Sindh Univ. Res. Jour. (Sci. Ser.) Vol. 48 (4) 919-924 (2016)

SINDH UNIVERSITY RESEARCH JOURNAL (SCIENCE SERIES)

Quantile Regression Analysis of Monthly Earnings in Pakistan

I. A. ARSHAD[++], U. YOUNAS, A. W. SHAIKH*, M. S. CHANDIO*

Department of Statistics, Allama Iqbal Open University, Islamabad.

Received 6th May 2016 and Revised 29th November 2016

Abstract: In this study, we empirically analyze the monthly earning distribution of Pakistan. The log of monthly earning is taken as a response variable, while education, experience, age, sex, marital status, nature of work, region, and the provinces are used as explanatory variables. Ordinary least square regression and quantile regression techniques are used to estimate the relationship among these variables. Quantile regression, instead of the point estimate of the conditional mean, can be used to estimate the whole distribution, especially the upper tail and lower tail which we are interested in. The comparison of OLS, and quantile regression shows that quantile regression can provide more informative estimation results. We also use quantile regression's equivariant property to transform our response variable from log to level.

Keywords: quantile regression, OLS regression, earnings distribution.

1. INTRODUCTION

Earning is the key motivating factor of human activities. According to the classical and modern economists, materialistic life is the function of earning and expenditure. According to (Robbins 1945) "economics is the study of human behavior adopted between unlimited end and scare means". Every human tried his best to spend his life with his limited resources. Earning and earning related issues are very important to study for researchers. The most important earning related issue is finding out the factor that influences the earning. In this study education, experience, age, sex, marital status, are taken as main factors towards monthly earning.

Koenker (1978) introduced quantile regression which models, conditional quantiles as functions of predictors. The quantile regression model is a natural extension of the linear regression model. While the linear regression model specifies the change in the conditional mean of the dependent variable associated with a change in the covariates (Wooldridge, 2013) the quantile regression model specifies changes in the conditional quantile. Since any quantile can be used, it is possible to model any predetermined location of the distribution. Ordinary linear regression estimates the average relationship between a set of regressors and the dependent variable based on the conditional mean $\mathbf{E(Y|X)}$ **(**Koenker and Bassett 1978**)**. In quantile regression, we use the conditional median$\mathbf{Q_q(Y|X)}$to estimate the relationship between dependent and independent variables, the quantile "$\mathbf{q}$" ϵ (1 , 0), where the median is the 50th quantile (Chen, *et al*., 2009).

It is a well-known fact that education, experience, age and marital status highly correlated with earning. Many researchers from different countries like Brazil, Portugal, Turkey, USA, UK, India, etc. (Tansel and Bodur 2012, Taylor 1999, Budria 2010 Filippaki *et al.,* 2012) had applied quantile regression. Sarwar and Sial(2012) analyzed the effect of education on different levels of log of earning distribution, using quantile regression technique and used Pakistan Standard Living Measurement (PSLM) dataset for the period of 2007-08. They found that education has a significant effect on the log of earning. The effect of education has witnessed an increasing trend from lower to upper quantile and log of earning also increased in different categories of education. Aslam *et al.,* (2014) analyzed the difference between the male and female income of workers in Punjab and used Pakistan Labor Force Survey (PLFS) data for the period of 2008-09. They used the median regression which is also called quantile regression when q =0. 5. They found males earn more income than females. Other predictors like age, area, marital status, different levels of education, and nature of work have also been found to be statistically significant. Davino *et al.,* (2014) explained in their book "Quantile Regression Theory and Application" that ordinary least square regression provides results on the basis of

[++]Corresponding author: I. A. ARSHAD, Email: irshad.ahmad@aiou.edu.pk
*Institute of Mathematics and Computer Science, University of Sindh, Jamshoroo

conditional mean; quantile regression expanded this idea of the whole conditional distribution of the dependent variable. Quantile regression provides location, scale, and shape shift information on the conditional distribution of the dependent variable. Quantile regression is a useful technique in the presence of heteroscedasticity and it is also useful to analyze behavior of the dependent variable at multiple location of the distribution. In the case when residuals were not normally distributed, quantile regression does not need this assumption. Hao and Naiman (2007) compared quantile regression models and ordinary least square model,in their book "Quantile Regression". They described quantile regression model's advantages over ordinary least square regression such as robustness, monotonic equivariance properties, which provide flexible estimates.

