PREDICTORS OF EDUCATIONAL ATTAINMENT IN INDONESIA: COMPARING OLS REGRESSION AND QUANTILE REGRESSION APPROACH

by

AGUNG SANTOSO

(Under the Direction of Jonathan Templin)

ABSTRACT

The current study applied quantile regression analysis to estimate the relationship between educational attainment and its predictors, and compared the results to parameter estimates using OLS regression. OLS regression is a regression technique that uses conditional mean as a solution to minimize the error variance. Predictors of educational attainment used were socioeconomic status represented by parent's education, parent's occupation, and family income, hours of study at school, intelligence, and students' employment. Results from the quantile regression analysis showed that for several variables parameter estimates were significant only for certain quantiles. Parameters for hours of study at school and family income were significant only for lower quantiles, while intelligence and managerial/professional were significant for higher quantiles. There were variables that had significant parameters on OLS but not on all quantile from quantile regression. Significance tests of difference between quantiles showed non-significant values. Therefore, an analysis to estimate scale and skewness shift were not reasonable to be conducted.

INDEX WORDS: quantile regression, Educational attainment, Indonesia, Socioeconomic status, Student employment, Learning hours, Parent's education, Parent's occupation, Family income, Intelligence, EBTANAS.

PREDICTORS OF EDUCATIONAL ATTAINMENT IN INDONESIA: COMPARING OLS REGRESSION AND QUANTILE REGRESSION APPROACH

by

AGUNG SANTOSO

BA., Gadjah Mada University, Indonesia 2001

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirement for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2008

© 2008

Agung Santoso

All Rights Reserved

PREDICTORS OF EDUCATIONAL ATTAINMENT IN INDONESIA: COMPARING OLS REGRESSION AND QUANTILE REGRESSION APPROACH

by

AGUNG SANTOSO

Major Professor :	Jonathan Templin
Committee:	Deborah Bandalos
	Seock-Ho Kim

Electronic Version Approved

Maureen Grasso
Dean of the Graduate School
The University of Georgia
December 2008

ACKNOWLEDGEMENTS

Scientific efforts have and will always be communal efforts. So did many individuals have contributed to current study.

I would like to thank my advisor, Dr. Jonathan Templin, for his support, encouragement, trust, and for introducing me to new thoughts. And also to Dr. Deborah Bandalos, Dr. Seock-Ho Kim and all other professors for their examples, teachings, and support.

I could not forget to thank my family for their support, trust, and love that have given me courage to face difficult times. And to dearest Fang, for her patience, love and encouragement that strengthen my motivation to finish my study.

Thanks to all of my friends for their help and kindness as I walked with you along the way, especially Hye-Jeong, Adeya, Young-soon, Flora Liu, Mushtaq, Ai jun, Fei ming, In Heok, and Maurice. Thank you for helping and teaching me many things that you may not even realize.

To Jesus, my Lord and my friend. My life has been a testimony of Your faithfulness, kindness and grace.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Education is one of many problems for the Indonesian people since the falling of Soeharto's regime in 1998. Many people, including practitioners, academicians, and politicians, have shown their concern regarding the educational process and educational attainment in Indonesia. These problems were reflected in reports about the educational attainment of Indonesian students as compared to students from other countries. The Programme for International Student Assessment (PISA) report showed that Indonesia ranks 50th out of 57 countries on students' achievement in mathematics, 49th of 57 in reading, and 50th of 57 in science (2006). The Trends in International Mathematics and Science Study (TIMSS) report showed that Indonesia ranks 34th of 45 countries on eight-grade student mathematics scores and 36th of 45 countries on science (International Association for Evaluation of Educational Achievement, 2003).

Problems related to education quality in Indonesia are attributed to many factors. One factor that is usually suspected as the biggest problem is the small education expenditure of government. The Indonesian government expenditure for education is only 9% of the national budget compared to 20% in Malaysia and 16% in Bangladesh (United Nations Statistics Division, 2008). Other factors thought to be the sources of the problem are the disparity between provinces regarding teachers, facilities and fund distribution, and economic hardships caused by the 1998 economic crisis in Asia.

Unfortunately, there are few research studies investigating the relationship of these factors with education quality represented by educational attainment in Indonesia. Discussions and government policies related to education quality were based mostly on discourse and theory or even "common sense." It is important, therefore, to conduct research related to this topic in order to give more empirically based evidence to factors predicting educational attainment. This master's thesis is dedicated to that purpose and is aimed at finding factors that predict educational attainment for students in Indonesia.

Research studies investigating predicting factors of educational achievement have been conducted in both developing and underdeveloped countries. The results of these research studies were not in agreement (Hanushek, 1995; Sirin, 2005; Velez, Schiefelbein, and Valenzuela, 1993; White, 1982). There were studies that gave evidence of the relationships between educational achievement with some factors, and then others that did not show enough evidence for those relationships. Some of these factors were even well known predictors of educational achievement, such as socioeconomic status, school facilities, and teachers' level education.

There are some explanations for the large variations in the results. First, it was suspected that there was actually no relationship between some of these factors and academic achievement (Hanushek, 1979, 1995). The variation of the degree of relationships between research studies was attributed to a problem of sampling error. Second, there were flaws in the studies that have been conducted related to the methods, including measures of response and explanatory variables (Velez et al., 1993). These variable results were also attributed to moderation by other variables included in the analysis such as minority status, grade level, and school location (Sirin, 2005).

All of these studies used moment-based statistical techniques such as the Pearson product moment correlations or regression analysis based on ordinary least square (OLS) estimation. Moment-based techniques use the conditional mean as their optimal value, a solution for the minimization problem posed by examining the negative log-likelihood function in each. In situations in which the model assumptions were not met or there were outliers, the conditional mean cannot accurately reflect the conditional distribution of the data. If the effect of predictors was different across varying percentiles of conditional distributions, then the effect of the predictors on the upper tail of the distribution may be cancelled out by the effect or predictors on the lower tail of the distribution, which in turn make the effects seem to be zero.

