LABOUR MARKET

PIDE Working Papers
No. 2024:1

PIDE
Pakistan Institute of Development Economics

Impact of Education Mismatch on Earnings: Evidence from Pakistan's Labor Market

Henna Ahsan

Pakistan Institute of Development Economics
Islamabad, Pakistan

E-mail: publications@pide.org.pk
Website: http://www.pide.org.pk
Fax: +92-51-9248065

Designed, composed, and finished at the Publications Division, PIDE.

PIDE Working Papers
No. 2024:1

Impact of Education Mismatch on Earnings: Evidence from Pakistan's Labor Market

Henna Ahsan
Pakistan Institute of Development Economics, Islamabad.

PAKISTAN INSTITUTE OF DEVELOPMENT ECONOMICS
ISLAMABAD
2024

C O N T E N T S

List of Tables

List of Figures

ABSTRACT

During the last 20 years developing countries like Pakistan heavily invested in their education sector to increase enrollment at primary, secondary and tertiary levels to boost their human capital. However, in the presence of poor governance institutions, stagnant labor markets and low educational quality, these additional years of schooling do not necessarily translate into enhanced human capital. It has been argued in literature that Human Capital Model based on Mincer Earning model produces biased results as a mismatch of education exists in the labor market. Therefore, present study investigated the impact of education mismatch on earnings by using the methodology of Duncan and Hoffman (1981). For this I used Pakistan Social Living Measurement PSLM (2019-20) data. Our results indicate that though over education yields positive returns, but these are less than adequate level of education. However, after controlling unobserved heterogeneity bias, over education has no positive value. The returns of over education from OLS model might be overestimated if overeducated workers have lower average ability levels. Moreover, results seem to support the job competition model where the earnings of the individuals are based on job characteristics and not on individual's education level.

Keywords: Mismatch Education, Earnings, Labor Market

1. INTRODUCTION

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

There has been a consensus since long that education increases productivity by raising earnings of the individuals. The Theory of Human Capital posits the same stance that each individual is paid according to his/her marginal product. Therefore, Schultz (1961), Becker (1964) and Mincer (1974) concluded that additional years of schooling increase earnings of the individual and so returns to education are positive. However, with the rapid expansion of education, it has been observed that there is also over education in some labour markets, that is education of some workers is beyond and above the level which is required to perform a specific task causing the mismatch of education in occupation (Rumberger, 1981; Hartog, 2000). Further, Bird (1975) predicted that opportunities for new college graduates declined in the labour market, especially during the recession years of 1975 and 1976, leading to a more widespread distrust regarding economic payoffs for the college graduates.

As noted by Pritchett (2001), developing countries like Pakistan during the past 20 years invested heavily in their education sector to increase enrollment at primary, secondary and tertiary levels in order to boost their human capital. However, in the presence of poor institutions, stagnant labour markets and low educational quality additional schooling years do not necessarily translate into enhanced human capital. Therefore, the basic earning model developed by Mincer (1974) that suggests a positive relationship between earnings and years of schooling of an individual may not hold true. Education mismatch leads to misallocation of human capital in labour market and penalises over educated individuals as they get low earnings than workers with similar education but whose education is in accordance with the job requirement. However, these over educated individuals tend to earn higher earnings than their co-workers who are not overeducated. Further, this educational mismatch also affects the undereducated individuals as these individuals receive low earnings when compared with adequately educated individuals (Duncan & Hofman, 1981; Groot & Maassen, 2000; Rubb, 2003 and McGuinness, 2006). Sial et al., (2019) suggested that an unregulated expansion of education without healthy growth of labour market is a warning sign for the policymakers as this leads to earning differentials, and hence income inequality in the labour market.

To analyse the impact of mismatch of education on earnings most of the studies have used the methodology of Duncan and Hofman (1981) which decomposes actual level of education in Mincer earning model into three components (adequate, over and under level of education) to estimate the returns of education (for example, Kiker et al, 1997; Clark et al., 2017 and Sial et al, 2019).

Despite a large number of empirical studies focusing on the impact of education mismatch in labour market on earnings there persist a number of problems that may cause bias in estimating this earning effect. The first is the sample selection bias. Recent studies (Nicaise, 2001; Cutillo & Di Pietro , 2006; Lee et al, 2016; Caroleo & Pastore, 2018) point out the issue of this bias in estimating over education as there lie 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 labour market. For instance, due to prevailing high unemployment rate, individuals may be compelled to take jobs requiring less schooling, rendering high chance of being overeducated if employed (Quintini, 2011; Lee et al, 2016). On the other hand, Ghignoni & Veraschagina (2014) found that unemployment is a voluntary choice; individuals with higher skills set and academic achievements will remain unemployed until they find a suitable job. Therefore, omitting individuals' decisions regarding participation in the labour market may cause a bias when estimating through simple OLS method.

Secondly many recent studies (Dolton and Vignoles, 2000; Kleibrink, 2016) point out that using over education in classical wage regressions depends on the assumption that equally educated individuals possess the same innate ability and, therefore, productivity; leading to unobserved 'heterogeneity bias'. On the other hand, it is possible that ability of individuals may vary even with the same level of educational achievement. Therefore, to ignore the impact of ability while analysing the earning model may cause bias known in literature as omitted variable bias (Baurer,2002; Kopri and Tahlin, 2009). It is widely known in the field of education that productivity is reflected not only by attainment but also through unobserved factors like "ability". Lee et al., (2016) and Bauer (2002) observe that significant increase in education attainment trend may characterise new workers with an increased heterogeneity biased. Unobserved heterogeneity does not only influence the educational attainment of individuals but also the extent they can make use of it in the labour market. Hence, each state — undereducation, overeducation, and being in an educational match — is the result of the decision process to find a suitable job, given the educational attainment. The question in which of these states' respondents end up is influenced by unobserved heterogeneity.