Following are the main objectives of this study:

1. To estimate the effect of various variables like education, experience, age, etc on the earnings distribution in Pakistan.
2. To see the behavior of earning distribution at multiple location (quantile).

2. MATERIALS AND METHODS

2.1 Data

We used the secondary data obtained from Pakistan Bureau of Statistics, about Pakistan Social and Living Standards Measurement (PSLM) survey for the year 2012-2013. The key variables of (PSLM) survey are demographic characteristics; education, health, employment, etc. Since the response variable is earning of the individuals, therefore individuals having age between 15 to 65 years are the target values for analysis.

2.2 Model Specification

This study attempts to estimate the impact of education, experience, age, sex, marital status, nature of work, region, and the provinces on the earnings distribution in Pakistan and our quantile regression model is

$$\begin{aligned}\mathbf{Q_q(y_i)} = {} & \boldsymbol{\beta_{q0}} + \boldsymbol{\beta_{q1}}\mathbf{age_{qi}} + \boldsymbol{\beta_{q2}}\mathbf{male_{qi}} \\ & + \boldsymbol{\beta_{q3}}\mathbf{urban_{qi}} + \boldsymbol{\beta_{q4}}\mathbf{mar_{qi}} \\ & + \sum_{\mathbf{j=5}}^{\mathbf{8}} \boldsymbol{\beta_{qj}}\mathbf{pro_{ji}} + \sum_{\mathbf{k=9}}^{\mathbf{17}} \boldsymbol{\beta_{qk}}\mathbf{edu_{ki}} \\ & + \sum_{\mathbf{m=18}}^{\mathbf{27}} \boldsymbol{\beta_{qm}}\mathbf{nat_{mi}} + \boldsymbol{\varepsilon_{qi}}\end{aligned}$$

We replace $\boldsymbol{\beta_q}$ for $\boldsymbol{\beta}$ to show the particular coefficients of Quantile regression model. $\boldsymbol{\varepsilon_{qi}}$, is an error term of Quantile regression model.

2.3 Model Building

Parameters for linear regression OLS are estimated by minimizing the sum of squared errors. This guarantees that the model is an optimum estimate of the expected conditional mean. In contrast, quantile regression obtains parameters by minimizing a weighted average of absolute errors. A detailed description of the quantile regression method is given in Koenker and Bassett (1978), Buchinsky (1998) and Hao and Naiman (2007). The quantile regression model has the following form;

$$\mathbf{y_{i=}}\mathbf{x_i'}\boldsymbol{\beta_q} + \mathbf{e_{qi}} \text{ and } \mathbf{Q_q(y_i|x_i)} = \mathbf{x_i'}\boldsymbol{\beta_q}$$

Where $\mathbf{X_i}$ is the matrix of independent variables and $\boldsymbol{\beta_q}$ is the vector of parameters and $\boldsymbol{\varepsilon_{qi}}$ is error term.$\mathbf{Q_q(y_i|X_i)}$Denotes the $\mathbf{q_{th}}$ conditional quantile$\mathbf{y_i}$ of given $\mathbf{X_i}$**.** Let suppose $\mathbf{X_i}$ and $\boldsymbol{\varepsilon_{qi}}$ are not correlated with each other. $\mathbf{Q_q(\varepsilon_{qi}|x_i) = 0}$ We use the conditional median $\mathbf{Q_q(Y|X)}$, to estimate the relationship between dependent and independent variables, where the median is the 50th percentile. The quantile $\mathbf{q}$,"$\mathbf{q} \in (1\,,0)$"is defined as that value of $\mathbf{Y}$ that splits the data into proportions $\mathbf{q}$ below and **(1 – q)** above.