The misspecification errors, as it was pointed out by Sirin (2005), related to exclusion of relevant variables, also tend to make the analysis biased. Sirin showed that the effect of socio-economic status (SES) can be different in different levels of other variables, e.g. grade level. It was also shown by Cade and Noon (2003) that the more variables that are excluded from the model, the more heterogeneous the error variances would be. This means that the results from the moment-based analyses can give incomplete and inaccurate information about the relationship between the response and outcome variables. For the results of the study to be accurate, what is needed is a statistical technique that can provide more information about the relationships between variables at varying locations of the distributions of the data. Another limitation in parameter estimation using OLS that was not investigated in the previous research is that of parameters estimated by OLS procedure being influenced by outliers. The existence of outliers violates one key model assumption: only one regression line is needed to represent the relationships for the whole distribution (Hao and Naiman, 2007). The outliers can alter the correlation coefficient or the regression parameters to be smaller or larger than the parameters

estimated when the outliers were not included in the data (Moore, 2007; Pedhazur, 1997). It was suggested that the outliers can be excluded from the analysis if they were, from thorough investigations, proven to be non-valid observations. But when the outliers are valid, it can give new insights about the nature of the data (Pedhazur, 1997). It means that a statistical technique is needed that will capture the outliers in the analysis and yet is less influenced by their presence.

An alternative technique that has more capability to solve some issues mentioned earlier is called quantile regression. Quantiles are values that give us information about location of a case in a distribution related to proportion of cases having smaller values (Koenker and F.Hallock, 2000). It was developed from a conditional median regression introduced by Boscovich in the 18th century, even before the idea of least squares regression estimators emerged (Koenker, 2005; Koenker and Bassett, 1978). Quantile regression was developed by applying estimation and minimization methods for the conditional median, which is quantile .5, and to other quantiles, rather than the conditional mean, as is done in OLS regression.

Quantile regression has some advantages over OLS regression. It provides information of location shift not only in terms of central tendency location but also other quantile locations (Hao and Naiman, 2007; Koenker, 2005; Koenker and Bassett, 1978). This means that we may have more than one regression line can be modeled, covering the whole conditional distribution including the outliers. For this reason, quantile regression may give us more information about relationship between variables, not only the relationship in term of location shift but also distributional shift including scale shift and skewness shift (Hao and Naiman, 2007). Furthermore, we do not need to assume certain characteristics of the data we used, especially homogeneity of error variances along explanatory variables and normality of error distribution (Hao and Naiman, 2007; Koenker and Bassett, 1978). Another advantage of using quantile

regression is its monotone equivariance property. Hao and Naiman (2007) explain that if we apply a monotone transformation to the outcome variable and then conduct a quantile regression analysis, the predicted values from this procedure will be approximately the same with predicted values from a procedure in which we conduct quantile regression first and then apply monotone equivariance to its prediction.

Applications of Quantile Regression are still limited to economics or environmental studies, but currently there are more and more studies using Quantile Regression as a data analysis tool. To date, few research studies in education have been conducted using Quantile Regression. One of them was conducted by Haile and Nguyen (2008) investigating the effect of family background and race on educational attainment in USA. They also tried to compare results from OLS regression and quantile regression and found that although results using quantile regression resembled OLS regression, the parameters varied across quantiles. For example, the parameter for race (black compared to white students) was -.849 (p <.01) using OLS regression and -1.128, -1.172, and -1.146 using quantile regression for $Q_{.5}$, $Q_{.75}$, and $Q_{.9}$ respectively. It can be seen that the OLS regression provided smaller values than those from quantile regression for median and higher quantiles. It means that the relationship between race and educational attainment were different across different quantiles of the conditional distribution.

The current research project will apply quantile regression on data from the Indonesia Family Life Survey 3 (Strauss et al., 2004) to estimate the relationship between educational attainment and some of its predictors. The term relationship is meant to be used in a broader sense: not only relationships in term of conditional locations but also conditional distributions.

The study will also compare information provided by this method to those provided by OLS regression to get the sense of how both methods provide different information about the data.

Predictors included in current research were those that have consistently shown a significant relationship with educational attainment and those still with contrasting results in previous research. Predictors that have been shown a significant relationship were intelligence (Chen, Lee, and Stevenson, 1996; Diseth, 2002; Laidra, Pullmann, and Allik, 2007; Rohde and Thompson, 2007) and hours of study at school (Gettinger, 1985; Gettinger and White, 1979; Wiley and Harnischfeger, 1974). Predictors considered to still have mixed results were SES (Hanushek, 1979; Sirin, 2005; White, 1982) and student employment (Cooper, Valentine, Nye, and Lindsay, 1999; D'Amico, 1984).

Educational attainment in the current study will be represented by student's score on EBTANAS (National Learning Evaluation) taken in the 6th grade. SES was measured by three indicators which were family income, parents' highest education, and parents' occupation. Student employment will be represented by how many hours of work students perform in a week. Hours of study at school was measured by hours of schooling while intelligence was measured using Raven's Standard Progressive Matrices.

Research Questions

There are several questions to be answered by the current study. They are:

1. What is the relationship between educational attainment and its predictors either using OLS regression or quantile regression methods?
2. Are there any differences in information given by OLS regression and quantile regression?

CHAPTER 2

LITERATURE REVIEW

Factors Predicting Academic Achievement

The second chapter presents a discussion of the factors predicting educational attainment. There are some predictors that have been acknowledged to have a close relationship to educational attainment in many studies that turn out to be controversial in other studies such as teacher experience (Hanushek, 1995), socioeconomic status (Sirin, 2005; White, 1982), etc. Other predictors were consistently reported to have a strong relationship with educational attainment such as intelligence (Velez, Schiefelbein, and Valenzuela, 1993).

