WAGE PENALTY FOR OVEREDUCATION AMONG UKRAINIAN YOUTHS: DOES IT EXIST?

by

Hanna Onyshchenko

A thesis submitted in partial fulfillment of
the requirements for the degree of

MA in Economic Analysis.

Kyiv School of Economics

2016

Thesis Supervisor: Professor Tom Coupe

Approved by _____
Head of the KSE Defense Committee, Professor Tymofiy Mylovanov

_____

_____

_____

Date _____

Kyiv School of Economics

Abstract

WAGE PENALTY FOR OVEREDUCATION AMONG UKRAINIAN YOUTHS: DOES IT EXIST?

by Hanna Onyshchenko

Thesis Supervisor: Professor Tom Coupe

This thesis examines the existence of a wage penalty for overeducation among Ukrainian youths (from 15 to 29 years). Overeducation refers to those individuals who claim that their university degree was neither required by law, nor useful to access their current job. In order to conduct such a research, we follow the methodology suggested by Nieto (2014), who estimated the overeducation, skills and wage penalty in Spain. We use the mean-approach to estimate the required years of education for job occupations and compare the wages of the overeducated individuals with their matched peers. Our results obtained from the two-step Heckman specification model found out insignificant positive effect on wage associated with overeducation, however suggest the substantial negative effect of undereducation. This means that the undereducated individuals get a lower wage compared to the individuals with the same degree employed at the work with required years of education matching to attained years of education. We also found out that return on each required year of education is insignificant comparing to return on each attained year of education, this points out the existence of educational mismatch.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

The author wishes to express gratitude to the thesis adviser, Prof. Tom Coupe for the valuable insights and suggestions that helped during the research; to Prof. Volodymyr Vakhitov for the help with data finding; to Prof. Olga Kupets for the significant contribution to this work through a number of consultations and propositions. The author also would like to thank all research workshop professors for their insightful comments.

I express great appreciation to Inna Lazutina for the help right when I needed help most.

Finally, I would like to thank my family for the constant support and motivation to do my best during the thesis development.

GLOSSARY

Educational-job mismatch refers to a situation of imbalance in which the level or type of education atteined does not correspond to labor market needs, and it can be a surplus as well as a lack of knowledge, abilities and competences**.**

Overeducation refers to those individuals who claim that their university degree was neither required by law, nor useful to access their current job.

Undereducation refers to those individuals who hold their job for which they are not educated enough, applicants may be hired based on the social network rather than their bilities and education.

Chapter 1

INTRODUCTION

Ukraine is the country with one of the highest share of people with tertiary education (Ukraine ranks 14th out of 140 countries in the 2015/2016 Global Competitiveness Index). At the same time, however, Ukraine has one of the lowest level of quality of tertiary education in the European region (46th), and even lower quality of job-training (74th). Average labor productivity and per capita income in Ukraine are persistently among the lowest in the region, and disparately below the level of peers across advanced economies with similar human capital facilities (Brown, 2004). There are several reasons for that, but skill and education-job mismatch is likely to be one of the more important reasons.

Since Ukraine's independence the number of schools with III and IV level of accreditation have more than doubled (from 149 to 325, Figure 1), and number of vocational schools decreased with the same volume (from 742 to 371, Figure 1). This process was accompanied with weak quality control of graduates' knowledge. UNESCO[1] in 2008 pointed out that Ukraine was not a part of any international learning assessments, and did not complete a national assessment of students' learning. Therefore, the proxy indicators are used to measure the quality of education. In the report, there was made an assumption, that lack of reliable assessment might lead to recession in educational quality, since there was no opportunity to compare the results at local, regional and national levels and, moreover, international level. Eventually, it was difficult to evaluate the national educational standards as well as to develop new programs.

[1] Education for All Global Monitoring Report, 2008 (UNESCO)

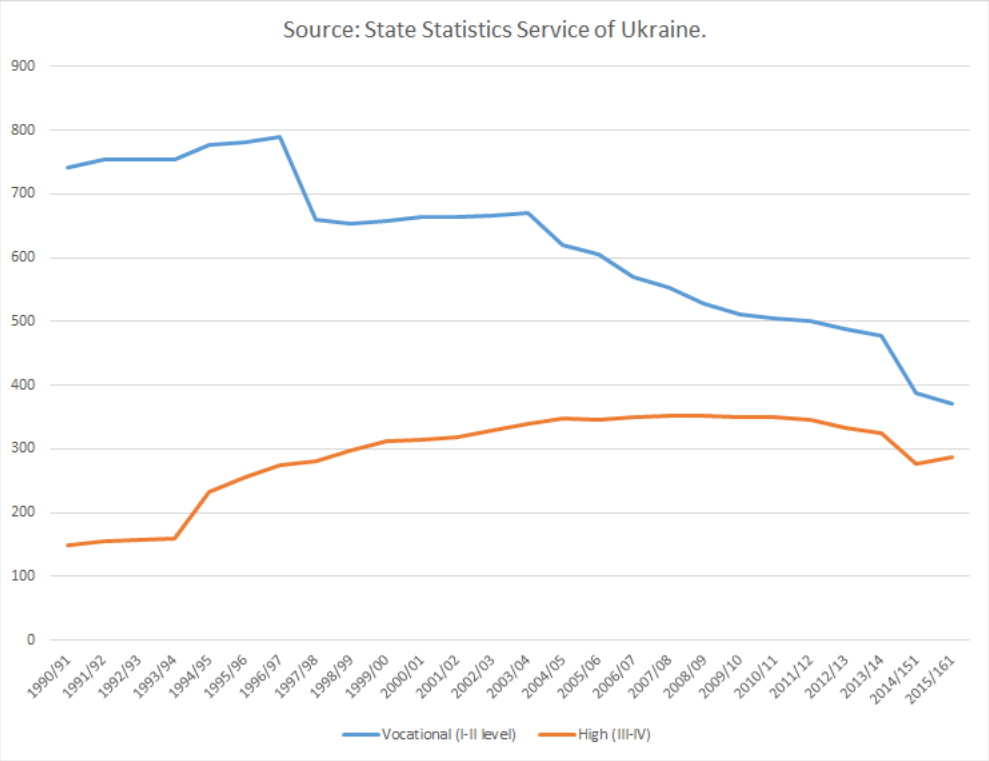

Figure 1: Number of schools in Ukraine (vocational vs high schools)

Indeed, out-dated and developed during Soviet times educational programs currently are often named as major problems of current Ukrainian higher education (KyivPost "The paradoxes of post-Soviet education" – Serhiy Kvit , Jun. 01, 2016).

Together with wide incidence of corruption in higher education (Osipian, 2015) the abundance of schools with III and IV level with uncontrolled educational quality probably cause to the oversupply of graduates with higher degree on the labor markets. This in turns may lead to lower wages and labor market distortion. When workers with higher education crowd out their peers with less years of education and occupy their jobs, in other words, so-called overeducation problem may raise, years of schooling required for the job is lower than the years of schooling completed. However, the situation may lead to even worse

consequences, Early School Leavers survey[2] conducted in 2010, noticed that "The school diploma does not provide the chance to find a job. They are trying to introduce 12th grade now, maybe that will help. (male, 17)."

In 1999 in the ICPS newsletter[3] the perspectives of overeducation in Ukraine were pointed out for the first time. Again this question was raised by Kupets (2015). Her recent studies revealed that "the education-job mismatch in Ukraine is high and quite persistent, with 39.7% of all employed people aged 15–70 years being overeducated for their jobs and 4% being undereducated in 2013." Kupets also noted the high incidence of overeducation among young university graduates (42.3% in 2004 and 39.6% in 2013).

Based on this, we may conclude that there is great discoordination between educational choices, real labor market needs and final occupational choices.

It is worth noting that education is fully provided by the government in Ukraine, Ukrainian spending on education are persistently high (from 5 to 6 percent of GDP on average) compared to Poland (3.5 percent) and USA (4.5 percent)[4]. More specifically, only this year government plan to spend 6 billion of hryvnia on vocational education and 20 billion on higher education, and this is a tremendous amounts for country with a GDP on the level of 1979 billion of hryvnia.[5]

The government should know whether these investments pay off. Especially from the point of view of ongoing educational reform, which was started in July 2014. This reform highlights the improvement of management efficiency in financing education as one of the top priorities. Under the macroeconomics prospective rate

[2] UNICEF Survey of Attitudes of Early School Leavers in Ukraine, 2010.
[3] Educational, U. N. (1999). International Centre for Policy Studies newsletter. Studies.
[4] Our own calculations based on WorldBank data.
[5] State Statistics Service of Ukraine.

of return on education is one of the leading indicators for overinvestment in education.

This MA thesis provides first available evidence of the impact of under-, overeducation on wages of youth people in Ukraine. We re-estimate the return on education based on newly available data from School to Work Transition Survey, conducted by International Labor organization, and compare it with the return on required education. Since in Ukraine there is no national standards of required education to occupy the job, we constructed it using the statistical approach suggested by Nieto (2014). We also compare the personal attitude of individuals about the relevance of their education with the relevance based on statistically determined required years of education. Finally, we find out the negative effect on wage associated with overeducation. The Heckit correction was used to control the possible sample selection bias, which arises from observing the wage of only employed people.

The thesis has the following structure: the Chapter 2 provides an overview of the literature about theoretical models behind the education mismatch and empirical approaches to applied to measurement the effect of under-, overeducatio. Chapter 3 describes the methodology of the research and possible limitations of the model. Chapter 4 focuses on the data description, main variables construction, and presents the analysis of the data. The Chapter 5 represents the empirical results of the research and discussion concerning advantages of the chosen methodological estimation method. Finally, the Chapter 6 follows with the conclusion and general policy recommendation together with considering the further direction of the research.

Chapter 2

LITERATURE REVIEW

At the beginning, we will present the origins of the overeducation problem, thereafter major theories behind the overeducation and empirical studies will be introduced. Finally, literature about the incidence of overeducation in Ukraine follows.

