Lahore Journal of Economics
Volume 29, Issue 2,
Winter 2024

Role of Education Mismatch in Shaping Earning Outcomes Across Different Employment Status in Pakistan

Henna Ahsan
Research Economist,
Pakistan Institute of Development Economics (PIDE),
Islamabad, Pakistan
henna.ahsan@pide.org.pk

Citation: "Ahsan, H. (2024). Role of Education Mismatch in Shaping Earning Outcomes Across Different Employment Status in Pakistan". *Lahore Journal of Economics*, *29* (2), 1–28. https://doi.org/10.35536/lje.2024.v29.i2.a1

Abstract: *This study contributes to the literature that highlights the penalties of education-occupation mismatch in terms of earnings across different employment statuses. Most existing literature analyzing the education-occupation mismatch has focused on paid employees, overlooking self-employed individuals, and has not controlled for sample selection bias and unobserved heterogeneity bias simultaneously. Therefore, the objective of this study is to analyze the impact of education mismatch on earnings across different employment statuses after correcting for both sample selection bias and unobserved heterogeneity bias. To achieve this objective, we applied the methodology of Duncan and Hoffman (1981) to the Pakistan Social and Living Standards Measurement (PSLM), 2019-20. Our results show that after controlling for unobserved heterogeneity bias and sample selection bias, over-education has no positive value for both paid employees and the self-employed. The returns from over-education based on the OLS model might be overestimated if overeducated workers possess lower average ability levels, whereas the returns of adequately educated individuals increase after correcting for the bias and are significantly higher for self-employed individuals compared to paid employees.*

Keywords: Education mismatch, earnings, labor market, sample selection bias, unobserved heterogeneity bias, Pakistan.

JEL Classification: E20, I20.

Paper type: Research paper

Received: 4th November 2024 Revised: 6th May 2025 Accepted: 16th May 2025

Copyright: The Lahore Journal of Economics is an open access journal that distributes its articles under the terms of the Creative Commons attribution-NonCommercial-Noderivatives license (http://creativecommons.org/licenses/by-nc-nd/4.0/): this licence permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way. With this Creative Commons license in mind, the Lahore Journal of Economics retains the right to publish an article upon successful completion of the submission and approval process, and with the consent of the author(s).

Conflict of interest:
The authors declare no conflict of interest.

Funding:
There is no funding for this research.

Role of Education Mismatch in Shaping Earning Outcomes Across Different Employment Status in Pakistan

1. Introduction

To what extent education plays a role in increasing an individual's earnings is an important question for policymakers and researchers. Although this question has existed for decades, it got due attention after the publication of the book "The Over-Educated American" by Freeman in 1976. The book highlighted the startling findings that the average earnings of high school and college graduates in the USA decreased by 16 to 40 percent between 1969 and 1974 (Freeman, 1976).

With the rapid expansion of education, it has been observed that overeducation also exists in some labor markets; that is, the education of some workers exceeds the level required to perform a specific task, causing an education-occupation mismatch (Rumberger, 1981; Hartog, 2000). Further, Bird (1975) predicted that opportunities for new college graduates declined in the labor market, especially during the recession years of 1975 and 1976, leading to a more widespread distrust regarding economic payoffs for college graduates.

As noted by Pritchett (2001), developing countries like Pakistan have invested heavily in their education sector over the past 20 years to increase enrollment at secondary and tertiary levels, to boost their human capital. However, in the presence of poor institutions, stagnant labor markets, and low educational quality, additional years of schooling do not necessarily translate into enhanced human capital. Therefore, the basic earning model developed by Mincer (1974), suggesting a positive relationship between earnings and years of schooling, may not hold true. Education mismatch leads to the misallocation of human capital in the labor market, penalizing overeducated individuals who receive lower earnings than workers with similar education but whose education aligns with job requirements. However, these overeducated individuals tend to earn higher earnings than their co-workers who are not overeducated (Duncan & Hoffman, 1981; Groot & Maassen, 1997; Rubb, 2003; and McGuinness, 2006).

Moreover, when people cannot find jobs that match their education, especially in developing countries where unemployment insurance policies are often lacking, they tend to start their own businesses, which may result in their education being underutilized. The literature indicates that self-employed workers are typically jack-of-all-trades, possessing expertise in various fields (Lazear, 2005), and perhaps that explains why they are often found working in areas unrelated to their educational background. Furthermore, many studies examine the education mismatch for paid employees while neglecting self-employed individuals, with the exception of the study by Bender and Roche (2013) conducted in the United States. However, in the context of developing countries like Pakistan, where unemployment is even more pronounced among well-educated individuals, it becomes essential to analyze the distinct impacts of education mismatch on earnings for both paid employees and self-employed individuals separately.

Moreover, despite the numerous empirical studies focusing on the impact of education mismatch in the labor market on earnings, several problems persist that may cause bias in estimating this earnings effect. The first is the sample selection bias. The issue of this bias in estimating overeducation lies in the clear differences in the attributes of unemployed and employed individuals. These different characteristics may affect individuals' choices to work and, thus, their outcomes in the labor market. For instance, due to the prevailing high unemployment rate, individuals may be compelled to take jobs that require less schooling, thereby increasing the likelihood of being overeducated if employed (Quintini, 2011; Lee et al., 2016). Therefore, omitting individuals' decisions regarding participation in the labor market may introduce bias when estimating using the simple OLS method.

Secondly, many recent studies (Dolton and Vignoles, 2000; Kleibrink, 2016) note that using overeducation in classical wage regressions relies on the assumption that equally educated individuals have the same innate ability and, consequently, productivity, which leads to unobserved 'heterogeneity bias.' Thus, overlooking the impact of ability while analyzing the earning model may yield biased results (Bauer, 2002; Kopri and Tahlin, 2009).

As previously mentioned, both heterogeneity bias and sample selection bias arise when analyzing the impact of education mismatch on earnings. However, it is generally challenging to control for both simultaneously, so most studies either account for sample selection bias or heterogeneity bias. The literature on controlling or correcting unobserved heterogeneity bias in conjunction with sample selection bias is limited. Therefore, to address this gap, our study examines the impact of education mismatch on earnings across different employment statuses in Pakistan while considering bias from both sample selection and heterogeneity perspectives. To evaluate the returns to education mismatch, the study employs the methodology proposed in the Duncan and Hoffman (1981) model. To mitigate the issue of sample selection bias, we utilize the methodology of Heckman (1979) and the Generalized Method of Moments (GMM) under the instrumental variable (IV) technique to address the heterogeneity bias associated with the observable educational variables included in the empirical specification. The study is based on data from the PSLM (2019-20).

The remainder of this study is organized as follows. Section 2 discusses the literature review regarding the impact of education mismatch on earnings, while Section 3 explains the methodology in detail. Section 4 covers the data and construction of variables, followed by the presentation of descriptive statistics and results of the empirical model in Section 5, with the study concluding in Section 6.