Socioeconomic Status (SES)

The relationship between SES and educational attainment was reported to have widely varying levels, from a very high to a non significant correlation (Haile and Nguyen, 2008; Halle, Kurtz-Costes, and Mahoney, 1997; Sirin, 2005; Velez et al., 1993; White, 1982). There was a tendency for the relationship between SES and educational attainment to be weaker when the units of analysis were students, and higher when the unit of analysis were the aggregate of students (e.g. school, district, etc.).

There were also various indicators used to represent socioeconomic status (SES). Indicators that were used consistently to represent SES were parent's education, parent's occupation and parent's or family income. These indicators were also reported to have stronger relationships with educational attainment among other indicators (Haile and Nguyen, 2008; Sirin, 2005; White, 1982).

Intelligence

Intelligence is consistently reported as having a significant relationship with educational attainment. Velez et al. (1993) conducted a meta-analysis on 5 studies that include intelligence in the regression model and found that in all of them intelligence has a significant relationship with educational attainment. The size of the effect, however, was not mentioned. Other studies reported correlations of .3 to .7 between intelligence and educational attainment (Chen, Lee, and Stevenson, 1996; Diseth, 2002; Rohde and Thompson, 2007). Therefore, it is expected that the relationship between intelligence and educational attainment in the current study will be significant.

Hours of study at School

Several studies have shown evidence of a relationship between hours of study at school and educational attainment. Wiley and Harnischfeger (1974) reported that there was a positive relationship between educational attainment and the amount of days in of schooling a year. Gettinger (1985) and Gettinger and White (1979) both found a significant relationship between hours of schooling and educational attainment. The relationship was reported to be stronger than the relationship between intelligence and educational attainment (Gettinger and White, 1979).

Student Employment

School-age children's employment has become a more serious problem in Indonesia since the economic crisis of 1998. The amount of elementary school-aged children employed outside the home has risen from 764,386 in 2006 to 2,854,123 in 2007 (Suara Pembaruan, 2008). Many of these children go to school in the morning then go to work after school.

Few research studies have investigated the relationship between employment and educational attainment, especially in elementary school. Studies that have been conducted have

been in high school or college and have shown a non-linear relationship between hours of work and employment. High school students who worked less than 20 hours a week would have more benefit from their work as the hours increased, while those who worked more than 20 hours would suffer from insufficient time for study (Brown and Steinberg, 1991; D'Amico, 1984; Mortimer and Finch, 1996). One study conducted on 6th grade and 12th grade students found that there were significant negative relationships between time spent working and several measures of educational attainment. Yet, the relationships were considered to be weak. The relationships between time spent working with achievement test scores, teacher assigned grade, and grade after controlling achievement score were -.12, -.17, and -.14 respectively (Cooper, Valentine, Nye, and Lindsay, 1999).

Quantile Regression

Quantile and Quantile Function

A quantile is a value that gives us information about the location of a case or a score in a group "… corresponds to a specified proportion of the sample or population" (Gilchrist, 2000, p. 1). A person's score on a test is said to be in the *p*-th quantile in his/her group if his/her score in the test is bigger than a proportion of *p* of his/her group and smaller than a proportion *(1-p)* of his/her group (Koenker and F. Hallock, 2000). The median is at the .5 quantile, because there is half of the group that have values bigger than the median, and half of the group that have values smaller than the median. The lower quartile is at .25 quantile and the higher quartile is at .75 quantile.

A function that gives us the value of a certain quantile is called a quantile function (QF) denoted as $Q^{(p)}$. For example, if a median of a group has a value of 50, it can also be said that

$Q^{(.5)}$ is 50. The quantile function is an inverse of cumulative distribution function (CDF) denoted as $F(x)$. The CDF can show us a proportion of a group that has a value equal to or smaller than a certain value of *x*. It can be formulated as:

$$F(x) = P(\mathrm{X} \leq \mathrm{x}). \quad (1)$$

The relationship between quantile function and CDF can be depicted in Figure 1 and denoted as

$$Q^{(p)} = F^{-1}(p). \quad (2)$$

For example, if $F(90) = .25$ then $Q^{(.25)} = 90$.

Quantiles as Solutions of Minimization Problems

It has been shown that the mean is a solution for a minimization problem. The arithmetic mean minimizes the mean of squared deviations (MSD) in a single distribution. To illustrate this point, data from 10 cases have been generated from a normal distribution with a mean of 10 and SD of 5. The mean of the sample was 10.052 and the standard deviation was 4.898. The MSD was counted for each value as it was used as the solution for minimization. The results were presented graphically in Figure 2.

The *x*-axis was the value of $\bar{y}$, while the y-axis represented the value of MSD produced if one used a certain value of *y* as the solution for minimization. From the curve, we could see that the mean of *y* (*y*=10.052) produced the smallest mean squared deviation (*MSD*=23.99) compared to other points in the distribution. The median (*y=10.782*) produced slightly higher mean square deviation (*MSD*=24.58).

The median is also a solution for a different minimization problem. It minimizes the mean of absolute deviations (MAD). For illustration, the means of different absolute deviations were counted from the same data mentioned above. The results were presented in Figure 3 with the *x*-axis as the median value, and *y*-axis as the value of absolute deviations using every point in

the data. It can be seen that the median (y=10.782) has the smallest value of mean absolute deviation (MAD=3.61) when compared with all other values. The mean of y (y=10.052) had a slightly higher value of MAD (MAD=3.68) compared to median.

