

Surplus Education and Earnings Differentials in Pakistan: A Quantile Regression Analysis

Maqbool H. Sial^{*}, Ghulam Sarwar^{}, and Mubashra Saeed^{***}**

Abstract

This study empirically investigates the effect of surplus education on the earnings distribution in Pakistan using quantile regression. The method of realized matches is used to measure the required level of education in each occupation from the Pakistan Social and Living Standards Measurement (PSLM) 2013-14 survey data. There is heterogeneity in returns to surplus education among overeducated workers. These returns are higher for workers at the upper half as compared to the lower half of the earnings distribution. Surplus education earns positive returns but less than the returns associated with the level of education required for jobs. Further, the difference in returns among the overeducated is higher than the difference in returns among workers who have the required education for the job. The findings imply that the surplus education factor is significant in explaining how education contributes in earnings differentials and inequality.

Keywords: Surplus education, earning inequality, labor markets, Pakistan.

JEL classification: I24, J31.

1. Introduction

Analyzing and identifying the underlying factors affecting earnings distributions has always been a source of heated debate among economists since the seminal work of Kuznets (1955). There are a range of factors that affect the distribution of earnings distribution including globalization, skill-biased technological change, labor market institutions, inflation and unemployment (Hoeller, Joumard, & Koske, 2014). Much of the current literature on earning inequality underlines education as a contributing factor towards earning inequality (Reis, 2017; Goel, 2017; Flinn & Mullins 2015; Sattinger & Hartog, 2013).

^{*} Professor, School of Business and Economics, University of Management and Technology, Lahore, Pakistan.

^{**} Assistant Professor, Faculty of Business and Economics, University of Sargodha, Pakistan.

^{***} MS Student, Department of Economics, The University of Lahore.

Policies that aim to expand the average schooling of individuals are expected to reduce earnings inequality. Equalization of educational attainment is therefore given as a reason for reducing income inequality (Budria, 2011). The supply of educated labor has risen in the last two decades in Pakistan. However, it is likely that the traditional structure of the labor market is not capable of absorbing the educated workers in jobs that require their attained education; thereby resulting in surplus education in the labor market. In such a situation, where demand for educated labor is stagnant, the Increase in the supply of educated labor can decrease returns to education (Pritchett, 2001). The existence of surplus education suggests that benefits to be gained from the expansion of education could potentially be lower than expected. Thus, an expansion in educational attainment maybe relatively unproductive for society and unprofitable from the individual's perspective (Duncan & Hoffman, 1981; Ordine & Rose, 2009).

Due to an increase in the relative supply of educated workers as compared to the skill requirements of the labor market for specific occupations, some workers take the jobs that require less education than they actually possess. These overeducated workers receive less wages than the correctly matched workers with the same level of education, but overeducated workers receive more wages than their undereducated co-workers, holding other characteristics constant (Niето and Ramos, 2017; Rubb, 2003; Dolton and Vignols, 2000). This evidence is a warning to policymakers that the unregulated expansion of education for some groups may increase earning differentials, hence income inequality.

The empirical literature on the nexus of surplus education and earnings has emerged with the seminal contribution of Duncan and Hoffman (1981). In their work, an extended version of the Mincer wage equation was proposed that allows a separate estimation of the returns to required education, surplus education and deficit education. They found that the returns to "surplus education" are substantially smaller than the returns to "required education". Subsequent studies (for example, Bauer, 2002; Cohn & Ng, 2000; Lee & Lee, 2016) also confirm their findings. The empirical research of Budria (2011), Budria and Moro-Egido (2009) and Hartog et al. (2001) has reported, using quantile regression, that returns to surplus education are heterogeneous.

In recent years, the returns to surplus education have attracted considerable attention among researchers and policymakers in developed countries due to its effect on earning inequality (for example Leuven &

Oosterbeek, 2011; Ordine & Rose, 2011; Slonimczyk, 2013). However, the contribution of surplus education towards earnings inequality is mostly ignored in developing countries such as Pakistan. Education, as a major determinant of earnings, contributes positively to earnings inequality in Pakistan. To date, the factors that are responsible for a positive link between education and earnings inequality are mostly unknown. Among many possible explanations, one is an educational mismatch (Martins & Pereira, 2004). But how education by its surplus nature contributes to earnings inequality, especially within a group of overeducated workers, is an unexplored area of research in Pakistan.

This study estimates a model where earnings inequality is related to educational mismatch in the labor market. It shows that earnings inequality may arise within occupations, because of the existence of heterogeneity in the education levels of workers employed in similar roles along with differential returns to required and surplus education. Data from the Pakistan Social and Living Standards Measurement Survey for 2013-14 is utilized, which collects comprehensive information on social, demographic and economic indicators from respondents.

The rest of the study is organized as follows. Section 2 discusses some of theories or frameworks presented in the context of surplus education. Section 3 outlines the empirical earnings function and discusses the methodology for its estimation. Further, the measurement of required, surplus and deficit education and the data are explained in section 3. Section 4 presents the empirical findings of the study. Concluding remarks are provided in section 5.

2. Theoretical Framework

There are a number of labor market theories that partially explain the observed mismatch between actual and required education in a particular job and the resulting earnings effect. Human capital theory implies that workers are paid according to their marginal productivity which is determined by their human capital (Becker, 1964). Human capital can be acquired through education, work experience and on-the-job-training. Attained education and work experience indirectly measure the worker's marginal productivity, and hence earnings as well. It is argued that when the supply of highly educated workers increases, employers adopt their production techniques in order to avail the advantage of a cheaper educated labor force. On the other side, productivity and earnings

are fixed in a particular job. Thus mismatched workers receive the earnings that are similar to those workers who are correctly matched in their jobs.

