©*The Pakistan Development Review*
46 : 4 Part II (Winter 2007) pp. 833–852

Returns to Education in Pakistan

SHABBAR JAFFRY, YASEEN GHULAM, and VYOMA SHAH

1. INTORDUCTION

There is an extensive empirical literature on returns to education that focuses both on developed and developing countries. Available literatures in developing countries compare the returns to academic education and vocational education [Nasir and Nazil (2000)], or seek to identify the impact of completing a given schooling cycle on earnings [Appleton (2001)]. The aim of this study is to contribute the literature by conducting a systematic analysis on returns to education and education inequality in Pakistan. In particular it asks to what extent inequality for different level of education vary across the wage distribution.

In order to address simultaneously the two issue of return to education and education inequality, study adopt a quantile regression framework. A characteristic of the wage and salary structure of most countries is that people with more education tend to receive higher remuneration than those with less [Colclough (1982)]. To do so, the paper has used data drawn from Labour Force Surveys, conducted by Government of Pakistan for the time period between 1990 and 2003, which contains eight different surveys, using methodology developed by Agrist, *et al.* (2006), where weighted least squares interpretation of Quantile Regression is used to derive an omitted variables bias formula and a partial quantile regression concept, similar to the relationship between partial regression and OLS. Estimation uses personal and household characteristics, occupational and employment characteristics in order to assess the education inequality. Empirical estimates indicate that education inequality is much higher for the middle level educates compare to educate that has less education or high level education and qualifications. The education level coefficients decrease when different sets of exogenous variables are introduced in the estimation equation. Analysis also suggests the existence of the education inequality across different areas and regions and over the time it has increased.

The rest of the paper structured as follow. Section 2 reviews the empirical literature done in this area, followed by representing data in Section 3. Methodology and results are discussed in Sections 4 and 5, respectively and paper concludes in Section 6.

Shabbar Jaffry <shabbar.jaffry@port.ac.uk>, Yaseen Ghulam, and Vyoma Shah are all based at the Portsmouth Business School, University of Portsmouth, U.K.

2. LITERATURE REVIEW

According to human capital hypothesis, it is widely argued that any investment in human capital has a pure productivity element [McMahon (1999)]. The traditional view of human capital theorists has been that schooling raises labour productivity through its role in increasing the cognitive abilities of workers. It has been shown that higher labour productivity is a positive function of the level of education received. This paper's review and subsequent analyses are based on this theoretical formulation about the relationship between years of schooling and wages.

Psacharopoulos' (1994) finds that returns to schooling (particularly for primary schooling) in least developed countries (LDCs) are high, but Bennell (1996) argues that with chronically low internal and external efficiencies at all educational levels in most Sub-Saharan Africa (SSA) countries, it seems highly implausible that rates of return to education are higher than in the advanced countries. Looking at returns country by country, it is certainly not the case that the level of returns to primary education is consistently higher than either secondary or higher education [Appleton, *et al.* (1999)]. There are also differences in returns to schooling within a country depending on the location of the individual in the wage distribution [Bauer, *et al.* (2002)]. Such evidence starts to emerge due to the recent econometric advances that are applied to different data sets to estimate earning functions [Arias, *et al.* (2001)]. The relationship between ability and returns can vary depending on the race and level of education of the individual as shown in the South African study by Mwabu and Schultz (1996).

Card (1999) reviews the existing theoretical and empirical literature that has been accumulated mainly using data sets from advanced economies. He also identified some of the outstanding econometric problems in the estimation of earning functions [Card (2001)]. These include, among others, the need to control for ability bias [Griliches (1977)].

Pereira and Martins (2004) has argued in their study that when more covariates are used in Mincer equation, which are depend on education, then the coefficient of the education should fall. And in meta-analysis on Portugal data they found that the coefficient decreases with all combinations of variables used and can drop to half of its size, especially when the sector of activity is one of the covariates used. The education-related choice of sector is an aspect that should reflect itself in over-education in the better paying sectors.

Dickerson, *et al.* (2001) has investigated the impact of trade liberalisation on wages and the returns to education in Brazil. They have argued that just using the pooled data for all available cross-section might lead to the bias result according to the theory developed by Deaton (1985) so to overcome this problem they have used pseudo-panel estimates for the returns to education and which shows that the returns are significantly lower than OLS estimates, signifying omitted ability bias in traditional cross-section estimated returns in developing countries. And on the basis of the evidence they have suggested that previous estimates of rates of returns for developing countries might be biased upwards, and perhaps to a considerable degree.

When it comes to the analysis of return to education in Pakistan, there is very little none of the existing studies has investigated the heterogeneity of returns to schooling at different point in the wage distribution. In the study, by Khan and Irfan

(1985) have analysed rate of return to education in Pakistan using Population, Labour Force and Migration Survey for 1979. Using standard earning functions authors found that private rates of returns to different level of education are low on an absolute level compare do an average of developing countries where these estimates exist. Also, their results confirms the earlier findings done by Handani (1977) and Guisinger, *et al.* (1984).

Nasir and Nazil (2000) has analysed the return to education using, technical training, school quality and literacy and numeracy skills by use of data based on PIHS for 1995-96. Where they have assumed that private schools to be provider of better quality education and have included dummy for private school in their model and they found that private schooling ahs positive, significant and substantial effect on individual earnings, a graduate of private school earns 31 percent higher than the graduate from the pubic school. From their estimation it wasn't clear that which level of education was acquired from private sector as the individucal may have acquired his half education in private and half in public. Akbari and Muhammed (2000) have argued in their study that Nasir and Nazil (2000) have used inappropriate specification of the earnings model as education quality itself affect the rate of return to schooling and hence should be incorporated in the earning model, accordingly. They have analysed the student-teacher ratio as educational quality predictor. Using years of schooling, years of labour force experience and student-teacher ration as independent variable they have shown that the marginal rate of return to education is only 5.71 percent. They also found that if one excludes the education quality then estimate yield marginal rate of return to education is 7.16 percent, which has an upward bias.

3. DATA

This study uses data drawn from the nationally representative Labour Force Survey (LFS) for Pakistan between 1990-91 and 2003-04, which was conducted by Federal Bureau of Statistics Government of Pakistan. The data collection for the LFS is spread over four quarters of the year in order to capture any seasonal variations in activity. The survey covers urban and rural areas of the four provinces of Pakistan as defined by the Population Census. The LFS excludes the Federally Administered Tribal Areas (FATA), military restricted areas, and protected areas of NWFP. These exclusions are not seen as significant since the relevant areas constitute about 3 percent of the total population of Pakistan.

The working sample, based on those who are engaged in wage employment and have positive earnings, comprises a total of 97,122 workers, once missing values and unusable observations are discarded over the time period. This includes variables such as pay, age, gender, level of education, occupational characteristics and employment status and household characteristics.