The idea of the median as the solution for a minimization problem has been generalized to other quantiles. For other quantiles to be solutions of minimization problems, different weights should be applied for values less and bigger than the quantiles. For example, for the quantile of .1 to be the solution, all values less than quantile .1 are weighted by 1-.1=.9, and values bigger than the quantile .1 are weighted by .1. The general notation of the solution is expressed as follows:

$$(Weighted)\ MAD = \frac{1-p}{n}\sum_{y_i<q}|y_i - q| + \frac{p}{n}\sum_{y_i>q}|y_i - q|. \quad (3)$$

The weighted MAD for quantile .1 was counted from previous data to illustrate this point. The results were graphically depicted in Figure 4. From this figure, we can see that quantile .1 (y=6.273) had the smallest value of MAD (MAD = 1.19) compared to other points. Hao and Naiman (2007) provide a proof of quantile as the solution of minimization problem using derivative of MAD.

Quantile Regression

The idea of quantiles as solutions to certain minimization problems was then applied on conditional distributions to investigate the relationship between an outcome variable and a set of response variables (Cade and Noon, 2003). In other words, this idea was extended to conditional quantile functions expressing "quantiles of the conditional distribution of the response variables as functions of observed covariates" (Koenker and F. Hallock, 2000, p. 2).

Quantile regression (QR) is similar to ordinary least squares regression (OLS-R) in a sense that both of them investigate relationships between variables and the rate of the outcome variable following response variables represented by regression coefficients denoted as β s. The main difference is that OLS-R chooses parameter values that have the least squared deviation from the regression line as the parameter estimates, expressed by

$$\min \sum_{i=1}^{n} (y_i - \hat{y})^2 \tag{4}$$

while QR chooses parameter values that have the least absolute deviation/error

$$\min \left[(1-p) \cdot \sum_{\hat{y}_i < \hat{y}_i^{(p)}} \left| y - \hat{y}^{(p)} \right| + (p) \cdot \sum_{\hat{y}_i \geq \hat{y}_i^{(p)}} \left| y - \hat{y}^{(p)} \right| \right] \tag{5}$$

(Hao and Naiman, 2007; Koenker, 2005). In Equation 4, $\hat{y}$ is the predicted value of y using regression line, while $\hat{y}^{(p)}$ in Equation 5 is the predicted value of y on quantile p.

OLS-R uses conditional mean $E(y_i \mid x_i)$ as the solution for minimization problem while QR uses conditional quantiles $Q^{(p)}(y_i \mid x_i)$. As a result, OLS-R will produce only one regression line, which is the regression line on the conditional mean,

$$E(y_i \mid x_i) = \beta_0 + \sum \beta_{ij} x_i \,, \tag{6}$$

while QR can produce more than one regression line, one for any quantile of interest:

$$Q^{(p)}(y_i \mid x_i) = \beta_0^{(p)} + \sum \beta_{ij}^{(p)} x_i \,. \tag{7}$$

Statistical Inference in Quantile Regression

There are several methods used to make inferences about parameter estimates in QR, including asymptotic distributions, Wald tests, ranks, and bootstrap methods. The bootstrap has been found to provide desirable results (Hahn, 1995; Koenker, 2005) especially when dealing with skewed distributions (Hao and Naiman, 2007). The bootstrap also facilitates the opportunity

to test additional hypotheses, like hypotheses related to difference between parameter estimates, scale shift or skewness shift (Hao and Naiman, 2007). Tests for parameter differences between multiple quantiles are conducted using Wald tests (Koenker, 2008).

The bootstrap is a method of estimating the sampling distribution of parameter estimates calculated from a sample drawn with replacement from and with size equal to the original data set (Hao and Naiman, 2007). This method provides reliable estimates of standard deviations for each parameter, especially when the distribution of the population cannot be identified as identically independently distributed (IID). The pair form (x_i, y_i) of this method provides simple and effective ways of drawing samples with replacement of pair (x_i, y_i) from the joint distribution of the original samples, with size *n* as large as *n* of the original sample. Each pair was drawn with the same probability of 1/*n* (Koenker, 2005).

Using the bootstrap may give two alternatives to make inferences about parameters. The first alternative is counting standard deviation of parameters and using it to obtain a t-value and its p-value of related parameters. Confidence intervals (CI) can be approximated using this method. The second alternative is by constructing 95% CI (or other CIs) using 97.5th percentile and 2.5th percentile of the samples of bootstrap estimates. If the CI captured the parameter, we can make the inference that the parameter is significant on α=.05 (Hao and Naiman, 2007).

Comparing Quantile Regression and OLS regression

Limitation of OLS regression

OLS-R is claimed to produce parameters with desirable characteristics - best, linear, unbiased estimators (BLUE). This means that parameters estimated using OLS-R have the smallest variance, model a linear relationship between response and outcome variables, and

resemble value of parameters in population. However, these characteristics only hold if there are not serious violations of the model assumptions or presence of influential outliers (Berry, 1993; Hao and Naiman, 2007; Pedhazur, 1997). Heteroscedasticity will make parameter estimates no longer BLUE (Berry, 1993), while the presence of influential outliers will cause the regression line to be leveraged in the outliers direction (Moore, 2007; Pedhazur, 1997). Furthermore, the presence of outliers also violates the one-model assumption that one regression line is sufficient to model the relationship between variables for the whole distribution (Hao and Naiman, 2007).

OLS-R still has inherent disadvantages even when procedures to overcome effects of violation assumptions and outliers are applied. Models suggested by OLS-R cannot be immediately extended to other locations in the distribution that may be more interesting to be investigated in other studies (Hao and Naiman, 2007). For example, the study of educational achievement focuses on over-achieving or under-achieving students.