According to the signaling model, in an imperfect labor market, education is used to identify more able, productive and ambitious workers. Therefore, individuals invest more in education with the hope that this will permit them to be distinguished from other applicants competing for the jobs (Spence, 1973).

The job competition model (Thurow, 1975) considers surplus education as demand driven and is a long-term problem in the labor market. It suggests that productivity and earnings depend upon job characteristics instead of individual characteristics. That is, Thurow's (1975), job competition model describes a labor market situation where workers compete for jobs based on their training costs. Employers consider overeducated workers more able and capable and hire them first to save training costs. In this way education serves as a proxy for training.

Matching theory and occupational mobility theory see an educational mismatch as transient. In matching theory (Jovanovic, 1979), a mismatch between required and actual education represents the quality of match between the job and the worker and results from imperfect information about jobs and job search costs. Workers with surplus education represent a poor match because their actual education implies that they are more able to obtain a better job. Over time they may obtain a higher-level job. According to occupational mobility theory, individuals with high levels of education may accept low or entry-level jobs while they gain specific human capital and experience through on-the-job training. It allows them to readily be promoted to higher level jobs (Sicherman & Galor, 1990).

Assignment theory (Sattinger, 1993) holds that an individual's marginal productivity and earnings are determined by both the productivity ceiling of jobs as well as human capital. In a dynamic economy, workers differ in attributes and jobs differ in their complexities. There exists an allocation problem in allocating differentiated jobs to heterogeneous workers. Therefore, educational mismatch will tend to be a lasting feature of the labor market.

Ordine and Rose (2009) present a theoretical framework where within group earnings, inequality is related to surplus education in the labor market. They specify possible theoretical mechanisms that lead to the

incidence of overeducation in the labor market. Within-group earnings inequality arises because of the existence of potential differences in the returns to education as well as heterogeneity in the productivity “signal” conveyed by the attained education of workers.

3. Methodology and Data

The details of the empirical earnings function and the measurement of required, surplus and deficit education are explained below. Further, this section provides information about the data, including the data source, sampling methodology, number of observations used in the regressions and summary statistics of the relevant variables.

3.1. Empirical Model

To quantify the effect of educational mismatch, i.e. over, required and under education on earnings distributions, the study adopts an extended Mincerian earnings function introduced by Duncan and Hoffman (1981). We begin with the basic Mincerian earnings function:

$$\ln Y_i = X_i \cdot \delta + \beta^a \text{Edu}_i^a + \varepsilon_i \quad (1)$$

Where Edu_i^a is years of attained education and β^a is return to attained education. Duncan and Hoffman (1981) decomposed the years of attained education (Edu_i^a) into years of required education for job (Edu_i^r), years of surplus or overeducation (Edu_i^s) and years of deficit or undereducation (Edu_i^d). Therefore, the following identity holds for years of attained education (Edu_i^a):

$$\text{Edu}_i^a \equiv \text{Edu}_i^r + \underbrace{\max(0, \text{Edu}_i^a - \text{Edu}_i^r)}_{\text{Edu}^s} - \underbrace{\max(0, \text{Edu}_i^r - \text{Edu}_i^a)}_{\text{Edu}^d} \quad (2)$$

where Edu_i^a was replaced in the Mincerian earnings function by these three components. The earnings function of Duncan and Hoffman (1981) is specified as follows:

$$\ln Y_i = X_i \cdot \delta + \beta^r \text{Edu}_i^r + \beta^s \text{Edu}_i^s + \beta^d \text{Edu}_i^d + \varepsilon_i \quad (3)$$

where $\ln Y_i$ is the natural logarithm of earnings the i^{th} individual, vector X_i includes a constant, characteristics of workers and other explanatory variables that affect earnings. These variables include experience, square of experience and dummies for gender, marital status, area, provinces and

industries. The description of these variables is given in Appendix Table A1. δ is the vector of parameters to be estimated and coefficients β^r, β^s and β^d are the returns to required, surplus and deficit education respectively. ε_i is the error term. This specification, with the variables of surplus, required and deficit education is referred to as the ORU (over, required and under education) model in the literature (Hartog, 1997).

An attractive feature of this earning function is that it fits in the human capital and job competition equation as special cases. By applying the restriction $\beta^r = \beta^s = -\beta^d$, equation (1) arrives at the standard human capital model, i.e. the returns to surplus, required and deficit education are equal. Equation (1) reduces to the Thurow (1975) job competition model by setting $\beta^s = \beta^d = 0$ where only required education is rewarded.

The study empirically tests where workers are located along the earnings distribution, conditional on their educational mismatch. Therefore, the study intends to estimate the earning function using the quantile regression approach. This technique allows us to estimate how the relationship changes between an explanatory variable and dependent variable along the conditional distribution of earnings.

OLS allows the effect of explanatory variable(s) to be estimated on the conditional mean of the dependent variable. In this approach one implicitly assumes that the marginal effect of independent variable(s) is constant over the distribution of the dependent variable. Therefore, it describes only a limited aspect of the statistical relationship between variables. In contrast, quantile regression estimates the effect of the explanatory variable(s) on a particular percentile of the dependent variable (Budria, 2011; Martins & Pereira, 2004; Hartog et al., 2001; Machado & Mata, 2001). Thus, quantile regression estimates provide a snapshot of the effect of independent variable(s) on the whole distribution of the dependent variable.

