0.002)	(0.001)
Female	-0.182***	-0.185***	-0.185***
	(0.005)	(0.005)	(0.005)
Age	0.046***	0.043***	0.044***
	(0.002)	(0.002)	(0.002)
Age squared/100	-0.052***	-0.049***	-0.050***
	(0.002)	(0.002)	(0.002)
Married	0.071***	0.069***	0.070***
	(0.007)	(0.007)	(0.007)
Majority ethnic	0.056***	0.057***	0.048***
	(0.011)	(0.011)	(0.011)
Urban	0.090***	0.074***	0.081***
	(0.006)	(0.006)	(0.006)
Contract	0.196***	0.146***	0.166***
	(0.008)	(0.008)	(0.008)
State enterprise	0.029**	-0.004	-0.000
	(0.013)	(0.013)	(0.013)
Foreign enterprise	0.164***	0.162***	0.168***
	(0.009)	(0.009)	(0.009)
_cons	1.677***	1.529***	1.711***
	(0.036)	(0.034)	(0.036)
N	38596	38596	38596
R^2	0.313	0.335	0.326

*Note: All models include six dummy variables for regions (Northern midlands and mountain areas, Red River Delta, North Central and Central Coastal area, Central Highlands, South East, and Mekong River Delta), 21 dummy variables for industries according to the Vietnam Standard Industrial Classification 2007 level 1. Standard errors are clustered by household in parentheses. * p < 0.10, ** p < 0.05, *** p < 0.01*

Second, we study whether returns to overeducation differ by age cohort by comparing the estimates of two groups: above and below 35 years old. Table 5 indicates that the return to required education of older group is almost twice as large as the corresponding return of young group. At a smaller pace, the returns to extra years of overeducation and undereducation also increase with age. For instance, the returns to overeducation for older and younger cohort are 1.1% and 2.3% respectively. Similarly, one extra year of undereducation drops wage by 1.5% and 3% for younger and older groups. This finding is consistent with the US (Tsai, 2010) and Thailand studies (Paweenawat and Vechbanyongratana, 2015) who find a higher wage penalty for younger cohort. Paweenawat and Vechbanyongratana (2015) explain this incidence by the oversupply of graduates, while Tsai (2010) shows that the evidence is in line with the career mobility theory. In the case of Vietnam, it is likely that the increase in labor supply with a relatively higher educational level results in the gap in returns to education among different age groups. On the one hand, it can be argued that while the labor supply has substantially increased, the tertiary participation rate is significantly lower than other countries in the region (Vietnam 16%, Malaysia 32%, Thailand 43% 2005). On the other hand, the quality of higher education is under the society requirement. Therefore, extra years of education of younger cohort may still receive a lower reward than that of older group. We observe the same pattern if the cutoff point is 30 or 40 years old and if workers under 25 years old, who may still acquire more education, are excluded from the sample (Table A-4 in the Appendix). However, this pattern is opposite to Estonia (Lamo and Messina, 2010) whose age pattern of overeducation is reverse.

Table 5: Heterogenous effect by individual characteristics

	Mean index			Mode index			
	Over educ.	Req educ.	Under educ.	Over educ.	Req educ.	Under educ.	N
Gender							
Male	0.020***	0.074***	-0.022***	0.028***	0.059***	-0.024***	22790
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	
Female	0.020***	0.070***	-0.025***	0.025***	0.055***	-0.033***	15806
	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	
Age cohort							
Age>=35	0.023***	0.092***	-0.028***	0.035***	0.073***	-0.033***	19598
	(0.003)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	
Age < 35	0.011***	0.049***	-0.015***	0.014***	0.038***	-0.020***	18998
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	
Contract							
Non-contract	0.006**	0.056***	-0.020***	0.015***	0.034***	-0.018***	19122
	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	
Contract	0.038***	0.085***	-0.026***	0.037***	0.078***	-0.035***	19474
	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)	(0.002)	
Urban							
Rural	0.014***	0.059***	-0.022***	0.021***	0.043***	-0.023***	23592
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	
Urban	0.028***	0.084***	-0.021***	0.032***	0.076***	-0.028***	15004
	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)	(0.002)	

*Note: All models include control variables. Standard errors are clustered by household in parentheses. * p < 0.10, ** p < 0.05, *** p < 0.01*

Next, we investigate whether returns to overeducation depend on workers' contract. Although the VHLSS does not specify the types of contract, we expect job with contract requires higher skills. In other

words, education is more beneficial for contract than non-contract jobs and overeducation may be more popular among contract workers. The estimation result in Table 5 confirms this prediction. On average, the returns to required schooling for contract workers is 3-4% higher than non-contract peers. In the same vein, having a contract is associated with higher returns to both over and undereducation. For example, for one additional year of surplus schooling, we observe a 4% and 1% increase in wages for contract and non-contract workers respectively.