OLS-R also assumes that response variables only affect the location shift of the conditional distribution, while response variables may affect other parameters of the distribution in some instances. This means that OLS-R provides limited information about the relationship between variables (Buchinsky, 1994; Eide and Showalter, 1999; Hao and Naiman, 2007; Koenker and F.Hallock, 2000). OLS-R may give inaccurate information about the nature of the relationship between variables. When heteroscedasticity occurs and the slope of the regression line on the conditional mean is zero, OLS-R or related approaches to overcome heteroscedasticity will suggest no relationship between variables, although there are relationships between variables on non-central locations or on other distributional parameters (e.g. scale, skewness; Cade and Noon, 2003).

Advantages of Quantile Regression over OLS regression

Quantile Regression can be a robust approach to regression with additional advantageous features. It is not sensitive to outliers on outcome variables because its estimation is based on the quantiles of scores instead of the score itself. Each score is bound to its location relative to its group, meaning that so as long as the extreme values do not change their location relative to its group, estimation of parameter in quantile regression will not be affected (Hao and Naiman, 2007; Koenker, 2005). Quantile regression estimates are also insensitive to heteroscedasticity for the same reason. Each score is bound to its quantile and the range of quantiles is stable across values of response variables, thus removing the need for procedures to overcome heteroscedasticity. The stochastic part of the QR model, the error term, is not based on a certain distribution function such as the normal distribution, so it does not need to assume normality of error distribution. It will make QR applicable to data that have very skewed error distributions (Gilchrist, 2000; Hao and Naiman, 2007).

Quantile Regression can provide a regression line with non-central locations because of its ability to examine the relationship between variables on any quantiles in a conditional distribution. This will enable researchers to conduct inequality studies involving non-central area of the conditional distribution (Cade and Noon, 2003; Hao and Naiman, 2007; Koenker, 2005).

Furthermore, this feature also enables QR to investigate the relationship between variables on other parameters of the conditional distribution such as scales and skewness. QR will provide information about how response variables affect the scale or skewness of the conditional distribution of the outcome variable (Hao and Naiman, 2007).

QR also has a feature of monotonic equivariance (Hao and Naiman, 2007) or "equivariance to monotone transformation" (Koenker, 2005, p. 39). This means that any

monotonic transformation of the data will not change the way the result will be interpreted. OLS-R only has a feature of equivariance of linear transformations:

$$E(c + ay \mid x) = c + aE(y \mid x). \tag{8}$$

QR has both equivariance of linear transformations and monotone transformations (Hao and Naiman, 2007):

$$Q^{(p)}(c + ay \mid x) = c + aQ^{(p)}(y \mid x) \tag{9}$$

$$Q^{(p)}(h(y) \mid x) = h\left(Q^{(p)}(y \mid x)\right). \tag{10}$$

For example, if we apply a log(y) transformation on OLS regression, then

$$E(\log(y) \mid x) \neq \log(E(y \mid x)). \tag{11}$$

But if we apply log (y) transformation to Quantile Regression,

$$Q^{(p)}(\log(y) \mid x) = \log\left(Q^{(p)}(y \mid x)\right). \tag{12}$$

Information Can Be Derived from Quantile Regression

There are at least three kinds of information that can be derived from QR about the relationship between variables: central (median) and non-central location shifts, parameter differences between quantiles, and shifts of other conditional distribution parameters. Relationships on central and non-central location shift and parameter differences between quantiles can be derived immediately from parameter estimates on median or quantile .5 and other quantiles. Information regarding scale and skewness shift should be developed based on parameters acquired.

One measure of distribution scale based on quantile is interquartile range (IQR). IQR is the difference between upper quartile (Q3) and lower quartile (Q1).

$$\begin{aligned} IQR &= Q_3 - Q_1 \\ &= Q^{(.75)} - Q^{(.25)}. \end{aligned} \tag{16}$$

Hao and Naiman (2007) proposed a method of estimating scale shift based on the IQR difference between two adjacent points, namely reference (R) and comparison (C=R+1). The difference between IQR for two adjacent points will provide information about the amount of increase of scale shift as a response variable increases one point. It was shown that for one predictor

$$SCS = \beta^{(.75)} - \beta^{(.25)}. \tag{13}$$

SCS will be zero if there is no scale shift across the values of a response variable. It will be negative if the scale of the conditional distribution of the outcome variable becomes smaller as the value of response variables increase and positive if the scale becomes larger as the value of response variables increase. This procedure will be applied for estimating the amount of scale shifts as one increase of a certain predictor assuming other predictors in the model are constant.

Quantile-based measures of skewness (QSK) can be expressed as a ratio of the upper spread to the lower spread of the conditional distribution and subtracted by 1 to center it to zero (Hao and Naiman, 2007):

$$\begin{aligned} QSK &= \frac{Q3 - m}{m - Q1} - 1 \\ &= \frac{\left(Q^{(.75)} - Q^{(.5)}\right)}{\left(Q^{(.5)} - Q^{(.25)}\right)} - 1 . \end{aligned} \tag{18}$$

The upper spread is defined by the difference between upper quartile and median, while the lower spread is defined by difference between median and lower quartile. The QSK will be zero if the distribution is symmetric. It will be negative if the distribution is skewed to the left and positive if the distribution is skewed to the right.