2.1. The emergence of the overeducation problem

In the 1970s, Freeman (1975,1976) and Berg (1970) argued that America's educational system produced too many college and higher graduates, which resulted in reduction of returns to education. Freeman, in particular, forecasted that this trend would continue in long-run. This strong statement turned a lot of social attention and laid the foundation for further discussion about overeducation's effect on wages.

A while later, Smith and Welch (1978) refuted Freeman, they found out that the experience rather than the educational level had much more influence the wage. Authors used the same data set as Freemen but expanded the time period by two years before the period covered and two years after (1965-1978). In order to eliminate the experience effect, they instead of comparing earnings of high school and graduates within the same age group, constructed the ratio with "new entrance" group of people, refers to fresh graduates entered the labor market. This helped them to look at the poor effect of education, consequently Smith and Welch reported increase in relative earnings of college graduates.

Later in 1981, Duncan and Hoffman published their first article, which became a fundamental work and started new subfield in the Labor Market theory: economics of overeducation. They conducted their research based on the individual level data, which was a new thing in the topic by itself, and compare three groups of workers, those who finished in the job corresponding to their education and that required the higher or lower level of education. Moreover, Duncan and Hoffman presented a new specification of the wage equation, which allowed to estimate separately the effect of the adequate number of years required for job, years of overeducation, and years of undereducation:

$$\ln w_i = \delta_r S_i^r + \delta_o S_i^o + \delta_u S_i^u + x_i'\beta + \varepsilon_i \qquad (1)$$

where w_i denotes the wage, S^r is years of education required for the job, S^o years of overeducation, and S^u years of undereducation; x_i' is a vector of control variables including experience and experience squared; $\delta_r, \delta_0,$ and δ_u are the returns to required years of education, years of overeducation, and years of underducation, respectively.

Freeman's inference goes in line with the neoclassical macroeconomics theory. Increasing supply of well-educated workers pushes their wages down, and at the same time decreases college the wage premium. The algorithm of the process could be the following: firms adopt their activities to the conditions, when cheap and affordable labor force is available. On the other hand, college graduates have to compete for the narrow amount of job places, which are adequate to their education, and consequently required wage level decreases.

Duncan and Hoffman present another view on how labor market works. They insisted that overeducation became the long-run problem only if there would be limited response of demand side on changes in supply, this means that change in

supply of workers with different level of education imply small effect on the skill composition of the labor market. Specifically, fixed pattern of skills needed for production technology implies no development in the production technology, which causes undersupply of job places for workers with above the average skills. Finally, based on the assumption that jobs are characterized by fixed productivity levels and fixed wages, Duncan and Hoffman marked that a person occupying the job below her skill level yields and produces similar to person with less education at the same job place, which is by itself the wasting of resources.

Later many scientists re-estimate Duncan and Hoffman's wage equation. For example, McGuinness (2007) pointed out that "[O]vereducation incurs significant wage costs on the individual and productivity costs on the economy that may well rise if a higher education participation continues to expand without corresponding increase in the number of graduate jobs".

Even though most scientists consent about existence of negative causal effect in wage-education relationships, there is no common point how resolve the problem. Some researchers propose to reduce the amount of schooling and resolve problem of overeducation in this way, why another propose to appoint overeducated workers to more demanding works. But it's still uncertain who will erect this more-demanding working places and who occupies them.

2.2 Major theories of overeducation

The general theory around overeducation ranges between two different theoretical frameworks: the human capital theory and the job competition model (Sloane 2003, McGuinnes 2006, Leuven and Osterbeek 2011).

Duncan and Hoffman's wage equation extends from the main human capital theory equation, Mincerian wage equation. In particular, it designates the Mincerian wage equation as:

$$\ln w_i = \delta_a S_i^a + x_i'\beta + \epsilon_i \qquad (2).$$

All the variables correspond to aforementioned variables of Duncan and Hoffman's wage equation. This model by itself is rejected by the data. This could be used by opponents of the model as the evidence of its inconsistency. However, improved by Sicherman and Galor (1990) model restores this human capital theory, which further evolved as career mobility theory. Based on the specific assumptions, the model suggests that a worker with inherent abilities should choose to begin his career at the job place with lower skill requirements in order to be promoted faster than peers.Furthmore, Sicherman (1990) regress the indicator of upward mobility on indicators of inadequate schooling (under-, overeducation), controlling for attained schooling, experience and others. As a result, he finds out the positive correlation for over-, underschooling indicators and negative for attained education. In addition, Sicherman detects that undereducated worker has higher chances to be promoted comparing to peers with similar level of schooling, which occupy job place corresponds to their education (the capacity of the effect is much more bigger than that of the overshooling dummy). It is suggested to explain this disconcerting fact by the correlation underschooling indicator with unobserved ability.

Another fundamental theory behind the overeducation problem is the assignment theory (Sattinger, 1993). Developed by Spence (1973) it is based on the assumption that return on education depends both on the person's human capital and the match to the job. Following this theory, workers' productivity is restricted by their

job characteristics. Consequently, it may be the case when overeducated worker underuses her skill and as a result becomes less productive.

The assignment theory is empirically checked. For this purposes personal workers' responses about whether they think that their skills fit well to their job were taken as a measure of skill mismatch. There are a lot of pieces of empirical research that followed the specification made by Verdugo and Verdugo (1989), who included dummy variables for both educational and skill mismatch in their studies (Green and McIntosh, 2007; Sánchez-Sánchez and McGuiness, 2013; Mavromaras et al. 2013). It is revealed that both overeducation and overskilling has some negative and statistically significant effect on earnings within the group of people with the same level of education, also the capacity of the overeducation effect being much higher comparing to the overskilling effect. Following the results obtained it could be concluded that wage penalty of overeducation is not accounted thefor by waste of workers' skills, which means that the assignment theory is not confirmed by the data.

2.3 Empirical studies

Regretfully, data describing the internal workers' abilities and skill level is very limited, meanwhile scientists developed many different empirical techniques in order to get rid the individual skill heterogeneity from the wage equation estimation.

Usage of panel data sets with the aim of considering all unobserved personal fixed effects (Korpi and Tåhlin, 2009; Tsai, 2010) which is one of the ways to resolve theproblem of heterogeneity. Following this approach, they reveal that the wage penalty phenomenon corresponding to overeducation drops drastically and may even fade out when fixed effect is controlled, suggesting that the problem of overeducation somehow is caused by the unwitnessed personal ability.

In place of longitudinal data approach, Chevalier (2003) turns his attention to cross-sectional data. He takes the difference between estimated and observed earnings as the proxy to unobserved worker's productivity, in wake of control for unobserved heterogeneity, wage penalty for overeducation was a little bit reduced. Following this approach, Chevalier and Lindley (2009) get the similar results. They predicted unobserved internal ability as the residual from a first-job equation, including all workers' observed characteristics. Authors claim that obtained residual might be used as a proxy for all time-invariant unobserved characteristics.

Furthermore, Chevalier (2003) develop a new method to link overeducation and worker's personal job satisfaction. He introduces two categories of overeducationbased on worker's personal attitude to their job place, such as 'apparent' overeducation referred to satisfied individuals, and 'genuine' overeducation composed by dissatisfied individuals. The results revealed that the 'genuine' group were forced to pay significantly large wage penalty compare to their 'apparent' peers.

Many other scientists apply the above approach (Green and Zhu, 2010; Korpi and Tåhlin, 2009; Levels et al.,2013) and all of them pointe out statistically significant effect of overeducation on wages while control for skill level.

Taking all previously mentioned into consideration, empirical studies do not appear to back the assignment theory up, because there is weak relationship between educational and skill mismatch.

Outside the theoretical discussion about the nature of wage penalty for overeducation, the problem of estimation accuracy was raised. Many scientists argued that simple OLS estimates might under/overestimate the wage penalty related to overeducation. There are three potential sources of bias associated with above problem: a) endogeneity; b) sample selection; c) measurement error.

As we mentioned earlier, the roblem of endogeneity refers to the problem of unobserved characteristics, such as preferences, motivation and skill level. Theproxy variables might address this issue Chevalier (2003), but what is the qualitative effect of endogeneity? Indeed, in case of controlling for unobserved skills and preferences, overeducation may cause greater wage penalty.

The measurement error has the opposite effect, during self-assessment individuals tend to overestimate their abilities, and report overeduction/ overskilling when actually they are not. In the other words, wage penalty should be lower than suggested in case of measurement error.

Things are much more complicated with sample selection bias, which potentially can generate ambiguous effect Nicaise (2001). Depending on the theoretical model applied, sample selection bias may be presented in different ways, as in the job assignment and the human capital model, the paramount effect of educational mismatch is high probability of unemployment (long unemployment time spell) and only aftermath wage penalty. After controlling for sample selection bias associated with unemployment, wage penalty for educational mismatch might be a lot higher. From the search theory point of view, people with substantial educational level prefer to wait for decent job offer, and they are more likely will get an adequate wage. Consequently, the wage penalty effect tends to be lower.

By far the majority of authors report that once controlling for endogeneity and sample selection; the job competition, job assignment and human capital models are supported by evidence associated with wage penalty for overeducation. For example, Cutillo and Di Pietro (2006) using the data of ISTAT survey on Italian graduates in 1998 revealed increased the wage penalty up to 40% in case of control for both endogeneity and sample selection bias.