As pointed out earlier, both heterogeneity bias and sample selection bias occur while analysing the impact of education mismatch on earnings. However, it is usually difficult to control both of them at the same time so most of the studies either control sample selection bias or heterogeneity bias. Literature regarding control or correction of unobserved heterogeneity with sample selection bias is not inclusive. In Pakistan limited work (Farooq, 2011; Sial *et al*., 2019; Khan et. 2022) has been done to analyse the impact of education mismatch on earnings. These studies did not incorporate the heterogeneity bias and sample selection bias.

Therefore, the objective of this study is to analyse the impact of education mismatch on earnings from Pakistan perspective by taking into account the biasedness from both sample selection and heterogeneity perspective. To analyse the returns to education mismatch, the study adopts the methodology proposed by Duncan and Hofman (1981) model. To handle the problem of sample selection bias I adopt the methodology of

Heckman (1979) and Genialised Method of Moments (GMM) under instrumental variable (IV) technique in order to address the heterogeneity bias for observable educational variables included in the empirical specification. The study is based on the data of Pakistan Social Living Measurement (PSLM) 2019-20.

The remainder of this study is organised as follow. Section 2 discusses the literature review regarding the impact of education mismatch on earnings, while Section 3 explains the methodology in detail. Data and construction of variables are explained in Section 4, while the descriptive statistics and results of econometric analysis are presented in Section 5 and the study is concluded in Section 6.

2. LITERATURE REVIEW

There are number of studies that explains the education mismatch and the impact of education mismatch on earnings. According to Job Competition Theory (JCT) by Thurow (1975), brings to light the institutional rigidities where an individual's marginal product and earnings depend on the job characteristics. Job allocation in the labour market depends on the availability of both workers and jobs. So, an excess supply of workers may cause workers with high skill set to settle for low jobs as their educational achievements only serve the basic purpose of getting them a job and do not yield any benefits beyond it. The Job Assignment Theory by Sattinger (1993) posits that there is an allocation problem in assigning heterogeneous workers to jobs differing in their complexity. Assignment theory, assumes workers with the same level of human capital are not equally productive; their productivity depends on the job to which they are matched. This implies that both actual and required education levels impact earnings. The other view from Job Search Model of Jovanovic (1979) explains the incidence of overeducation when an individual with high skills starts in a job that is below his or her ability level but over time, he gets job according to their skills.

Therefore, to analyse the impact of education mismatch on earnings different methods have been used and the most popular method that analyses this mismatch is proposed by Duncan and Hoffman (1981). The study reported that overeducated workers earn higher returns than co-workers who are not overeducated, but lower returns than workers with similar education who work in jobs that require the level of education they possess. Further, undereducated workers receive lower earnings when compared to their coworkers having required level of education.

Duncan and Hoffman, (1981) found that in the USA average returns for adequately educated workers was 6 to 10 percent per year and the returns for over-educated workers were estimated between 2.9 to 4.7 percent. The study further found that each additional year of under education reduces earnings on average by 4 percent for under educated individuals. The results were endorsed later by various studies (for example, Groot and Massan, 2000 ; Rubb, 2003; Groeneveld and Hartog, 2004 and Clark et al., 2017). These findings on the returns of over education are consistent with the Assignment Model which explain wage differentials both on the basis of workers' education level and job characteristics (Sattinger, 1993).

2.1. Sample Selection Bias

Heckman (1976) argues that earnings are only observed for employed workers who are not randomly selected, therefore a selectivity bias can arise when estimating earnings equations. Moreover, job search theoretical model considers unemployment is largely a voluntary choice. People usually accept a job which offers wage higher than their reservation wage. Highly skilled individuals prefer to remain unemployed and wait until they find a job that offers them their best expected wage. Contrary to these, less skilled individuals usually wait less and accept the first job offer they receive though it may render them over educated. Therefore, because of this selection bias observed earnings may not be a true reflection of earnings distribution across the whole population.

Many studies like Nicaise (2001), Linsley (2005), Cutillo & Di Pietro (2006), Lee et al. (2016), and Caroleo & Pastore (2018) point out the existence of sample selection bias in the estimation of over-education based on difference in the characteristics of the unemployed and employed workers. Lee et al. (2016) found that after controlling for the sample selection bias the estimated coefficients for the years of education in the earning function for over education increase by approximately 0.2 to 0.5 percentage points for Korean labour market compared to the estimates that do not cater for sample selection bias. This indicates that once we can control the selection bias arising from non-employment, the estimated returns to over education increase. According to the job search theoretical model, unemployment is a voluntary choice and the most skilled graduates prefer to remain unemployed waiting for the best job offer they can get. If employed, they would be less likely to experience overeducation. In this case, once controlling for the selection bias arising from considering non-employment, the wage penalty of those experiencing the educational mismatch might be lower.

Heckman sample selection procedure can be applied as a screening tool to choose among different theoretical interpretations available for over education. The corrected estimates suggest that OLS largely understate the true effect of over-education on labour market earnings. A understates the OLS based coefficient indicates a higher ability level of the un-employed individuals, and this is in line with the expectations of Job Search model (Nicaise, 2001; Lee et al, 2016). Conversely, an overestimate the OLS coefficient suggests a lower ability level of the un-employed individuals, and this is in line with the theory of job competition model, and job assignment model, where the unemployment is high and hence dominated by involuntary component (Caroleo & Pastore, 2018).