Hao and Naiman (2007) proposed a measure of skewness shift (SKS) obtained by taking the ratio between QSK on comparison points to QSK on reference points. SKS would be

in percent of skewness change relative to the reference point. In this case, the QSK will not be subtracted by 1 to avoid division by zero. The ratio will be subtracted by 1 to center it to zero:

$$SKS = \frac{\left(\beta^{(.75)} + \alpha^{(.75)}\right) - \left(\beta^{(.5)} + \alpha^{(.5)}\right) \Big/ \alpha^{(.75)} - \alpha^{(.5)}}{\left(\beta^{(.5)} + \alpha^{(.5)}\right) - \left(\beta^{(.25)} + \alpha^{(.25)}\right) \Big/ \alpha^{(.5)} - \alpha^{(.25)}} - 1 \quad . \tag{19}$$

SKS will show negative values if the skewness shifts to be more left skewed (if the conditional distribution on reference point is already left skewed) or less right skewed (if the conditional distribution on reference point is right skewed) as the value of the response variable increases. It will show positive values if the skewness shifts to be more right skewed or less left skewed as the value of the response variable increases

Unfortunately this procedure was only applied to situation in which there was only one categorical response variable with only two categories. So researchers have tried to apply some strategies for the procedure to be relevant with current studies:

1. Apply the procedure only for one predictor at a time. There would be one measure of skewness shift (SKS) for each predictor. This strategy holds only if we assume that there was no interaction between response variables.
2. For categorical variables or dummy variables, Equation 19 can be applied directly as it was proposed. SKS from this procedure would be interpreted as percent increase/decrease of skewness relative to the reference group, group labeled as 0, holding other variable constant.
3. For continuous variables, the modification of Equation 19 (Equation 20) will be applied assuming that the reference point is 0:

$$SKS = \frac{\left(\beta^{(.75)} \cdot C + \alpha^{(.75)}\right) - \left(\beta^{(.5)} \cdot C + \alpha^{(.5)}\right) \Big/ \alpha^{(.75)} - \alpha^{(.5)}}{\left(\beta^{(.5)} \cdot C + \alpha^{(.5)}\right) - \left(\beta^{(.25)} \cdot C + \alpha^{(.25)}\right) \Big/ \alpha^{(.5)} - \alpha^{(.25)}} - 1 \quad . \tag{20}$$

The comparison points (C) would be some points of interest of a predictor's distribution such as median, first and third quartile, and so on. SKS produced after applying these strategies will be interpreted as percent of skewness shift of the conditional distribution of the outcome variable, on a comparison point relative to the reference point of a response variable, holding other variables constant.

SKS can give us important information related to whether the conditional distribution has changed its skewness as the value of the response variable increase. For example, if parent's education affected the skewness of the conditional distribution of educational attainment, in an extreme situation, the skewness of the conditional distribution can change from right skewed, for students whose parents' only finished elementary education, to left skewed, for students whose parents' had a higher education. Even though there is no difference between the two groups, the parents' education still has an effect on educational attainment: the amount of students who have higher achievement on the first group are smaller than the amount of students who have higher achievement on the second group. The illustration can be depicted graphically on Figure 5.

CHAPTER 3

PROCEDURE

Variables

Outcome Variables

The outcome variable in the current study is educational attainment, as measured using The Evaluasi Belajar Tahap Akhir /EBTANAS (National Final Exam). EBTANAS was given when a student was in 6th grade. The EBTANAS consisted of five subtests: moral education, science, social science, math, and literacy. Each subtest was scored from 0 to 10. Student's EBTANAS score is the sum of subtest scores, thereby ranging from 0 to 50.

Response Variables

The first response variable used was social economic status (SES). SES will be represented by family income, highest parents' education, and parent's occupation. Family income will be represented by salary obtained by all family members, including parents and children who have worked but are still living with their parents. Size of family also had an important role because a larger family will decrease the amount of resources per member (Haile and Nguyen, 2008). Researchers have decided to use the average salary in a million rupiahs (SALARY) to represent family income. The average was taken by dividing total salary earned by parents and children by the number of children in the family and their parents.

Parents' education will be represented by the highest education earned by parents. Parents' education will be categorized into two categories: 0 if the highest education attended was primary school and 1 for high school and above (1) (PARENT.EDU.RECODE) based on

classification of primary education and higher education by Department of Education (Departemen Pendidikan Nasional, 2008).

Parents' occupation will be the highest level of occupation of both parents (OCCUPATION). The levels of parents' occupation in current research followed categories made by Haile and Nguyen (2008) and one category of farmer. Each level will be a dummy variable. There would be three dummy variables included in the model, management / professional (MANA.PROF), skilled non-manual (SKILL.NONMAN), and skilled manual (SKILL.MAN). The last category was not included to hinder multicolinearity.

Intelligence (IQ) was measured using Raven's Coloured Progressive Matrices. Student's score on intelligence will be the number of correct items of students' work. Hours of study at school (HOURS.STD) were the amount of hours the student used to learn in school.

The last response variable was hours of work to earn for money each week (CHLD. WRK.HRS), measured by the amount of hours of the last week they work.

Data

Data for the current research project were taken from The Third Wave of the Indonesian Family Life Survey (IFLS3) (Strauss et al., 2004), conducted as a collaboration effort between Research And Development Corporation (RAND) and Center for Population and Policy Studies (CPPS) of the University of Gadjah Mada. IFLS study itself has been conducted four times: IFLS1, IFLS2, IFLS2+, and IFLS3. The third wave was taken on 2000.

The purpose of the IFLS was "to provide data for studying behaviors and outcomes" (Strauss et al., 2004, p. 1). It contains many variables on individual, household and community levels, from socio-economic variables, education, health issues including contraceptive use, and

migration. It was a longitudinal study that began with IFLS1 and most of the respondents in the next study were re-contacted from the previous study. The survey was divided into two larger studies; a survey to cover individual and household issues and another to cover community issues.

There were 38,823 individuals interviewed in the household and individual survey, 37,173 of them were interviewed directly while 1,260 were interviewed using proxy interviews, originating from 10,435 households. The data were taken in 13 out of 26 provinces in Indonesia. The distribution of data can be seen in Figure 6. Data to be used in the current research were data from children who have taken EBTANAS, consisting of 754 cases.