The quantile regression model was proposed by Koenker and Basset (1978). Quantile regressions estimate the relationship between independent variable(s) and the θ^{th} conditional percentile of the dependent variable. The quantile regression model is written as:

$$Y_i = X_i' \beta_\theta + u_{\theta i} \text{ with } Quant_\theta(Y_i | X_i) = X_i' \beta_\theta \quad (4)$$

$$(i = 1, 2, 3, \dots, n)$$

Where β is the vector of parameters, X is the vector of independent variables and u_θ is the disturbance term. $Quant_\theta(Y_i|X_i)$ represents the θ_{th} conditional percentile of the dependent variable (Y) given X . By variation of θ , the entire distribution of the dependent variable can be traced out. That is, the marginal effect of X is not necessarily identical across different percentiles of the conditional distribution of Y . Quantile regression estimates are interpreted likewise as OLS estimates. This study intends to estimate ORU earning functions at multiple deciles. Thus, it implements a simultaneous-quantile regression. Simultaneous-quantile regression obtains the variance–covariance matrix of the estimators (VCE) via bootstrapping according to the procedure described by Buchinsky (1998). Simultaneous-quantile regression can estimate a coefficient at different percentiles simultaneously which allows to test for the equality of coefficients across percentiles.

The study estimates the ORU earnings functions at nine different deciles using the overall sample of workers. The prevalence of the gender gap in educational attainment is obvious in Pakistan. Generally, the analysis based on the overall sample of workers is deficient in estimating the returns to education due to structural differences in gender outcomes. Therefore, a gender-based analysis is also carried out, and the ORU earning function is estimated for samples of male and female workers separately.

3.2. Measurement of Over, Under and Required Education

Given the data on the years of attained education and occupations of the workers, the variable of required education is measured using the realized matches method to measure the required education that was proposed by Kiker et al. (1997). Once a measure of required education is established, variables of surplus and deficit education are measured as:

$$\begin{aligned} Edu_i^s &= Edu_i^a - Edu_i^r \text{ if } Edu_i^a > Edu_i^r \\ &= 0, \text{ otherwise} \end{aligned} \quad (5)$$

$$\begin{aligned} Edu_i^d &= Edu_i^r - Edu_i^a \text{ if } Edu_i^r > Edu_i^a \\ &= 0, \text{ otherwise} \end{aligned} \quad (6)$$

Therefore, $Edu_i^a = Edu_i^r + Edu_i^s - Edu_i^d$ must hold.

All education components are computed in years of formal education. In this study, required education in a given occupation is measured as the

mode (i.e. most frequent) years of education of workers in that occupation. Only workers, whose attained education deviates from years of required education, are considered as mismatched. Therefore, for matched workers $Edu_i^a = Edu_i^r$ or $Edu_i^s = Edu_i^d = 0$. A worker has surplus years of education if his/her attained education is above the mode value of education in a particular occupation. Conversely, a worker has deficit years of education if his/her attained education falls below the mode value of education in his/her occupation. Workers are classified as correctly matched if their attained education equals the modal value within a specific occupation.

3.3. Data Source and Sample

The data is taken from the Pakistan Social and Living Standard Measurement (PSLM) survey for 2013-14. The PSLM survey data is collected by the Pakistan Bureau of Statistics. The survey is conducted at the provincial level for alternate years using a two-stage stratified sample design. At the first stage, enumeration blocks in urban areas and villages in rural areas are taken as primary sampling units (PSUs). At the second stage, 12 households from the urban domain and 16 households from the rural domain within each PSU were selected. These households within the PSUs were considered for the selection of secondary sampling units. The survey covers a sample of 17,989 households distributed over 1307 PSUs in the four provinces of Pakistan. It collects comprehensive information on individual and household characteristics including earnings, education, demographic and economic indicators. In this study, the sample is restricted to the employed workers aged 15-65 years who reported their occupation. Observations with missing values were dropped from the sample.

Table 1 presents the summary statistics of variables for the overall sample as well as for the split male and female samples. The overall sample consists of 15,366 individuals, among whom the vast majority, 13,927, are male. Table 1 reports the means and standard deviations of all variables used in the regressions. The average of log monthly earning of the males is higher than that of female workers. The required education for jobs is considerably different for males and females; average years of required education are higher for female than male workers, while average years of surplus education are lower among female workers than male workers. Further, the sample data shows the average deficit years of education is almost the same for female as for male workers. Note that there are no women employed in a number of occupations, including mining and quarrying, extraterritorial organizations, water supply, sewerage, waste management, construction, administrative and support activities.