2. Literature Review

There are several studies that explain the education mismatch and its impact on earnings. Job Competition Theory (JCT) by Thurow (1975) sheds light on the institutional rigidities that affect an individual's marginal product and earnings based on job characteristics. Job allocation in the labor market relies on the availability of both workers and jobs. Consequently, an excess supply of workers may force those with high skill sets to accept lower-level jobs, as their educational achievements primarily serve the basic purpose of securing employment without providing additional benefits. The Job Assignment Theory by Sattinger (1993) suggests that there is an allocation problem in assigning heterogeneous workers to jobs with varying complexity. This theory assumes that workers with the same level of human capital do not necessarily have equal productivity; rather, their productivity depends on the jobs to which they are matched. This indicates that both actual and required education levels can influence earnings. The Job Search Model of Jovanovic (1979) offers another perspective, explaining the occurrence of overeducation when an individual with high skills starts in a job below his or her ability level, but eventually finds a position that aligns with their skill set.

Therefore, to analyze the impact of education mismatch on earnings, different methods have been used. The most popular method that examines this mismatch is proposed by Duncan and Hoffman (1981).

The study reported that overeducated workers earn higher returns than their co-workers who are not overeducated, but lower returns than individuals with the same education working in jobs that match their education. Furthermore, undereducated workers receive lower earnings compared to their coworkers who have an adequate level of education.

Duncan and Hoffman (1981) found that in the USA, the average returns for adequately educated workers were 6 to 10 percent per year, while the returns for over-educated workers were estimated to be between 2.9 and 4.7 percent. The study further discovered that each deficit year of undereducation reduces earnings by an average of 4 percent for undereducated individuals. These results align with a recent study by Lasso-Dela-Vega et al. (2023), which found that adequately educated individuals in the Spanish economy earn 5%, over-educated individuals earn 2.4% for each additional year of education, whereas under-educated individuals face a wage penalty of 2.9% for each deficit year of education. Various other studies also support these findings (for example, Groot and Massan, 1997; Sial et al., 2019; Rubb, 2003; Groeneveld and Hartog, 2004; Clark et al., 2017; Takeuchi, 2023; Sulaimanova, 2022). These findings on the returns of over-education are consistent with the Assignment Model, which explains wage differentials based on workers' education levels and job characteristics (Sattinger, 1993).

2.1 Sample Selection Bias

Heckman (1979) argues that earnings are only observed for employed workers who are not randomly selected; therefore, selection bias can arise when estimating earnings equations. Moreover, the job search theoretical model considers unemployment to be largely a voluntary choice. People typically accept a job that offers a higher wage than their reservation wage. Highly skilled individuals prefer to remain unemployed, waiting until they find a job that provides their best expected wage. In contrast, less skilled individuals tend to wait less and accept the first job offer they receive, even if it renders them overeducated. Thus, due to this selection bias, observed earnings may not accurately reflect the earnings distribution across the entire population.

Many studies, including Nicaise (2001), Cutillo & Di Pietro (2006), Lee et al. (2016), and Caroleo and Pastore (2018), highlight the existence of sample selection bias in estimating overeducation based on the differences in characteristics between unemployed and employed workers. Lee et al. (2016) found that after controlling for sample selection bias, the estimated coefficients for years of education in the earnings function for overeducation increased by approximately 0.2 to 0.5 percentage points in the Korean labor market, compared to estimates that do not account for sample selection bias. This suggests that once we control for the selection bias arising from non-employment, the estimated returns to overeducation increase. According to the job search theoretical model, unemployment is often viewed as a voluntary choice, and the most skilled graduates may prefer to remain unemployed while waiting for the best job they can secure. If they were employed, they would be less likely to experience overeducation. Thus, once controlling for the selection bias arising from considering non-employment, the wage penalty for those experiencing educational mismatch might be lower.

The Heckman sample selection procedure can be used as a screening tool to choose among different theoretical interpretations of overeducation. The corrected estimates suggest that OLS largely understates the true effect of overeducation on labor market earnings. This may indicate a higher ability level of unemployed individuals, which aligns with the expectation,s of the Job Search model (Nicaise, 2001; Lee et al, 2016). Conversely, an overestimate of the OLS coefficient suggests a lower ability level of unemployed individuals, which is consistent with the job competition model and job assignment model, where unemployment is high and dominated by an involuntary component (Caroleo & Pastore, 2018).

2.2 Unobserved Heterogeneity Bias

Analyzing the impact of education mismatch on earnings by assuming that education-job mismatches are an exogenous phenomenon is problematic due to the issue of 'unobserved heterogeneity.' Bias may arise from the presence of unobserved factors that correlate with education mismatches in the labor market and earnings. This is known as omitted variable bias. For example, workers with higher skills or ability may have higher levels of educational attainment; conversely, unobserved ability or skills included in error terms also impact individuals' earnings. These unobserved factors are likely to cause bias (Allen and Van der Velden, 2001; Dolton and Silles, 2008).

A substantial body of literature on managing omitted variable bias employs IV techniques (Korpi and Tåhlin, 2009; Caroleo and Pastore, 2018; Kleibrink, 2016; Lee et al, 2016; Tran et al., 2023) and fixed-effect models (Bauer, 2002; Dolton and Silles, 2008; Tsai, 2010). Additionally, many studies, such as Allen and Van der Velden (2001) and Kleibrink (2016), directly incorporate ability or skills in OLS estimations to control for ability bias in the specification.

Dolton and Silles (2008) analyzed the impact of overeducation on earnings in the United Kingdom (UK) economy, controlling for omitted ability bias using a fixed effects model and addressing measurement error with the instrumental variable (IV) technique. The results from the fixed effect model indicate that educated individuals face greater penalties in terms of earnings, and the OLS estimates are somewhat biased upwards, suggesting a negative correlation between overeducation and ability. Furthermore, to mitigate bias from both omitted ability and measurement error, the fixed effect IV method is employed, estimating that over-education reduces earnings by 35–40 percent. However, they posited that over-educated graduates may still earn more than coworkers whose education matches the occupation. Therefore, for the UK economy, they concluded that over-education provides some benefits.

Similarly, based on Verdugo and Verdugo's model, Njifen and Smith (2024) used several types of regression models to examine both the heterogeneity in returns to education resulting from education-job mismatch and the selection of different types of models. Their results show that overeducation is associated with a wage penalty, while undereducation leads to a wage premium.

Some studies also attempted to address unobserved heterogeneity bias through instrumental variable (IV) techniques, such as Kopri and Tahlin (2009) for the Swedish economy, Kleibrink (2016) for Germany, and Lee et al. (2016) for Korea. All these studies found that their results come with notable caveats and provide limited support for the compensation hypothesis, which suggests that overeducated individuals do not receive positive returns on the extra years of schooling beyond their required job level. They concluded that years of overeducation have no effect on pay, and only the type of job holds real importance.

3. Empirical Model

To quantify the effect of educational mismatch—specifically, overeducation, adequate education, and undereducation—on earnings, the study adopts an extended Mincer (1974) earnings function as introduced by Duncan and Hoffman (1981). For this purpose, we began with the basic Mincer earnings function:

$$\log Y_i = \alpha + \beta\, E_i + \delta\, X_i + \; + \boldsymbol{E}_i \tag{1}$$

Where Y_i is the log of earnings of the ith individual, E_i is years of education attainment level, and the vector X_i includes characteristics of workers and other explanatory variables that affect earnings. Duncan and Hoffman (1981) decomposed the total years of education attainment (E_i) into adequate education for occupation (E_i^a), years of overeducation (E_i^o) and years of under education (E_i^u).

To measure the adequate level of education in each occupation, a statistical method provided by Kiker et al. (1997) has been used. In this method, adequate education is measured by mode of education as classified by the International Standard Classification of Education (ISCED) through a 3-digit level of the International Standard Classification of Occupation (ISCO). Each individual with educational attainment exactly equal to what is required for an occupation is classified as "adequately educated" (E_i^a).