The survey for household was conducted using twelve (12) books of items. Each book covered certain information of specific aspects of individuals or households. Most of the books were assigned for adult and household issues, such as marriage, household income, occupation, highest level of education, etc. There was one book to cover issues related to children such as education and health and another book to get information of respondents' intellectual capacity.

Computer Program

The current study will use R (R Development CoreTeam, 2008) to analyze the data, using a package for quantile regression developed by Koenker (2008). Researchers also used SPSS 16 (SPSS, 2007) for data preparation and exporting data from SAS data file to a .csv file so it can be read by R.

Descriptive Statistics

Descriptive statistics for each variable can be seen in Table 1. Educational attainment has a mean of 32.74, median of 32.46 and standard deviation of 5.570. It can be seen that the students' educational attainment tend to be high because the mean and median are located near the maximum score possible.

There are only 36 students who worked after school. It can be seen that at least 95% of the students in the data did not work. The mean of working hours, including students who did not work, is 1.344 hours a week. The mean of working hours only for students who worked is 28.139. There are 19 students who worked more than 20 hours a week, with mean of 45.21.

The mean of hours of study at school is 34.02 with standard deviation of 9.784. It is interesting that the distribution is greatly varied. Five percents of the students studied at school for only 6 hours a week or less, while another five percents studied at school for more than 40 hours a week.

There are 71 students whose parents worked in managerial/professional job, while 150 and 295 of the students have parents who worked in skilled manual and skilled non manual job respectively. The rest of the students have parents who worked as farmers.

The amount of students whose parents have attended only elementary school is 512. The other 242 students have parents who attended a higher than elementary school education.

The income per capita has a right skewed distribution, the mean and the median have a large difference about 700 thousands rupiahs. The values of the very high quantiles are deviated greatly from the mean and the median: about 5 and 9 million rupiahs for quantile .95 and .99 respectively.

Students' intelligence has the mean of 10.36, median of 11 and standard deviation of 2.356. The distribution is skewed to the left as it can be seen that 50% of the students have very high scores while 50% others are distributed from about 3 to 11.

Correlation coefficients between variables can be seen in Table 2. Correlation coefficients between variables mostly have lower values, although they are significant. The highest correlation is between managerial/professional and parent's education (r=.440, p<.05). The correlation between skilled manual and skilled non manual is also high (r=-.400, p<.05), but does not have meaningful interpretation because the two variables are dummy variables for parents' occupation. The other significant relationships, ignoring the direction of the relationship, have coefficients between .076 to .334. Based on the correlation matrix, it is suspected that there would be no multicolinearity issues in the analysis.

CHAPTER 4

RESULTS AND DISCUSSION

Assumption Check and Diagnostic for OLS regression

It has been mentioned earlier that OLS regression will give estimates which have BLUE characteristics only if the data meet its assumptions. For that reason, it is necessary to check whether there are any violations of the assumptions or influential outlier.

Homogeneity Assumption

Homogeneity assumption was checked by plotting residuals and predicted values. If the plots form a megaphone-like pattern, we concluded that the heterogeneity assumption was violated. The plots are shown in Figure 7.

Based on the plots, it can be concluded that there is no violation of homogeneity assumptions, although there are some potential outliers observed.

Linear Relationship between Variables

The linear relationship assumption was investigated using the same plots to check the homogeneity assumption. It can be seen that there are no patterns suggested that the relationship between variables is not linear.

Errors term are distributed normally

Normal distributions of error term can be confirmed from the Q-Q plot of the residuals, as it can be seen in Figure 8. From the plot, it can be seen that there is no violation of the normality of error distribution. It can be confirmed also that there are some potential outliers to be checked.

Influential Outliers

Previous plots showed that there are some potential outliers in the data that should be checked. To serve this purpose, researchers plotted standardized residuals against Leverage and Cook's *D*. Plots of standardized residuals against Leverage can be seen in Figure 9, while plots of standardized residuals against Cook's *D* can be seen in Figure 10.

Based on the plots, it can be seen that there are some potential outliers. The researcher decided to exclude potential outliers from the data and run the analysis. The results from these analyses were compared to see whether there were obvious differences between them. If there were negligible differences between the two analyses, the results from the first analysis would be used. Otherwise, both results would be presented when they were compared with the quantile regression method.

The parameter estimates of both analyses are presented in one table in Table 1, while *R*-squares and *F*-values are presented in Table 2.

From Table 1, it is shown that the differences between two analyses are negligible. The results from the first analysis were in accord with the second analysis especially for significant parameters. Significant parameters in the first analysis were also significant in second analysis

The *R*-square and *F* values between results from the two analyses had small differences. Given the results, researcher concluded that the potential outliers did not have a strong influence on the analysis and use all of the data in the analysis.

Comparison of OLS regression and Median Regression

Parameter estimates from OLS regression, OLS regression excluding outliers and median regression are presented in Table 3. More comprehensive tables including information about standard errors and *t*-values were presented in Appendix A.

The results from median regression resembled those from OLS regression, whether the outliers were included or excluded. The differences between parameter values from three analyses were considered small. Almost all of the significance tests for each parameter were similar between OLS regression and median regression.

Small parameter differences between median regression and OLS regression were due because the assumptions of OLS regression were not violated and there were no influential outliers. The conditional distribution was also normal and symmetrical which makes the mean and median have relatively similar values.

The only noticeable difference was on the significance level of Skilled Manual, representing a comparison of student's attainment between those for whom one of their parents work in a skilled manual occupation and farmer/forestry. Parameter estimates for this variable from OLS regression analyses were significant (b=1.266 and b= 1.362, p<.05), although results from median regression analyses (QR using quantile .5) were not significant (b=1.037, p>.05). It is worth noting that the parameters of OLS regression had larger values than the median regression parameters. Given the results, it could be said that the conditional distribution of educational attainment in the population could be slightly skewed to the right. More discussion on this issue is provided next section.