Table 1 depicts the means and standard deviations of selected variables for overall, as well as for urban and rural areas. There is a clear difference in average characteristics between urban and rural areas. On average, the wages and number of hours worked are higher in urban area, whilst the experience and numbers of job holders in a household are higher in rural areas.

Table 1

Means and Standard Deviations of Selected Variables[1]

Characteristic	Overall		Urban		Rural	
	Mean	Std.Dev.	Mean	Std.Dev.	Mean	Std.Dev.
Real Hourly Wage (in PKR)[2]	2.73	0.76	2.85	0.77	2.54	.699
Prior Potential Experience[3]	21.23	13.38	20.62	13.24	22.15	13.53
Number of Hours worked in a year	2532.72	613.49	2535.78	600.91	2528.06	632.07
Number of Job Holders in a household	2.18	1.34	2.17	1.30	2.19	1.40
Number of Observation	97122	97122	58550	58550	38572	38572

4. METHODOLOGY

The methodology adopted to estimate return to education is consistent with that of Angrist, *et al.* (2006). A key methodological issue is that the LFSs are only cross-sectional, while ideally, one would like to have a panel of individuals or households that can be traced through time, in order to investigate the changing wage structure and returns to education. In addition, estimation with the cross-section data can be seriously affected by unobserved individual heterogeneity. However, this problem can be circumvented, or at least mitigated, by tracking cohorts as suggested by Deaton (1985), and estimating relationships based on cohort means.

Starting with a simple model, suppose that base panel regression equation could be written as:

$$y_{it} = x_{it}'\beta_t + \alpha_i + \varepsilon_{it}, \qquad t = 1,.....,T,$$

where i = index individuals and t = time periods. Unfortunately, in the LFSs, the same individuals are not observed in subsequent surveys. Hence we do not have a genuine panel data available to estimate such an equation. In such circumstances, the approach first developed by Deaton (1985) proceeds as follows. Define a set of C cohorts, based on a district in a province say, such that every individual i is a member of one and only one cohort for each t. Averaging over the cohort members:

$$\bar{y}_{ct} = \bar{x}_{ct}'\beta_t + \bar{\alpha}_{ct} + \varepsilon_{ct}, \qquad c = 1,.....,C,$$

where $\bar{y}_{ct}$ is the average of the y_{it} for all members of cohort c at time t. this is a so-called 'pseudo-panel'. The 'cohort fixed effects', $\bar{\alpha}_{ct}$, will, in fact, vary with t since they comprise different individuals in each cohort c at time t, but can be treated as constant if the number of individuals per cohort is large. Estimation can then proceed with the standard fixed-effects estimator on the cohort means, thus eliminating any unobserved differences between individual cohorts.

[1]In addition to these variables we have used education levels, regions, occupations, marital status dummies. We have also used dummies for different employment status, gender and area.

[2]The real hourly wage is calculated as weekly income/number of hours worked per week and then deflated with GPI (General Price Index) for that particular year.

[3]Experience has been computed as: age-6-years of education.

Deaton (1985), argues that there is a potential measurement error problem arising from using $\overline{y}_{ct}$ as an estimate of the unobservable population cohort mean and an adjustment based on errors-in-variables techniques is therefore needed. However, researchers typically ignore this if the number of observations per cohort is reasonably large. Moreover, Verbeek and Nijman (1992) suggest that when the cohort size is at least 100 individuals, and the time variation in the cohort means is sufficiently large, the bias in the standard fixed-effects estimator will be small enough that the measurement error problem can be safely ignored. Although, this issue will be considered in the analysis, given the size of the LFSs, suitably chosen cohorts should fulfil this size criterion, hence this is the approach used in this paper.

The construction of the pseudo-panel data is undertaken by computing cohort or cell means in each available cross-section, where the cells are defined by the four-digit district codes, age of the individual, provinces and the type of industry in which the individual is working.[4] Thus in total, it results in a group between 6000 and 8000 approximately, in each pseudo-panel for each cross-section. Next we present the methodology, which is used in the paper according to the pooled as well as the pseudo panel method in estimation of return to education.

For the calculation of return to education at different level the paper uses the methodology used by Matrins and Pereira (2004) with the approximation properties illustrated by Angrist, *et al.* (2006). An ordinary least squares (OLS) regression is based on the *mean* of the conditional distribution of the regression's dependent variable. This approach is used because one implicitly assumes that possible differences in terms of the impact of the exogenous variables along the conditional distribution are unimportant.

However, this may prove inadequate in some research agendas. If exogenous variables influence parameters of the conditional distribution of the dependent variable other than the mean, then an analysis that disregards this possibility will be severely weakened [Koenker and Bassett (1978)]. Unlike OLS, quantile regression models allow for a full characterisation of the conditional distribution of the dependent variable.

In a wage equation setting, the quantile regression model cab be written as:

$$\ln w_i = x_i\beta_\theta + u_{\theta i} \text{ with } Quant_\theta(\ln w_i \mid x_i) = x_i\beta_\theta$$

where x_i is the vector of exogenous variables and β_θ is the vecor of parameters. $Quant_\theta(\ln w \mid x)$ denotes the θth conditional quantile of the ln *w* given *x*. The θthe regression quantile, 0<θ<1, is defined as a solution to the problem:

$$\min_{\beta\in R^k}\left\{\sum_{i:\ln w_i \geq x_i\beta}\theta\left|\ln w_i - x_i\beta_\theta\right| + \sum_{i:\ln w_i \geq x_i\beta}(1-\theta)\left|\ln w_i - x_i\beta_\theta\right|\right\}$$

This is normally written as:

$$\min_{\beta\in R^K}\sum\rho_\theta(\ln w_i - x_i\beta_\theta)$$

[4]We choose to use the four-digit district codes, age, provinces, education level and industry type to allow for unobserved differences between these similar individuals such as differences in the quality of their education, their skills and attitudes etc. to be controlled via fixed effects.

where $\rho_\theta(\varepsilon)$ is the check function defined as $\rho_\theta(\varepsilon) = \theta\varepsilon$ if $\varepsilon \geq 0$ or $\rho_\theta(\varepsilon) = (\theta - 1)\varepsilon$ if $\varepsilon < 0$.

This problem does not have an explicit form but can be solved by linear programming methods. The least absolute deviation (LAD) estimator of β is a particular case within this framework. This is obtained by setting θ=0.5 (the median regression). The first quartile is obtained by setting θ=0.25 and so on. As one increased θ from 0 to 1, one traces the entire distribution of *y*, conditional on *x*.