Table 1: Summary statistics

Variables	Overall		Male		Female	
	Mean	SD	Mean	SD	Mean	SD
Log Monthly Earnings	9.310	0.943	9.424	0.796	8.211	1.423
Required Education	8.279	3.571	8.133	3.346	9.698	5.060
Surplus Education	1.582	2.287	1.637	2.305	1.058	2.027
Deficit Education	1.109	1.968	1.100	1.959	1.200	2.054
Experience	20.521	12.141	20.978	12.186	16.097	10.730
Experience Squared	568.48	613.18	588.551	624.172	374.185	450.364
Area	0.457	0.498	0.450	0.497	0.533	0.499
Gender (Male)	0.906	0.291	-	-	-	-
Marital Status	0.713	0.453	0.730	0.444	0.547	0.498
KPK	0.183	0.387	0.188	0.391	0.136	0.342
Sindh	0.269	0.444	0.276	0.447	0.203	0.402
Balochistan	0.081	0.273	0.086	0.280	0.038	0.192
Ind1	0.004	0.062	0.004	0.065	0	0
Ind2	0.169	0.374	0.158	0.365	0.268	0.443
Ind3	0.012	0.107	0.013	0.112	0.001	0.026
Ind4	0.005	0.068	0.005	0.071	0	0
Ind5	0.109	0.312	0.121	0.326	0	0
Ind6	0.180	0.384	0.198	0.398	0.013	0.111
Ind7	0.071	0.257	0.078	0.269	0.001	0.037
Ind8	0.018	0.134	0.020	0.140	0.003	0.053
Ind9	0.011	0.106	0.012	0.109	0.003	0.059
Ind10	0.012	0.108	0.012	0.109	0.008	0.087
Ind11	0.005	0.071	0.006	0.074	0.001	0.026
Ind12	0.009	0.095	0.010	0.099	0.001	0.026
Ind13	0.005	0.072	0.006	0.076	0	0
Ind14	0.048	0.214	0.052	0.223	0.008	0.087
Ind15	0.084	0.278	0.059	0.235	0.333	0.471
Ind16	0.027	0.162	0.022	0.145	0.080	0.271
Ind17	0.002	0.039	0.002	0.041	0.001	0.026
Ind18	0.036	0.186	0.030	0.172	0.087	0.282
Ind19	0.008	0.089	0.006	0.078	0.026	0.158
Ind20	0.000	0.018	0.000	0.019	0	0
Sample Size	15366		13927		1439	

Source: Authors' calculations from PSLM 2013-14.

4. Results and discussion

This section presents estimates of the effect of educational mismatch on earnings distributions using quantile regression on the overall sample. In the second part, estimates of the effect of educational mismatch on earnings distributions are presented for males and females separately.

4.1. Analysis of overall sample

The study tests for alternative specifications to Duncan and Hoffman's (1981), specification in equation (1) to verify that this specification fits the sample data better. Table 2 provides the values of F-statistics for the hypothesis together with p-values. The null hypothesis that returns to surplus, required and deficit education are equal, in other words the standard Mincer earnings function and the job competition model that hypothesize that only required education is rewarded, is rejected by the sample data. This allows the estimation of an earning function that estimates separate returns for surplus, required and deficit education. That is, the Duncan and Hoffman's (1981), specification in the ORU model is correct for the overall sample.

Table 2: Empirical specification test

Mincer earnings function	112.47
$H_0: \beta^r = \beta^s = -\beta^d$	(0.00)
Job Competition Model	918.18
$H_0: \beta^s = \beta^d = 0$	(0.00)

Note: p-values in parenthesis.

Source: Authors' calculations.

The study estimated the ORU earning function by using the quantile regression for the overall sample. The earning function was estimated at nine different deciles: $\theta = .10$, $\theta = .20$, $\theta = .30$, $\theta = .40$, $\theta = .50$, $\theta = .60$, $\theta = .70$, $\theta = .80$ and $\theta = .90$. The estimated coefficients of surplus education (β^s), required education (β^r) and deficit education (β^d) at the nine deciles are reported in Table 3.¹ These coefficient estimates illustrate the percentage change in earnings resulted from an additional year of surplus, required or deficit education at the estimated decile. The OLS estimates of β^r , β^s and β^d are also provided. In the last panel, the Wald F-test for the equality of coefficients across deciles demonstrates that differences in the rate of return to the surplus, required and deficit education are significantly different at different deciles.

The rate of returns to an additional year of required education, surplus education or deficit education is different across the earning distribution. That is, there are earning differentials among workers with

¹Complete results are available on request.

surplus education or deficit education. Returns to required education exhibit a decreasing trend until $\theta=0.20$ and an upward trend is observed after that. Returns to required education increase from 9.3 percent at the lowest decile to 12.5 percent at the highest. The returns to surplus education approximately double as we move from the bottom to the top of the earning distribution. These returns reveal an increasing trend from the lowest to the highest decile. The findings are similar to that found by Ramos (2011) for Spain and Hartog et al. (2001) for Portugal.

Table 3: Returns to Surplus, Required and Deficit Education

Quantiles	Required Education	Surplus Education	Deficit Education
$\theta=0.10$	0.0934*** (19.67)	0.0593*** (10.13)	-0.0622*** (-12.72)
$\theta=0.20$	0.0922*** (25.22)	0.0608*** (16.56)	-0.0630*** (-17.18)
$\theta=0.30$	0.0966*** (32.68)	0.0670*** (14.23)	-0.0711*** (-23.28)
$\theta=0.40$	0.101*** (26.77)	0.0737*** (17.79)	-0.0712*** (-22.89)
$\theta=0.50$	0.104*** (33.09)	0.0773*** (21.99)	-0.0739*** (-23.57)
$\theta=0.60$	0.110*** (34.75)	0.0814*** (29.36)	-0.0763*** (-23.55)
$\theta=0.70$	0.109*** (35.55)	0.0861*** (23.57)	-0.0723*** (-22.45)
$\theta=0.80$	0.114*** (42.72)	0.0904*** (26.70)	-0.0743*** (-24.43)
$\theta=0.90$	0.125*** (35.55)	0.104*** (27.27)	-0.0749*** (-27.21)
OLS	0.118*** (43.39)	0.0825*** (29.44)	-0.0793*** (-24.22)
Wald F statistics	9.12 [0.00]	19.06 [0.00]	2.34 [0.02]
$H_0: \beta_{.10} = \beta_{.20} = \dots = \beta_{.90}$			