2.2. Unobserved Heterogeneity Bias

To analyse the impact of education mismatch on earnings by assuming that education-job mismatches are an exogenous phenomenon, which may not be true due to issue of ‘unobserved heterogeneity’. The bias may arise due to presence of unobserved factors that are correlated with mismatch of education in labour market and earnings. This is called omitted variables bias. For example, the workers with higher skills/ability may have higher levels of education attainment, whereas on the other side ability/skills that are not observed and, hence included in error terms, also effect the earnings of the individuals. So, these unobserved factors are likely to cause bias (Allen and Van der Velden, 2001; Dolton and Silles, 208).

A large body of literature on managing omitted variable bias uses IV technique methods (Robst, 1994 ; Groot and Maassen Van Den Brink, 1997; Korpi and Tåhlin, 2009; Caroleo and Pastore, 2018; Kleibrink, 2016 and Lee et al, 2016) and fixed-effect model (Bauer, 2002; Dolton and Silles, 2008; Tsai, 2010). Moreover, many studies, like Allen and Van der Velden (2001), Sohn (2010) and Kleibrink (2016), use ability or skills directly OLS estimation to control ability bias in the specification.

Dolton and Siles (2008) address the impact of over-education on earnings for United Kingdom (UK) economy by accounting for omitted ability bias using fixed effect model and addressing measurement error by instrument variable IV technique. The result of fixed effect model shows that over educated face more penalty in terms of earnings and OLS estimates are somewhat upward biased, suggesting that over-education and ability are negatively correlated. Moreover, to control bias both from omitted ability bias and measurement error fixed effect IV method is used, where the estimate suggests that overeducation reduces earnings by 35–40 percent. However, they suggested that over-educated graduates may still get higher earnings than coworkers whose education is exactly match in occupation. Therefore, for UK economy they infer that overeducation is somehow beneficial.

Some studies also attempted to address unobserved heterogeneity bias through instrument variable IV technique like Kopri and Tehlin (2009) for Swedish economy Kleibrink (2016) for German and Lee et al. (2016) for Korea. All these studies found that results come with some notable caveats and provide little support for compensation hypothesis that over educated does not get positive returns on their extra year of schooling which is more than their required level of their job. They concluded that year of overeducation have no effect on pay and only type of job has real importance.

From Pakistan perspective Farooq (2015) analyse the impact of education mismatch on earnings via controlled the unobserved heterogeneity for over education of graduates by using the workers' job satisfaction. The study divided overqualified into two categories: 'apparent' overqualified, workers who are satisfied over their mismatch; and 'genuine' overqualified, consisting of dissatisfied graduate workers. The regression results are consistent with the earlier studies (Chevalier, 2003; Chevalier & Landley, 2006) that genuine overqualified face more wage penalties as compared to apparently overqualified.

3. EMPIRICAL MODEL

To quantify the effect of educational mismatch, i.e. over, adequate and under education on earnings, the study adopts an extended Mincer (1974) earnings function introduced by Duncan and Hoffman (1981). I started with the basic Mincer earnings function:

$$\log Y_i = \alpha + \beta\, E_i + \delta\, X_i + \varepsilon_i \qquad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (1)$$

Where Yi is the log of earnings of the ith individual, E_i is years of education attainment level and the vector Xi 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 over education (E_i^o) and years of under education (E_i^u).

To measure adequate level of education in a given occupation statistical method provided by Kiker et al. (1997) has been used. In which adequate education is measured by mode of education is classified as International Standard Classification of Education (ISCED) in given occupation classify by 3-digit level of International Standard Classification of Occupation (ISCO). Each individuals having educational attainment exactly equal to what is required against an occupation are classified as Adequately educated (E_i^a).

Whereas over educated and under educated is defined as ,

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

Therefore, the following identity holds;

$$E_i = E_i^a + \text{Max}(0, E_i - E_i^a) – \text{Max}(0, E_i^a - E_i)$$

Duncan and Hoffman (1981) replace education attainment replaced in the Mincer earnings function by these three components (adequately educated, over educated, under educated) as separate variables with potentially different values of the three-regression coefficient. Accordingly, the earnings function of Duncan and Hoffman (1981) is specified as follows.

$$\log Yi = \alpha + \beta_a\ E_i^a + \beta_o\ E_i^o + \beta_u\ E_i^u + Xi \cdot \delta + \varepsilon i \qquad \ldots \qquad \ldots \qquad \ldots \qquad (2)$$

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

The parameters β_a, β_o *and* β_u are the returns to adequate, over and under education respectively. The usual finding is that the over educated workers earns more than the co-workers who have adequate level of education within a given occupation. Since education raises productivity, it is expected that β_o is > 0. However, over educated workers are expected to earns less than the adequately educated workers i.e., the workers whose education is well matched according to their occupation requirement, so $\beta_a > \beta_o$. On the other hand, undereducated workers suffer a wage penalty compared to co-workers whose education exactly matches with the required educational level, therefore $\beta_u < 0$.

The control variables include experience and experience square, as it has been assumed that there is inverse U-shaped relationship between experience and earnings of individual. Secondly, gender earning gap is high especially in developing countries like Pakistan as has been observed by Sabir & Aftab (2007) that in general male earn more than female. Therefore, to capture the earning differential across gender, gender dummy variable is used that takes the value equal to one for male workers. Moreover, while setting wages employers also take into account individuals' other credentials as indicators of efficiency and productivity. These credentials include of human capital include specialised formal trainings and or on-the-job trainings. Many studies support the same notion that trainings result in raising wages of the workers (Winkelmann ,1994; Dearden,2006). Therefore, I have assumed that individual who attend any vocational training will be more productive in terms of their earnings.