According to Angrist, *et al.* (2006)'s theorems QR implicityly provides a weighted minimum distance approximation to the true linear CQF. It is therefore useful to compare the QR fit to an explicit minimum distance (MD) fit similar to described by this authors. The MD estimator for QR is the sample analog of vector $\tilde{\beta}(\tau)$ that solves

$$\tilde{\beta}(\tau) = \arg\min_{\beta \in R^d} E\left[(Q_\tau(Y \mid X) - X'\beta)^2\right] = \arg\min_{\beta \in R^d} E\left[\Delta_\tau^2(X, \beta)\right]$$

In other words, $\tilde{\beta}(\tau)$ is the slope of the linger regression of $Q_\tau(Y|X)$ on *X*, weighted only by the probability mass function of *X*, π(x). In contrast to *QR*, this *MD* estimator relies on the ability to estimate $Q_\tau(Y|X)$ in a nonparametric first step by, which, as noted by Chamberlain (1994), may be feasible only when X is low dimensional, the sample size is large and sufficient smoothness of $Q_\tau(Y|X)$ is assumed.

At end, quantile regressions provide snapshots of different points of a conditional distribution. They therefore constitute a parsimonious way of describing the whole distribution and should bring much value-added if the relationship between the regressors and the independent variable evolves across its conditional distribution.

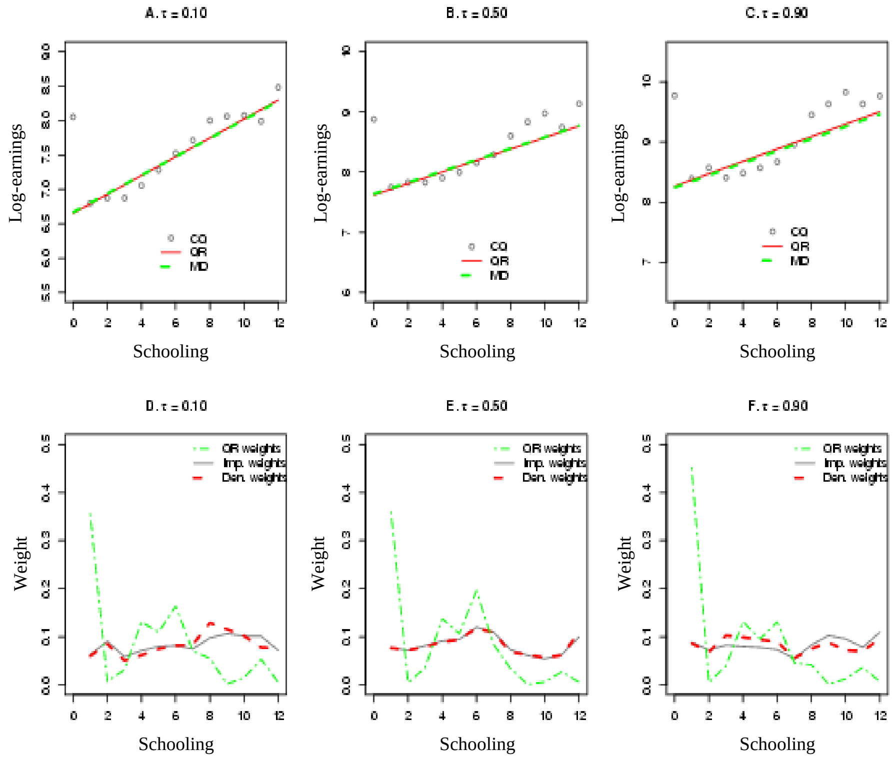
This flexibility has so far been precluded in the returns-to-education literature. In doing so, it has left unaddressed the possible impact of schooling upon inequality, through its within-levels inequality component. If the schooling-related earnings increment were the same across the wage distribution, the schooling would not impact upon within-levels wage inequality as distributions of wages conditional on different levels of schooling would differ only on their locations and not on their dispersions.

However, it may be the case that these dispersions do indeed vary across educational levels, thus resulting in an impact of schooling upon the wage distribution, through its within-level channel. This is the possibility the paper tests, by using quantile regression.

5. RESULTS

The nature of *QR* approximation property is illustrated in Figure 1 [Angrist, *et al.* (2006)]. Panel A-C plot a nonparametric estimate of the conditional quantile function $Q_\tau(Y \mid X)$, along with the linear QR fit for the 0.10, 0.50 and 0.90 quantiles, where *X* includes only schooling variable. Here, discreteness of schooling and large set of LFS data gives advantage to compare *QR* fits to the non-linear *CQFs* computed at each point in support of *X*. the figure has been drawn from the pooled data, which contains eight *LFS* surveys over fourteen years. Figure 1 plots *MD* fit (as explained in methodology) with a dashed line. The *QR* and *MD* regression lines are close, as predicted but they are not identical. To further investigate the *QR* weighting function, panel D-F in Figure 1 plot

Fig. 1. Conditional Quantile Function and Weighting Schemes in LFS



the overall *QR* weights against the regressor *X*. the panels also show estimates of the importance and their density approximations. The importance weight and the actual density weights are fairly close.

Table 1 (in Appendix) represents the overall return to education for different level of education, using different set of variables where findings suggests that the model with all different set of variables is the best fit model according to the R-squared and the Hausman test. So, in carrying out the further analysis, the study uses that model, which includes the personal and household characteristics, as well as employment status and the occupation. Table 1 depicts that the education coefficients are almost significant in all the models and the coefficient value decreases from raw return education after introducing different set of variables. The coefficient of age and experience shows substantial increases in wage with each additional year. The concavity of age-earnings profile is evident from the negative and significant coefficient of experience squared. The negative and significant coefficient of gender (–0.565) and regional dummies (–0.138) strengthens *a priori* expectation that females earn less than males and earnings are lower in rural areas as compared to urban area. These estimates are consistent with earlier studies [Khan and Irfan (1985), Arshaf and Asharf (1993) and Nasir and Nazil (2000)].