The terms "overeducated" and "undereducated" are defined as:

$$\mathrm{E_i^o} = \begin{cases} \mathrm{E_i} - \mathrm{E_i^{\mathit{a}}}, \text{if } \mathrm{E_i} > \mathrm{E_i^{\mathit{a}}} \\ 0, \text{otherwise} \end{cases} \quad \text{and} \quad \mathrm{E_i^u} = \begin{cases} \mathrm{E_i^{\mathit{a}}} - \mathrm{E_i}, \text{if } \mathrm{E_i^{\mathit{a}}} > \mathrm{E_i} \\ 0, \text{otherwise} \end{cases}$$

Therefore, the following identity holds:

$$\mathrm{E_i} = \mathrm{E_i^{\mathit{a}}} + \mathrm{Max}\,(0, \mathrm{E_i} - \mathrm{E_i^{\mathit{a}}}) - \mathrm{Max}\,(0, \mathrm{E_i^{\mathit{a}}} - \mathrm{E_i})$$

Duncan and Hoffman (1981) replaced educational attainment in the Mincer earnings function with three components (adequately educated, overeducated, undereducated) treated as separate variables, each having potentially different values of the three regression coefficients. Consequently, the earnings function proposed by Duncan and Hoffman (1981) is specified as follows:

$$log\ Y_i = \alpha + \beta_a\, E_i^a + \beta_o\, E_i^o + \beta_u\, E_i^u + X_i \cdot \delta + \varepsilon_i \tag{2}$$

That is, $\beta_a > \beta_o$ and $\beta_u < 0$.

The parameters β_a and β_u represent the returns to adequate education, overeducation, and undereducation, respectively. The common finding is that overeducated workers earn more than their co-workers who possess an adequate level of education within a given occupation. Since education enhances productivity, it is anticipated that $\beta_o > 0$. However, overeducated workers are expected to earn less than adequately educated workers, meaning those whose education closely matches their occupation, so $\beta_a > 0$. Conversely, undereducated workers experience a wage penalty compared to co-workers whose education exactly aligns with the required educational level; therefore, $\beta_u < 0$.

The control variables include experience and experience squared, as it is assumed that there is an inverse U-shaped relationship between experience and the earnings of individuals. Secondly, the gender earnings gap is particularly high in developing countries like Pakistan, as noted by Sabir and Aftab (2007), who observed that, in general, males earn more than females. Therefore, to capture the earning differential across genders, a gender dummy variable is used that takes the value of one for male workers. Moreover, while setting wages, employers also consider individuals' other credentials as indicators of efficiency and productivity. These credentials include human capital, specialized formal training, and on-the-job training. Many studies support this notion, indicating that training leads to higher wages for workers (Winkelmann, 1994; Dearden et al., 2006).

Another important factor contributing to earnings is the individual's geographical location. Major cities provide better earning and learning opportunities because of their advanced infrastructure. Additionally, high competition in these major cities drives workers to improve their skills and level of competence (Glaeser & Maré, 2001). Thus, it is assumed that individuals from major cities have better opportunities to earn compared to those living in smaller cities. The description of these variables is provided in the next section.

To reduce bias in sample selection, we applied the Heckman Model (1979), a two-stage sample selection model, to address the non-random sampling issue. Heckman's model consists of a system of two equations. In the first equation, called the selection model, an individual's decision to participate in the labor market depends on the difference between the wage offer and the reservation wage. Thus, a Probit regression equation is utilized to construct the Inverse Mills Ratio (IMR) for correcting earnings equations for selection bias. For this, we need at least one variable that influences selection but is excluded from the earnings model. Aslam (2009) and Yusnandar (2020) used household demographic variables as instruments in the Heckman selection model, such as the number of children under 7, adults over 60, marital status, and unearned income. Moreover, Comola and Luiz de Mello (2009) considered whether the reference individual's attendance at school serves as a selection variable to determine their participation in the labor market. Therefore, I have included the number of dependents in the household and whether the reference individual is currently receiving education as independent variables for the selection model, i.e., whether the individual is working or not.

To address the problem of unobserved heterogeneity bias related to the observable educational variables included in the empirical specification, we applied the instrumental variable (IV) technique. To apply the IV technique, the instruments must fulfill two conditions: first, each instrumental variable, denoted by z, must be uncorrelated with the error term (exogeneity), that is $\mathrm{Cov}(z, \varepsilon)=0$; and second, the instruments should have a high sample correlation with the endogenous explanatory variable (relevance), that is $\mathrm{Cov}\ (z,\ \mathrm{E}_{\mathrm{i}}^{a},\ \mathrm{E}_{\mathrm{i}}^{\mathrm{o}}, \mathrm{E}_{\mathrm{i}}^{\mathrm{u}}) \neq 0$.

In the presence of three endogenous variables, $\mathrm{E}_{\mathrm{i}}^{a}$, $\mathrm{E}_{\mathrm{i}}^{\mathrm{o}}$ and $\mathrm{E}_{\mathrm{i}}^{\mathrm{u}}$, in Equation (2), we need at least three instruments. Our first two instruments are father's education and the mother's education, as education positively affects the individual's years of education. Moreover, parents' education may not directly affect an individual's earnings, so it is expected to be uncorrelated with the error term. As in the case of developing countries like Pakistan, parents have a great influence on children's upbringing, especially on their education, so it is likely that children of educated parents are strongly encouraged to attain higher levels of education.

For the third instrument, we followed Lee et al. (2006) and utilized the macroeconomic variable, specifically the labor market conditions during the year when individuals were 15 years old. To assess the state of the labor market, we used the unemployment rate that prevailed when a respondent was 15, as this is the time when individuals decide to either pursue further education or enter the workforce. A high unemployment rate at this point is likely to lead individuals to remain in school if they are able to afford it, as the labor market may not be providing suitable job opportunities.

We also took the square of these three instruments to analyze the non-linear impact of these instruments on education choice variables. So, we have six instruments for three endogenous variables appearing as 'independent variables' (adequate education, overeducation, and undereducation), and this is often called the over-identified case. In this situation, we applied GMM under the IV estimation technique. As GMM covers the problem of unobserved heterogeneity with minimum standard error as compared to two-stage least squares (2SLS).

4. Data and Variable Construction

The data for this study are taken from *PSLM 2019-20*, conducted by the Pakistan Bureau of Statistics. The analysis is conducted for earners between the ages of 15 and 60 years. In this data set, occupation classification is available at three and even four digits. This level of disaggregation is expected to give unbiased and relatively accurate results when one wants to measure the mismatch of education and the impact of this mismatch on earnings in the labor market by occupation through the realized method. The construction of various variables used in the study is explained in Table 1, and descriptive statistics of these variables are given in Appendix Table 4.

Table 1: Variable Construction

Log Earnings	=	Log of monthly earnings of individual *i* For self-employed individuals PSLM provides the net earnings of a given occupation of individuals after eliminating all rental cost of capital and land. Moreover, we take those individuals who have positive earnings.
Education Attainment Level	=	Education attainment level defined by International Standard Classification of Education (ISCED)[1] of individual *i*.
Adequate Level of Education	=	Most frequent year of education defined by (ISCED) in each 3-digit level of occupation defined by (ISCO) a given sample measured by mode method.[2]
Over Educated	=	Years of education levels defined by ISCED that is above from adequate level of education in a given occupation, 0 otherwise.
Under Educated	=	Years of education levels defined by ISCED that is below from adequate level of education, 0 otherwise.
Gender	=	1 for Male, 0 for Female
Experience	=	Experience of individual *i* measure through potential experience, that is year of age minus five years assuming that experience starts after five years of schooling.
Married	=	A dummy variable that takes the value equal to one if the individual is married, and zero otherwise.