Notes: t-statistics in parentheses and p-values in brackets, ***p<0.001

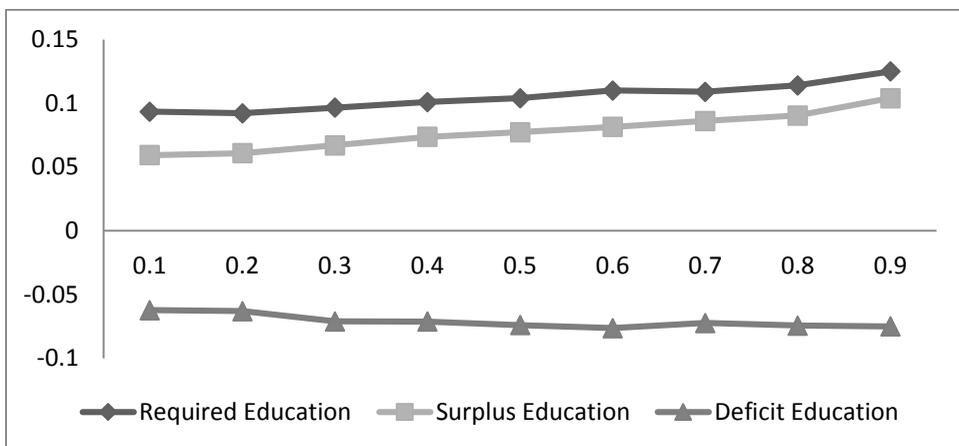
Source: Authors' calculations.

The earning differential between surplus and required education is lower at the upper deciles compared to bottom deciles of the earnings distribution. Returns to years of education below the job requirements are negative indicating a penalty for undereducated workers. This is consistent with the findings of Budria (2011) and Verdugo and Verdugo (1989). The penalty of deficit education grows slowly up to $\theta=0.60$ then remains consistent at the upper deciles.

In line with the studies of Budria (2011), Hartog et al. (2001) and Duncan and Hoffman (1981), this study also found that surplus education brings positive returns but less than those associated to required education. The study based on OLS by Abbas (2008) also found superior returns to required education as compared to surplus education in Pakistan. OLS estimates also verify that returns to a year of required education are higher than the returns to a year of surplus education. The average pay-off to required education and surplus education is 11.8 percent and 8.3 percent respectively. On average, the penalty of pay for deficit years of education is 7.9 percent. OLS estimates show differences in the returns to surplus and required education. This implies that an educational mismatch is a contributing factor to earning differential/inequality between workers with surplus and required education.

The estimated coefficients of surplus education (β^s), required education (β^r) and deficit education (β^d) are plotted for each decile in Figures 1.

Figure 1: Returns to Surplus, Required and Deficit Education



Source: Authors' compilation using data from Table 3.

4.2. Gender-Based Analysis

The coefficient of the gender dummy was statistically significant in estimated earning functions which use overall sample data. This shows that there are structural differences in the earnings of male and female workers. Hence, the study extends the analysis in order to estimate returns to the educational mismatch separately for male and female workers. To this end, the ORU earning function is estimated for the sub-samples of men and

women. Table 4 reports the F-statistics together with p-values for the test of the basic Mincerian model (versus the alternative ORU specification). These results reject the null hypothesis of the basic Mincerian earnings function and job competition model for both male and female samples.

Table 4: Empirical Specification Test

	Male	Female
Mincer earnings function	59.14	20.20
$H_0: \beta^r = \beta^s = -\beta^d$	(0.00)	(0.00)
Job Competition Model	799.40	121.27
$H_0: \beta^s = \beta^d = 0$	(0.00)	(0.00)

Source: Authors' calculations.

Note: p-values in parenthesis.

On the basis of this finding, separate pay offs to surplus, required and deficit education should be estimated. In doing so, the ORU earning function was estimated at nine different deciles for both samples by including the same control variables listed in the Appendix Table A1 except for the gender dummy. Only the coefficients of surplus, required and deficit education are given in Table 5 for male and female workers separately. The OLS estimates of β^r , β^s and β^d are also provided by gender.

In the last panel of Table 5, Wald F statistics for the joint equality of coefficients across deciles and p-values are displayed. It shows that differences in the rate of return to surplus and required education are statistically significantly different across the deciles for the male sample. This indicates that the rate of returns to an additional year of required education, or surplus education is different for male workers as we move along the earnings distribution.

For male workers, returns to required years of education increases from 8.3 percent at the lowest decile to 11.5 percent at the highest decile. It shows that the returns differential, between workers who have required years of education to perform their job but located at two extreme deciles of the earning distribution, amounts to 3.2 percentage points. Returns to surplus education amount to 5.2 percent and 10.2 percent at the 0.10 and 0.90 quantiles respectively. The superior returns to required education as compared to surplus education are also obvious. OLS estimates show that average returns to required and surplus education are 10 percent and 7.6 percent respectively. The average pay penalty for males who are undereducated is 6.95 percent. This indicates that educational mismatch is

related to the earnings differential between male workers who have surplus and required education.