Table 2

Comparison of CQF and QR-Based Interquantile Spread

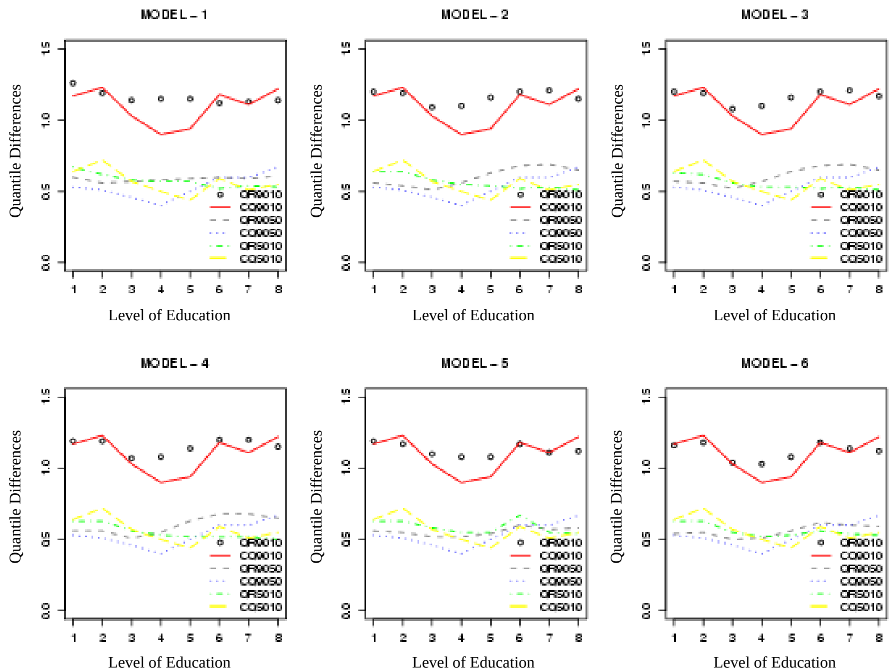
A. POOLED ESTIMATION							
	LFS	Model1	Model2	Model3	Model4	Model5	Model6
Interquatile Spread	Obs.	97122	97122	97122	97122	97122	97122
90–10	CQ	1.25	1.25	1.25	1.25	1.25	1.25
	QR	1.31	1.27	1.26	1.25	1.27	1.23
90–50	CQ	0.59	0.59	0.59	0.59	0.59	0.59
	QR	0.64	0.61	0.63	0.62	0.61	0.59
50–10	CQ	0.66	0.66	0.66	0.66	0.66	0.66
	QR	0.67	0.65	0.63	0.66	0.63	0.63
B. PSEUDO ESTIMATION							
	LFS	Model1	Model2	Model3	Model4	Model5	Model6
Interquatile Spread	Obs.	47344	47344	47344	47344	47344	47344
90–10	CQ	1.10	1.10	1.10	1.10	1.10	1.10
	QR	1.19	1.16	1.16	1.15	1.14	1.12
90–50	CQ	0.51	0.51	0.51	0.51	0.51	0.51
	QR	0.58	0.57	0.58	0.57	0.55	0.54
50–10	CQ	0.59	0.59	0.59	0.59	0.59	0.59
	QR	0.61	0.59	0.58	0.58	0.59	0.58

Model-1 Education, Experience, Experience^2, Female.
Model-2 Model-1 + Occupational Dummies.
Model-3 Model-1 + Employment Status Dummies.
Model-4 Model-1 + Occupational and Employment Status Dummies.
Model-5 Model-1 + Household Characteristics and Marital Status.
Model-6 Model-5 + Occupational and Employment Status Dummies (Full Model).

Also of interest is the ability of QR to track changes over time in quanitle-based measures of conditional inequality. Before analysing changes over time, the paper describes the overall conditional inequality using six different models. The row labelled *CQ* in Table 2 panel A shows nonparametric estimates of the average 90-10 quantile spread conditional on different set of endogenous variable as explained above. Quanile regression estimates match with *CQ* estimates also perfectly with Model 6. So, it is the best-fit model as well. The conclusion is same from the pseudo panel data as well, which is depicted in Table 2 panel B.

The fit is not as good, however, when averages are calculated for specific groups, as reported in Figure 2. These results highlight the fact that *QR* is only approximation. Figure 2 shows the quantile difference for different models at specific level of education. The *CQ* lines in Figure 2 are identical for all the models at different quantile interval as *CQ* is the descriptive wage differential for that interval which will remain constant in different models. As seen from the figure, the highest conditional inequality is in quantile 90-10 for the education group having post-graduate degree, while lowest is found in education group who has done Matriculation but less than Intermediate. The findings are also similar for the uanile spread 90-50. although, for the quantile spread 50-10, education group having done primary found to have highest conditional inequality, while

Fig. 2. Inequality for Different Models at Different Education Levels



having intermediate but not completing degree found to have lowest inequality in this quantile spread. A, noted from all the results, findings obtained from pseudo panel are fairly same as obtained from pooled data. So, paper uses estimates obtained from pseudo panel data for further analysis.

The analysis has been categorised according to different provinces, regional area, gender and the individual's working industry to get insight of the education inequality in different areas. Table 3 shows the overall inequality for provinces, Punjab, Sindh, NWFP and Balochistan, gender, Male and Female, area of living, Urban and Rural, and basic industries, Agriculture *and Fishing*; *Mining and Quarrying*; *Manufacturing*; *Electricity*, *Gas and Water Supply*; *Construction*; *Wholesale and Retail Trade*, *Hotels and Restaurants*; *Transport*, *Storage and Communication*; *Financial Intermediation and Community*, *Social and Personal Services*, which are classified by Pakistan Standard Industrial Classification. As depicted in table, Punjab has the highest conditional inequality across all the quantiles while Balochistan has the lowest conditional inequality in all quantiles spread compare to other provinces. According to finding from PSLM (Pakistan Social and Living Measurement Survery) 2004-05, Sindh has highest literacy rates, the education inequality is higher in Punjab according to the papers estimates, which could be due to the reason of migration as more people migrate to Punjab compare to all other provinces in search of better jobs or opportunity. In case of area of living and gender, rural area and female found to have more conditional inequality compare to urban area and male, respectively. The discrepancies at the industry level persist ranging from Agriculture with highest inequality 1.21 and Mining at0.80 for the quantile spread 90-10.

Table 3

Comparison of CQF and QR-Based Interquantile Spread for Different Categories

		Interquantile Spread					
		90-10		90-50		50-10	
Category	Obs.	CQ	QR	CQ	QR	CQ	QR
		A. Provinces					
Punjab	22178	1.13	1.14	0.52	0.53	0.62	0.60
Sindh	10481	0.99	0.99	0.45	0.49	0.54	0.50
NWFP	9483	1.04	1.09	0.49	0.53	0.55	0.56
Balochistan	5202	0.80	0.84	0.37	0.42	0.43	0.42
		B. Gender					
Male	44687	1.08	1.06	0.50	0.51	0.58	0.55
Female	2657	1.41	1.53	0.59	0.64	0.83	0.88
		C. Area of Living					
Urban	26400	1.01	1.04	0.47	0.50	0.54	0.54
Rural	20944	1.15	1.13	0.52	0.52	0.63	0.60
		D. Industries					
Agriculture	3210	1.21	1.30	0.62	0.63	0.60	0.67
Mining	719	0.80	0.99	0.35	0.47	0.45	0.52
Manufacturing	6267	1.15	1.18	0.53	0.54	0.62	0.64
Electricity, Gas and Water	3390	0.86	0.90	0.38	0.42	0.48	0.48
Construction	7440	1.04	1.01	0.48	0.48	0.56	0.54
Trade and Restaurants	5188	0.91	0.97	0.42	0.45	0.49	0.51
Transport	5520	0.99	1.05	0.45	0.49	0.55	0.56
Financial Intermediaries	2105	0.93	1.05	0.46	0.52	0.46	0.53
Social Services	13505	1.01	1.07	0.47	0.50	0.54	0.57

Findings according to different level of education, for quantile spread, overall results suggest Punjab having the highest differential in all quanitle spread for different level of education. As, depicted in Figure 3 the quantile inequality is less up to having done intermediate but not having degree compare to have degree or further education. Balochistan is exclusion in this as in Balochistan inequality rate is very low compare to other provinces especially for the education group who has degree in Agriculture, Medicine or Engineering.