[1] ISCED 0 = Illiterate, ISCED 1=Primary education, ISCED 2 = Lower Secondary education, ISCED 3 = Secondary education, ISCED 4= Post secondary and Non tertiary education, ISCED 5=Tertiary education, ISCED 6= Graduate and equivalent, ISCED 7= Postgraduate.

[2] The Mode method allows abrupt changes to be captured whereas that the Mean method changes gradually and may produce classification errors (Wen & Maani, 2022).

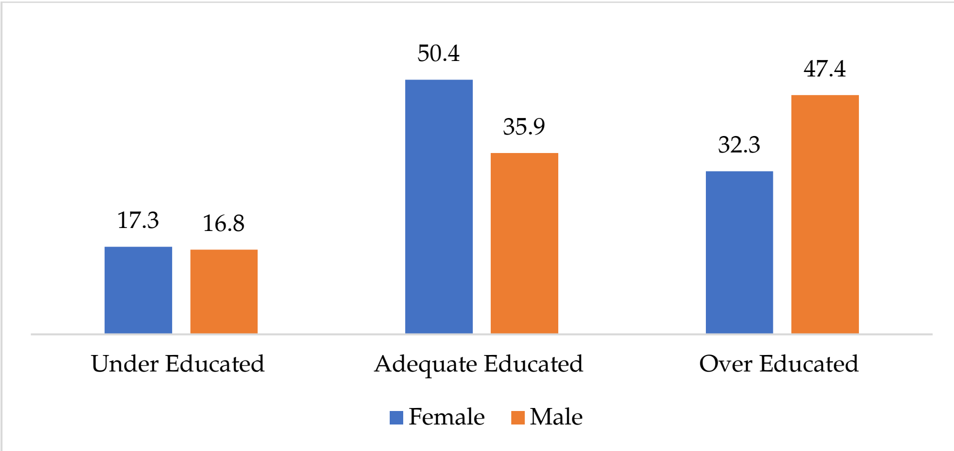
Employment Status	=	A set of two dummy variables. Paid-employee dummy that takes the value equal to one if the individual belongs to paid employee, and zero otherwise. Self-employed dummy that takes the value equal to one if the individual belongs to self-employed, and zero otherwise. The self-employed is set as the reference category.
Big Cities	=	A dummy variable that takes the value equal to one if the individual belongs to big city defined by PSLM, and zero otherwise.
Industry	=	A set of dummy variables. = 1 for Agriculture and Mining, 0 otherwise. = 1 for Construction, 0 otherwise. = 1 for Manufacturing, Electricity and Water Supply, 0, otherwise. = 1 for Retail Trade and Transportation, 0 otherwise. Other services (accommodation and food services, information and communication, financial and insurance activities, professional, scientific and technical activities etc.) as the reference group.
Father's Education Level	=	Father education level is measured by ISCED.
Mother's Education Level	=	Mother education level is measured by ISCED
Unemployment rate	=	Unemployment rate of Pakistan in the year when the individual was at aged 14 years.
Total number of dependents	=	The number of individuals who are not doing any job at household level.
Currently Taking Education	=	1 if individual is currently taking education and 0 otherwise.

5. Descriptive Statistics and Results

The data show that about 63.05 percent of Pakistanis are having educational mismatch, with 46.2 percent over-educated and 16.8 percent under-educated. Figure 1 presents the mismatch in education by gender, which shows that the percentage of over-educated males is higher than that of over-educated females. A possible explanation is that, due to their lower participation, women do not face as much competition in the labor market as do men.

Secondly, compared to men, women tend to have lower economic responsibilities but greater domestic responsibilities, as is often seen in traditional societies. This increases the opportunity cost of labor force participation for women, which may lead to an unwillingness to work among educated women if they cannot find a well-paid job or one that matches their level of education.

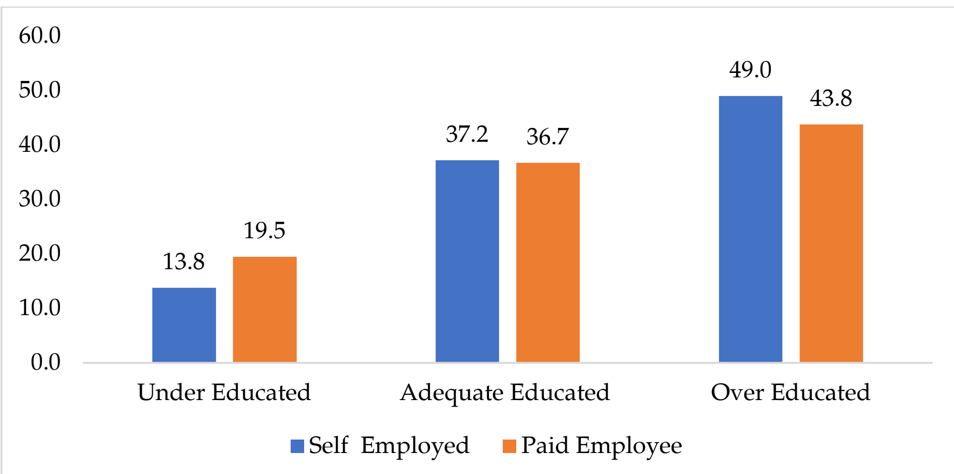
Figure 1: Mismatch of Education by Gender



Source: Author

Referring to employment status, undereducation is observed to be higher for paid employees (19.5%) compared to self-employed individuals (13.8%), as shown in Figure 2. On the other hand, the self-employed are more over-educated than paid employees.

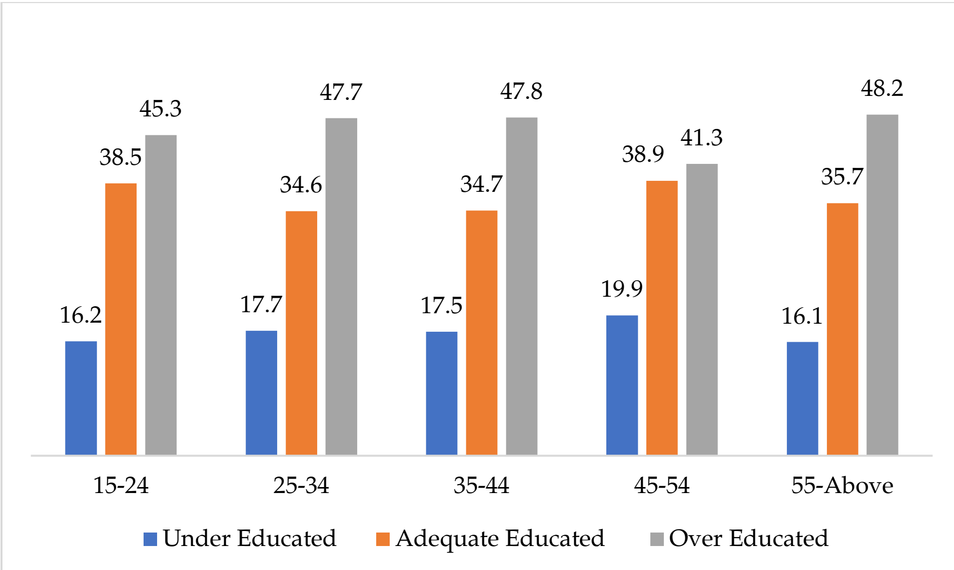
Figure 2: Mismatch of Education by Employment Status



Source: Author

Figure 3 illustrates the mismatch in education by age group. The overeducation rate is relatively low for the middle-aged group, particularly those aged 45-54. An interesting fact is that this group also experiences more instances of undereducation compared to other age groups, which may indicate that, with experience, these individuals have acquired advanced skills that could be more important than their educational credentials.