Finally, we analyze the variation of return to education mismatch in the urban and rural area which is motivated by fundamental differences between the rural and urban labor markets. Table 5 shows that the return to undereducation and overeducation are almost the same for both rural and urban workers. An extra year of undereducation is correlated with a wage increase of 2%. Similarly, for each year of overeducation, wages increase by 1.5-3%. However, there is a considerable difference in return to required education as expected. For adequately educated workers, the return to a year of schooling in urban area is 8%, while this rate in rural area is 4-6%.

6.3 Within and between-individual variation

The major threat to the OLS estimators is omitted variable bias. Fixed effect model is a better alternative since it can potentially control for individual time-invariant heterogeneity. However, the identification of the model must depend on substantial within-individual variation. Previous studies suggest that the within-individual variance is considerably smaller than the between-individual number especially for the mean index (Bauer, 2002; Pecoraro, 2011). This observation is confirmed in this study as shown in Table 6. For the years of acquired education, the within standard deviation is 0.69, six times smaller than the between standard deviation, implying that the difference in years of education between individuals is much larger than that within one individual. In addition, since the within variation is near zero, workers are not likely to change their qualification over two years. In the same vein, for the required years of education, we observe a significantly lower within-individual variation than the between-variation. Notably, the gap between the two measures of standard deviation for the mode index is also higher than previous studies of Bauer (2002) and Pecoraro (2011) who only count on the mode index for their analysis. Therefore, we may not be able to identify the fixed effects.

Table 6: Large gap between the within and between-individuals variation

	Mean	Overall Std. Dev.	Between Std. Dev.	Within Std. Dev.
Log of hourly wage	3.16	0.56	0.50	0.25
Year of schooling	10.48	4.46	4.41	0.69
Mean index				
Year of surplus schooling	1.06	1.61	1.46	0.68
Year of required schooling	10.47	3.36	3.27	0.79
Year of deficit schooling	1.06	1.86	1.76	0.62
Mode index				
Year of surplus schooling	0.95	1.85	1.59	0.95
Year of required schooling	10.75	3.86	3.66	1.22
Year of deficit schooling	1.23	2.15	1.98	0.81

Note: Balanced panel data 2014-2016. Number of individuals: 10533. Number of observations: 21066

If we expect that workers, in general, are not likely to advance their education once they enter the labor market and they do not change jobs in a short period, the within-individual variation should be considerably less than the between-individual variation. For instance, in the study of Bauer (2002), the within and between standard deviation of years of required schooling for the male sample are 1.02 and 1.65 respectively. Meanwhile, these figures for years of overeducation are 1.79 and 0.41. Due to that nature, no study so far has been able to credibly capture the return to overeducation using fixed effect model. Thus, using panel methods to estimate the impact of overeducation on wages remain a challenge for future research.

6.4 Data inconsistencies and robustness check

The construction of panel data reveals some mismatches in individuals' years of schooling between two survey rounds. As can be seen from Table 7, some workers (13%) experience a reduction in their qualification while others (4%) witness an increase of more than two years of schooling. It also shows that nearly 70% of respondents do not change their qualification, while 12% of the sample experience a 1-2-year increase in their schooling. Since we do not expect a decrease or an unreasonable increase of years of education, those observations should be wrongly reported. Thus, data inconsistencies suggest that the true within-variation of the years of education could be even smaller.

Table 7: Mismatch in years of education

Change in years of schooling	No. of individuals	%
Decrease	1,416	13.44
Unchanged	7,331	69.60
Increase 1-2 years	1,316	12.49
Increase > 2 years	470	4.46
Total	10,533	100

Instead of interviewing each household member separately, the survey collects information from the most knowledgeable member. Therefore, the years of education are likely to be affected by the reporting errors. If we assume the reporting error is unintentional, the measurement error is possibly classical because they are independent of years of schooling together with respondent and family characteristics. When the independent variable suffers from classical measurement errors, OLS estimators will be biased towards zero, implying that the actual return to overeducation and required education would be higher than our previous results. The literature suggests two solutions for this issue: obtaining new data and using another measure as an instrument for this error measure. However, the data constraint does not allow us to do so. Therefore, we can test whether the OLS estimation result is robust when we attempt to correct the 2016 data based on the previous survey round.