Table 5: Returns to Surplus, Required and Deficit Education of Male and Female

Quantiles	Required Education		Surplus Education		Deficit Education	
	Male	Female	Male	Female	Male	Female
$\theta=0.10$	0.0826*** (17.91)	0.233*** (11.60)	0.0523*** (13.54)	0.0633 (1.89)	-0.0602*** (-11.73)	-0.206*** (-7.39)
$\theta=0.20$	0.0840*** (31.42)	0.219*** (11.39)	0.0558*** (22.91)	0.0894*** (4.35)	-0.0614*** (-14.87)	-0.218*** (-11.98)
$\theta=0.30$	0.0852*** (31.82)	0.230*** (9.79)	0.0606*** (16.36)	0.108*** (5.92)	-0.0663*** (-23.86)	-0.200*** (-9.84)
$\theta=0.40$	0.0905*** (29.85)	0.234*** (12.95)	0.0703*** (19.90)	0.111*** (9.88)	-0.0657*** (-24.10)	-0.190*** (-12.16)
$\theta=0.50$	0.0942*** (37.17)	0.232*** (17.65)	0.0749*** (20.92)	0.118*** (7.74)	-0.0687*** (-26.54)	-0.201*** (-10.35)
$\theta=0.60$	0.0970*** (40.28)	0.232*** (15.62)	0.0770*** (22.02)	0.135*** (10.01)	-0.0684*** (-24.41)	-0.175*** (-7.84)
$\theta=0.70$	0.101*** (32.92)	0.231*** (13.37)	0.0824*** (22.33)	0.141*** (7.43)	-0.0677*** (-17.18)	-0.163*** (-10.30)
$\theta=0.80$	0.105*** (28.35)	0.226*** (12.77)	0.0882*** (25.05)	0.151*** (4.82)	-0.0689*** (-18.52)	-0.155*** (-10.16)
$\theta=0.90$	0.115*** (36.13)	0.213*** (11.60)	0.102*** (19.23)	0.150*** (5.77)	-0.0670*** (-13.25)	-0.136*** (-6.19)
OLS	0.0998*** (38.33)	0.228*** (14.80)	0.0761*** (28.31)	0.113*** (8.23)	-0.0695*** (-21.91)	-0.177*** (-12.35)
Wald F	35.28	0.30	11.34	1.08	1.25	1.50
statistics	[0.00]	[0.97]	[0.00]	[0.38]	[0.27]	[0.15]
$H_0: \beta_{.10} = \beta_{.20} = \dots = \beta_{.90}$						

Source: Authors' calculations.

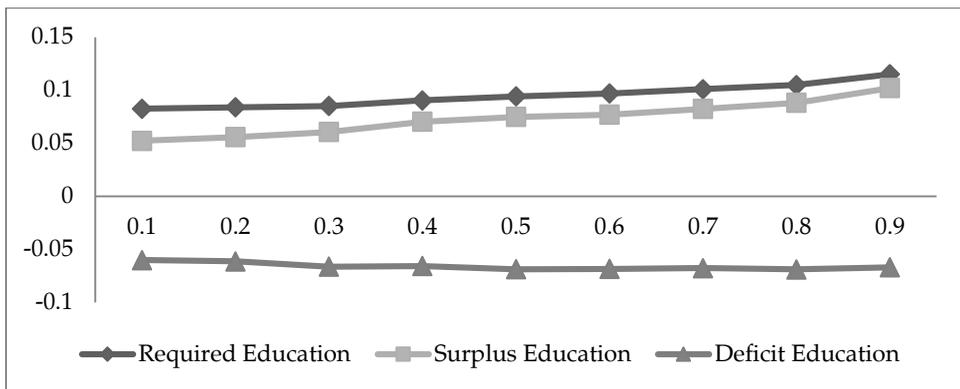
Note: t-statistics in parentheses and p-values in brackets, *p<0.05, **p<0.01, ***p<0.001.

For the female sample, Wald F-statistics show that differences in the rate of return to surplus, required and deficit education are not statistically significant. It suggests that rate of returns to an additional year of required, surplus or deficit education is the same for female workers across the earnings distribution. However, OLS estimates show that average returns to surplus and required education amounts to 11.3 and 22.8 percentage points respectively. This indicates that returns to surplus education are approximately half of returns to those of required education. On average, the pay penalty for females is 17.7 percent for those who are undereducated. Similar to male workers, educational mismatch is a source of earning differential between female workers who have surplus and required education.

These findings are similar to Hartog et al. (2001) for Portugal, McGuinness and Bennett (2007) for Northern Ireland and Santos and Sequeira (2013) for European countries.

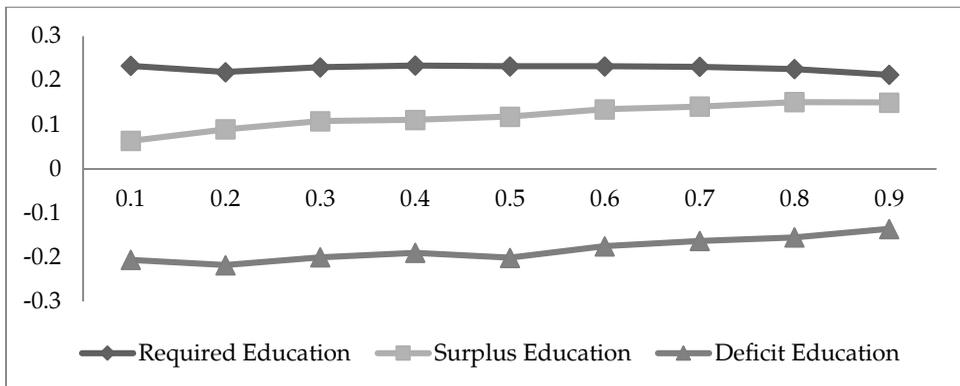
The estimated coefficients of surplus, required and deficit education are plotted against the decile numbers in Figure 2 and Figure 3 for males and females respectively. The surplus and required education affects earnings differently at different points of the earning distribution for male workers. The pattern of returns for female workers is different from those found for the males.

Figure 2: Returns to Required, Surplus and Deficit Education (for Male Workers)



Source: Authors' compilation using data from Table 5.