Figure 3: Mismatch of Education of Age Group



Source: Author

5.1 Empirical Results

Table 2 presents the estimation results for paid employees based on Equation (2) proposed by Duncan and Hoffman (1981). We find that the returns to adequate education listed in column 1 of Table 2 are higher than the returns of the attained education level in the Mincer Earnings Model, as shown in column 1 of Table 5 in Appendix A. This indicates the presence of an education mismatch. The returns of education mismatch in column 1 demonstrate that overeducation yields positive returns, though these are lower than the returns for adequate education. Additionally, the negative returns for undereducation remain remarkably stable across countries and datasets over time (Hartog, 2000).

Table 2: Returns of Education Mismatch on Earnings for Paid Employee

Explanatory Variables	(1) OLS without IMR	(2) OLS with IMR	(3) GMM without IMR	(4) GMM with IMR
Adequate Schooling	0.0984***	0.108***	0.182***	0.177***
	(0.00139)	(0.00187)	(0.00708)	(0.00864)
Over Schooling	0.0670***	0.0786***	-0.0138	-0.0367
	(0.00171)	(0.00226)	(0.0833)	(0.106)
Under Schooling	-0.0858***	-0.0940***	-0.207***	-0.227***
	(0.00209)	(0.00234)	(0.0430)	(0.0491)
Male	0.338***	0.342***	0.426***	0.434***
	(0.00734)	(0.00734)	(0.0195)	(0.0211)
Married	0.0285***	0.0531***	0.0283***	0.00853
	(0.00438)	(0.00539)	(0.00821)	(0.0183)
Experience	0.0173***	0.0175***	0.0137***	0.0104***
	(0.000753)	(0.000753)	(0.00182)	(0.00288)
Experience squared	-0.000177***	-0.000178***	-0.000142***	-9.89e-05**
	(2.10e-05)	(2.10e-05)	(3.61e-05)	(4.79e-05)
Big Cities	0.0353***	0.0549***	0.0102*	-0.0161
	(0.00364)	(0.00442)	(0.00602)	(0.0221)
Agriculture Mining	-0.157***	-0.161***	0.0807*	0.107**
	(0.00754)	(0.00756)	(0.0423)	(0.0494)
Manufacturing	-0.0195***	-0.0195***	0.200***	0.229***
	(0.00564)	(0.00563)	(0.0265)	(0.0336)
Construction	-0.0359***	-0.0358***	0.232***	0.265***
	(0.00599)	(0.00598)	(0.0368)	(0.0447)
Retail Trade & Transport	-0.0613***	-0.0598***	0.118***	0.138***
	(0.00564)	(0.00564)	(0.0203)	(0.0250)
Inverse Mills Ratio		1.556***		-1.920
		(0.199)		(1.552)
Constant	3.451***	3.360***	3.293***	3.442***
	(0.0109)	(0.0160)	(0.0629)	(0.137)
Observations	29,414	29,414	29,414	29,414
R-squared	0.306	0.307		
Hausman endogeneity test			334.971***	379.888***
Hansen J statistics p-value			0.205	0.137

Source: Author.
Note: Standard errors are given in parentheses. The parameters significant at 10%, 5% and 1% levels of significancearne indicated by *, **, and *** respectively.

Our findings indicate that the estimated returns from a year of adequate education are 9.8 percent, while for each additional year of overeducation, these returns amount to 6.7 percent. Meanwhile, undereducated workers face a wage penalty of approximately 8.5 percent for each year of deficit compared to those with an adequate level of education.

Column 2 presents the results of the Duncan and Hoffman model after controlling for sample selection bias. The estimated coefficients of the inverse Mills ratio (Heckman's λ) are statistically significant, indicating that there is a sample selection problem in our model. The positive sign implies that there is a positive selection effect on earnings, suggesting that unobserved variables increase the probability of selection and lead to earnings higher than the average earnings observed in a simple OLS model.

After controlling for sample selection bias in column 2, the estimated coefficients for overeducation and adequate education improve compared to the corresponding estimates in column 1 of Table 2. This indicates that OLS regression estimates are downward biased, although the difference between the OLS and Heckit estimates is small. The increase in the coefficients for overeducated and adequately educated individuals in the Heckit model suggests that including unemployed individuals in the labor market improves earnings, indicating that these individuals possess high skill levels and are involuntarily unemployed. Moreover, our results are consistent with those of previous studies, including Cutillo and Di Pietro (2006), Caroleo and Pastore (2018), and Lee et al. (2016).

Column 3 presents the impact of educational mismatch on earnings after correcting for unobserved heterogeneity bias using the GMM-IV technique. Column 4 presents the results of earnings by incorporating both heterogeneity bias and sample selection bias. Our findings on overeducation confirm the direction of the IV approach, clearly rejecting the human capital compensation hypothesis, which posits that overeducated individuals have positive and significant earnings. These results are also consistent with Robst (1994), Tahlin (2009), and Kleibrink (2016). The results for undereducation, after accounting for heterogeneity bias and sample selection bias in columns 3 and 4, show a positive but statistically insignificant effect. One of the most interesting findings is that adequately educated individuals receive greater returns when we incorporate heterogeneity and sample selection bias, rather than estimating it through the OLS model presented in Table 2. The results of adequate schooling increase from 9.8 percent (Column 1, Table 2) to 17.7 percent (Column 4, Table 2) after considering sample selection bias and heterogeneity bias. It has been argued that overeducated individuals may have low levels of innate ability, as studies assessing ability find that ability and overeducation are indeed negatively correlated (Leuven & Oosterbeek, 2011). Furthermore, the results seem to support the job competition model, where individuals' earnings are based on job characteristics rather than their education level, indicating that overeducation holds no economic value. This finding contrasts with the job assignment model, where wages are determined by both workers' education levels and job characteristics.

Furthermore, the rapid expansion of higher education over a short period led to a heterogeneity of skills in the labor market, as many of the newly established universities could not provide the quality of skills demanded by employers. A significant gap exists between the supply and demand for education in certain areas, with challenges including access, equity, and quality, particularly the weak curriculum models that are poorly aligned with employer requirements. Moreover, due to the deteriorating economic conditions, labor markets cannot absorb the large number of skilled workers that Pakistan produces each year, as unemployment rates among graduates rose from 6% to 17% between 2004 and 2017 (Ahsan & Khan, 2023).

To determine whether the IV analysis yields reliable estimates, the instruments must be valid. For this, the Hansen J test is used to check if the p-value is greater than 0.05, which allows us to accept the null hypothesis that the instruments are valid. Additionally, the results of the first-stage Heckman model in Table 6 and the first-stage GMM model under the IV technique are provided in Table 7 for paid employees and Table 8 for self-employed individuals.