We first re-estimate the model with 10533 observations that can be matched with the VHLSS 2014 to form the benchmark result for comparison. The estimation results are shown in column 1a and 2a of Table 8 for the mean and the mode measures of required education respectively. After that, we correct the year of schooling using the VHLSS 2014 data and re-estimate the model. For those whose years of schooling in 2016 is greater than theirs in 2014 by 1-2 years, the information is checked with whether they were at school in either 2014 or 2016. If they were not taking part in any class in two survey years, the 2016 years of education is corrected as it was in 2014. For those who experience a decrease or an increase of more

than two years in their educational level, it is uncertain which information is true given all survey information. We hence correct them as they were in 2014. The estimation results for this "corrected" sample are presented in column 1b and 2b of Table 8. Finally, we drop all observations that include illogical information, implying that the sample contains respondents who do not change their years of education or experience less than three-year increases. We re-estimate the model and the results are presented in column 1c and 2c of Table 8.

The estimation result shows that the lowest coefficients on years of overeducation occur to the second sample – the "corrected" since the modification reduces the years of schooling for almost 20% of the sample. For the third sample that is the most reliable, the absolute value of the coefficient on required and undereducation is larger than the benchmark model. Hence, it confirms that the coefficients in the benchmark model are biased towards zero as we predict. The estimated results from all three samples are quite similar and confirm the wage penalty of education mismatch. Therefore, it can be concluded that the mismatch in years of education between two survey rounds does not affect the estimation of wage effects of overeducation.

Table 8: Robustness check

Log of hourly wage	(1a) Benchmark	(1b) Corrected	(1c) Drop	(2a) Benchmark	(2b) Corrected	(2c) Drop
Year of overeducation	0.024***	0.020***	0.024***	0.035***	0.030***	0.034***
	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)	(0.003)
Year of required education	0.076***	0.076***	0.079***	0.062***	0.060***	0.066***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Year of undereducation	-0.026***	-0.026***	-0.029***	-0.028***	-0.029***	-0.031***
	(0.003)	(0.003)	(0.004)	(0.002)	(0.002)	(0.003)
N	10533	10533	7822	10533	10533	7822
R^2	0.340	0.339	0.356	0.330	0.328	0.346

*Note: a. 2016 sample, b. "Corrected" 2016 sample using 2014 data, c. 2016 sample dropping illogical observations. 1. Mean index, 2. Mode index. Standard errors in parentheses. * p < 0.10, ** p < 0.05, *** p < 0.01*

7. Discussion and Conclusion

The study confirms that overeducation, as commonly defined in the literature, also exists in Vietnam. The proportion of overeducated workers varies across different measures of required education. However, similar patterns are observed: overeducation is more popular among younger cohort for both female and male groups. This situation can be explained by several reasons. First, the young generation has better access to higher education. Since 2000, the rapid establishment of higher education institutions across the country together with the improvement of living standards allow more people to attend universities and colleges. The remarkable increase in such a short period may lead to the oversupply of graduates. Hence, job mismatch is expected to be more prevalent among younger cohorts. Second, even in the absence of oversupply, since the development of higher education is not parallel with the quality of education, the low job competencies of graduates require firms to hire overqualified employees to reduce training cost. Thus, people tend to take surplus years of education to compete for jobs according to job competition theory.

Finally, according to the career mobility theory, recent graduates are likely to accept underqualified jobs because they expect a higher chance of promotion in the future.

The OLS estimation results using Duncan and Hoffman specification are consistent with the literature, confirming the earnings gap among overeducated, adequately educated, and undereducated employees. Our estimates are consistent across different samples although there are data inconsistencies between two survey rounds. Specifically, overeducated workers receive a wage penalty for job mismatch, implying that their wages are smaller than correctly matched peers with the same level of education although they earn more than their colleagues. Conversely, undereducated workers are paid lower than their co-workers but receive a higher wage than matched workers with the same qualification. Therefore, the OLS estimation results do not confirm the human capital and job competition hypotheses, while supporting the assignment theory. Accordingly, overeducation is persistent if job allocation does not change. If we assume overeducation arises from the poor matching ability of the labor market, government's policies are needed to facilitate this process. However, the literature also suggests that the earning gaps between overeducated and adequately educated workers disappear when individuals' heterogeneity is included. Based on that result, it can be argued that overeducation could be individuals' rational choices to compensate for their lower unobservable characteristics. In this case, the government's intervention is unnecessary.

When estimating the wage return of educational mismatch by different age cohorts, we find that the return to overeducation increases with the age groups. In other words, a higher wage penalty is associated with a lower age. One explanation for this pattern is the oversupply of university and college graduates. This would imply that the government should reduce the number of places in graduate education (since the number of newly recruited students for each sector are assigned by the central government). Although this interpretation might be true for some occupations such as teachers and polices for which there have been reports of oversupply of graduates, this is not likely the case for the whole labor market because firms still struggle to find suitable candidates. Also, Vietnam has lower participation rates in tertiary education than in surrounding countries. Therefore, we are more convinced that a reduction in the quality of education resulting in younger workers having insufficient job competencies contribute to the low return to overeducation for young workers. Although we cannot test this prediction using the VHLSS data, quality improvements for higher education could be the first and foremost approach to address this issue.