2.3 Overeducation in Ukraine

The question of wage penalty for overducation in Ukraine has not been considered explicitly yet. Olga Kupets (2016) investigated the dynamics of educational-job mismatch incidence in Ukraine over the last 10 years and the determinants of under-, overeducation at the regional level. She pointed out the presence of "*hysteris of habitus*" effect, when people perceive higher education as a guarantee of better living conditions. This phenomenon leads to the overeducation appearance. At the same time, the author took note about job polarization, the employment structure in Ukraine is severely skewed toward elementary occupations, which composes 23% of total employment in 2013 (based in Labor Force Survey). Kupets (2016) calls attention to deterioration of middle-level job places and increasing the number of graduates with tertiary education, which means strengthening competition in low-skilled labor market, and decreasing probability for graduates to get decent jobs. Based on Labor Force Survey, she reveals rather high and stable level of education-job mismatch, 39,7% of employed people under age of 15-70 reported overeducation, meanwhile only 4% claimes themselves as undereducated in 2013. Moreover, Kupets (2016) claimed the much more incidence of the phenomena in the regions. Libanova et al. (2014) using recent School-to-Work transition survey (2013) reports similar results. To our knowledge, no extensive analysis of wage penalty phenomena focusing on the Ukraine case made, as well as no analysis was made at Ukrainian STWS. Current work tries to close this gap and estimate the quantitative effect of overeducation on wage on both waves of School-to-Work Transition Survey (2013, 2015).

Based on the presented theoretical and empirical studies, we can hypothesize that the significant negative effect of overeducation on wages among youth in Ukraine. In addition, a few important issues should be kept in mind during the empirical analysis of the wage penalty. Firstly, the endogeneity could be a source of biased estimates. Secondly, problem of sample selection bias needs to be addressed.

C h a p t e r 3

METHODOLOGY

In order to check the phenomena of wage penalty for education mismatch, we make use of the nonparametric approach, which does not require any specific assumption about the distribution of data observed. We compare earnings distribution of matched and mismatched education graduates, based on answers of self-assessment. For this purpose, kernel density estimators were built and non-parametric density equality test was used to prove the difference between two density functions.

Following the Silverman (1998) kernel density estimator is:

$$\widehat{f}\,(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right) \qquad (3),$$

where n is the sample size, h is smoothing parameter (bandwidth), x – mean and K(u) is Epanechnikov Kernel[6].We used Kolmogorov-Smirnov test of the equality of continuous, one-dimensional probability distributions to strictly compare the distribution. The null hypothesis of the test is the equality of two univariate density functions denoted by f_1 and f_2. Since these functions are not given explicitly, kernel estimators give their estimation and next statistics is derived:

$$S_p = \frac{1}{2}\int (f_1^{\frac{1}{2}} - f_2^{\frac{1}{2}})^2\, dx \qquad (4).$$

[6] $K(u) = \frac{3}{4}(1-u^2)1_{\{|u|\leq 1\}}$ is Epanechnikov kernel and satisfies the assumptions:

1. Non-negative real-valued integral function;
2. $\int_{-\infty}^{+\infty} K(u)\, du = 1$;
3. $K(-u) = K(u)$ for all values of u.

If the assumption that density functions of earnings distribution are different holds, it makes sense to measure the quantitative effect of education mismatch on wages.

In an effort to estimate the effect, we use an alternative version of a traditional wage equation (Mincer, 1974) proposed in literature: the ORU specification (Duncan and Hoffman, 1981), the Verdugo and Verdugo (1989) specification.

It is to be recalled that the traditional Mincerian wage equation is:

$$\ln w_i = \alpha + \delta_a S_i^a + \beta x_i' + \epsilon_i \qquad (5a),$$

where $\ln w_i$ is natural logarithm of the monthly wage of worker i; S_i^a denotes the number of years of attained education; x_i' is a vector of control variables including dummy for year, experience and experience squared, which describes personal, professional and regional characteristics of worker. Moreover, we add dummy variable that takes value of 1 if the worker has participated in any kind of trainings over the last 12 months before the survey and 0 otherwise. As usual, ϵ_i stands for error term with zero mean and constant variance.

According to the traditional wage model, formal education is a proxy of individual's human capital. At the same time, the problem of endogeneity, in particular omitting individual's ability and skills level, is addressed. In the light of absence of the reliable measure for worker's skill level, we will try to control for ability variable in all models.

We add the proxy to the ability variable instead the omitted one:

$$\ln w_i = \alpha + \delta_a S_i^a + \gamma_w will_i + \beta x_i' + \epsilon_i \qquad (5b),$$

where $will_i$ is a dummy variance that takes value of 1 if the worker is willing to continue further education or participate in vocational trainings and 0 otherwise. Fouarge et al. (2010) suggested that low-educated workers should be less interested in further training and education, because of personality traits (locus of control, exam anxiety, and openness to experience). In addition, Harvey (2000) suggests that willingness to learn and continue learning are as important triggers of high internal ability as intellect, ability to find things out and willingness to take a risk.We assume that a strong desire to professional development may be proxy to internal educational ability of the worker.

One of the most common specifications of Mincerian equation is ORU (Over-Required_Under-educated) invented by Duncan and Hoffman (1981). They distinguish general years of attained education variable onto three pieces: S^r - years of required education, S^o - years of overeducation and S^u - years of undereducation. The transformation could be denoted as: $S^a = S^r + S^o - S^u$. According to this logic, rest of variables are defined by the rule:

- $S^o = S^a - S^r$ if the worker is overeducated and 0 otherwise, and
- $S^u = S^r - S^a$ if the worker is undereducated and 0 if otherwise.

Finally, ORU specification the form of the following equation:

$$\ln w_i = \alpha + \beta_r S^r + \beta_o S^o + \beta_u S^u + \gamma x_i' + \epsilon_i \qquad (6a).$$

All other variables correspond to the above traditional Mincerian wage equation (1a). Effect of under-, overeducation on the worker's wage is made through comparison with their well-matched peers. Literature suggests next relation between coefficient $\beta_r > \beta_o > |\beta_u|$.

In the same way as before, we include dummy variable, which explains worker's willingness to participate in the further education and trainings.

$$\ln w_i = \alpha + \beta_r S^r + \beta_o S^o + \beta_u S^u + \delta_w will_i + \gamma x_i' + \epsilon_i \qquad (6b),$$

$will_i$ is the same as in equation (1b). Chevalier and Lindley (2009) marked that $\beta_o = \beta_u = 0$ if omitted abilities fully capture the wage's effects of educational mismatch.

Another way to count the effect of education mismatch was introduced by Verdugo and Verdugo (1989), hereafter V&V. They propose instead of discrete values of under-, overeducation out the dummy variables in Mincerian wage equation. Specifically, OE is a dummy variable for the worker to be overeducated, and similarly UE is a dummy for the worker to be undereducated.

ORU specification is modified into:

$$\ln w_i = \alpha + \beta_a S^a + \beta_o OE + \beta_u UE + \gamma x_i' + \epsilon_i \qquad (7a),$$

other variables follow the initial equation. Using the same logic, coefficients near dummy variables describe the average effect of lack or excess years of education comparing workers with their well-matched peers. Typically, $\beta_0 < 0$ and $\beta_u > 0$, which means that undereducated workers, contrary to overeducated, benefit from wage premium compare to their peers with the adequate educational level.

We, as usual, deeper the model by adding the omitted ability variable to control for heterogeneity:

$$\ln w_i = \alpha + \beta_a S^a + \beta_o OE + \beta_u UE + \delta_w will_i + \gamma x_i' + \epsilon_i \qquad (7b).$$

All the variables are defined above in the paper. If β_0 and β_u turn out to be statistically insignificant, then wage is fully determined by attained years of education and willingness to study.

Interestingly, that Nieto (2014) find out no evidences showing that worker's skills heterogeneity fully describes the effect of educational mismatch in Spain. She reported permanent statistically significant effect of overeducation on wage.

Though unobserved heterogeneity problem is still hard to resolve, problem of selection bias may be addressed with help of Heckman two step specification (Heckman, 1979). Recall, that for wage equation we use only the employed individuals, but education missmatch arise at first time as a problem to find a job, and only after that wage penalty takes place. At the initial stage we estimate the probability to be employed using a probit equation.[7] Hereupon, we include obtained probit estimation onto wage equation as explanatory variable (control for selection bias). If the coefficient near this control variable will be statistically significant, sample selection bias will take place, that is the omitted information about probability to be employed distorts the previous results. We apply Heckman two step specification to all the modification of Mincerian wage equation for the purpose to find out the most accurate model. We choose Heckman two step specification because maximum-likelihood estimation may be time consuming with large datasets.

[7]We include marital status, number of children in the family into probit model as well as traditional variables that also determine the wage, i.e. experience, attained aducation, speciality etc.

Chapter 4

DATA OVERVIEW

4.1 Data source and variables

For this paper we used data from the School to Work Transition Surveys, I and II waves, conducted in 2013 and 2015 by the International Labor organization. The School to Work Transition Survey gathered the information about the Ukrainian youth at the age 15-29 years and described their personal characteristics and transition paths from school to work, and it also includes "longitudinal information about transition within the labor market" (ILO website). The pooled data constructed from the SWTS (2013, 2015) consists of 6728 observation of Ukrainian youth. Questionnaires of both waves of survey are almost the same, which helped us easily build the dataset. Following the methodology of survey, all the respondents were randomly assigned. After eliminating respondents with missing values and non-finished education, the sample was restricted to 1775 employed and unemployed graduates.

Wage, required years of education, highest level of attained education and triggers of under-, overeducation were chosen as the central variables for analysis. For the purposes of the paper, we used the following methods to construct these variables appropriately.

4.1.1. Wage

We construct the variable, which indicates the part time position for those individuals, who reported hours period less than 160 hours per month (40 hours per week are required for full-time position in Ukraine) and wage period less than 1 month.

Following the wage distribution (Table 1), we have decided to top-code the wage with unreasonable high or law values: replace the wage values after 99-th percentile by value 7001 and below 2-th percentile by 220 (104 out of 2006 values were changed).

Table 1: Wage distribution by percentiles

variable	N	p1	p5	p10	p50	p90	p95	p99	min	max
wage	1707	200	1000	1200	2000	4000	5000	8000	50	1707

4.1.2 Required years of education

In contrast to USA [8]in Ukraine there is no standardized educational requirement to job occupation. This means that individual can pretend to any job with any level of educations. Required level of education is one of the central variable, which we use for analysis. It is needed to construct the objective under-, overeducation triggers and we have included it into ORU and V&V regressions as exogenous variable.