Figure 3: Returns to required, surplus and deficit education (for female workers)



Source: Authors' compilation using data from Table 5.

There are some factors that may explain why earnings vary from person to person in the labor market even if they have similar observed characteristics. Discrimination is a key explanation of why labor market outcomes differ among people with similar characteristics, such that heterogeneous opportunities are offered to similar workers that result in earning differentials. Other factors include compensating differences in earnings, efficiency wage, unionization of workers and most importantly (unobserved) natural ability and efforts that determine the productivity of workers.

5. Conclusion

The present study empirically investigates the effect of surplus education on the earning distribution in Pakistan. Data from the Pakistan Social and Living Standards Measurement Survey for 2013-14 is utilized for the analysis. Duncan and Hoffman's (1981), earning function is estimated using quantile regression. The study tests for two alternative specifications: A basic Mincer earnings function hypothesizes that returns to surplus, required and deficit education are equal, and the hypothesis of the job competition model is that only required education is rewarded. Both are rejected by the sample data.

For the overall sample, quantile regression estimates show that returns to an additional year of surplus education is different across the earnings distribution. That is, there exists earning differentials among workers who have surplus education. The returns to surplus education are positive but less than returns to required education. The earnings differential between overeducated workers and workers who have required education for the job is lower at the upper deciles compared to the bottom quantiles. OLS estimates show the difference in the returns to surplus and required education. This implies that educational mismatch contributes in earning differential/inequality between workers with surplus and required education.

The gender-based analysis indicates that rates of returns to an additional year of required education, or surplus education varies for male workers along the earnings distribution. Earning differences within overeducated workers are found to be higher than those of workers who have required education for the job. Returns to surplus education are higher for workers at the upper tail of the earnings distribution. The returns to surplus education are almost double at the ninth decile compared to the first decile. That is, there is a difference in returns to

education for male workers who have surplus education. For the female sample, the difference in the rate of return to surplus education across the earning distribution is not statistically significant. It suggests that rates of return to an additional year of surplus education is the same for all overeducated female workers. Therefore, surplus education does not contribute further to earning differentials among female workers who are overeducated. Further, OLS estimates show that the returns to surplus and required education are different for male workers. This implies that educational mismatch is related to the earnings differential between male workers with surplus and required education. Similar to male workers, educational mismatch is a source of the earnings differential between female workers who have surplus and required education.

The findings of the study show that there is a difference in return to education for male workers with surplus education. Also, the returns to surplus and required education are different for male and female workers. An important policy implication arises from these findings as households and public resources extensively invest in education. It is necessary to make sure that the educational programs are useful and their benefits are being distributed equitably. Further, the surplus education phenomenon highlights the imperfections in the labor market that limit the absorptive capacity of society to fully utilize and compensate the workers with surplus years of education. In this respect, changes in the supply of educated workers and the skill demand of the labor market should be in order.

References

- Abbas, Q. (2008). Over-education and under-education and their effects on earnings: Evidence from Pakistan, 1998–2004. *SAARC Journal of Human Resource Development*, 4, 109-125.
- Bauer, T. (2002). Educational mismatch and wages: A panel analysis. *Economics of Education Review*, 21(3), 221-229.
- Becker, G. (1964). Human Capital, Nueva York. *National Bureau of Economic Research*, 8(299), 493-517.
- Buchinsky, M. (1998). Recent advances in quantile regression models: A practical guideline for empirical research. *Journal of Human Resources*, 33(1), 88-126.
- Budria, S. (2011). Are educational mismatches responsible for the 'inequality increasing effect' of education? *Social Indicators Research*, 102(3), 409-437.
- Budria, S., & Moro-Egido, A. I. (2009). The overeducation phenomenon in Europe. *Revista Internacional de Sociología*, 67(2), 329-345.
- Cohn, E., & Ng, Y. C. (2000). Incidence and wage effects of overschooling and underschooling in Hong Kong. *Economics of Education Review*, 19(2), 159-168.
- Dolton, P., & Vignoles, A. (2000). The incidence and effect of overeducation in the U.K graduate labour market. *Economics of Education Review*, 19(2), 179-198.
- Duncan, G. J., & Hoffman, S. D. (1981). The incidence and wage effects of overeducation. *Economics of Education Review*, 1(1), 75-86.
- Flinn, C., & Mullins, J. (2015). Labor market search and schooling investment. *International Economic Review*, 56(2), 359-398.
- Goel, M. (2017). Inequality between and within skill groups: The curious case of India. *World Development*, 93, 153-176.
- Government of Pakistan, (2015). Living Standards Measurement Survey (PSLM) 2013–14 National. Statistics Division, Pakistan Bureau of Statistics, Islamabad.