The other important factor that contributes in earnings is the geographical location of the individual. Big cities offer better earning and learning opportunities due to their advanced and developed infrastructures. Further, high competition in these big cities compels workers to enhance their skills and competency level. Therefore, it is assumed that individual who belong to big cities have better opportunities to earn as compared to those living in small cities. The description of these variables is provided in the next section.

To reduce the bias sourced from sample selection, I applied Heckman Model (1979) two stage sample selection model to address the non-random sampling issue. Heckman's model considers a system of two equations. In the first equation, called selection model individual chooses either to work in labour market or not depends on the difference between wage offer and the reservation wage. Thus, Probit regression equation is used to construct the Inverse Mills Ratio (IMR) for the purpose of correcting earning equations for selection bias. For this we need at least one variable that matters for selection but is excluded from earnings model. So, I have taken marital status as independent variable for selection model i.e., whether the individual is working or not. It is generally assumed that if a male (female) individual is married then there is more (less) chance that he (she) would accept be working as being married increases his responsibilities of earnings (homecare).

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

In the presence of three endogenous variable $\mathrm{E_i^{\mathit{a}}}$, $\mathrm{E_i^o}$ and $\mathrm{E_i^u}$ in equation (2) we need at least three instruments. Korpi and Tåhlin (2009) applied four instruments that relate to the respondents' youth: place of residence, the number of siblings, economic problems, and family disruption. In a similar study Kleibrink (2016) also made use of instrumental approach namely number of siblings, biological parents and macro-economic variables like unemployment rate in the country when the individual is in his/her 15 years of age.

Our first instrument is parents' education, as parents' education positively affects the individual's years of education. Moreover, the parent education may not directly affect individual's earnings, so it is expected to be uncorrelated with error term. As in case of developing countries like Pakistan parents has 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. Lee et al., (2016) also constructed instrument for his endogenous variable over educated, by taking the dummy where respondent has more education than his mother, in order to analyse the impact of education mismatch on earnings for Korean's labour market. Similarly Card (1993), Ashenfelter and Zimmerman (1997), and Aslam (2009), among many others, use parental education as instruments to analyse the impact of education on earnings distribution. However latest literature argues that parental education should be taken as a control variable while measuring the impact of education on earnings instead of using it as an instrument. As Gong et al (2022) maintained that parental education has a positive correlation with dependent variable i.e earnings which shows that the individual whose father is more educated may have an unobservable networking which helps in their earnings beside education. Moreover, literature also supports the parental education as an instrument of education variable arguing that a good instrument should be correlated with the endogenous regressor for reasons that can be verified and explained, and uncorrelated with the outcome variable for reasons beyond its effect on the endogenous regressor. Further due to limited information available in secondary data, that can provide the good information for individual education outcome, we have used parental education as an instrument and to check whether the IV analysis yields reliable estimates or instrument is valid we will apply the Hansen J test.

Our second instrument is the average level of education attained (defined by ISCED) in the enumeration block where the individual lives. The reason for using this instrument as explained in Bhatti et al. (2018) is to have the combined effect of several commonly used exogenous factors, which include social environment, distance from school, number of educational institutes in e area and general trend towards education.

For the third instrument I followed Lee et al. (2006) and used the macroeconomic variable, specifically the condition of labour market in the year when individual had age of 15 years. To measure the labour market condition, I used the unemployment rate prevalent at the time a respondent was 15 years of age, because at this stage individuals decide either to continue education or enter the labour market. A high unemployment rate at this time is likely to make individuals to stay in the schooling system if they can afford as the labour market may not be offering suitable job opportunities.

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

4. DATA AND VARIABLE CONSTRUCTION

The data for this study are taken from *PSLM 2019-20* conducted by Pakistan Bureau of Statistics. The analysis is done for the earners whose age is between 25 to 60 years assuming that most of the individuals complete their sixteen years of education at the age of 25 years and retire at the age of 60 years. The study uses PSLM data 2019-20. 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 mismatch of education and the impact of this mismatch on earnings in labour market by occupation through realised method.

The construction of various variables used in the study is explained in Table 1 and descriptive statistics of these variable are given in Appendix Table 3.

5. DESCRIPTIVE ANALYSIS

The data shows that about 54.8 percent Pakistanis are having educational mismatch with 40 percent over educated and 14 percent under educated. Figure 1 shows that the prevalence of over education is higher in urban areas than in rural areas, which indicates that quite a large number of workers in urban areas possess education more than what is required by the labour market and, hence, cannot get work according to their skills and knowledge. Further Pakistan is witnessing a continuous rural to urban migration because urban areas are much more advanced and equipped with superior infrastructure offer hope for better economic opportunities. However, extensive migrations may crowd the labour market and as a result increased population may have reduced job and business opportunities, compelling individuals to work as overeducated.

Table 1

Variable Construction

Log Earnings	=	Log of monthly earnings of individual i.
Education Attainment Level	=	Education attainment level defined by International Standard Classification of Education (ISCED) of individual i.
Adequate Level of Education	=	Most frequent year of defined by International Standard Classification of Education (ISCED) in each 3-digit level of occupation defined by International Standard Classification of Occupation (ISCO) a given sample measured by mode method.
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.
Training	=	A dummy variable that takes the value equal to one if the individual has ever attended any vocational training during the last year, and zero otherwise.
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 define 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.
Parents Education Level	=	Average of father and mother education attainment level defined by ISCED.
Education at Stratum	=	Average year of education level at enumeration block.
Unemployment rate	=	Unemployment rate of Pakistan in the year when the individual was at aged 15 years.
Marital Status	=	A dummy variable that takes the value equal to one if the individual is Married, and zero otherwise.