The $\mathbf{q_{th}}$quantile regression estimator $\widehat{\boldsymbol{\beta_q}}$ minimizes over $\boldsymbol{\beta_q}$ the objective function

$$\mathbf{Q}(\boldsymbol{\beta_q}) = \sum_{\mathbf{i:y_i \geq x_i'\beta}} \mathbf{q}\left|\mathbf{y_i} - \mathbf{x_i'}\boldsymbol{\beta_q}\right| + \sum_{\mathbf{i:y_i < x_i'\beta}} (\mathbf{1}-\mathbf{q})\left|\mathbf{y_i} - \mathbf{x_i'}\boldsymbol{\beta_q}\right|$$

This objective function is not differentiable; it minimizes through linear programing with the simplex method.

2.4 Retransformation

If the dependent variable **"y"** is in the form of natural log, then quantile regression gives the marginal effect in log **"y",** if we want to compute the marginal effect on levels Davino, *et al.* (2014) suggested retransformation property.

In ordinary least square regression

$\mathbf{log}\,[\,\mathbf{E}\,(\mathbf{y}|\mathbf{x})] \neq \mathbf{E}\,[\mathbf{log}(\mathbf{y})|\mathbf{x}]$But in quantile regression

$$\mathbf{Q_q}[\,\mathbf{log}(\mathbf{y})|\mathbf{x}] = \mathbf{log}\,[\mathbf{Q_q}(\mathbf{y}|\mathbf{x})]$$
$$\mathbf{Q_q}(\mathbf{log\,y}|\mathbf{x}) = \mathbf{X'}\boldsymbol{\beta_q}$$
$$\mathbf{Q_q}(\mathbf{y}|\mathbf{x}) = \mathbf{exp}[\mathbf{Q_q}(\mathbf{logy}|\mathbf{x})$$
$$\mathbf{exp}\big[\mathbf{Q_q}(\mathbf{logy}|\mathbf{X})\big] = \mathbf{exp}(\mathbf{X'}\boldsymbol{\beta_q})$$

The marginal effect of "y" in levels is

$$\frac{\partial \mathbf{Q_q}(\mathbf{y}|\mathbf{X})}{\partial \mathbf{x_j}} = \mathbf{exp}(\mathbf{X'}\boldsymbol{\beta_q})\,\boldsymbol{\beta_{qj}}$$

3. RESULTS AND DISCUSSIONS

Quantile regression coefficients are estimated at 5 different levels, namely 5th, 25th, 50th, 75th, and 95th antile

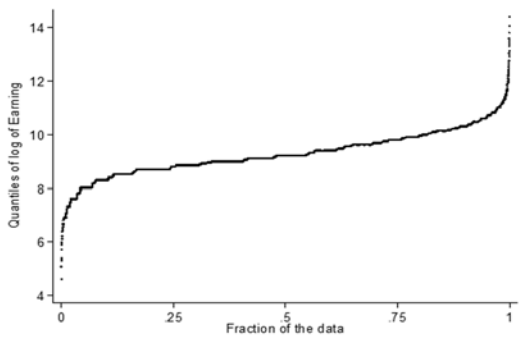

Fig.1. Log of Earning by Quantile

(Fig. 1). Indicates log of earning by quantiles, y-**axis** indicates the values of log of earning and x- axis shows the quantile of the dependent variable. We sort the data from smallest to the largest, value of the monthly log of earning is very low at lower quantiles, and very high at upper quantiles. The distribution of log of earning looks approximately normal between 10^{th} and 90^{th} quantiles. Fig. 1 also indicates that both upper and lower tails have extreme values.

Table 1. OLS coefficients and QR coefficients at five different quantiles

.