There are some limitations to our study. First, we cannot interpret the effect of educational mismatch on wage as causal because of potential omitted variable bias. Individual heterogeneity could explain a large part of wage penalties according to the literature, and a fixed effect model could be a better alternative. However, given our dataset, the fixed effect model cannot be identified due to low within-individual variation. This low variation is more severe because of data inconsistencies between two surveys. Second, the measure of human capital has some weaknesses. The years of schooling captures the quantity rather than the quality of education which would better reflect the incidence of educational mismatch. Furthermore, the realized matches measures of overeducation raise concerns of mismeasurement because two different incidences of overeducation yield the same estimates of wage return (Leuven and Oosterbeek, 2011). Future studies should address these two key issues.

References

Bauer, T. K. (2002) 'Educational mismatch and wages: A panel analysis', *Economics of Education Review*, 21(3), pp. 221–229. doi: 10.1016/S0272-7757(01)00004-8.

Becker, G. S. (1964) *Human Capital: A theoretical and empirical analysis, with special reference to education*. National Bureau of Economic Research.

Büchel, F. and van Ham, M. (2003) 'Overeducation, regional labor markets, and spatial flexibility', *Journal of Urban Economics*, 53(3), pp. 482–493. doi: 10.1016/S0094-1190(03)00008-1.

Caroleo, F. E. and Pastore, F. (2017) 'Overeducation at a Glance. Determinants and Wage Effects of the Educational Mismatch Based on AlmaLaurea Data', *Social Indicators Research*, pp. 1–34. doi: 10.1007/s11205-017-1641-1.

Chua, K. and Chun, N. (2016) 'In Search of a Better Match: Qualification Mismatches in Developing Asia', *Asian Development Bank Economics Working Paper Series*, No. 476. doi: http://dx.doi.org/10.2139/ssrn.2811502.

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

Cohn, E. and Ng, Y. C. (2000) 'Incidence and wage effects of overschooling and underschooling in Hong Kong', *Economics of Education Review*, 19(2), pp. 159–168. doi: 10.1016/S0272-7757(99)00006-0.

Dang, H. A. H. and Glewwe, P. W. (2017) 'Well Begun, but Aiming Higher: A Review of Vietnam's Education Trends in the past 20 Years and Emerging Challenges', *Journal of Development Studies*, pp. 1–25. doi: 10.1080/00220388.2017.1380797.

Doan, T., Le, Q. and Tran, T. Q. (2017) 'Lost in Transition? Declining Returns to Education in Vietnam', *European Journal of Development Research*, pp. 1–22. doi: 10.1057/s41287-017-0080-9.

Duncan, G. J. and Hoffman, S. D. (1981) 'The incidence and wage effects of overeducation', *Economics of Education Review*. doi: 10.1016/0272-7757(81)90028-5.

Freeman, R. (1976) 'The overeducated american'. Academic press.

Groot, W. (1993) 'Overeducation and the returns to enterprise-related schooling', *Economics of Education Review*, 12(4), pp. 299–309. doi: 10.1016/0272-7757(93)90064-N.

Groot, W. (1996) 'The incidence of, and returns to overeducation in the UK', *Applied Economics*, 28(10), pp. 1345–1350. doi: 10.1080/000368496327895.

Groot, W. J. N. and Maassen van den Brink, H. (1997) 'Allocation and the returns to overeducation in the United Kingdom', *Education Economics*. Edited by A. S. E. RI and F. E. en Bedrijfskunde, 5, pp. 169–183.

Groot, W. and Maassen Van Den Brink, H. (2000) 'Overeducation in the labor market: A meta-analysis', *Economics of Education Review*, 19(2), pp. 149–158. doi: 10.1016/S0272-7757(99)00057-6.

GSO (2017) *Report on Labour Force Survey*.

Hartog, J. (2000) 'Over-education and earnings: Where are we, where should we go?', *Economics of Education Review*, 19(2), pp. 131–147. doi: 10.1016/S0272-7757(99)00050-3.

Hung, C. Y. (2008) 'Overeducation and undereducation in Taiwan', *Journal of Asian Economics*, 19(2), pp. 125–137. doi: 10.1016/j.asieco.2008.02.001.

Joona, P. A., Gupta, N. D. and Wadensjö, E. (2014) 'Overeducation among immigrants in Sweden: incidence, wage effects and state dependence', *IZA Journal of Migration*, 3(1). doi: 10.1186/2193-9039-3-9.

Kiker, B. F., Santos, M. C. and de Oliveira, M. M. (1997) 'Overeducation and undereducation: Evidence for Portugal', *Economics of Education Review*, 16(2), pp. 111–125. doi: 10.1016/S0272-7757(96)00040-4.