Fig. 1. Mismatch of Education by Region

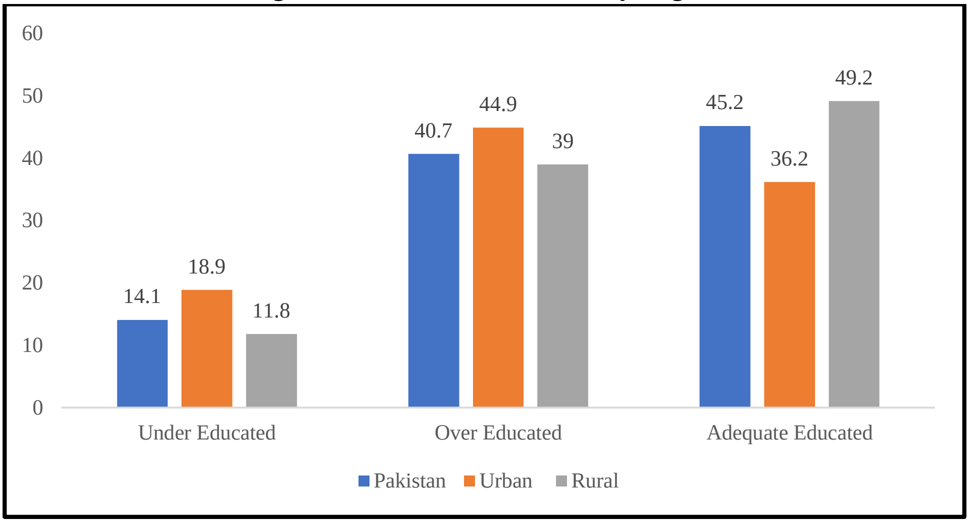
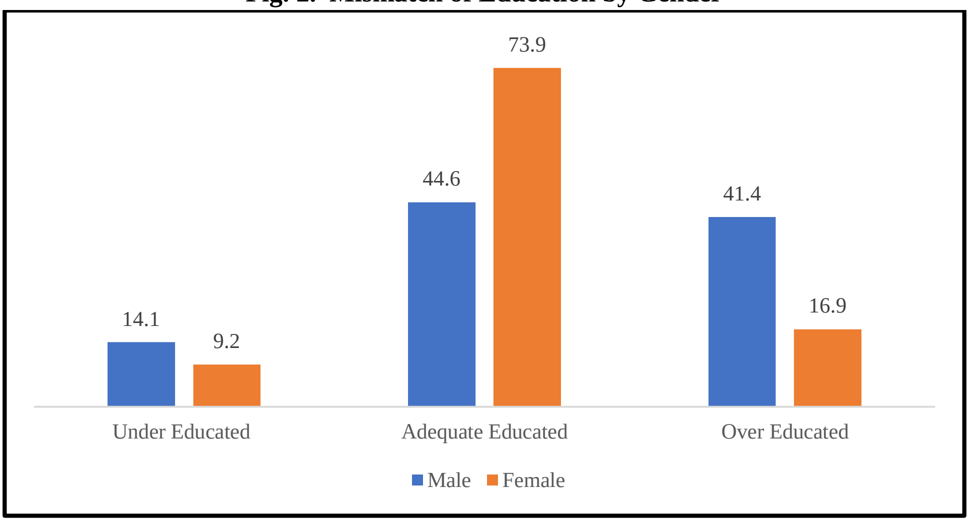


Figure 2 presents the mismatch of education by gender, which shows that the percentage of overeducated males is more than twice the percentage of overeducated females. A possible reason is that because of lower participation rate of women do not face as much competition in the labour market as faced by men.

Fig. 2. Mismatch of Education by Gender



Secondly, compared to men, women tend to have lower economic responsibilities but higher responsibilities at home as is seen in traditional societies. This increases the opportunity cost of labour force participation for women, which may cause an unwillingness to work among educated women if they do not find a well-paid job or a job not matching their level of education as is seen in the case of Italy (Cutillo & Di Pietro, 2006) and Ghana (Herrera and Merceron, 2013).

5.1. Empirical Results

It has been observed that without addressing the problem of sample selection bias and unobserved heterogeneity the results estimated through simple OLS model may give bias results. Therefore, in order to address the sample selection bias coming from the individual's decision to work, I adopted two-stage sample selection model proposed by Heckman's (1979). The results of Heckman sample selection model have been given in Appendix A Table 5. These estimates contribute in constructing the sample selection bias variable called Inverse Mills Ratio IMR (λ). Where dependent variable is dichotomous showing whether individual accepts to work/employed or unemployed and our instrument variable is marital status and independent variables for sample selection model are male, highest level of education, training, experience and its square and the city to which individual belongs.

The second source of bias is due to unobserved heterogeneity. In order to handle the unobserved heterogeneity, I made use of instrumental variables (IV) technique. The IV estimator is less efficient than OLS when explanatory variables are exogenous as IV estimates can have very large standard errors. Therefore, it is useful to have a test of endogeneity of an explanatory variable that shows whether IV is necessary or not before incorporating the unobserved heterogeneity bias. In order to check the endogeneity, the Hausman test is used and the p-value of Hausman test in table 2 rejects the null hypothesis that our regressors (adequate educated, under educated and over educated) are exogenous therefore there is a need to address heterogeneity bias.