In order to construct the required years of education variable, we generated the variable profession, which is the unique identification of job position at each particular industry. Than we took the average years of obtained years of schooling

[8] https://en.wikipedia.org/wiki/Education_in_the_United_States

for each profession and defined them as years of required education. This method was used by Nieto (2014), she used mean to determine the required years of education, and after that compare the results.

4.1.3 Highest level of obtained education

We have converted the highest level of formal education attained by individual from the categorical variable form (like Master degree) into actual years of schooling using the definition presented in the law of Ukraine "About the higher education". We repeated the same transformation to obtain the father's and mother's years of schooling, but instead of current classification we used the USSR standards.

4.1.4 Under-, overeducation triggers

In addition to the existing triggers of under-, overeducation constructed based on the respondents' self-assessment, we also developed the triggers based on statistical information. Individual is considered to be:

- Undereducated if $highestlevel\_comp < req\_year - SD$, where highestlevel_comp – years of obtained education, req_year - required level of education for occupied job and SD is the standard deviation of years of obtained education for this particular occupied job; and

- Overeducated if $highestlevel\_comp > req\_year + SD$.

Years of under- and overeducation were constructed according the methodology explained in chapter "Methodology" of this paper. In addition, we defined triggers of under-, overeducation based on statistics.

For our analysis we used all the data available and put no restrictions on the time from graduation, that is our results may mix the determinants either permanent or transient under-, overeducation. Unfortunately, now there is no decent data on Ukrainian 5 years after graduates, which allows us to separate these two phenomena.

4.1.5. Experience

We defined experience as the difference between the year, when survey was conducted, and year of first employment experience (or when the person started looking for a job).

4.2. Data description

The SWTS contained a lot of questions about the respondents' individual and employment characteristics, sample of 2015 year also include questions about the choice of field of knowledge to study at the university. Individual characteristics include questions about the sex, age, language spoken, marital status, household financial situation, family background. Employment questions are represented by monthly wage, firm size, type of contract, industry specification and job position. The most valuable information is about the educational level and job occupation; it helps us to proceed the variable construction process mentioned in the section 4.1 of this Chapter.

Our sample is consisting of equal parts of male and females, among them urban population is prevailing (Table 2). Since SWTS considers youth only 44 percent of sample are employed people, and roughly speaking, only half of individuals has finished educational process. You can find extended descriptive statistics in the Table 5 (Appendix 1).

Table 2: Short descriptive statistics

Variable	N	mean	sd	min	max
d2015	1775	0.1487	0	0.3559	0
language	1763	0.4583	0	0.4984	0
age	1775	24.2580	25	3.3837	15
female	1775	0.5544	1	0.4971	0
city	1775	0.7352	1	0.4413	0
attained education, y	1775	14.5414	14	2.3384	9
employed	1775	0.8394	1	0.3672	0
exper, y	1775	5.1882	5	3.3365	0
match	1455	0.7911	1	0.4066	0
matchS	1775	0.6625	1	0.4729	0
required years of education	1775	14.5518	14.06	1.1412	12.68
OE	1755	0.0632	0	0.2434	0
UE	1755	0.1120	0	0.3155	0
OES	1775	0.1521	0	0.3592	0
UES	1775	0.1854	0	0.3886	0
wage	804	2303.9730	2000	1416.158	100
wage_t	804	2283.3130	2000	1294.563	200

In the focus of our interest are variables Match and MatchS, because they explain the matching between the job occupied and attained educational level. Near 80 percentage of employed people reported that their education corresponds to their job, however, the matching rate based in statistics significantly reduce this matching rate to the 66 percent. This fact goes in line with world findings (Chevalier, 2003), people usually hesitate to tell that they are undereducated because they afraid to harm their ego. The same story is with overeducation, report that you are overeducated means that you waste your time and money or you are not smart enough to find a decent job. Figure 2 depicts the average worker's perception about the attained education by educational level (individual sample weights considered) and its counterpart average statistical matching level.

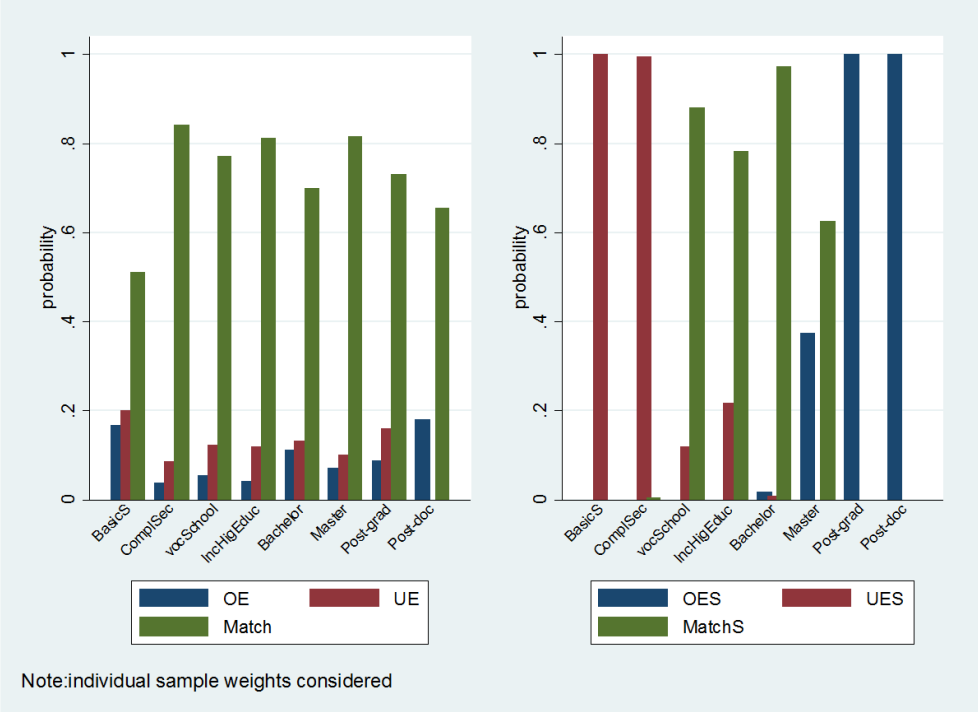

Figure 2: Relevance of attained education (by educational level)

As can be seen from the part of the Figure 2 based on the individuals' self-assessment, education matches to the job occupied at the most cases, only post-doctoral graduates reported the significant overeducation and workers with complete basic school education noted undereducation. Interestingly, that there are workers with any level of education, who considered themselves as under- or overeducated. This means that educational mismatch exists at any educational level. At the same time, statistically estimated indicators of mismatch (right part of the Figure 2) suggest that bachelor degree is the optimal educational level in sense of "fitness" to job. Statistics shows that there are is the trend the following trend: succeeding degrees after bachelor are resulted in overeducation with the probability more than 40 percent; and school-level degrees produce the undereducation with the probability more then 80 percent.

Educational mismatch distribution by job occupation based on statistics doesn't diverge significantly from distribution based on questionnaire responses. It just enhance the existing effect. Service and sales worker, craft-workers, and clerical support workers reported significant overeducation in both cases (Figure 3). These findings are consistent with Kupets (2016) and may be explained by substantial polarization of Ukrainian labor market, there are many offers to highly qualified specialists and low-level service sphere workers, but there are few propositions to their middle-skilled peers. This leads to high competition at the labor market and forced overeducation.

You can find distribution of education mismatch by gender and field of attained education in Figure 5 and Figure 6 respectively in Appendix 2. It makes sense to point out, that graduates of Social and Behavioural science (Sociology and cultural studies, Psychology, Political sciences) departments have highest level of overeducation (by statistics) and at the same time reported the highest level of unemployment at the sample (Figure 7 in Appendix 2).

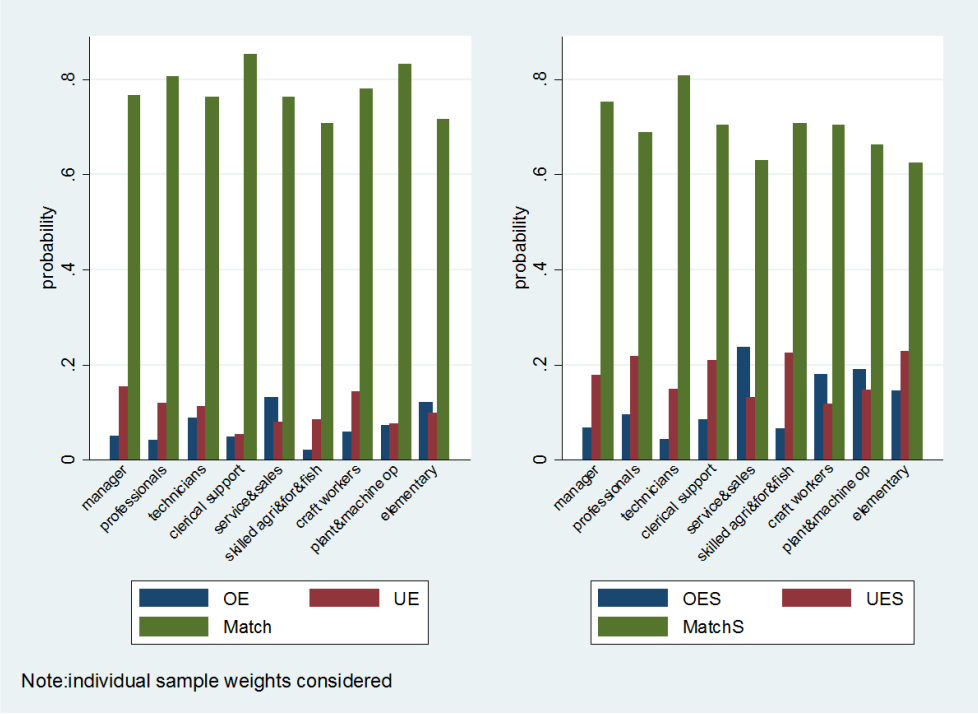

Figure 3: Relevance of attained education (by job occupation)

Chapter 5

EMPIRICAL RESULTS

5.1. Testing the difference in wage

We used Stata program for empirical part of work. For the purpose of research development, firstly we tested the impact of education-job mismatch on wages qualitatively. In effect, we compared the distributions of monthly wage for two groups of respondents: matched and mismatched. This groups were determined by respondents' self-assessment and by statistical approach described in the Chapter 3 and 4 of this paper.