Fig. 3. Education Inequality in Different Provinces

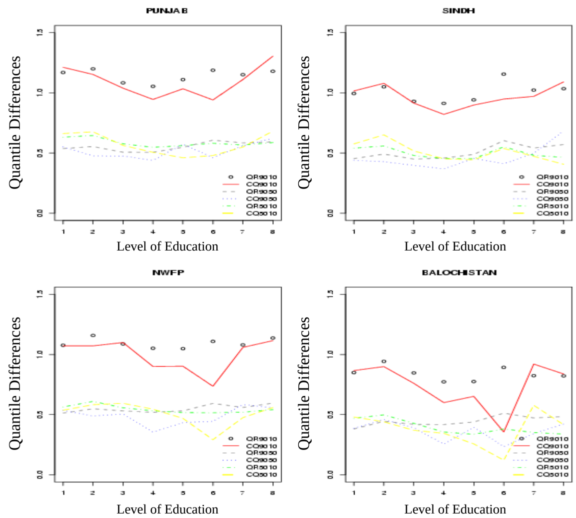
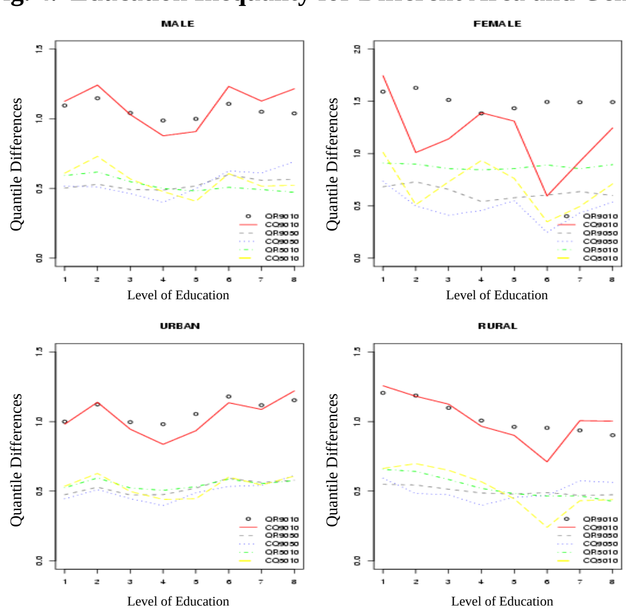


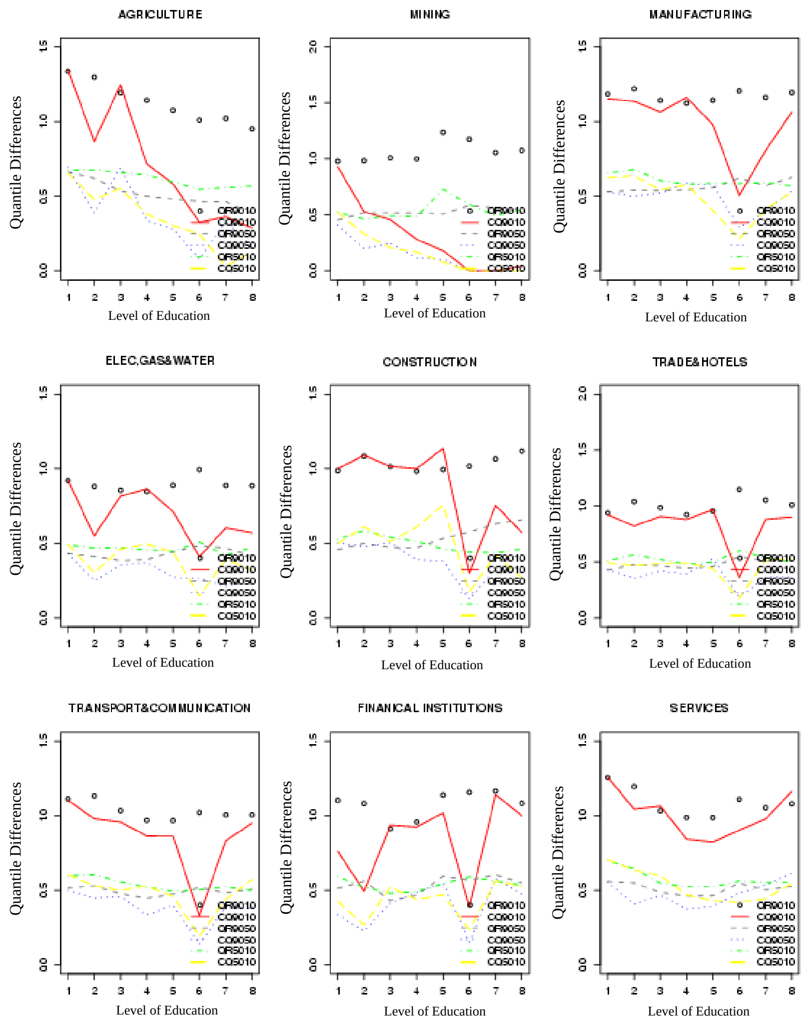
Figure 4 shows the conditional inequality at different level of education for male and female as well as for urban and rural area. Female found to have higher inequality at all the education level compare to male as in Pakistan female literacy ratio is only 40 percent (PSLM, 2004-05), so not many female acquiring high level of education which rises the inequality at different level of education. Observing conditional inequality for urban and rural, urban found to have higher inequality. Rural found to have decaying line of conditional inequality as person who acquires higher qualification migrate to urban areas.

Fig. 4. Education Inequality for Different Area and Gender



Categorising into different industries, *Community, Social and Personal Services* found to have overall highest conditional inequalities between all quantile spread at all different education level compare to all other industrial sectors. Analysis does not include Mining industry due to having less number of observations, but just for the knowledge it's represented in Figure 5, shown below.

Fig. 5. Education Inequality for Different Industries



Agriculture found to have declining line from No Formal Education to having postgraduate degree as person having higher qualification is less likely to find in this industry. Having degree in Agriculture, Medicine or Engineering found to lave lowest conditional inequality for all the industries except for the Service sector. Electricity, Gas and Water and Trade and Hotels have the lowest conditional education inequality across all the level of education.