Variable	OLS	QR-05	QR-25	QR-25	QR-75	QR-95
male	0.73***	1.06***	1.01***	0.74***	0.45***	0.33***
urban	0.15***	0.16***	0.14***	0.12***	0.13***	0.16***
married	0.19***	0.34***	0.20***	0.14***	0.11***	0.09***
age	0.02***	0.01***	0.01***	0.02***	0.02***	0.02***
Edu2	0.09***	0.08***	0.09***	0.08***	0.07***	0.08***
Edu3	0.22***	0.19***	0.17***	0.19***	0.21***	0.24***
Edu4	0.40***	0.34***	0.32***	0.37***	0.39***	0.40***
Edu5	0.57***	0.33**	0.48***	0.57***	0.61***	0.62***
Edu6	0.66***	0.58***	0.61***	0.60***	0.59***	0.66***
Edu7	0.89***	0.78***	0.83***	0.83***	0.83***	0.88***
Edu8	1.30***	1.20***	1.17***	1.24***	1.25***	1.40***
Edu9	1.59***	1.57***	1.55***	1.48***	1.47***	1.40***
pro1	0.01	-0.04*	-0.00	0.02**	0.02*	0.00
pro3	-0.02**	0.04*	-0.01	-0.04***	-0.06***	-0.05***
pro4	0.23***	0.35***	0.28***	0.22***	0.18***	0.06***
Nat1	0.44***	0.53***	0.32***	0.38***	0.46***	0.59***
Nat2	0.33***	0.31***	0.37***	0.38***	0.29***	0.16***
Nat3	0.19***	0.13***	0.19***	0.25***	0.21***	0.11***
Nat4	0.17***	0.47***	0.22***	0.19***	0.11***	-0.10**
Nat5	0.01	0.24***	-0.04**	-0.05***	-0.02	-0.04*
Nat6	0.00	-0.04	-0.16***	-0.07**	-0.01	0.25***
Nat8	0.11***	0.46***	0.11***	0.05***	0.02	-0.04
Nat9	-0.08***	0.24***	-0.06***	-0.13***	-0.18***	-0.32***
_cons	7.45***	6.15***	7.01***	7.56***	8.08***	8.65***
$R^2/PseudoR^2$	0.451	0.2121	0.2343	0.2907	0.3396	0.331

Note:*Indicates p<0.05, **Indicates p<0.01, ***Indicates p<0.001

The OLS estimated coefficient of male is 0.728, and p-value is 0.0000. The male coefficient has a positive impact on log of earning. It means male log of earning is 72.8% more than female on average keeping other variables as constant. These results are given in the first column of table 1, Quantile Regression (QR) estimated coefficients for male shows that they have a positive effect on log of earning. But we can see in figure 2 and table 1 the magnitude of the effect is decreasing as quantile increases. The OLS estimated coefficient of urban is 0.149, and p-value is 0.0000. The urban coefficient has a positive impact on log of earning. It shows that urban log of earning is 14.9% more than rural on average keeping other variables as constant. Quantile Regression (QR) coefficients for urban shows that they have a positive effect on log of earning. In table .1 the magnitude of the effect is increasing as quantile increases. For example, urban coefficient is 0.16 for 5th quantile its mean urban log of earning is 16% more than rural for 5th quantile, while it will decrease log of earnings by 0.125 for 50th quantile and 0.157 for 95th quantile. All quantile regression estimated coefficients are statistically significant for this variable.

Both Ordinary Least Squares (OLS) and Quantile Regression (QR) results indicate that married variable has a positive effect on log of earning. The OLS estimated coefficient is 0.187. The married variable effect is decreasing as quantile increases for the 5th quantile 0.338, for 50th quantile 0.14 and for 95th quantile 0.092.

Age influenced the log of earning positively both OLS and quantile regression estimated coefficients statistically significant. It indicated that age positively correlated with log of earning. OLS regression and median (50th quantile) regression coefficients provide the similar results. We can **(Fig. 2)** the magnitude of the effect is increasing as quantile increases.

Education plays an important role in earning. The effect of education on log of monthly earning is positive. The effect of log of earnings increased at each level of education as we move from lower to upper quantiles. The results showed that log of earnings higher for MS/PhD individuals as compared to all other education levels across all quantiles of log of earning distribution. Both ordinary least square regression and quantile regression results showed that log of earnings is higher for (legislator, senior officials, and managers) as compared to all other nature of work levels across all quantiles of log of earning distribution. OLS and quantile regression results indicates that log of earnings is higher for Baluchistan as compared to all other provinces across all quantiles of log of earning distribution.