Figure 4 depicts the density plots for groups determined by each of the methods named above. In the Figure 4 you can clearly see the negative effect of education-job mismatch on wage, the wage density function curves estimated for the mismatched respondents are shifted to the left, that is the probability of finding a lower-paid job is higher for the overeducated and undereducated graduates. In the Figures 8 and 9 at the Appendix 2 you can find the wage density function curves by groups estimated for each year in the sample separately. Notably that, in 2015 the density functions observed for matched and mismatched group based on statistical approach are almost the same. It could be explained by unobserved for this research distortion in labor market.

We have run the Kolmogorov-Smironov test to compare the distributions explicitly, the results for the whole sample comparison are presented in the Table 3. We rejected the null-hypothesis about distributions equality for matched and mismatched group (p-value is less than 0.1) based on both individuals' responses and determined by statistical approach. The equality of wage distribution functions

within each year separately (Table 9 and 10 Appendix 3) was barely rejected for 2013-year at the same way as for the whole sample, but fail to reject for 2015-year.

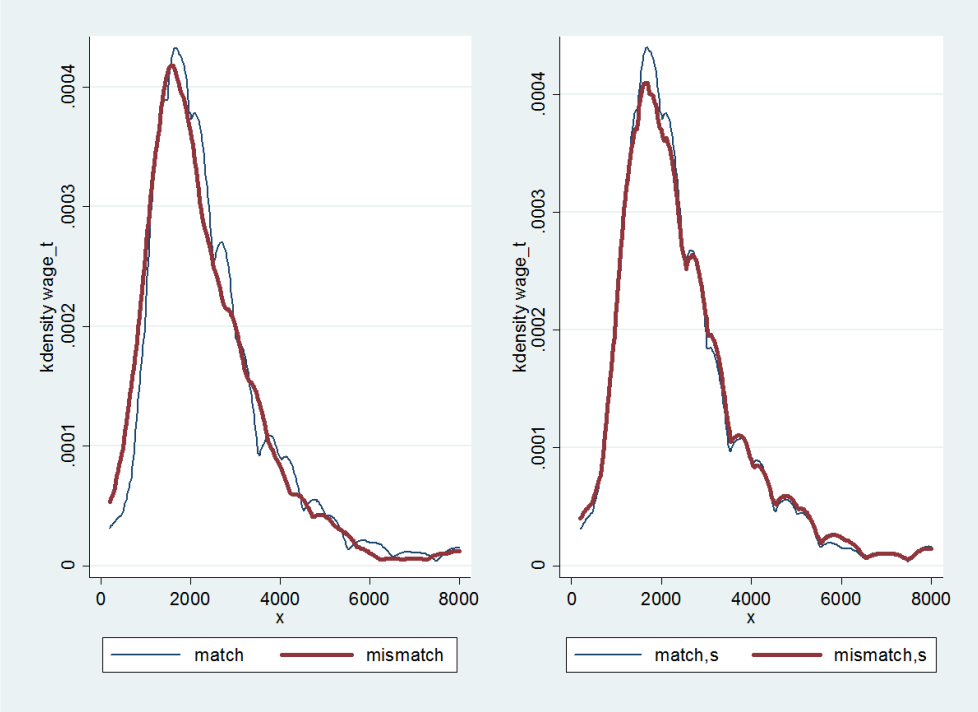

Figure 4: Wage density function (perception vs statistics) by matched and mismatched groups

Notably that for youth males the probability to get lower wage is higher the for their peer female (Figure 10, Appendix 2).

Table 3: Two-sample Kolmogorov-Smirnov test for equality of distribution functions (perception vs statistics)

Based on perception			
Smaller group	D	P-value	Corrected
0:	0.0827	0.026	
1:	-0.0018	0.998	
Combined K-S:	0.0827	0.052	0.044
Based on statistics			
Smaller group	D	P-value	Corrected
0:	0.0980	0.031	
1:	-0.0231	0.824	
Combined K-S:	0.0980	0.061	0.051

5.2. Estimation of the wage penalty for overeducation

Following the approach suggested by Nieto (2014) we controlled for possible sample selection problem in all the specification explained in the Chapter 2. We have used Heckman two step specification model. Recall, that for wage equation we use only the employed individuals, but education mismatch arises at first time as a problem to find a job, and only after that wage penalty takes place. At the initial stage we estimated the probability to be employed using a probit[9] equation (results are presented in Table 6 Appendix 1). We used this specification for all further estimation. The invers mills ratio were added into the wage equation as explanatory variable[10]. We

[9] We included into the probit equation gender, experience, experience squared, years of attained education, number of children, marital status as explanatory variables.

[10]Suggested by Nieto (2014) variables – number of children, marital – were used as exclusion restrictions; they do not influence the wage but could affect the probability to be employed.

did not use one step Heckman specification, because it was time consuming on our dataset.

For our analysis we considered two cases depends of how mismatch triggers were defined: based on individuals' responses or estimated statistically. We would like to focus on discussion of results obtained from specifications with statistically defined under- and overeducation. We could not apply the ORU specification to years of under-, overeducation defined by self-assessment, because they take negative values. However, you can find the comparison of the results obtained from V&V specifications applied to under-, overeducation triggers defined by both approach (see Table 7 in Appendix 1).

Table 4 presents the results from the estimation of simple Mincerian wage equation and more complex its specifications equations (5a) and (5b), the ORU models (6a) and (6b), and finally Verdugo & Verdugo specifications (7a), (7b).

The simplest Mincerian wage equation shows the positive relationship between the wage and years of obtained education, coefficient near years of education is statistically significant and each additional year of education produce 2.5 percent increase in wage. After that we extended model to (5a) and (5b) specification, and positive relationship between education and wage were saved but coefficient near years of education became statistically insignificant. This finding is in line with the results of previous research Coupe and Vakhitova (2011).

We would like to build the main discussion around the variables related to human capital. Nevertheless, it is important to note some interesting results obtained from estimation that differ from those which suggested by literature. First of all, gender matters, but in contrast, females get by more than 20 percent higher wage then males. It could be explained by proactive position prevailing among Ukrainian youth females, on average female has more years of experience. In addition,

Ukrainian language spoken youth have lower salary by 16 percent compare to their Russian-spoken peers. This may be caused by dominance of urban population in the sample. It is well-known fact that Russian language dominates Ukrainian[11] in the cities across the country. Furthermore, mills ratio coefficient is negative and significant for the most of specifications. That proves the rationality to use Heckman correction of sample bias, missing of the probability to be employed would cause the bias results (see the results of raw estimation in Table 8 Appendix 1).

Back to the human capital variables, we found that willingness to continue education or participate in trainings is statistically significant and implies negative effect on wage. Willingness to continue education influence the size of under-, overeducation triggers, but doesn't change their significance.

In terms of ORU specification (6a), we found that return on required education is negative, but statistically insignificant, it may be explained by existence of educational mismatch or the low relevance of obtained education among the employers. At the same time, each year of overeducation produces insignificant benefit to wage, while undereducation conversely produce 4.5 percent penalty to the worker, compare to the well-educated workers in the same job. The coefficients near under-, overeducation do not follow the order suggested in literature (overeducation have to imply greater effect than undereducation), however the sign of the effects is right. Including the variable "willingness to countinue education" slightly change the size of coefficients, but doesn't resolve the problem of omitted "ability", since overeducation is insignificant in statistical terms.

[11] https://en.wikipedia.org/wiki/Russian_language_in_Ukraine

The most interesting results provides the Verdugo&Verdugo specification, which strictly reject the initial hypothesis about the existence of wage penalty for overeduction, it indicates the 12 percent penalty for been undereducated, overeducation is still insignificant in this specification.

Table 4: Estimated wage equation functions

lwage_t	Mincer equation			ORU		V&V	
Variable	simple	(5a)	(5b)	(6a)	(6b)	(7a)	(7b)
d2015		-0.150**	-0.198***	-0.135**	-0.208***	-0.151**	-0.209***
		(0.064)	(0.061)	(0.064)	(0.062)	(0.064)	(0.062)
attained education	0.024***	0.019	0.018				
	(0.012)	(0.012)	(0.012)				
part time position	-0.2415541**	-0.224***	-0.206***	-0.211***	-0.211***	-0.219***	-0.212***
	0.0480	(0.045)	(0.045)	(0.045)	(0.045)	(0.045)	(0.045)
req.years				-0.024	-0.006	-0.006	-0.007
				(0.019)	(0.019)	(0.019)	(0.019)
UE (year)				-0.050***	-0.045**		
				(0.018)	(0.017)		
OE (year)				0.012	0.011		
				(0.018)	(0.018)		
UE (dummy)						-0.128**	-0.127**
						(0.051)	(0.051)
OE (dummy)						0.059	0.048
						(0.064)	(0.064)
Willing to continue education			-0.162***		-0.152***		-0.154***
			(0.049)		(0.049)		(0.049)
city		0.034	0.031	0.04	0.038	0.036	0.032
		(0.043)	(0.043)	(0.043)	(0.043)	(0.043)	(0.043)
female		0.254***	0.249***	0.230***	0.237***	0.236***	0.235***
		(0.036)	(0.037)	(0.038)	(0.038)	(0.038)	(0.038)
age		0.033***	0.029***	0.033***	0.030***	0.033***	0.029***
		(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
have 1 child		0.082*	0.096**	0.056	0.062	0.061	0.065
		(0.047)	(0.046)	(0.055)	(0.053)	(0.054)	(0.053)
language		-0.167***	-0.167***	-0.167***	-0.167***	-0.165***	-0.165***
		(0.043)	(0.043)	(0.043)	(0.043)	(0.043)	(0.043)
Regions	No	Yes	Yes	Yes	Yes	Yes	Yes
Firm size	No	Yes	Yes	Yes	Yes	Yes	Yes
Mills ratio	-0.046	-0.392*	-0.400**	-0.416**	-0.421**	-0.411**	-0.416**
	(0.112)	(0.201)	(0.202)	(0.203)	(0.203)	(0.203)	(0.203)
Constant	8.055***	6.782***	6.798***	7.128***	7.131***	7.115***	7.118***
	(0.162)	(0.286)	(0.288)	(0.356)	(0.356)	(0.356)	(0.356)
Observations	1775	1775	1775	1775	1775	1775	1775

Standard errors in parentheses. *Statistically significant at the 10% level. ** Statistically significant at the 5% level. ***Statistically significant at the 1% level. Individual sample weights considered.