Fig. 6. Overall Education Inequality for the Time Period of 1990 to 2003

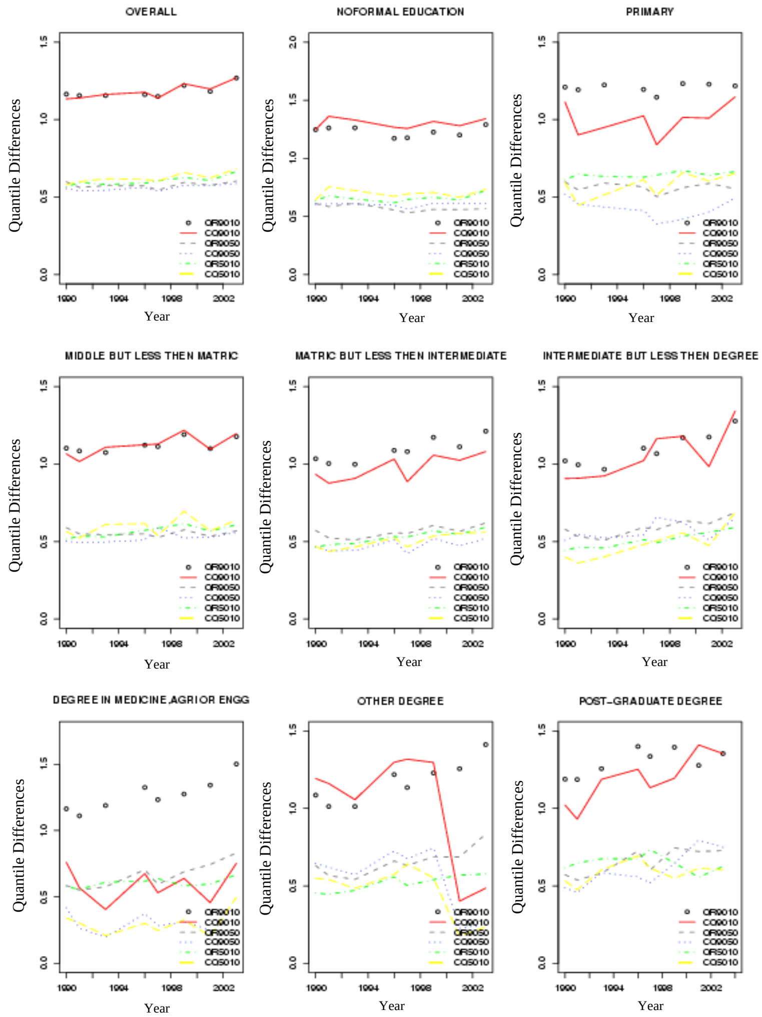
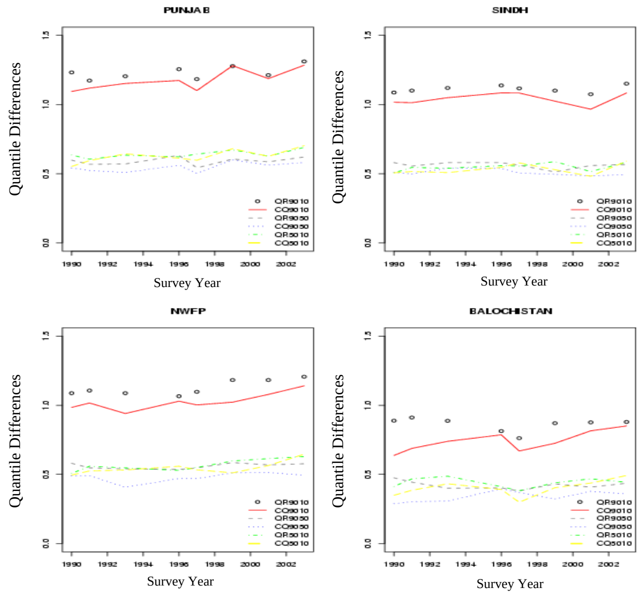


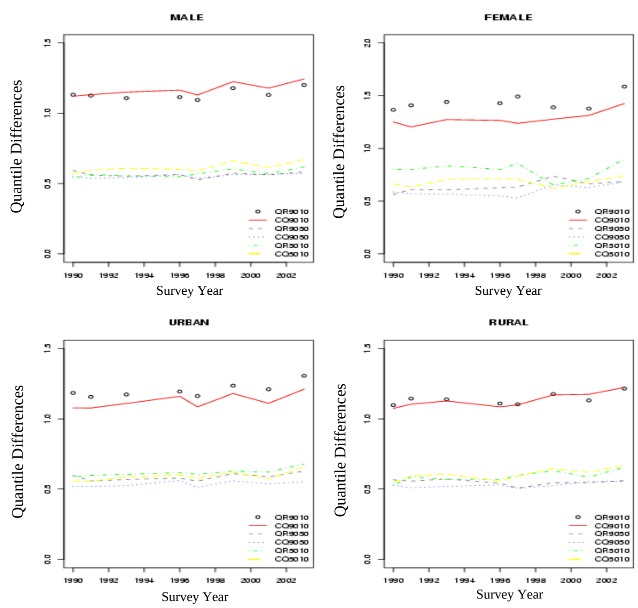
Figure 6 shows nonparametric estimates of average quantile spread over the time period of 1990 to 2003. The spread increased from 1.13 to about 1.17 from 1990 to 1996, and then to 1.26 from 1996 to 2003. Figure 6 documents some important substantive findings, apparent in both the CQ and QR estimates. The overall figure shows that conditional inequality increasing in the upper half as well as lower half of the distribution. The increase in conditional inequality is much higher for person who has done matriculation but not intermediate or who has done intermediate but not have the degree or having post graduate degree compare to other level of educations. There is very small increase in conditional inequality for the education group who has done primary or who had Degree in Agriculture, Medicine or Engineering. Figure of Degree in Agriculture, Medicine or Engineering shows the wide gap between the line of CQ and QR, which is due to less number of observation at this education level which leads to bias QR approximation.