Kler, P. (2005) 'Graduate overeducation in Australia: A comparison of the mean and objective methods', *Education Economics*, 13(1), pp. 47–72. doi: 10.1080/0964529042000325207.

Korpi, T. and Tåhlin, M. (2009) 'Educational mismatch, wages, and wage growth: Overeducation in Sweden, 1974-2000', *Labour Economics*, 16(2), pp. 183–193. doi: 10.1016/j.labeco.2008.08.004.

Lamo, A. and 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), pp. 1086–1099. doi: 10.1016/j.econedurev.2010.06.002.

Leuven, E. and Oosterbeek, H. (2011) 'Overeducation and mismatch in the labor market', *Handbook of the Economics of Education*. Elsevier Amsterdam, 4, pp. 283–326.

Lindley, J. and McIntosh, S. (2009) 'A panel data analysis of the incidence and impact of over-education'. Department of Economics, University of Sheffield.

McGuinness, S. (2006) 'Overeducation in the graduate labour market', *Journal of Economic Surveys*, 20(3), p. 387–432. doi: 10.1111/j.0950-0804.2006.00284.x.

McGuinness, S. and Pouliakas, K. (2017) 'Deconstructing Theories of Overeducation in Europe: A Wage Decomposition Approach☆', in *Skill Mismatch in Labor Markets*. Emerald Publishing Limited, pp. 81–127.

Mehta, A., Felipe, J., Quising, P. and Camingue, S. (2011) 'Overeducation in developing economies: How can we test for it, and what does it mean?', *Economics of Education Review*, 30(6), pp. 1334–1347. doi: 10.1016/j.econedurev.2011.06.004.

MOET (2017) *Số liệu thống kê giáo dục đại học năm học 2016 - 2017 (Statistics on higher education 2016-2017)*. Available at: https://moet.gov.vn/thong-ke/Pages/thong-ko-giao-duc-dai-hoc.aspx?ItemID=5137 (Accessed: 3 January 2018).

Ng, Y. C. (2001) 'Overeducation and undereducation and their effect on earnings: evidence from Hong Kong, 1986–1996', *Pacific Economic Review*. Wiley Online Library, 6(3), pp. 401–418.

Nielsen, C. P. (2011) 'Immigrant over-education: Evidence from Denmark', *Journal of Population Economics*, 24(2), pp. 499–520. doi: 10.1007/s00148-009-0293-0.

Nieto, S. and Ramos, R. (2017) 'Overeducation, Skills and Wage Penalty: Evidence for Spain Using PIAAC Data', *Social Indicators Research*, 134(1), pp. 219–236. doi: 10.1007/s11205-016-1423-1.

Paweenawat, S. W. and Vechbanyongratana, J. (2015) 'Wage consequences of rapid tertiary education expansion in a developing economy: The case of Thailand', *Developing Economies*, 53(3), pp. 218–231. doi: 10.1111/deve.12078.

Pecoraro, M. (2011) *Estimating the Returns to Educational Mismatch with Panel Data: The Role of Unobserved Heterogeneity*. Available at: https://ssrn.com/abstract=2293903 or http://dx.doi.org/10.2139/ssrn.2293903.

Pouliakas, K. (2012) 'The skill mismatch challenge in Europe', *Employment and social developments in Europe*.

Quinn, M. A. and Rubb, S. (2006) 'Mexico's labor market: The importance of education-occupation matching on wages and productivity in developing countries', *Economics of Education Review*, 25(2), pp. 147–156. doi: 10.1016/j.econedurev.2005.01.003.

Rubb, S. (2003) 'Overeducation in the labor market: a comment and re-analysis of a meta-analysis', *Economics of Education review*. Elsevier, 22(6), pp. 621–629.

Sattinger, M. (1993) 'Assignment Models of the Distribution of Earnings', *Journal of Economic Literature*, 31(2), pp. 831–880. doi: 10.2307/2728516.

Sheridan, G. (2010) 'Viet Nam: Preparing the Higher Education Sector Development Project (HESDP)', *Technical Assistance Consultant's Report, Prepared for the Higher Education Sector Development Project–Developing New Model Universities (NMUs) in Vietnam, Vietnam Higher Education Sector Analysis, Project*, (4207901).

Sicherman, N. (1991) '"Overeducation" in the Labor Market', *Journal of Labor Economics*, 9(2), pp. 101–122. doi: 10.1086/298261.

Sicherman, N. and Galor, O. (1990) 'A Theory of Career Mobility', *Journal of Political Economy*, 98(1), pp. 169–192.

doi: 10.1086/261674.

Thurow, L. C. (1975) *Generating Inequality: Mechanisms of Distribu-Lion in the US Economy*. Basic Books.

Tsai, Y. (2010) 'Returns to overeducation: A longitudinal analysis of the U.S. labor market', *Economics of Education Review*, 29(4), pp. 606–617. doi: 10.1016/j.econedurev.2010.01.001.