The impact of control variables on earnings for paid employees reveals one of the interesting results regarding the gender wage gap. It has been observed that after controlling for sample selection bias and unobserved heterogeneity bias, the gender wage gap increases. The estimates from the simple OLS model in Table 2, column 1, show that males earn approximately 33.8 percent more than females before correcting for bias. However, after correcting for sample selection bias and unobserved heterogeneity bias, males earn approximately 43.4 percent more than females, which increases the gender wage gap by 9.6 percentage points (43.4 - 33.8). Moreover, the increased gender earning gap after correcting for unobserved heterogeneity bias may indicate that men are perceived as more capable and possess more ability than women.

Turning to experience (and experience squared), it presents an inverse U-shaped relationship for all workers and aligns with the literature. It is also observed that experience pays off more for paid employees when we control for sample selection bias and unobserved heterogeneity bias. The results are consistent with the study conducted on

Pakistan's economy by Bhatti et al. (2018), which shows that after adjusting for unobserved heterogeneity, each additional year of labor market experience contributes to an increase in monthly wages. Training appears to have no economic value, as it does not have a statistically significant impact on earnings. Regarding big cities, they provide better opportunities for employment and earnings due to their developed infrastructure. Our results indicate that people living in big cities earn significantly more than those residing in small cities, though the magnitude of the earning differential between big cities and other locations is relatively small.

Further examining the impact of industries on earnings, our results confirm that individuals in the services sector earn more than those in other industries. However, when we account for sample selection bias and unobserved heterogeneity, the results are reversed, showing that individuals in the agriculture and mining sectors, followed by the construction sector, earn more compared to those in the services sector. This may indicate that industries such as agriculture, mining, and construction often require physical labor and specialized skills, which can command higher wages due to the associated risks, skills, and efforts. The services sector may encompass a broader range of job types, some of which may not necessitate specialized skills, resulting in lower average earnings.

Table 3 presents the results of education mismatch on earnings for self-employed individuals. The estimated coefficients for adequately educated, overeducated, and undereducated individuals, as shown in column 1 of Table 3, are statistically significant and comparable in magnitude to those of paid employees reported in Table 2.

Table 3: Returns to Education Mismatch on Earnings for Self Employed

Explanatory Variables	(1) OLS without IMR	(2) OLS with IMR	(3) GMM without IMR	(4) GMM with IMR
Adequate Schooling	0.0939***	0.104***	0.251***	0.236***
	(0.00270)	(0.00650)	(0.0322)	(0.0349)
Over schooling	0.0686***	0.0785***	-0.0660	-0.157
	(0.00271)	(0.00658)	(0.0718)	(0.167)
Under Schooling	-0.0738***	-0.0836***	-0.207	-0.286
	(0.00421)	(0.00729)	(0.259)	(0.245)
Male	0.400***	0.439***	0.286***	0.139
	(0.0160)	(0.0285)	(0.0393)	(0.232)
Married	0.0234***	0.0425***	0.0389*	-0.0146
	(0.00742)	(0.0138)	(0.0200)	(0.0838)
Experience	0.0170***	0.0196***	0.0122***	0.000277
	(0.00123)	(0.00202)	(0.00381)	(0.0191)
Experience square	-0.000181***	-0.000190***	-0.000142**	-5.78e-05
	(2.85e-05)	(2.90e-05)	(7.17e-05)	(0.000160)
Big Cities	0.0626***	0.0751***	0.0574***	0.0150
	(0.00656)	(0.0100)	(0.0104)	(0.0663)
Agriculture Mining	-0.0917***	-0.0922***	0.313*	0.333***
	(0.0122)	(0.0122)	(0.174)	(0.0886)
Manufacturing	-0.0234*	-0.0235*	0.201**	0.214***
	(0.0122)	(0.0122)	(0.0855)	(0.0495)
Construction	-0.0567***	-0.0567***	0.217*	0.222***
	(0.0143)	(0.0143)	(0.125)	(0.0661)
Retail Trade & Transport	0.00545	0.00567	0.141***	0.155***
	(0.0102)	(0.0102)	(0.0356)	(0.0359)
Inverse Mills Ratio		0.103*		-0.346
		(0.0624)		(0.534)
Constant	3.463***	3.280***	3.416***	4.141***
	(0.0211)	(0.113)	(0.217)	(1.154)
Observations	12,683	12,683	12,683	12,683
R-squared	0.220	0.220		
Hausman endogeneity test			112.228***	114.354***
Hansen J statistics p-value			0.0849	0.1092

Source: Author.
Note: Standard errors are given in parentheses. The parameters significant at 10%, 5%, and 1% levels of significance are indicated by *, **, and *** respectively.

The result of the Inverse Mills Ratio is statistically insignificant at the 10% level of significance, which may again indicate that individuals who are not working are more capable than those who are self-employed. Meanwhile, the Hausman test suggests that there is an issue of endogeneity, and the education variables are not exogenous.

One of the interesting findings is that when we estimate the impact of education mismatch on earnings using the OLS model, the returns for adequately educated self-employed individuals are slightly lower than those for adequately educated paid employees, with a difference of only 0.45 percentage points (9.39-9.84). However, after correcting for unobserved heterogeneity bias and sample selection bias as presented in column 4 of Tables 2 and 3, the returns for adequately educated self-employed workers surpass those of paid employees by 5.9 percentage points (23.6-17.7). This indicates that self-employed individuals are more capable and possess greater abilities, and once we control for the effect of ability, adequately educated self-employed individuals earn more than their adequately educated counterparts in paid employment. Our results align with Bender and Roche (2013), who found that well-matched self-employed men earn more than well-matched salaried men. The literature also suggests that self-employed individuals are often 'jack of all trades,' possessing skills in many areas (Lazear, 2005), which may explain why they can work more efficiently than paid employees in fields that align well with their educational backgrounds.

6. Conclusion

The main contribution of this study is analyzing the impact of education mismatch on earnings across employment status by correcting sample selection bias and unobserved heterogeneity bias in Pakistan, adopting the methodology of Duncan and Hoffman (1981) and utilizing the PSLM 2019-20 data. It has been observed that overeducated individuals earn more than their coworkers in certain occupations that require less education; however, they earn less when compared to their adequately educated counterparts in other occupations, while undereducated individuals experience negative returns. However, after controlling for both unobserved heterogeneity and sample selection bias, overeducation holds no significant economic value. Therefore, measuring the returns of education while ignoring unobserved heterogeneity and sample selection bias yields biased results.

Our results confirm that only a sufficient or "adequate" number of years of education is necessary for a job, and additional years do not contribute significantly. This indicates that education beyond what is required does not prove productive for the individual. The implications of overeducation support the Job Competition Theory, which is a demand-side theory where marginal productivity is considered a fixed characteristic of a particular job, unrelated to the worker's characteristics.

Furthermore, our findings suggest that low-ability individuals may find better job opportunities by investing in higher levels of education; however, this may result in increased unemployment and over-education in the labor market. Thus, the government should work to increase jobs that require lower levels of education, as this strategy may discourage unnecessary pursuit of higher education.