Fig. 7. Education Inequality Over the Years for Different Provinces



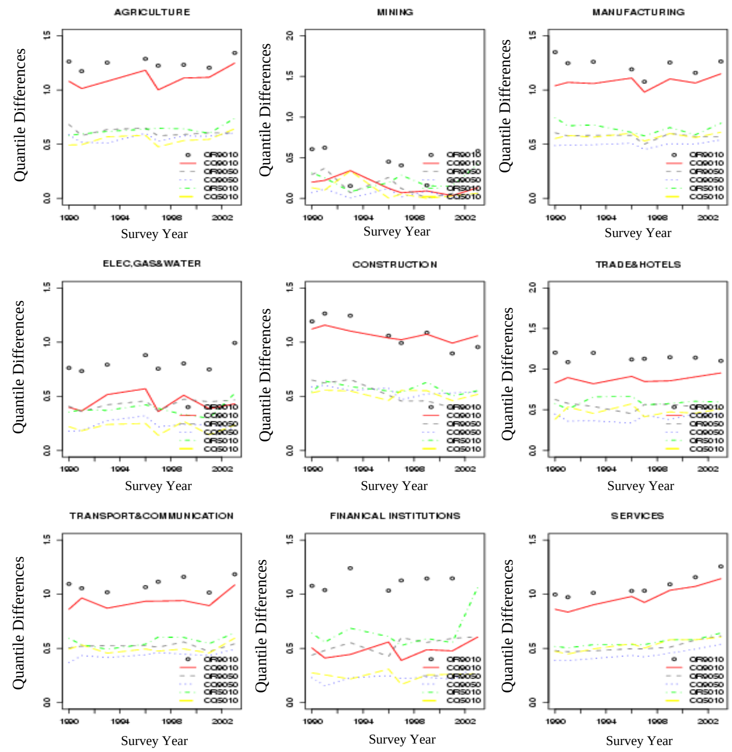
The conditional inequality estimates for different provinces is depicted in Figure 7 where Punjab and Balochistan found to have highest increase in conditional inequality over the year, from 1.09 to 1.17 and from 0.64 to 0.79 over 1990 to 1996 and then to 1.28 and to 0.85 for year 1996 to 2003, respectively. Sindh found to have more or less stable inequality as it found to have increase by only 0.7 over fourteen years of time period, which is almost less than half increase compare to other provinces of Pakistan.

Fig. 8. Education Inequality for Gender and Area of Living over the Year



For female, the conditional inequality increase is slightly more compare to increase for male but the inequality remain higher for female compare to male. Female's inequality is increased from 1.25 to 1.43 from 1990 to 2003 while it is 1.12 to 1.24 for male for the same time period. This is drawn in Figure 8. Urban and rural found to have almost same sort of increasing trend in conditional inequality, urban being more or less similar in inequality term compare to rural. The inequality has increased from 1.08 to1.21 and 1.07 to 1.22 for urban and rural over the time period of 1990 to 2003.

Fig. 9. Education Inequality for Different Sectors between 1990 and 2003



In the different industry sector, service sector found to have upward line in conditional inequality, which also shows the increasing trend over the year compare to all other industries. Financial Institutions and Trade and Hotels found to have minimal increase in conditional inequality as drawn in Figure 9. Construction sector found to have decrease in conditional inequality till year 1998 but increasing there after. Declining trend also found for the sector Electricity, Gas and Water and it also has the lowest conditional inequality among all the sectors. Agriculture and Transport and Communication shows increase in conditional inequality from 1.08 to 1.25 and from 0.86 to 1.09, respectively, over 1990 to 2003. The results strongly endorse the existence of education inequality in Pakistan, which also found to be increasing over the time in different provinces and different sectors. Inequality also exists for having same level of education across the wage distribution and which is quite high at middle education compare to have no education and having high qualification.

6. CONCLUSION

This paper uncovers evidence that education inequality in Pakistan exhibit substantial heterogeneity across the income distribution. Due to lack of data previous studies are lacking in observing role of variables on earnings over the time. As LFS provides information on different level of schooling for each time period used in this study, this paper not only identifies the education in equality but it also measures the trend of education in equality over the time. The paper uses quantile regression approach developed by Angrist, *et al*. (2006), which captures the correction bias for omitted variables. The empirical estimates would appear to suggest that the inequality for the quantile spread 90-10 is much higher compare to other quantile spread in the distribution and it also found to increasing over the year from 1.13 to 1.26 for the time period of 1990 to 2003. it also documents the existence of education inequality in different regions, provinces, gender and industry.

Punjab found to have more education inequality compare to all other provinces, while Balochistan has the lowest inequality. Punjab’s education inequality is due to the migrations in this province. Female found to have more inequality compare to male and the inequality gap between male and female is quite higher compare to all other categories, female’s inequality is increased from 1.25 to 1.43, while it’s 1.12 to 1.24 for male over the time period of 1990 to 2003. over the time, inequality trend found to be almost similar for urban and rural area, but when analysed at different level of education rural found to have decaying line for the high level of education compare to urban area. In industry, Services sector found to have highest increase in the inequality over the time and also for the different level of education it has the high inequality compare to other.

For, different level of education, conditional inequality has increased for both upper half and lower half of the distribution and the increase in conditional inequality is much higher for person who has done matriculation or intermediate or having degree or postgraduate degree compare to have less education or no education. Having degree in Agriculture, Medicine or Engineering found to have less inequality compare to all other education level.

The main policy implication from the findings is requirement of narrowing the disparities between the education inequality for male and female which is quite high and even within the female category the inequality is quite high between upper half and lower half, this requires not only an increase in the budgetary allocation for female education but also its optimal utilisation.

APPENDIX

Table 1

OLS Estimation Results for Different Specifications

	Model-1	Model-2	Model-3	Model-4	Model-5	Model-6	Model-7
prim	0.107**	0.193**	–0.026	–0.003	–0.011	–0.004	–0.009
	–17.29	–32.82	–1.49	–0.18	–0.62	–0.23	–0.55
middle	0.259**	0.366**	0.013	0.032	0.025	0.038	0.016
	–33.59	–50.31	–0.5	–1.22	–0.95	–1.44	–0.63
matric	0.427**	0.556**	0.062	0.071*	0.069*	0.124**	0.052
	–64.07	–87.76	–1.91	–2.21	–2.18	–3.83	–1.65
inter	0.594**	0.767**	0.140**	0.145**	0.152**	0.217**	0.116**
	–65.73	–90.48	–3.57	–3.75	–3.95	–5.56	–3.05
profess	1.297**	1.465**	0.628**	0.667**	0.663**	0.668**	0.558**
	–76.89	–94.66	–11.81	–12.7	–12.7	–12.58	–10.77
uni	0.954**	1.122**	0.379**	0.400**	0.403**	0.444**	0.331**
	–84.35	–107.03	–8.28	–8.85	–8.95	–9.72	–7.43
pgrad	1.066**	1.269**	0.457**	0.492**	0.493**	0.494**	0.397**
	–91.23	–117.11	–8.79	–9.57	–9.64	–9.51	–7.83
exper		0.048**	0.013**	0.013**	0.014**	0.012**	0.014**
		–93.1	–3.91	–3.97	–4.41	–3.75	–4.36
exper2		–0.001**	–0.001**	–0.001**	–0.001**	–0.001**	–0.000**
		–69.82	–64.79	–60.45	–59.5	–55.07	–48.53
female		–0.523**	–0.538**	–0.548**	–0.547**	–0.539**	–0.565**
		–81.63	–85.37	–86.81	–86.52	–81.33	–85.32
age			0.031**	0.027**	0.025**	0.029**	0.022**
			–9.7	–8.65	–8.11	–9.11	–7.04
pubpriv			–0.118**	–0.080**	–0.101**	–0.114**	–0.099**
			–24.24	–16.66	–20.94	–23.75	–20.65
rural			–0.172**	–0.176**	–0.165**	–0.160**	–0.138**
			–41.76	–44.03	–40.64	–38.87	–33.53
spedu						0.041**	0.037**
						–40.91	–38.38
h616						–0.008**	–0.005**
						–6.26	–4.31
hun1665						0.022**	0.018**
						–14.08	–11.44
heun65						–0.025*	–0.029*
						–1.99	–2.37
hhfem						0.180**	0.185**
						–7.91	–8.35