Tsang, M. C., Rumberger, R. W. and Levin, H. M. (1991) 'The impact of surplus schooling on worker productivity', *Industrial relations: a journal of economy and society*. Wiley Online Library, 30(2), pp. 209–228.

UNESCO (2016) *Education Financing in Viet Nam, 2009-2013*.

Verdugo, R. R. and Verdugo, N. T. (1989) 'The Impact of Surplus Schooling on Earnings: Some Additional Findings', *The Journal of Human Resources*, 24(4), p. 629. doi: 10.2307/145998.

Voon, D. and Miller, P. W. (2005) 'Undereducation and overeducation in the Australian labour market', in *Economic Record*. doi: 10.1111/j.1475-4932.2005.00247.x.

Appendix A Supplement tables

Table A-1: Conversion table of years of education

Formal education		Vocational training	
Completed level	*Years of schooling*	*Completed level*	*Years of schooling*
Primary Lower secondary Upper secondary school	Actual years of schooling answered by the respondents (AE)		
		Short term vocational training	AE+1
		Long term vocational training/Professional training	AE+2
College	15	Vocational college	AE+3
University	16		
Master	18		
PhD	22		

Table A-2: OLS estimation result using the 2014 data

Log of hourly wages	(1) Base model	(2) Mean index	(3) Mode index
Years of schooling	0.039***		
	(0.001)		
Years of overeducation		0.022***	0.021***
		(0.002)	(0.002)
Years of required education		0.077***	0.074***
		(0.002)	(0.002)
Years of undereducation		-0.027***	-0.031***
		(0.002)	(0.001)
Female	-0.178***	-0.180***	-0.186***
	(0.006)	(0.005)	(0.005)
Age	0.042***	0.040***	0.040***
	(0.002)	(0.002)	(0.002)
Age squared/100	-0.048***	-0.045***	-0.045***
	(0.003)	(0.003)	(0.003)
Married	0.091***	0.087***	0.086***
	(0.007)	(0.007)	(0.007)
Majority ethnic	0.020*	0.021*	0.019
	(0.012)	(0.012)	(0.012)
Urban	0.089***	0.072***	0.076***
	(0.006)	(0.006)	(0.006)
Contract	0.182***	0.138***	0.150***
	(0.009)	(0.009)	(0.009)

Log of hourly wages	(1) Base model	(2) Mean index	(3) Mode index
State enterprise	0.065***	0.031**	0.034***
	(0.012)	(0.012)	(0.012)
Foreign enterprise	0.140***	0.137***	0.146***
	(0.010)	(0.010)	(0.010)
_cons	1.623***	1.480***	1.493***
	(0.038)	(0.038)	(0.038)
N	36827	36827	36827
R^2	0.329	0.351	0.349

*Note: All models include six dummy variables for regions (Northern midlands and mountain areas, Red River Delta, North Central and Central Coastal area, Central Highlands, South East, and Mekong River Delta), 21 dummy variables for industries according to the Vietnam Standard Industrial Classification 2007 level 1. Standard errors are clustered by household in parentheses. * p < 0.10, ** p < 0.05, *** p < 0.01*

Table A-3: Required years of education by years

Code GSO	Occupation description	2014			2015		
		N	*Mean*	*Mode*	*N*	*Mean*	*Mode*
11	Leaders in the Communist Party	52	14.5	16	60	15.0	16
15	Leaders in local People's Councils and People's Committees	277	14.4	16	269	14.8	16
16	Leaders in Mass Organizations	243	13.2	16	200	13.7	16
17	Leaders in humanitarian organizations; organizations for other particular benefits	13	14.5	16	-	-	-
18	Leaders in major organizations (groups, general corporations)	77	15.7	16	99	15.9	16
19	Leaders in small organizations (companies, businesses, and enterprises, small schools)	291	15.6	16	290	15.2	16
21	High-level experts in natural sciences and technology	500	16.1	16	543	16.0	16
22	High-level experts in healthcare	210	15.8	16	245	15.8	16
23	High-level experts in education and training	1,630	15.9	16	1,599	15.9	16
24	High-level experts in business and management	1,556	16.0	16	1,727	16.0	16
25	High-level experts in IT and communication	176	15.7	16	170	16.0	16
26	High-level experts in legal, cultural and social affairs	373	15.8	16	371	15.9	16
31	Technicians in science and technology	395	14.3	14	301	14.4	14
32	Technicians in healthcare	491	13.9	14	432	14.1	14
33	Average-level experts in business and management	574	14.4	14	647	14.3	14
34	Average-level experts in legal, cultural and social affairs	342	13.9	14	303	14.2	14