5.3. Weaknesses and Limitations of the research

During the research we have used the data from SWTS in order to estimate the effect of under-, overeducation on wage. The sample we used provides extensive information about individuals' school to work transition process. However, similar to many surveys conducted in Ukraine, it gives no opportunity to measure personal abilities of respondents. Missing the personal ability leads to the omitted variable bias and this makes the obtained results less robust.

Another problem is great variety of the job occupation presented in the sample, these caused the high standard deviation, when we determined the required years of education to get the particular job.

Chapter 6

CONCLUSION

This thesis examines the existence of a wage penalty for overeducation among Ukrainian youths (from 15 to 29 years). Overeducation refers to those individuals who claim that their university degree was neither required by law, nor useful to access their current job. We combined two waves of SWTS in 2013, 2015 into the pooled data. The research was done by using the modifications of Mincerian wage equation.

In order to conduct such a research, we follow the methodology suggested by Nieto (2014), who estimated the overeducation, skills and wage penalty in Spain. We use the mean-approach to estimate the required years of education for job occupations and compare the wages of the overeducated individuals with their matched peers.

Our results obtained from the two-step Heckman specification model suggest a that there is positive insignificant effect of overeducation on wage, and at the same time substantial negative effect of undereducation. This means that the under reducated individuals get a lower wage compared to the individuals with the same degree employed at the work with required years of education matching to attained years of education. We also found out that return on each required year of education is insignificant comparing to return on each attained year of education, this points out the existence of educational mismatch and low relevance of education among the employers.

During the research we explored that the educational level, associated with the lowest rate of people with education-job mismatch, is bachelor degree. Recall that

under the macroeconomics prospective rate of return on education is one of the leading indicators for investment in education. Based on this and negative effect of undereducation on wage, we may suggest that aforementioned educational level appears to be the most investment attractive from the government point of view.

The main weakness of the estimation is the unobserved ability factors. Missing the personal ability leads to the omitted variable bias and this makes the obtained results less robust.

WORKS CITED

Berg, I., 1970. Education and Jobs: The Great Training Robbery. Praeger.

Brown, J. David, and John S. Earle., 2006. Job Reallocation and Productivity Growth in the Ukrainian Transition. Comparative Economic Studies48(2) (2006): 229-251.

Chevalier, A., 2003. "Measuring Overeducation", Economica, 70 (3), 509–531.

Chevalier, A. and Lindley, J., 2009. "Overeducation and the skills of UK graduates", Journal of the Royal Statistical Society: Series A, 172, 307-337.

Coupé, Tom and Vakhitova, Hanna, 2011: Recent Dynamics of Returns to Education in Transition Countries, Kyiv School of Economics Discussion Paper No. 39.

Cutillo, A. and G. Di Pietro, 2006, "The effects of overeducation on wages in Italy: a bivariate selectivity approach", International Journal of Manpower, 27(2): 143-168.

Duncan, G., Hoffman, S., 1981. The Incidence and Wage Effects of Overeducation. Econ. Educ. Rev.1(1), 75–86.

Fouarge, D. Schils, T. & de Grip, A., 2010. Why do Low-Educated Workers Invest Less in Further Training? Discussion Paper No. 5180. (September). Forschungs institute zur Zukunft det Arbeit Institute for the Study of Labor. Germany: Bonn, 2–8.

Freeman, R., 1975. Overinvestment in College Training? J. Hum. Resour. 10,287–311.

Freeman, R., 1976. The Overeducated American. AcademicPress, New York.

Green, F. and McIntosh, S., 2007. "Is There a Genuine Under-utilisation of Skills Amongst the Overqualified?", Applied Economics, 39 (4), 427-439.

Green, F. and Zhu, Y., 2010. "Overqualification, Job Dissatisfaction and Increasing Dispersion in the Returns to Graduate Education", Oxford Economic Papers, 62 (4), 740–763.

Harvey, L., 2000. New realities: The relationship between higher education and employment. Tertiary Education & Management, 6(1), 3-17.

Heckman, J., 1979. "Sample Bias as a Specification Error," Econometrica, Vol. 47, pp. 153-167.

Korpi, T. and Tåhlin, M., 2009. "Educational mismatch, wages, and wage growth: Overeducation in Sweden, 1974–2000", Labour Economics. 16, 183–193.

Kupets, Olga, 2015. Skill mismatch and overeducation in transition economies. IZA World of Labor 2016: 224.

Kupets, Olga, 2016. Education-job mismatch in Ukraine: Too many people with tertiary education or too many jobs for low-skilled?. Journal of Comparative Economics 44(2016): 125-147.

Leuven E. and H. Oosterbeek, 2011. "Overeducation and Mismatch in the Labour Market", in E.A. Hanushek, S. Machin and L. Woessmann, Handbook of the Economics of Education,vol. 4: 283-326.

Levels, M., van der Velden, R. and Allen, J., 2013. "Educational mismatches and skills: new empirical tests of old hypotheses", RM/13/062.

Libanova, Ella, Cymbal, Akeksandr, LarysaMarchenko, Iryna, Iarosh, 2014. Labor Market Transition of Young Women and Men in Ukraine. International Labor Office, Geneva.

Mavromaras, K., McGuinness, S., O'Leary, N., Sloane, P. and Wei, Z., 2013 "Job Mismatches and Labour Market Outcomes: Panel Evidence on University Graduates", Economic Record, 89, 382–395.

McGuinness, S., 2006. "Overeducation in the labour market", Journal of Economic Surveys, 20(3): 387-418.

McGuinness, S., 2007. How Biased Are the Estimated Wage Impacts of Overeducation? A Propensity Score Matching Approach. Appl. Econ. Lett. 15, 145–149.

Mincer, J., 1974. Schooling, Experience, and Earnings, New York: NBER Press.

Nicaise, I., 2001. "Human capital, reservation wages and job competition: Heckman's lambda reinterpreted", Applied Economics, Vol. 33 No. 3, pp. 309-15.

Nieto, S., 2014. Overeducation, skills and wage penalty: Evidence for Spain using PIAAC data. IREA–Working Papers, IR14/011.

Osipian, A., 2015. Higher education corruption in Ukraine: Opinions and estimates. International higher education, (49).

Sánchez-Sánchez, N. and McGuinness, S., 2013. "Decomposing the impacts of overeducation and overskilling on earnings and job satisfaction: an analysis using REFLEX data", Education Economics, in press. http://dx.doi.org/10.1080/09645292.2013.846297

Sattinger, M., 1993. "Assignment models of the distribution of earnings", Journal of Economic Literature, 31(1): 851-80.

Sicherman, N.,1991. "Overeducation in theLaborMarket ".J. LaborEcon.9(2),101.

Sicherman, N.,Galor,O., 1990. A Theory of Career Mobility. J.Polit. Economy, 98(1), 169–192.

Silverman, B.W., 1998. Density Estimation for statistics and data analysis. New York: Chapman & Hall/CRC, (Chapter 3).

Sloane, P., 2003. "Much Ado about Nothing? What does the Overeducation Literature really Tells us?", in F.A. Büchel, F., A. de Grip and A. Mertens (op. cit.).

Smith, J., Welch,F., 1978.The Overeducated American? A Review Article. Technical Report, RandCorporation.

Spence, M.,1973.Job Market Signalling.Q.J.Econ.87,355–374.

Tsai, Y., 2010. "Returns to overeducation: A longitudinal analysis of the U.S. labor market", Economics of Education Review, 29, 606–617.

Verdugo, R. and Verdugo, N., 1989. "The impact of surplus schooling on earnings: some additional findings", Journal of Human Resources, 24 (4), 629-643.

APPENDIX 1

Table 5: Extended descriptive statistics

variable	N	mean	p50	SD	min	max
dist	1775	12.30529	12	7.029223	1	26
language	1775	0.4733701	0	0.4993276	0	1
age	1775	22.15562	22	4.324679	15	29
female	1775	0.5544	1	0.5000345	0	1
age of married	1775	11.01077	17	10.98851	0	29
spouse's actv.	1775	1.744656	0	2.088471	0	7
#of children	1775	0.5303933	0	0.7224308	0	5
father_edu	1775	12.17939	12	1.952548	0	22
mother_edu	1775	12.50429	12	2.06163	0	22
work_study	1775	3.431027	4	1.002905	1	4
job position	1775	4.717817	5	2.503572	1	9
industry	1775	9.142155	7	5.325416	1	20
firm size	1775	3.580424	4	1.897727	1	7
contract dur.	1775	0.9297405	1	0.2556348	0	1
wage	804	2443.901	2000	1440.422	50	15000
perception	1775	1.399103	1	0.8205516	1	4
planeduw	1775	2.063801	2	0.6743732	1	3
training_w~k	1775	2.714909	3	0.5901891	1	3
fam_disc	1775	0.8351031	1	0.3711455	0	1
city	1775	0.7214625	1	0.4483126	0	1
employed	1775	0.4426278	0	0.4967344	0	1
unemployed	1775	0.0792212	0	0.2701037	0	1
d2015	1775	0.4759215	0	0.499457	0	1
exper	1775	5.549693	5	3.332966	0	18
exper2	1775	41.90503	25	43.04586	0	324
profess	1775	68.83943	77	35.20088	1	138
Match	1775	0.7813039	1	0.4134333	0	1
OE	1775	0.0745085	0	0.2626418	0	1
UE	1775	0.1079683	0	0.3103939	0	1
MMatch	1775	0.2186961	0	0.4134333	0	1
req_year	1775	13.94093	13.34457	1.105413	11	17.33333
UE_y	1775	0.0294569	0	1.610733	-9.655427	9.344573
OE_y	1775	-0.0131221	0	1.593435	-9.344573	9.655427
SD	1775	2.014277	2.120842	0.3970484	0	5.656854
OES	1775	0.1387458	0	0.3457335	0	1
UES	1775	0.1584368	0	0.3652054	0	1
MatchS	1775	0.7028173	1	0.4570869	0	1
MMatchS	1775	0.2971827	0	0.4570869	0	1
UE_y_S	1775	0.4760631	0	1.160712	0	9.344573
OE_y_S	1775	0.511318	0	1.338994	0	9.655427
part_time	1775	0.0827517	0	0.2755753	0	1
lwage	804	7.649086	7.600903	0.5798367	3.912023	9.615806
wage_t	804	2435.521	2000	1274.355	1000	7001
lwage_t	804	7.680957	7.60090	0.473418	6.90775	8.85380