Table 2 presents the estimation results of above equation (2) proposed by Duncan and Hoffman (1981), where years of attained education are decomposed into three components: years of over, adequate and under educated with other control variables like male, experience, training, geographical variable and industry dummies. Column (1) estimates the impact of education mismatch on earnings by OLS, whereas column 2 presents the results after incorporating the sample selection model with Heckman Model, column 3 give results after correction of unobserved heterogeneity bias with GMM and column 4 presents the result of education mismatch on earnings after correction of sample selection biased and unobserved heterogeneity bias.

We find that the return of adequate education given in column1 table 2 are higher than the returns of attained education level in mincer earning model given in column 1 of table 4 Appendix A. It shows the presence of educational mismatch. Returns of education mismatch in column 1 show that over education has positive returns but these are lower than the returns to adequate education, and negative returns for under education are remarkably stable across countries and datasets over the time (Hartog, 2000). Our finding points out that estimated returns from a year of adequate education are 10.3 percent, whereas that for each addition year of over education these are 6.0 percent. Meanwhile, undereducated workers suffer a wage penalty of approximately 6.9 percent for each year of deficit, compared to those with the required level of education. This indicates that overeducated individuals who are working in jobs/businesses that demand less education than their actual education receive higher earnings than their coworkers with adequate education by approximately 6 percent. As education improves the productivity positively without regard for nominal requirements of education, therefore years of over schooling raises earning of individual. However, overeducated gets low returns as compared to well-matched education by (10.3- 6) approximately 4.3 percentage points.

Table 2

Returns of Education Mismatch on Earnings

Explanatory Variables	(1) Duncan and Hoffman Model	(2) Duncan and Hoffman Model with IMR	(3) GMM- IV Technique	(4) GMM with IMR
Adequate Education	0.103***	0.0996***	0.228***	0.226***
	(0.000661)	(0.000747)	(0.0125)	(0.0134)
Over Education	0.0637***	0.0594***	0.0205	0.0289
	(0.000746)	(0.000844)	(0.0438)	(0.0481)
Under Education	-0.0834***	-0.0794***	0.139	0.223
	(0.00117)	(0.00123)	(0.136)	(0.158)
Male	0.419***	0.330***	0.520***	0.337***
	(0.00295)	(0.00868)	(0.0289)	(0.0507)
Experience	0.0148***	0.0115***	0.0224***	0.0150***
	(0.000335)	(0.000454)	(0.00240)	(0.00354)
Experience Square	-0.000169***	-0.000122***	-0.000362***	-0.000254***
	(5.25e-06)	(6.76e-06)	(5.60e-05)	(7.19e-05)
Training	0.00328	0.00960*	0.00406	0.0128
	(0.00549)	(0.00552)	(0.0226)	(0.0242)
Paid Employee	-0.0984***	-0.0978***	-0.103***	-0.0965***
	(0.00186)	(0.00186)	(0.0119)	(0.0134)
Big Cities	0.00260***	0.00241***	0.00157***	0.00117**
	(0.000124)	(0.000125)	(0.000490)	(0.000540)
Agriculture Mining	-0.154***	-0.154***	0.439***	0.489***
	(0.00307)	(0.00307)	(0.0765)	(0.0900)
Manufacturing	-0.0182***	-0.0176***	0.413***	0.442***
	(0.00297)	(0.00297)	(0.0430)	(0.0499)
Construction	-0.0739***	-0.0735***	0.484***	0.529***
	(0.00321)	(0.00321)	(0.0685)	(0.0804)
Retail Trade and Transportation	-0.0414***	-0.0415***	0.213***	0.225***
	(0.00269)	(0.00269)	(0.0248)	(0.0277)
Inverse Mills Ratio		-0.255***		-0.399***
		(0.0235)		(0.101)
Constant	3.562***	3.739***	2.875***	3.158***
	(0.00657)	(0.0176)	(0.141)	(0.142)
Observations	146,866	146,866	19,372	19,372
R-squared	0.338	0.338		
F-stat				
Hausman endogeneity test			201.56	176.88
Hansen J statistics p-value	5766	5376	480.5*** 0.108	490.0*** 0.078

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

Column2 presents the results of Duncan and Hoffman model after controlling the sample selection bias. The coefficients of inverse mills ratio (Heckman's λ) estimated are statistically significant which indicates that there is sample selection problem in simple in our model. Whereas negative sign implying that there exists a negative selection effect on earnings that indicates there are unobserved variables which increase the probability of selection and lower the earnings than average earnings observed in simple OLS model.

After controlling for sample selection bias in column 2 the estimated coefficients of the over education and adequate education is less, as compared to the corresponding estimates in columns (1) in Table 2. This indicates that OLS regression estimates are upward biased, although the differences between the OLS and Heckit estimates are small. Our findings are aligned with the previous studies addressing sample selection bias in the estimation of educational mismatch (Nicaise, 2001; Cutillo & Di Pietro, 2006; and Caroleo & Pastore, 2018). The decrease in coefficient of over educated and adequate educated in Heckit model explains that when we add the unemployed individuals in labour market it decreases earnings suggesting that these unemployed individuals are of a lower skill level and the individuals who are more competent are doing the jobs. Our results support and are consistent with the assumptions of the job competition model and assignment theory. As in job assignment model mismatches probably exist as the skill requirements of the assigned position may not fit well with those acquired by the workers. So, when they are allowed to work as over educated it reflects the lower ability of these workers who accept a job offer.