3.1 Retransformation Coefficients

In this study logarithm transformation used on monthly earning because it reduced the skewness problem. A distribution that is symmetric or nearly so is often easier to handle and interpret than a skewed distribution. The logarithmic transformation also used to produce approximately equal spreads**.** The results of quantile regression give marginal effect for log of monthly earning. We can also estimate the marginal effect on earning, not a log of earning. The equivariance property of quantile regression is used. We estimate the marginal effect of 50th quantile or median regression in levels results are given in **(Table .2).**

Table 3.2 Retransformation Coefficients

Variable	QR_50	AME QR50
Male	0.7389	9625
Urban	0.1249	1627
Married	0.1397	1820
Age	0.0169	220
Edu2	0.0801	1044
Edu3	0.1903	2479
Edu4	0.3674	4786
Edu5	0.5683	7403
Edu6	0.6036	7862
Edu7	0.8280	10786
Edu8	1.2442	16208
Edu9	1.4803	19283
pro1	0.0238	310
pro3	-0.0425	-554
pro4	0.2241	2919
Nat1	0.3778	4922
Nat2	0.3782	4926
Nat3	0.2535	3303
Nat4	0.1937	2523
Nat5	-0.0503	-655
Nat6	-0.0673	-877
Nat8	0.0496	647
Nat9	-0.1289	-1678
Constant	7.5617	98501

We can also find a marginal effect for any quantile in levels. The logarithmic transformation is a no decreasing function that is only applied when the response variable is positively skewed. This transformation technique does not hold in OLS regression. The Average Marginal Effect (AME) estimated as