Continued—

Table 1—(*Continued*)

married						0.051**	0.043**
						–8.16	–7.1
widow						–0.068**	–0.065**
						–4.23	–4.23
divorced						–0.041	–0.036
						–1.18	–1.07
ychild						–0.060**	–0.055**
						–12.65	–12.02
tech			–0.107**	–0.141**	–0.100**		–0.093**
			–14.32	–19.41	–13.72		–12.88
wcjob			0.068**	0.153**	0.044**		0.026**
			–7.65	–25.95	–5.06		–3.05
cpwork				–0.151**	–0.132**		–0.125**
				–28.32	–24.1		–23.04
pwork				–0.093**	–0.087**		–0.074**
				–16.3	–14.7		–12.59
pfapp				–1.016**	–1.056**		–1.040**
				–55.65	–57.65		–57.29
clerks			–0.163**		–0.158**		–0.147**
			–16.91		–16.68		–15.69
servwrk			–0.156**		–0.179**		–0.181**
			–17.16		–20.04		–20.4
sagfwk			–0.251**		–0.249**		–0.240**
			–19.6		–19.78		–19.27
crftwk			–0.143**		–0.098**		–0.100**
			–17.48		–11.85		–12.22
eleocc			–0.219**		–0.203**		–0.200**
			–27.88		–25.82		–25.58
Constant	7.692**	7.082**	7.422**	7.420**	7.563**	7.140**	7.566**
	–2287.59	–990.16	–319.18	–332.07	–329.01	–325.04	–323.73
Observation	97102	97102	97102	97102	97102	97102	97102
R-squared	0.19	0.33	0.37	0.38	0.39	0.36	0.4

Absolute value of t statistics is below the coefficient value.
* significant at 5 percent; ** significant at 1 percent.
Model-1 Only Educational Dummies.
Model-2 Educational Dummies, Experience, Experience^2,Female.
Model-3 Model-2 + Occupational Dummies.
Model-4 Model-2 + Employment Status Dummies.
Model-5 Model-2 + Occupational and Employment Status Dummies.
Model-6 Model-2 + Household Characteristics and Marital Status.
Model-7 Model-6 + Occupational and Employment Status Dummies (Full Model).

REFERENCES

Angrist, J., V. Chernozhukov, and I. Fernandez-Val, (2006) Quantile Regression under Misspecification, with An Application to the U.S. Wage Structure. *Econometric* 74:2, 539–563.

Appleton, S. (2001) Education, Incomes and Poverty in Uganda in the 1990s. School of Economics, University of Nottingham. (CREDIT Research Paper No. 01/22.)

Arias, O., K. F. Hallock, and W. sosa-Escudero (2001) Individual Heterogeneity in the Returns to Schooling: Instumental Variables Quantile Regression Using Twins' Data. *Emprical Economics* 26, 7–40.

Ashraf, J. and B. Ashraf (1993) Estimating the Gender Wage Gap in Rawalpindi City. *The Journal of Development Studies* 29:2, 365–76.

Bauer , T. K., P. J. Dross, and J. P. Haisken-DeNew (2002) Sheepskin Effects in Japan. (IZA Discussion Paper No. 593.)

Bennell, P. (1996) Rates of Return to Education: Does the Conventional Pattern Prevails in SubSaharan Africa? *World Development* 24:1, 183–200.

Card, D. (2001) Estimateing the Returns to Schooling: Progress on Some Persistent Econometric Problems. *Econometrica* 69:5, 1127–60.

Card, D. (1999) The Casual Effect of Education on Earnings. In O. Ashenfelter and D. Card (eds.) *Handbook of Labour Economics,* Vol.3A. Amsterdam: Elsevier Science, NorthHolland and 1801–63.

Chamberlain, G. (1984) Panel Data. In Z. Griliches and M. Intriligator (eds.) *Handbook of Econometrics*, Vol 2. Amsterdam: North-Holland, 1247–1318.

Colclough, C. (1982) The Impact of Primary Schooling on Economic Development: A Review of the Evidence. *World Development* 10:3, 167–85.

Deaton, A. (1985) Panel Data from a Time-series of Cross-sections. *Journal of Econometrics* 30, 109–126.

Griliches, Z. (1977) Estimate the Returns to Schooling: Some Econometric Problems. *Econometrica* 45:1, 1–22.

Guisinger, S. E., J. W. Henderson, and G. W. Scully (1984) Earnings, rate of Returns to Education and Earning Distribution in Pakistan. *Economics of Education Review* 3:4.

Hamdani, K. (1977) Educaiton and the Income Differentials: An Estimation of Rawalpindi City. *The Pakistan Development Review* 16:2.

Khan, S. R. and M. Irfan (1985) Rate of Returns to Education and Determinats of Earnings in Pakistan. *The Pakistan Development Review* 34:3&4.

Koenker, R. and G. Bassett (1978) Regression Quantiles. *Econometrica* 46, 33–50.

McMohan, W. W. (1999) Education and Development: Measuring the Social Benefits. Oxford: Oxford University Press.

Mwabu, G. and Schultz (1996) Education Returns across Quantiles of the Wage Function: Alternative Explanations for Returns to Education by Race in South Africa. *American Economic Review* 86:2, 153–58.

Nasir, Z. M. and H. Nazil (2000) Education and Earnings in Pakistan.

Pereira, P. T. and P. S. Martins (2004) Returns to Education and Wage Equations. *Applied Economics* 36, 525–531.

Psacharopoulos, G. (1994) Returns to Investment in Education: A Global Update. *World Development* 22:9, 1325–44.

Verbeek, M. and T. Nijman (1992) Testing for Selectivity Bias in Panel Data Models. *International Economic Review* 33:3, 681–703.