Table 6: Heckman's specification first step. Determinants to being employed

Variables	(1)
age	0.021**
	(-0.009)
female	-0.035
	(-0.034)
highestlevel_comp	0.024***
	(-0.009)
married	0.057
	(-0.065)
nb_of_children	-0.016
	(-0.033)
exper	0.074***
	(-0.019)
exper2	-0.005***
	(-0.001)
Regional dummy	Yes
Constant	-0.311*
Observations	1771

Standard errors in parentheses. *Statistically significant at the 10% level. ** Statistically significant at the 5% level. ***Statistically significant at the 1% level. Individual sample weights considered.

Table 7: V&V estimates by two approaches

	V&V	
lwage	Perception	Statistics
d2015	-0.213***	-0.209***
	(0.06)	(0.062)
req_year	0.049**	0.050**
	(0.021)	(0.021)
UE	-0.038	
	(0.063)	
OE	-0.002	
	(0.075)	
UES		-0.127**
		(0.051)
OES		0.048
		(0.064)
yes_cont_educ	-0.149***	-0.154***
	(0.049)	(0.049)
part_time	-0.211***	-0.212***
	(0.044)	(0.045)
city	0.015	0.032
	(0.043)	(0.043)
female	0.224***	0.235***
	(0.038)	(0.038)
age	0.028***	0.029***
	(0.007)	(0.008)
child1	0.094**	0.065
	(0.045)	(0.053)
language	-0.186***	-0.165***
	(0.04)	(0.043)
Regions	Yes	Yes
Firm size	Yes	Yes
Mills ratio	0.445**	0.416**
	(0.208)	(0.203)
Constant	7.279***	7.118***
	(0.357)	(0.356)
Observations	911	911

Standard errors in parentheses. *Statistically significant at the 10% level. ** Statistically significant at the 5% level. ***Statistically significant at the 1% level. Individual sample weights considered.

Table 8: Estimated wage equation functions without control for sample selection bias (based on self-assessment)

	Mincer		**V&V**	
	(5a)	(5b)	(7a)	(7b)
d2015	-0.14	-0.085	-0.116	-0.065
	(0.087)	(0.088)	(0.085)	(0.087)
highestlevel_comp	0.014	0.014		
	(0.01)	(0.009)		
UE			-0.018	-0.02
			(0.082)	(0.081)
OE			-0.07	-0.044
			(0.054)	(0.056)
req_year			-0.024	-0.022
			(0.017)	(0.017)
yes_cont_educ		-0.155***		-0.143***
		(0.045)		(0.045)
part_time	-0.189***	-0.199***	-0.181***	-0.190***
	(0.049)	(0.049)	(0.049)	(0.049)
city	0.024	0.026	0.034	0.036
	(0.04)	(0.04)	(0.04)	(0.04)
female	0.267***	0.261***	0.242***	0.238***
	(0.037)	(0.037)	(0.041)	(0.041)
age	0.023***	0.020***	0.026***	0.023***
	(0.007)	(0.007)	(0.007)	(0.007)
language	-0.212***	-0.209***	-0.197***	-0.195***
	(0.04)	(0.039)	(0.039)	(0.039)
child1	0.072*	0.075*	0.082*	0.085*
	(0.043)	(0.043)	(0.043)	(0.043)
Constant	6.877***	7.011***	7.389***	7.487***
	(0.211)	(0.215)	(0.306)	(0.308)
Region	Yes	Yes	Yes	Yes
Firm size	Yes	Yes	Yes	Yes
R-Square	0.207	0.218	0.204	0.213
Observations	798	798	793	793

Standard errors in parentheses. *Statistically significant at the 10% level. ** Statistically significant at the 5% level. ***Statistically significant at the 1% level. Individual sample weights considered.

APPENDIX 2

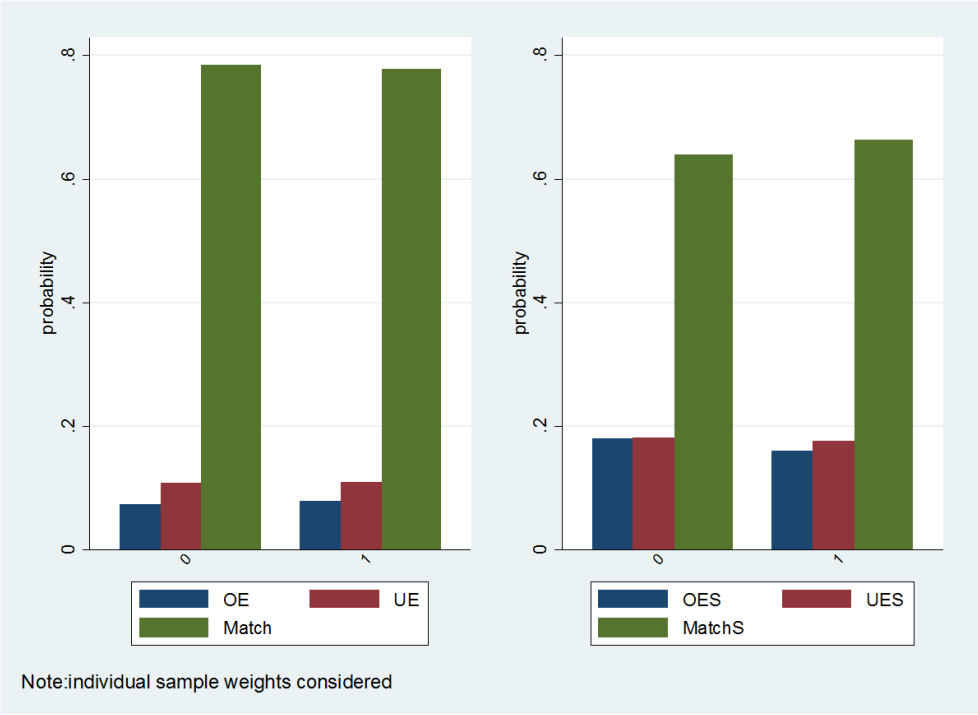

Figure 5: Relevance of attained education (by gender)

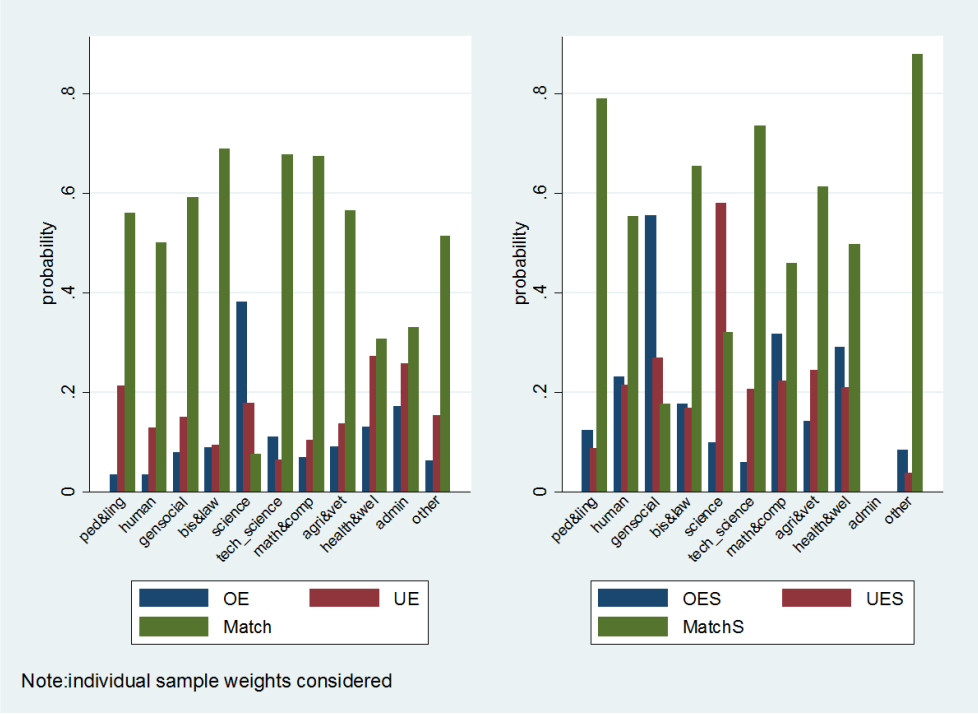

Figure 6: Relevance of attained education (by field of attained education)