Column 3 presents the impact of education mismatch on earnings after correcting the unobserved heterogeneity bias through GMM-IV technique. Whereas column 4 presents the results of earnings by incorporating both heterogeneity bias and sample selection bias. Our results for over education confirm the direction of the findings of the IV approach, clearly rejecting the human capital compensation hypothesis that over educated has positive and significant earnings and these findings are also consistent with (Robst, 1994; Tahlin 2009; Kleibrink 2016). The result of under education after cooperating the heterogeneity bias and sample selection bias in column 3 and 4 is positive but statistically insignificant. Whereas one of the most interesting findings is that adequate educated get more returns when we incorporate heterogeneity and sample selection bias rather than estimate it through OLS model given in table 2. As the results of adequate schooling are increased from 10.3 percent (column1 table 2) to 22.6 percent (column 4 table 2) after incorporating sample selection bias and heterogeneity bias. It has been argued that over educated individuals may have low levels of innate ability as the studies that asses ability measures find that ability and over schooling are indeed negatively correlated (Leuven & Oosterbeek, 2011). Over-educated workers may be less able in terms of skill development hence their lower wages may reflect their lower ability and productivity (McGuiness, 2006). Hence, the returns of over education from OLS model might be overestimated. However, if overeducated workers have lower average ability levels and once unobserved heterogeneity and sample selection bias are controlled, most researchers find that the wage penalty associated to over education further increases (Cutillo and Di Pietro, 2006). Moreover, results seem to support the job competition model where the earnings of the individuals are based on job characteristics and not on individual’s education level and in this case over education has no economic value and these are contrary to job Assignment Model where wages are determined by both workers education level and job characteristics.

To address whether the IV analysis yields reliable estimates the instruments should be valid. For this Hansen J test is used where p-value is greater than 0.05, so we accept the null hypothesis that instruments are valid.

The result of control variables like male suggest that male earns relatively[1] 52 percent more than female in column 1. Results for gender biasedness are consistent with Sarwar and Sial, 2012. Lower wages for females can also result from missing continuity in their working experience as they have to look after their homes and children as well (Ahsan and Idrees, 2014). However, this earnings gap has been decreased when we correct the problem of sample selection bias by 39 percent given in column 2. This shows that women who are unemployed are more skilled and educated, so when I corrected this sample selection bias this reduces the gender earning gap. Similarly, after correcting the unobserved heterogeneity bias and sample selection bias the gender earning gap has been decreased by 12 percentage points (52.04 – 40.07). Once the selection bias and unobserved heterogeneity bias is corrected, gender wage gap decreases. These results indicate that without selectivity and unobserved heterogeneity results are upward biased. It indicates the perceived discrimination against females in the labour market. The females who are unemployed are more skilled and capable. The coefficients of the experience variable and its square term suggest an inverted U-shaped relationship between experience and earnings. The term of quadratic experience is included to capture the concavity of the earnings profile. Assuming all other variables are held constant, it takes 40 to 43 years of experience approximately depending on model to reach an individual's earnings at a maximum level. Moreover, training seems to have no economic value as it is not statistically significant. However, our available data only captures details of trainings done by the individuals in last 12 months so it does not help us to find the impact of any trainings beyond one year period. There are studies from Pakistan like Nasir & Nazli (2010) which show that more than one year of training significantly increases earnings of the individual.

Earnings of paid employees are relatively 10 percent less than the earnings of self-employed workers. The earnings differentials with respect to employment status are statistically significant and are consistent with the conclusions drawn in Kurosaki (2001). Whereas Kurosaki (2001) analyse the earnings gap for rural region of Pakistan, where as our analyses is based on for whole economy. Moreover, our findings are consistent with Martin (2013) who also concluded that in case of Germany self-employed earns more than wage employed.

Moreover, big cities provide more opportunities in terms of employment and earnings, due to their developed infrastructures, so our results show that people living in big cities earn significantly more than those living in small cities. Turning to the impact of Industries on earnings, our results confirm that individuals who belong to services sector earn more than those working in other industries. It has been argued that during the past few decades Pakistan's economic structure has moved quite rapidly towards services sector, especially at the cost of agricultural sector, services sector not only provides better employment opportunities but also provides more earnings than other trades.

---

[1]Here our dependent variable is log of earnings and independent variable is categorical variable so we cannot interpret it as slope at a point or the rate of change in log Y per unit change in independent variable. Since independent variable is not a continuous variable, derivative with respect to categorical variable does not exist. Hence, we take relative change of categorical variable, for male with reference to female variable relative change is as fellow.

$$\frac{\widehat{Y_f}-\widehat{Y_m}}{Y_f}100 = \left(e^{\beta}-1\right)100$$

6. CONCLUSION

The main contribution of this study is to analyse the impact of education mismatch on earnings by correcting sample selection bias and unobserved heterogeneity bias in Pakistan by adopting the methodology of Duncan and Hoffman (1981) and using the data of PSLM 2019-20. It has been observed that over educated individuals earn more than their coworkers in given occupation that requires less education, however earn less when compared to their adequate educated counterparts working in other occupations. Further under educated face negative returns. However, after controlling both unobserved heterogeneity and sample selection bias the over education has no significant economic value. Therefore, measuring returns of education while ignoring the unobserved heterogeneity and sample selection bias gives biased results.

Our results ascertain that only adequate years of education are really required for a job payoff, and additional years of education do not have any significant contribution. This means that education over and above than what is required does not prove to be productive in individual level.

The results of over education support the Job competition theory which is a demand-side theory, where marginal productivity is taken as a fixed characteristic of a particular job and is not related to the characteristics of the worker characteristics. Moreover, our results suggest that low-ability individuals may see better job opportunities by investing more in acquiring high level of education however this may lead to more unemployment and over education in the labour market. So, government should try to increase jobs which require low levels of education as this strategy may stop people from unnecessary pursuit of higher of education.