$$\left[\frac{1}{N}\sum_{i=1}^{N}\exp\left(x_i'\beta_q\right)\right]\beta_{qj}$$

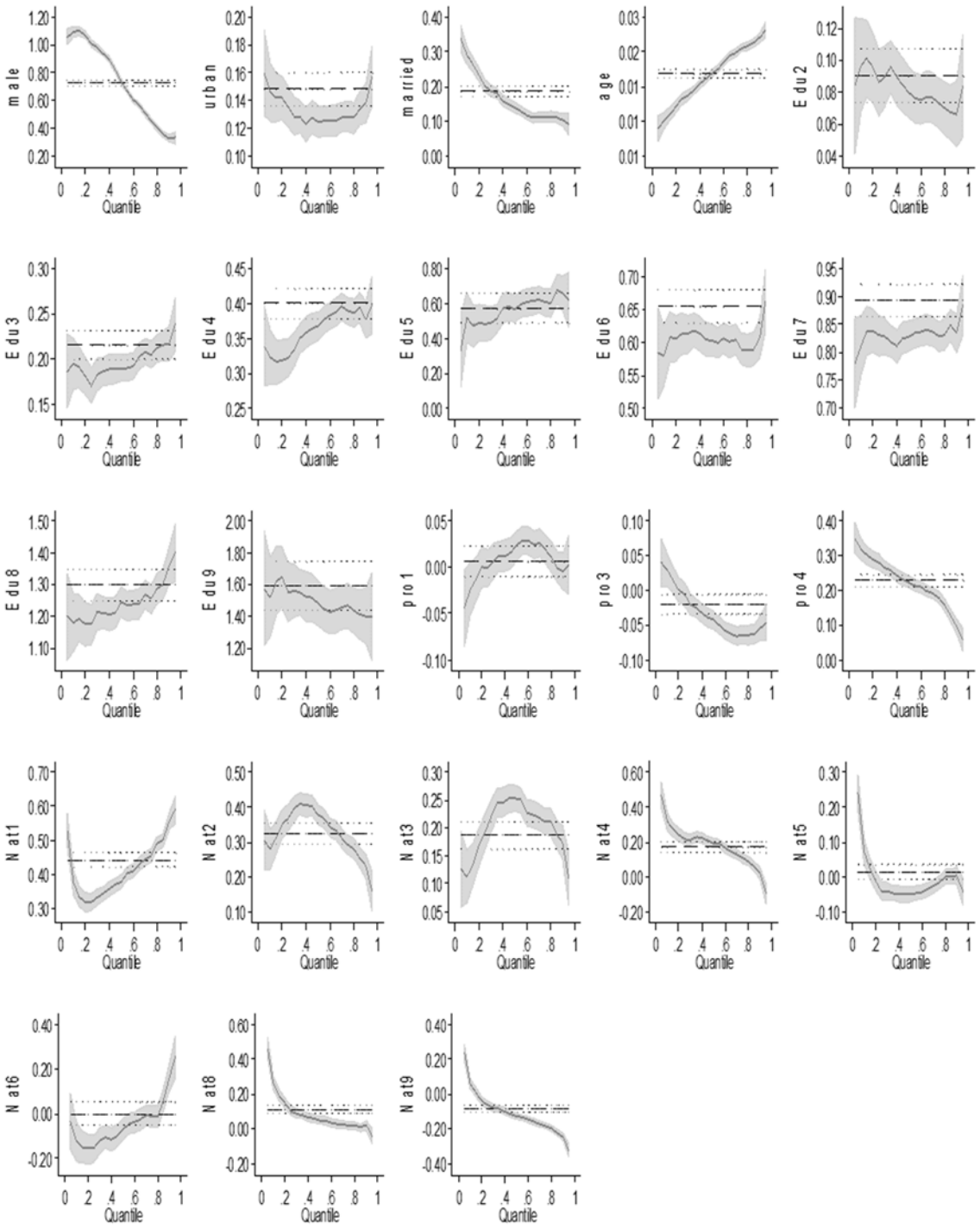

Fig. 2 OLS and QR Estimates

In 2 the marginal effect of male is 0.7389. The marginal effect in levels is 9626. Male monthly earning is 9626 more than female on the conditional median. The marginal effect of married is 0.1397. Married marginal effect in level is 1820. The other variable coefficients also interpret like this. The retransformation coefficients property is accurate only if the conditional quantile function is correctly defined (Buchinsky, 1998).

4. CONCLUSIONS

The results showed that ordinary least square regression gives only information on the conditional expectation, quantile regression provided the results on the whole conditional distribution of the earnings. In the presence of normality, the OLS estimator is more efficient, but the distributions of log of earning, education, age and nature of work are skewed so the precision of the QR estimates is much greater than OLS. Two important assumptions of OLS regression homoscedasticity and residuals follow normal distribution are mostly not meet in real life data sets. QR provides the results that are not affected with heteroscedasticity because QR provides a method for modeling the rates of change in the log of earning at multiple points of the distribution when such rates of change are different, and no parametric distribution assumption is required for the error distribution

REFERENCES:

Aslam, M., A. Saeed, S. Altaf, (2014). Median Regression Analysis of Gender-wise Income Gap in Punjab, Pakistan. Economy, 1(1): 15-19.

Buchinsky, M., (1998). Recent Advances in Quantile Regression Models. Journal of Human Resources, 33 (1): 88-126.

Budria, S., (2010). Schooling and the Distribution of Wages in the European Private and Public Sectors. Applied Economics, 42(8): 1045-1054.

Cameron, C. A., K. P. Trivedi, (2009). Micro Using Stata. Stata press Texas.

Chen, Y. M., L. F. Lin, K. C. Chang, (2009). Relations between health care expenditure and income: an application of local quantile regressions. Applied Economics Letters, 16(2): 177-181

Davino, C., M. Furno, D. D. Vistocco, (2014). Quantile Regression: theory and applications. John Wiley & Sons, Ltd.

Filippaki, K. A., E. Mamatzakis, F. Pasiouras, (2012). A quantile regression approach to bank efficiency measurement. Munich Personal RePEc Archive (MPRA).

Hao, L.X.,D.Q. Naiman, (2007). Quantile Regression. Sage Publications.

Heckman, J., (1979). Sample selection bias as a specification error. 47: 153–161.

Koenker, R., G. Bassett, (1978). Regression Quantiles. 46(1): 33-50.

Koenker, R., (2005). Quantile Regression. Cambridge University Press, Cambridge.

Onyedikachi, O. J., (2015). Robustness of Quantile Regression to Outliers. American Journal of Applied Mathematics and Statistics, 2(3): 86 – 88.

Robbins, L., (1945). An essay on the nature and significance of science. Macmillan and Co. Ltd. Manchester

Sarwar, G., H. S. Sial, (2012). Education and Distribution of Earnings in Pakistan. World Applied Sciences Journal, 17(6): 679-683.

Tansel, A., B. F. Bodur, (2012). Wage Inequality and Returns to Education in Turkey: A Quantile Regression Analysis. Review of Development Economics, 16(1): 107-121.

Taylor, J. W., (1999). A Quantile Regression Approach to Estimating the Distribution of Multipored Returns. Journal of Derivatives, 7(1): 64-78.