- Hartog, J. (1997). On returns to education: Wandering along the hills of ORU land. Keynote speech on the AEA-Conference 1997. *AEA (ed.)*.
- Hartog, J., Pereira, P. T., & Vieira, J. A. (2001). Changing returns to education in Portugal during the 1980s and early 1990s: OLS and quantile regression estimators. *Applied Economics*, 33(8), 1021-1037.
- Hoeller, P., Joumard, I., & Koske, I. (2014). Reducing income inequality while boosting economic growth: Can it be done? Evidence from OECD Countries. *The Singapore Economic Review (SER)*, 59(01), 1-22.
- Jovanovic, B. (1979). Job-matching & the theory of turnover. *Journal of Political Economy*, 87(5), 972-990.
- Kiker, B. F., Santos, M. C., & De Oliveira, M. M. (1997). Overeducation and undereducation: Evidence for Portugal. *Economics of Education Review*, 16(2), 111-125.
- Koenker, R., & Bassett Jr, G. (1978). Regression Quantiles. *Econometrica*, 46(1), 33-50.
- Kuznets, S. (1955). Economic growth and income inequality. *The American Economic Review*, 45(1), 1-28.
- Lee, H., Lee, J. W., & Song, E. (2016). Effects of educational mismatch on wages in the Korean labor market. *Asian Economic Journal*, 4(30), 375-400.
- Leuven, E., & Oosterbeek, H. (2011). Overeducation and mismatch in the labor market. In *Handbook of the Economics of Education*, 4, 283-326. Elsevier.
- Machado, J. A., & Mata, J. (2001). Earning functions in Portugal 1982-1994: Evidence from quantile regressions. *Empirical Economics*, 26(1), 115-134.
- Martins, P. S., & Pereira, P. T. (2004). Does education reduce wage inequality? Quantile regression evidence from 16 countries. *Labour Economics*, 11(3), 355-371.
- McGuinness, S., & Bennett, J. (2007). Overeducation in the graduate labour market: A quantile regression approach. *Economics of Education Review*, 26(5), 521-531.

- Nieto, S., & Ramos, R. (2017). Overeducation, skills and wage penalty: Evidence for Spain using PIAAC data. *Social Indicators Research*, 134(1), 219-236.
- Ordine, P., & Rose, G. (2009). Overeducation and instructional quality: A theoretical model and some facts. *Journal of Human Capital*, 3(1), 73-105.
- Ordine, P., & Rose, G. (2011). Inefficient self-selection into education and wage inequality. *Economics of Education Review*, 30(4), 582-597.
- Pritchett, L. (2001). Where has all the education gone? *The World Bank Economic Review*, 15(3), 367-391.
- Ramos, R. (2011). Returns to schooling, overeducation and wage inequality: new evidence. In *Actas de las IX Jornadas de Economía Laboral: Santiago de Compostela, 7 y 8 de julio, 2011* (p. 68). Asociación Española de Economía del Trabajo.
- Reis, M. C. (2017). Educational mismatch and labor earnings in Brazil. *International Journal of Manpower*, 38(2), 180-197.
- Rubb, S. (2003). Over education in the labor market: A comment and re-analysis of a meta-analysis. *Economics of Education Review*, 22(6), 621-629.
- Santos, M. S. & Sequeira, T. N. (2013). Skills mismatch and wage inequality: Evidence for different countries in Europe. *Technological and Economic Development of Economy*, 19(S1), S425-S453.
- Sattinger, M. (1993). Assignment models of the distribution of earnings. *Journal of Economic Literature*, 31(2), 831-880.
- Sattinger, M. and Hartog, J. (2013). Nash bargaining and the wage consequences of educational mismatches. *Labour Economics*, 23, 50-56.
- Sicherman, N., & Galor, O. (1990). A theory of career mobility. *Journal of Political Economy*, 98(1), 169-192.
- Slonimczyk, F. (2013). Earnings inequality and skill mismatch in the US: 1973–2002. *Journal of Economic Inequality*, 11(2), 163-194.
- Spence, M. (1973). Job market signaling. *The Quarterly Journal of Economics*, 87(3), 355-374.

Thurow, L. C. (1975). *Generating inequality-mechanisms of distribution in the U.S. Economy*. New York: Basic Books.

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

Appendix

Table A1: Description of variables

Variables	Description
Log Monthly Earnings	= Natural logarithm of monthly income
Required Education	= Most frequent years of education within a specific occupation
Surplus Education	= Number of years of education above required education
Deficit Education	= Number of years of education below required education
Experience	= $\max(0, \text{Age} - \text{years of completed education} - 5)$
Experience Squared	= Squared of experience
Area	= 1 for urban, 0 otherwise
Gender	= 1 for male, 0 otherwise
Marital Status	= 1 for currently married, 0 otherwise
Provincial Dummies*	
KPK	= 1 if the worker form KPK, 0 otherwise
Sindh	= 1 if the worker form Sind, 0 otherwise
Balochistan	= 1 if the worker form Baluchistan, 0 otherwise
Industrial Dummies**	
Ind1	=1 for Mining and quarrying, 0 otherwise
Ind2	=1 for manufacturing, 0 otherwise
Ind3	=1 for Electricity, gas, steam and air conditioning supply, 0 otherwise
Ind4	=1 for Water supply; sewerage, waste management and remediation activities, 0 otherwise
Ind5	=1 for Construction, 0 otherwise
Ind6	=1 for wholesale and retail trades; repair of motor vehicles and motorcycles, 0 otherwise
Ind7	=1 for transportation and storage, 0 otherwise
Ind8	=1 for accommodation and food services activities, 0 otherwise
Ind9	=1 for information and communication, 0 otherwise
Ind10	=1 for financial and insurance activities, 0 otherwise
Ind11	=1 for real estate activities, 0 otherwise
Ind12	=1 for professional, scientific and technical activities, 0 otherwise
Ind13	=1 for administrative and support services activities, 0 otherwise
Ind14	=1 for public administration and defense; compulsory social security, 0 otherwise
Ind15	=1 for education, 0 otherwise
Ind16	=1 for human health and social work activities, 0 otherwise
Ind17	=1 for arts, entertainment and recreation, 0 otherwise
Ind18	=1 for other services activities, 0 otherwise
Ind19	=1 for activities of households as employers, 0 otherwise
Ind20	=1 for activities of extraterritorial organizations and bodies, 0 otherwise

Notes: *Punjab is omitted category; **Agriculture, forestry and fishing is omitted category.