Code GSO	Occupation description	2014			2015		
		N	*Mean*	*Mode*	*N*	*Mean*	*Mode*
35	Technicians in IT and communication	128	14.5	14	98	14.6	15
36	Average-level expert in education and training	1,149	15.0	16	1,072	15.0	15
41	General officers and desk-based officers	516	14.3	16	547	14.7	16
42	Customer service staff	198	13.5	16	271	14.1	16
43	Data and input enumerators	252	12.9	16	316	12.9	16
44	Other office assistants	610	12.1	12	514	12.3	12
51	Personal service staff	629	10.2	12	755	10.3	12
52	Sales staff	1,488	11.0	12	1,748	11.1	12
53	Personal care staff	139	10.4	12	158	10.2	12
54	Security service staff	824	10.5	9	769	10.5	9
61	Skilled worker in agriculture	381	7.6	9	347	7.6	9
62	Skilled worker in forestry, fisheries and hunting	259	6.3	9	232	7.1	9
63	Worker in agriculture, fisheries, hunting and collection of farm products for self-subsidy	70	7.1	9	79	7.6	9
71	Construction-related workers (except electricians)	3,669	8.2	9	3,900	8.2	9
72	Metalsmiths, mechanics and other workers related	888	10.2	9	1,067	10.2	9
73	Handcrafters, and printing-related workers	305	9.3	9	338	8.9	9
74	Electricians and electronics workers	539	12.3	12	575	12.1	12
75	Workers in food-processing, woodwork, garment making, and other handicrafts, and other workers related	3,522	9.0	9	3,606	9.2	9
81	Operators of fixed machines and equipment	2,281	9.7	9	2,934	9.8	9
82	Machine assembling workers	286	11.4	12	451	11.4	12
83	Vehicle drivers and operators of moving equipment	1,483	10.7	13	1,727	10.7	13
91	Cleaners and domestic helps	577	7.4	9	588	7.5	9
92	Low-skilled workers in agriculture, forestry and fisheries	3,921	5.3	5	3,773	5.3	0
93	Workers in mining, construction, industry, and transport	4,465	7.4	9	4,421	7.3	9
94	Assistants in food preparation	193	9.2	9	218	8.7	9
95	Street-based and sales-related workers	385	7.7	9	362	7.6	9
96	Waste collectors and other low-skilled workers	472	7.8	9	474	8.2	9
	Total	**36,829**			**38,596**		

Table A-4: Return to education mismatch by different age cohorts

	Mean index			Mode index			
	Over educ.	Req educ.	Under educ.	Over educ.	Req educ.	Under educ.	N
Age cohort							
Age>=35	0.023***	0.092***	-0.028***	0.035***	0.073***	-0.033***	19598
	(0.003)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	
Age<35	0.011***	0.049***	-0.015***	0.014***	0.038***	-0.020***	18998
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	
Age cohort							
Age>=30	0.023***	0.087***	-0.025***	0.032***	0.069***	-0.030***	25720
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	
Age<30	0.004*	0.036***	-0.013***	0.007***	0.026***	-0.017***	12876
	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	
Age cohort							
Age>=40	0.023***	0.097***	-0.031***	0.036***	0.077***	-0.035***	14256
	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)	
Age<40	0.015***	0.058***	-0.019***	0.019***	0.045***	-0.024***	24340
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	
Age cohort							
Age>=35	0.023***	0.092***	-0.028***	0.035***	0.073***	-0.033***	19598
	(0.003)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	
25<=Age<35	0.015***	0.059***	-0.014***	0.018***	0.046***	-0.021***	12593
	(0.003)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	

*Note: All models include control variables. Standard errors are clustered by household in parentheses. * p < 0.10, ** p < 0.05, *** p < 0.01*

Appendix B Literature review

Table B-1: Summary of studies using realized matches measures

Study	Year – Country	OE (M/F)	UE	Obs	Occupation	Specification	Control variables	Return on AE	Return on RE	Return on OE	Return on UE
Groot (1993)	1983 – The Netherlands	16	21.8	1057	7		Gender, IQ, Experience	5.5		-7.4	2.6
Cohn and Khan (1995)	1985 – The US	13	12	3588	7 (professional, managerial, sales, clerical, precision, operatives, laborers)	Duncan and Hoffman	Potential experience (age-education-5), county unemployment rate, regional dummies (northeast, south, west), government sector dummy, occupational dummies, marital status, weeks unemployed in 1984, and hours worked in 1984, firm tenure, union membership, race, disability		8.4	5.9	-4.4
Groot (1996)	1991 – The UK	13/10	10/8	4126	Three-digit standard occupational classification	Duncan and Hoffman + interaction with age	Experience, experience squared, tenure, tenure squared, ten dummy variables for industry, eight dummy variables for region, dummy variables for ethnicity and marital status, and a dummy variable for gender	4.5/5.6	7.9/9.4	-2.6/-3.0	12.2/10.9
Groot and Maassen van den Brink (1997)	1991 – The UK	15/8	9/4	4606	Two-digit level = 90 occupations	Duncan and Hoffman (OLS, IV)	Experience, experience squared, days of tenure, tenure squared, eight dummy variables for industry, four dummy variables for firm size, and a dummy variable for private sector	4.5/5.6	7.7/8.9	-2.4/-3.3	5.9/4.8
Kiker, Santos and de Oliveira (1997)	1991 – Portugal	10.9/6.9 – 25.5/25.3	5.3/4.7 – 16/18.7	30336	Three-digit level	Duncan and Hoffman – Modal + Mean	Experience, experience squared, tenure, tenure squared, interaction between education and tenure/exp, monthly working hours		8.1/6.7	4.5/5.6	-5.2/-6.8