The findings of our study highlight a serious problem in Pakistan's educational system and its connection to the labor market. Unchanged educational policies over the past two decades, which have not accounted for changing market demands, may have contributed to this education mismatch in the labor market. The mismatch in education is a common feature of Pakistan's labor market, with up to 60 percent of workers being either overqualified or underqualified. Therefore, there is an urgent need to promote basic education and skills development to help reduce undereducation. Furthermore, improving university-industry linkages by introducing internship programs and other practical learning opportunities into the tertiary education curriculum can also help to narrow the gap between supply and demand in the labor market. Finally, focusing on quality alongside quantity should be the priority for those in leadership positions, as only then can individuals achieve optimal returns for the years invested in their education.

References

Ahsan, H., & Khan, M. J. (2023). Disaggregating the Graduate Unemployment in Pakistan (No. 2023: 100). Pakistan Institute of Development Economics.

Allen, J. and Van der Velden, R. (2001), Educational mismatches versus skill mismatches: Effects on wages, job satisfaction, and on-the-job search. Oxford Economic Papers 3, 434-452.

Bauer, T. (2002). Educational mismatch and wages: a panel analysis. Economics of Education Review.21, 221–229.

Bender, K. A., & Roche, K. (2013). Educational mismatch and self-employment. Economics of Education Review, 34, 85-95.

Bhatti, H, S., Aslam, M., & Bourdon, J. (2018). Market returns to education in Pakistan, corrected for endogeneity bias. The Lahore Journal of Economics, 23(1),79-96.

Bird, C. (1975). The Case Against College. New York: David McKay.

Caroleo, F.E & Pastore, F. (2018). Overeducation at a glance. Determinants and wage effects of the educational mismatch based on AlmaLaurea data. Social Indicator Research, 37(3), 999-1032.

Clark, B., Joubert, C., and Maurel, A. (2017). The Career Prospects of Overeducated Americans. Journal of Labor Economics, 6(1), 1–29. doi: 10.1186/s40172-017- 0053-4.

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

Dearden, L., Reed, H., & Van Reenen, J. (2006). The impact of training on productivity and wages: Evidence from British panel data. Oxford Bulletin of Economics and Statistics, 68(4).

Dolton, P., & Silles, M. (2008). The effects of over-education on earnings in the graduate labor market. Economics of Education Review, 27 ,125–139.

Dolton, P., & Vignoles, A. (2000). The incidence and effects of overeducation in the U.K. Graduate labor market. Economics of Education Review, 19, 179-198.

Duncan, G. J. and S. D. Hoffman (1981) The Incidence and Wage Effects of Overeducation. Economics of Education Review, 1(1), 75-86.

Freeman, R. (1976). The Overeducated American, Academic Press, New York, NY, USA.

Glaeser, E. L., & Maré, D. C. (2001). Cities and skills. Journal of Labor Economics, 19(2), 316-342.

Groeneveld, S. and J. Hartog (2004), Overeducation, wages and promotions within the firm, Labor Economics, 11: 701-714.

Groot, W., & Maassen Van Den Brink, H. (1997). Allocation and the returns to over-education in the UK. Education Economics, 5(2), 169-183.

Hartog, J. (2000). Overeducation and earnings: Where we are and where we should go. Economics of Education Review,19(2),131–147.

Jovanovic, B. (1979). Job matching and the theory of turnover. Journal of Political Economy, 87 (5, Part 1), 972–990.

Heckman, J. J. (1979). Sample selection bias as a specification error. Econometrica, 47(1):153–61.

Kiker, B., Santos, M. C., & de Oliveiria, M. M. (1997). Overeducation and undereducation: Evidence for Portugal. Economics of Education Review, 16(2), 111–125.

Kleibrink, J. (2016). Inept or badly matched? Effects of educational mismatch in the labor market. Labour, 30(1), 88–108.

Kopri,T., Tahlin, M.(2009). Educational mismatch, wages, and wage growth: Overeducation in Sweden, 1974-2000.Labor economics, 16(2), 183-193.

Lasso-Dela-Vega, E., Sánchez-Ollero, J. L., & García-Pozo, A. (2023). Effects of educational mismatch on wages across industry and occupations: sectoral comparison. International Journal of Manpower, 44(9), 237-255.

Lazear, E. P. (2005). Entrepreneurship. Journal of Labor Economics, 23(4), 649– 680.

Lee, H., Lee, J.W. and Song, E. (2016), Effects of educational mismatch on wages in the Korean labor market, Asian Economic Journal, 30 (4), 375-400.

Leuven, E., & Oosterbeek, H. (2011). Overeducation and mismatch in the labor market. In E. A. Hanushek, S. Machin, & L. Woessmann (Eds.), Handbook of the economics of education, 4 pp. 283–326).

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

Mincer, J, A. (1974). Schooling, Experience, and Earnings, NBER Books, National Bureau of Economic Research, Inc, number minc74-1, June.

Nicaise, I. (2001). "Human capital, reservation wages and job competition: Heckman's lambda reinterpreted", Applied Economics,33(3):309-315

Njifen, I., & Smith, P. (2024). Education-Job mismatch and heterogeneity in the return to schooling: Evidence from Cameroon. Journal of the Knowledge Economy, 15(1), 860-887.

Pritchett, L. (2001). Where has all the education gone? The World Bank Economic Review, 5 (3), 367-391.

Quintini, R. (2011). Right for the job: over-qualified or under-skilled?, OECD Social, Employment and Migration Working Papers 120, OECD Publishing.

Robst, J. (1994) Measurement error and returns to excess schooling. Applied Economics Letters, 1, 142–144.

Rubb, S. (2003). Overeducation in the labor market: a comment and re-analysis of a meta-analysis. Economics of Education Review, 22(6),621-629.

Rumberger, R. W. (1981). The rising incidence of overeducation in the U.S. labor market. Economics of Education Review, 1(3), 293-314.

Sabir, M. & Aftab, Z. (2007). Dynamism in the gender wage gap: Evidence from Pakistan. The Pakistan Development Review, 46 (4), 865–882.

Sattinger, M. (1993). Assignment Models of the Distribution of Earnings, Journal of Economic Literature, 31(2): 831-88.

Sial, M.H., Sarwar, G and Saeed, M. (2019). Surplus education and earnings differentials in Pakistan: A Quantile regression analysis. The Lahore Journal of Economics, 24(2), 93–114.

Sulaimanova, B. (2022). The impact of job-education mismatch on earnings: Empirical evidence from Kyrgyzstan with focus on generation and gender difference. OSCE Academy in Bishkek, June, 11, 1-47.

Takeuchi, M. (2023). Impact of vertical and horizontal mismatches on earnings among highly educated employees in Japan. International Journal of Educational Research Open, 5, 100270.

Thurow L. C. (1975). Generating Inequality, New York: Basic Book.

Tsai, Y. (2010). Returns to overeducation: A longitudinal analysis of the U.S. labor market. Economics of Education Review,29(4), 606-617.

Tran, T. Q., Vu, N. B. T., & Van Vu, H. (2023). Does job mismatch affect wage earnings among business and management graduates in Vietnam?. Research in International Business and Finance, 65, 101982.

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

Winkelmann, R. (1994). Training, earnings and mobility in Germany. Discussion Paper No. 982, Centre for Economic Policy Research.