Findings revealed in our study highlight a serious problem in the Pakistan's educational system and its link to the labour market. Unchanged educational policies in past two decades, that did not keep in consideration the changing market demands, may have led to this education mismatch in labour market. Mismatch of education is a common feature of the Pakistan's labour market, with up to 50 percent of workers working as over or under educated. Hence there is a dire need to promote the basic education and skills development that might help to reduce under education. Further, improvement of university industry linkage by introduction of internship programs and other practical learning opportunities in the tertiary education curriculum can also help to narrow down the gap between the supply and demand in the labour market. Finally, with focusing on quantity of education quality should also be the focus of people at helm as only then individuals can have optimum returns for the number of years invested in their education.

APPENDIX A

Table 3

Descriptive Statistics

Variables	Mean	Std. Dev.
Monthly Earning	26017.85	44223.97
log Monthly Earning	4.25	0.37
Male	0.92	0.28
Highest Education (years)	5.06	4.48
Education Attainment Level (Grade ISCED)	1.73	1.88
Over Educated (years)	0.87	1.27
Under Educated (years)	0.29	0.82
Adequate educated (years)	1.21	1.91
Training	0.02	0.14
Paid Employee	0.54	0.50
Experience	28.70	11.63
Married	0.88	0.31
Agriculture and Mining	0.25	0.44
Manufacturing	0.15	0.35
Construction	0.14	0.34
Retail trade and Services	0.25	0.43
Other Services	0.21	0.41
Sample Size	146,866	

Table 4

Returns of Education via Mincer Earning Model

Variables	(1)
Grade ISCED	0.0845***
	(0.000572)
Male	0.405***
	(0.00297)
Experience	0.0146***
	(0.000338)
Experience Square	-0.000159***
	(5.30e-06)
Training	0.00438
	(0.00554)
Paid Employee	-0.0963***
	(0.00187)
Big Cities	0.00270***
	(0.000125)
Agriculture Mining	-0.218***
	(0.00284)
Manufacturing	-0.0729***
	(0.00283)
Construction	-0.136***
	(0.00302)
Retail Trade Transportation	-0.0719***
	(0.00266)
Constant	3.615***
	(0.00654)
Observations	146,866
R-squared	0.324

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

Table 5

First Stage of Heckman Selection Model

Variables	Coefficient
Male	0.993***
	(0.0116)
Married	0.406***
	(0.0163)
Grade ISCED	0.0251***
	(0.00120)
Training	-0.112***
	(0.0314)
Big Cities	0.00441***
	(0.000923)
Experience	0.0395***
	(0.00207)
Experience Square	0.000557***
	(3.11e-05)
Constant	-0.502***
	(0.0328)
Observations	159,955

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

REFERENCES

Ahsan, Henna & Idrees, M. (2014). Impact of health on earnings: Individual and district level analysis for Pakistan. (MPRA Paper).

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

Ashenfelter, O., & Zimmerman, D. J. (1997). Estimates of the returns to schooling from sibling data: Fathers, sons and brothers. *Review of Economics and Statistics*, *79*(1), 1–9.

Aslam, M. (2009). Education gender gaps in Pakistan: Is the labour market to blame? *Economic Development and Cultural Change*, *57*(4), 747–784.

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

Becker, G, S. (1964). Human capital: A theoretical and empirical analysis, with special reference to education. New York: National Bureau of Economic Research; distributed by Columbia University Press.

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.

Card, D. (1993). Using geographic variation in college proximity to estimate the return to schooling Cambridge, MA: National Bureau of Economic Research. (NBER Working Paper No. 4483).

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 Labour 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 labour market. *Economics of Education Review*, *27*, 125–139.

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

Duncan, G. J. & Hoffmann, S. D. (1981). The incidence and wage effects of overeducation. *Economics of Education Review*, *1*(1), 75–86.

Farooq, S. (2015). Job mismatches in Pakistan: Is there some wage penalty to graduates? *Pakistan Development Review,54*(2), 147–164.

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

Groeneveld, S. & Hartog, J. (2004). Overeducation, wages and promotions within the firm. *Labour Economics*, *11*, 701–714.

Gong, B. (2019). Like father like son? Revisiting the role of parental education in estimating returns to education in China. *Review of Development Economics*, *23*(1), 275–292.

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

Groot, W., & Maassen Van Den Brink, H. (2000). Overeducation in the labour market: A meta-analysis. *Economics of Education Review*, *19*, 149–158.

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 labour market. *Labour*, *30*(1), 88–108.

Kopri, T. & Tahlin, M. (2009). Educational mismatch, wages, and wage growth: Overeducation in Sweden, 1974-2000. *Labour Economics*, *16*(2), 183–193.

Kurosaki, T. (2001). Effects of human capital on farm and non-farm productivity in rural Pakistan. World Bank, Washington, DC. (Working Paper).

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

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

Nasir, Z. M. & Nazli, H. (2010). Education and earnings in Pakistan. (Working Papers & Research Reports).

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

Pritchett, L. (2001). Where has all the education gone? T*he 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 labour 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. labour 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.

Schultz, T. W. (1961). Investment in human capital. *The American Economic Review*, *51*(1), 1–17.

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

Thurow L. C. (1975). *Generating inequality*. New York: Basic Book.

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

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

Pakistan Institute of Development Economics
Post Box No. 1091, Islamabad, Pakistan

www.pide.org.pk