Study	Year – Country	OE (M/F)	UE	Obs	Occupation	Specification	Control variables	Return on AE	Return on RE	Return on OE	Return on UE
Cohn and Ng (2000)	1986, 1991 – Hong Kong	37/31 (1991)	28/23 (1991)	179889/ 120777 (1991)	Two-digit level	Duncan and Hoffman – **Modal**	Potential experience, interaction between education and experience, dummy variables for being married and one-digit industry codes		13/15	5/4	-4/-5
Ng (2001)	1986, 1991, 1996 – Hong Kong	13.6/ 13.7 – 27.9/ 23.4 (1996)	13.6/12.7 – 35.5/33.9 (1996)	211712/ 154864 (1996)	Two-digit level	Duncan and Hoffman – **Modal** + Mean	Experience, dummy variables for being married and one-digit industry codes		15/18	5/4	-5/-7
Bauer (2002)	1984-1998 – German	12/10	10/15	13364 obs of 1824 males and 5273 obs of 922 females	Two-digit level (- job < 10 obs)	Duncan and Hoffman – **Panel**	Experience and experience squared, tenure and tenure squared, a dummy for marital status, three dummies for firm size, two dummies for region, five dummies for industry and year dummies		10.7/12.5	9.0/5.2	-10.0/-11.5
Kler (2005)	1996 – Australia	46/38		3403/3228	Two-digit ASCO coded occupation	Duncan and Hoffman	Industries, working hours, government and other sectors, experience, gender, marital status, fields of study, mother language, indigenous Australian		11/6	8.3/7.3	
Voon and Miller (2005)	1996 – Australia (Pop Census)	15.8/ 13.6	13.7/18.5	28219/14550	44	Duncan and Hoffman	Experience and experience squared, married, government employment, oversea born, mother language		18.2/14.9	6.6/5.3	-3.2/-3.4
Quinn and Rubb (2006)	1987-1999 – Mexico	17.2	19.4	4945	(-job < 10 obs.)	Duncan and Hoffman	Exchange rate, experience, experience in US, marital status, region dummies, state and year dummies	6.3	8.5	4.3	-3.0

Study	Year – Country	OE (M/F)	UE	Obs	Occupation	Specification	Control variables	Return on AE	Return on RE	Return on OE	Return on UE
Hung (2008)	1997-2002 – Taiwan	17.3	14.1	1606	9	Duncan and Hoffman	Experience, tenure, male, married		11.1	6.5	-5.8
Tsai (2010)	1979-2005 – US	22	9	83,449 obs. of 14,611 individuals	Two-digit level (-job < 10 obs.)	Duncan and Hoffman	Tenure, tenure squared, experience, and experience squared. Workers' age, race, sex, marital status, and number of children under age 18, disability		11.4	-3.8	5.7
Nielsen (2011)	1995-2002 – Denmark	16.3		43702	30	Duncan and Hoffman – Modified RM app.	Age, age squared/100, experience in Denmark, experience squared/100, number of children, marital status and year dummies. Immigrant equations include YSM (years since migration), YSM2 and ethnicity dummies		7.7	2.3	-1
Joona, Gupta and Wadensjö (2014)	2001-2008 – Sweden	11.9/ 12.4	16.3/1 1.2	5M	113 (-100 workers, military personnel, self-employed)	Duncan and Hoffman	Age, age squared, marital status, having young children, years since migration (for immigrants), sector (five categories), municipality, birth region for immigrants (six categories) and year dummies		7.8/6.4	6/4.7	-3.3/-3.1
Nieto and Ramos (2017)	2012 – Spain	14.1	17.1	1928		Duncan and Hoffman	Variables related to workers' human capital: years of education (derived from levels of education), experience, experience squared, non-formal education, and 10 plausible values test scores in literacy. Gender, age, nationality, type of contract (full-time/part-time), contact term, sector, economic activity and 17 regions.		9.8	7.2	-8.0