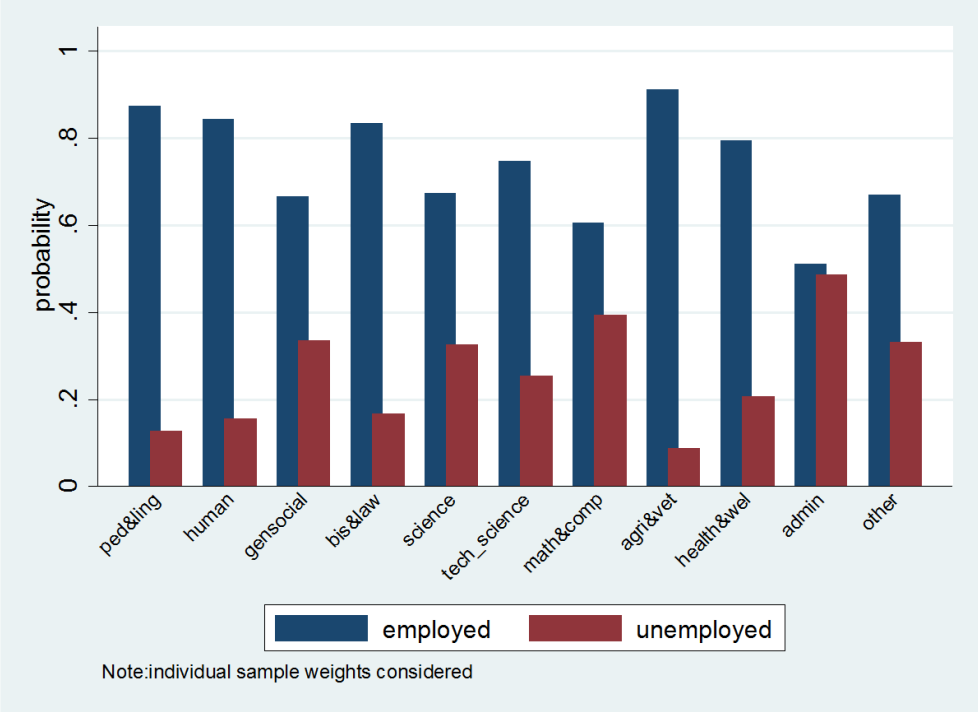

Figure 7: Employment vs Unemployment distribution (by field of attained education)

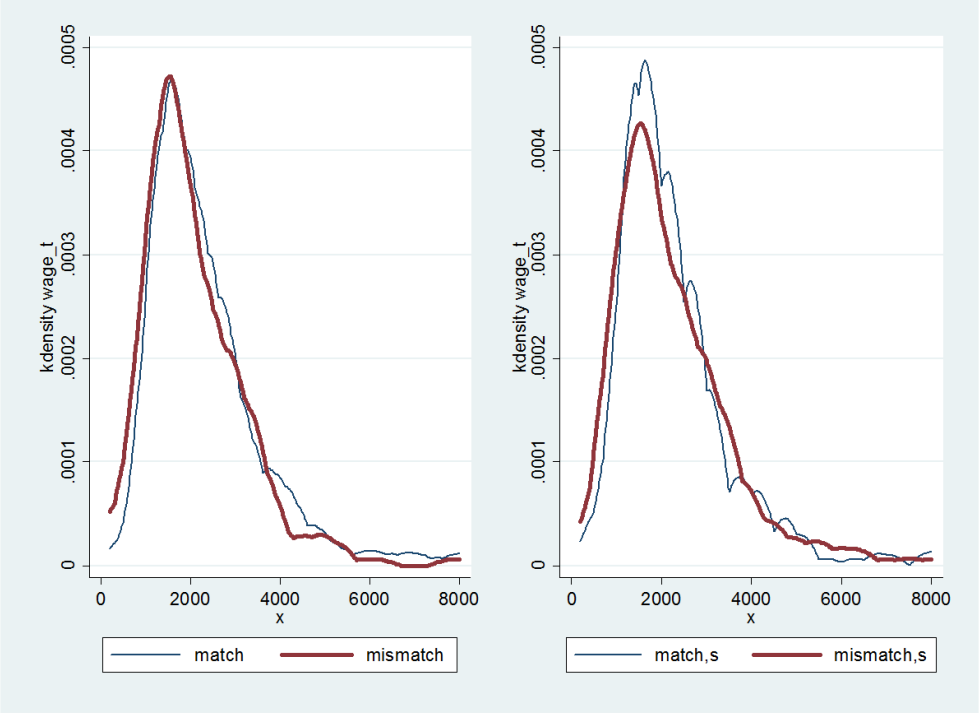

Figure 8: Wage density function 2013 (perception vs statistics) by matched and mismatched groups

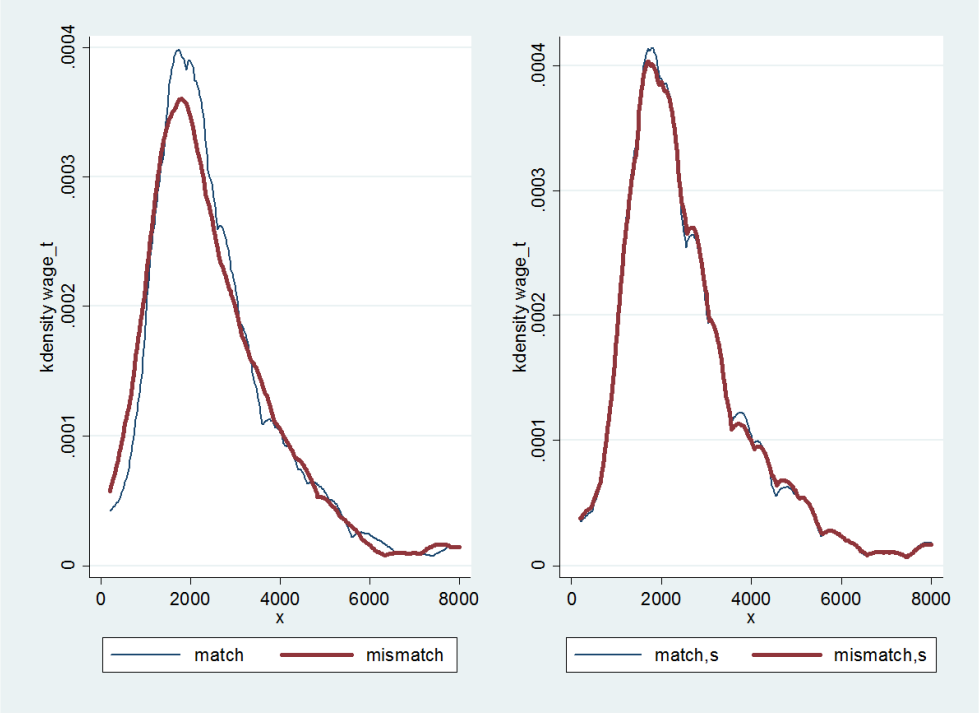

Figure 9: Wage density function 2015 (perception vs statistics) by matched and mismatched groups

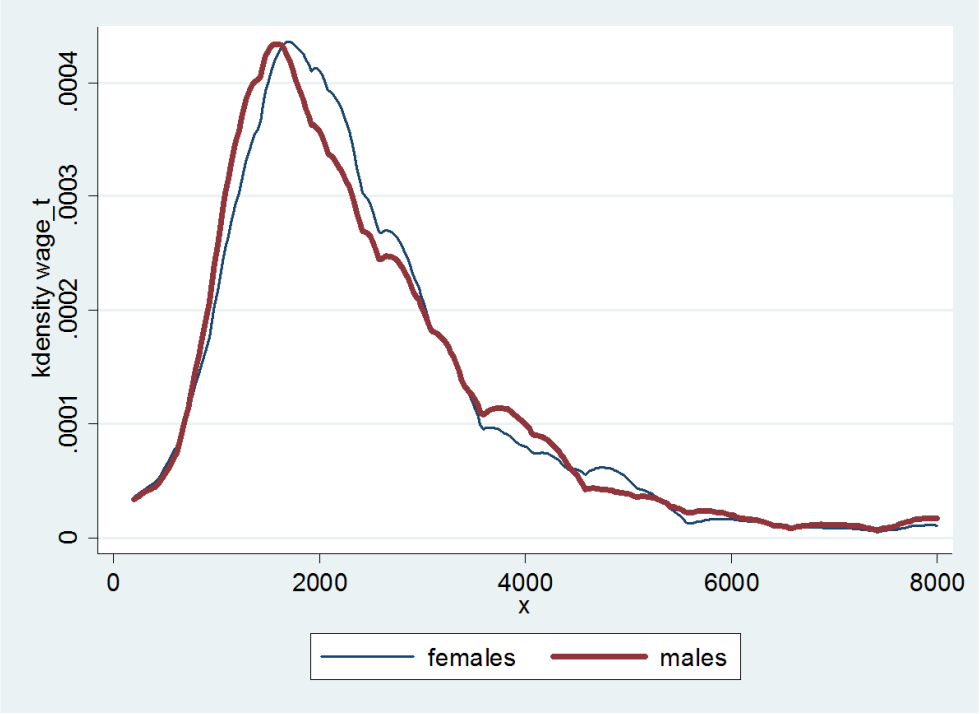

Figure 10: Wage density function (perception vs statistics) by gender

APPENDIX 3

Table 9: Two-sample Kolmogorov-Smirnov test for equality of distribution functions (perception vs statistics), 2013

Based on self-assessment			
Smaller group	D	P-value	Corrected
0:	0.1043	0.061	
1:	0.0000	1.000	
Combined K-S:	0.1043	0.122	0.101

Based on statistics			
Smaller group	D	P-value	Corrected
0:	0.0874	0.092	
1:	-0.0306	0.747	
Combined K-S:	0.0874	0.184	0.159

Table 10: Two-sample Kolmogorov-Smirnov test for equality of distribution functions (perception vs statistics), 2015

Based on self-assessment			
Smaller group	D	P-value	Corrected
0:	0.0632	0.331	
1:	-0.0318	0.756	
Combined K-S:	0.0632	0.639	0.598

Based on statistics			
Smaller group	D	P-value	Corrected
0:	0.1440	0.356	
1:	-0.0863	0.691	
Combined K-S:	0.1440	0.681	0.602