Appendix

Table 4: Descriptive Statistics

Variable	Mean	Std. Dev.
Monthly Earning	5836.59	19733.300
Log Monthly Earning	9.54	0.824
Grade ISCED	1.34	1.721
Total Dependents	5.24	3.000
Currently Taking Education	27%	0.446
Married	15%	0.361
Male	66%	0.475
Paid Employee	52%	0.500
Experience	11.76	6.318
Agriculture and Mining	28%	0.447
Manufacturing	16%	0.367
Construction	15%	0.356
Retail Trade and Transportation	24%	0.429
Other Services	17%	0.376
Father ISED	1.47	1.753
Mother ISED	0.61	1.270
Unemployment rate	3.07%	1.887
Sample size	56310	

Source: Author's calculations.

Table 5: Returns of Education via the Mincer Earning Model

VARIABLES	(1) Paid Employee	(2) Self Employed
Grade ISCED	0.0888***	0.0745***
	(0.000666)	(0.000989)
Male	0.365***	0.457***
	(0.00341)	(0.00524)
Experience	0.0165***	0.0140***
	(0.000394)	(0.000599)
Experience Square	-0.000192***	-0.000152***
	(6.44e-06)	(9.02e-06)
Married	0.00135	0.00804
	(0.00603)	(0.0105)
Big Cities	0.00232***	0.00354***
	(0.000134)	(0.000247)
Agriculture Mining	-0.273***	-0.158***
	(0.00370)	(0.00515)
Manufacturing	-0.0584***	-0.0567***

VARIABLES	(1) Paid Employee	(2) Self Employed
	(0.00296)	(0.00617)
Construction	-0.134***	-0.0916***
	(0.00307)	(0.00773)
Retail Trade Transportation	-0.0839***	-0.0182***
	(0.00313)	(0.00509)
Constant	3.529***	3.545***
	(0.00708)	(0.0118)
Observations	78,886	67,980
R-squared	0.416	0.247

Source: Author's calculations.
Note: Standard errors are given in parentheses. The parameters significant at 10%, 5%, and 1% levels of significance are indicated by *, **, and ***, respectively.

Table 6: First Stage of the Heckman Selection Model

VARIABLES	(1) Paid Employee	(2) Self Employed
Total Dependents	0.0375***	0.0184***
	(0.00885)	(0.00258)
Currently taking Education	0.150	-0.147***
	(0.130)	(0.0441)
Grade ISCED	0.119***	0.163***
	(0.0167)	(0.00641)
Male	0.0290	0.517***
	(0.0877)	(0.0321)
Married	0.179***	0.251***
	(0.0620)	(0.0222)
experience	0.00944	0.0391***
	(0.00938)	(0.00372)
Experience square	-0.000225	2.67e-05
	(0.000256)	(9.74e-05)
Big Cities	0.208***	0.210***
	(0.0477)	(0.0187)
Constant	1.681***	-1.562***
	(0.125)	(0.0454)
Observations	29,774	26,536

Source: Author's calculations.
Note: Standard errors are given in parentheses. The parameters significant at 10%, 5%, and 1% levels of significance are indicated by *, **, and *** respectively.

Table 7: First Stage Result of GMM for Paid Employees

VARIABLES	(1) Over Education	(2) Adequate Education	(3) Under Education
Male	0.352***	-0.869***	-0.356***
	(0.0285)	(0.0374)	(0.0228)
Married	0.373***	0.265***	-0.187***
	(0.0168)	(0.0220)	(0.0134)
Experience	-0.134***	-0.0815***	0.0553***
	(0.00288)	(0.00378)	(0.00231)
Experience Square	0.00197***	0.00155***	-0.000669***
	(8.42e-05)	(0.000110)	(6.73e-05)
Big Cities	-0.0549***	0.142***	0.0740***
	(0.0146)	(0.0191)	(0.0117)
Agriculture Mining	0.648***	-2.720***	-0.885***
	(0.0267)	(0.0350)	(0.0214)
Manufacturing	0.659***	-1.933***	-0.474***
	(0.0203)	(0.0266)	(0.0163)
Construction	0.838***	-2.571***	-0.743***
	(0.0204)	(0.0267)	(0.0163)
Retail Trade Transportation	0.437***	-1.594***	-0.304***
	(0.0208)	(0.0273)	(0.0167)
Father ISCED	0.129***	0.112***	0.0233**
	(0.0123)	(0.0161)	(0.00983)
Mother ISCED	0.00398	0.222***	0.0134
	(0.0187)	(0.0245)	(0.0149)
Unemployment rate	0.204***	-0.280***	0.0996***
	(0.0109)	(0.0143)	(0.00871)
Father ISCED square	-0.0214***	0.0226***	-0.00616***
	(0.00260)	(0.00340)	(0.00208)
Mother Education Square	-0.00903**	-0.0183***	-0.00329
	(0.00436)	(0.00572)	(0.00349)
Unemployment rate Square	0.0324***	0.0430***	-0.0171***
	(0.00145)	(0.00190)	(0.00116)
Constant	1.337***	4.494***	0.586***
	(0.0382)	(0.0500)	(0.0305)
Observations	29,414	29,414	29,414
R-squared	0.186	0.490	0.133

Source: Author's calculations.
Note: Standard errors are given in parentheses. The parameters significant at 10%, 5%, and 1% levels of significance are indicated by *, **, and *** respectively.

Table 8: First Stage Result of GMM for Self-Employed

VARIABLES	(1) Over Education	(2) Adequate Education	(3) Under Education
Male	-0.297***	0.569***	0.0887**
	(0.0587)	(0.0654)	(0.0388)
Married	0.543***	0.200***	-0.168***
	(0.0264)	(0.0294)	(0.0174)
Experience	-0.149***	-0.0560***	0.0368***
	(0.00450)	(0.00501)	(0.00297)
Experience Square	0.00223***	0.00116***	-0.000385***
	(0.000110)	(0.000123)	(7.27e-05)
Big Cities	0.0618**	-0.0201	-0.0287*
	(0.0245)	(0.0273)	(0.0162)
Agriculture Mining	0.926***	-2.241***	-0.677***
	(0.0411)	(0.0459)	(0.0272)
Manufacturing	0.437***	-1.205***	-0.307***
	(0.0438)	(0.0488)	(0.0290)
Construction	0.494***	-1.582***	-0.462***
	(0.0508)	(0.0566)	(0.0336)
Retail Trade Transportation	0.206***	-0.681***	-0.0831***
	(0.0371)	(0.0414)	(0.0245)
Father ISCED	0.182***	0.0629***	-0.0590***
	(0.0202)	(0.0225)	(0.0134)
Mother ISCED	0.0247	0.123***	-0.0264
	(0.0346)	(0.0386)	(0.0229)
Unemployment rate	0.165***	-0.0880***	0.0299**
	(0.0179)	(0.0199)	(0.0118)
Father ISCED square	-0.0181***	0.0205***	0.00771***
	(0.00443)	(0.00494)	(0.00293)
Mother Education Square	0.00182	0.00139	-0.000163
	(0.00872)	(0.00971)	(0.00576)
Unemployment rate Square	0.0271***	0.0161***	-0.00574***
	(0.00232)	(0.00259)	(0.00154)
Constant	2.189***	2.047***	0.235***
	(0.0758)	(0.0845)	(0.0501)
Observations	12,683	12,683	12,683
R-squared	0.204	0.303	0.112

Source: Author's calculations.
Note: Standard errors are given in parentheses. The parameters significant at 10%, 5%, and 1% levels of significance are indicated by *, **, and *** respectively.