Comparison to Other Quantiles

Regression parameters for other quantiles were also estimated. The parameters of several important quantiles can be seen in Table 4. Parameters for the OLS regression and quantile regression on the whole distribution and their 95% confidential intervals are depicted graphically in Figure 11. Parameter estimates for quantile regression are represented by black dots, while its confidential intervals are represented by gray area. Parameter estimates for OLS regression are represented by red line, while its confidential intervals are represented by a red dashed-line. Graphical representation of the parameters provides a more comprehensive image of relationship between variables on all quantiles.

The intercept cannot be interpreted meaningfully because some of the variables did not reasonably have zero values, like income per capita. Children's hours of work did not have significant parameters on all quantiles ($-.008 \leq b \leq .050$, $p > .05$) and on conditional mean ($b = .001$, $p > .05$). It could be said that variation of educational attainment was not following variation of how many hours children worked, or there was no relationship between educational attainment and hours of children worked.

Hours of study at school gave us interesting results. The parameters for Quantile Regression on quantile .05 and .1 were significant ($b^{.05} = .105$, $p < .05$; $b^{.1} = .105$, $p < .05$) but those from OLS regression ($b = .0327$, $p > .05$) and Quantile Regression on all other quantiles ($-.013 \leq b \leq 058$, $p > .05$) were not significant. This means that hours of study at school was related to students' attainment only in lower quantiles. In other words, hours of study can predict educational attainment only for students who had very poor performances. For students who performed better than 25% of population, hours of study at school was not a good predictor.

Students whose parents' occupation was in the managerial / professional class had significantly higher attainment compared to students whose parents' were in farmer class only on middle to higher quantiles ($2.070 \leq b \leq 3.667$, $p<.05$). The result from OLS regression resembled estimates from higher quantiles ($b=3.150$, $p<.05$). But for students who have a very low attainment, on quantile .05 and .1, the difference were not significant ($b^{.05}=.691$, $p<.05$; $b^{.1}=1.308$, $p<.05$).

There was no difference of educational attainment between students whose parents were in a skilled manual class and farmer, on all important quantiles ($-.066 \leq b \leq 1.567$, $p>.05$). The result for OLS regression was there was a significant differences between two groups ($b=1.266$, $p<.05$). Figure 11 sheds some light on these differences. It could be seen that the parameters for Quantile Regression on most of the quantiles were not significant except for those from quantiles bigger than .5 and lower than .75. Parameter estimates for these quantiles were significantly different from zero. As it has been discussed in the previous section, it could be an indication that the conditional distribution of educational attainment in the population was slightly skewed to the right. When a distribution is skewed to the right, the mean value will be pulled to the higher quantiles. In this case, the parameters on the conditional mean performed differently from the conditional median but resembled with parameters of conditional quantiles that are slightly higher than the median. This however was a tentative conclusion for more evidence was needed to justify it.

A significant difference of educational attainment between students whose parents were in the skilled non manual class and those who were farmers was only identified at quantile .25 based on the table ($b^{.25}=1.096$, $p<.05$). Figure 11 gives information that significant difference between two classes also identified approximately between quantile .55 to .70. OLS regression

and quantile regression on other quantiles provided non-significant parameters. This means that only on limited quantiles was the difference of educational attainment between both groups was significant.

Income per capita had significant parameters on almost all quantiles ($.587 \leq b \leq .349$, $p < .05$) and on the OLS regression parameters ($b = .390$, $p < .05$). The parameters were non-significant only on higher quantiles, that is quantile .75 and higher ($.170 \leq b \leq .399$, $p > .05$). This means that income per capita was a significant predictor of educational attainment only for students who performed at the average and lower level. For students who performed well or very well, Q.75 and above, income seemed to have no significant effect on achievement.

The highest level of parent's education did not have a significant affect in OLS regression ($b = .618$, $p > .05$) or in any quantiles ($-.198 \leq b \leq 1.548$, $p > .05$). It can be said that parents' education did not have any relationship with education achievement. It should be noted that parent's education in the current research was a dummy variable consisting of two categories: primary school and high school or higher. So this result can be interpreted as there was no difference on educational attainment between students whose parents attended only primary school and those parents attended high school or higher education.

As found in previous research studies (Rohde and Thompson, 2007; Velez, Schiefelbein, and Valenzuela, 1993), intelligence had a significant relationship with educational attainment in OLS regression ($b = .277$, $p < .05$) and almost all quantiles ($.235 \leq b \leq .350$, $p < .05$) except for very low quantiles ($b^{.05} = .081$, $b^{.10} = .052$, $p > .05$). From Figure 11, it could be seen that the relationship between educational attainment and intelligence was not significant approximately on quantiles .2 and lower. It means that for students with very low achievement levels, intelligence was not a good predictor of students' attainment.

Differences Between Parameters on Certain Quantiles

The first step to check whether there was a scale and skewness shift on the outcome variable along response variables was to check whether there was a significant difference between certain quantiles. Parameters that were to be tested were those in median, quantile .25, and quantile .75. These quantiles were important to count interquartile range, as a measure of distribution scale, and quantile-based skewness.

Tests for significant differences between these quantiles were conducted using the Wald test. The results are shown in Table 5. The first row of the table was the test of differences between parameters on Q.75 and Q.25, while on the second row was the test of differences between parameters on Q.75, median and Q.25.

There was no significant difference between parameters on quantiles .75 and .25 (F (8, 1500) =1.9227, p>.05) and between parameters on quantiles .75, median and .25 (F (16, 2246) =1.5590, p>.05). This means that there was no evidence that suggested that the parameters were not parallel. Furthermore, there was no evidence suggesting scale and skewness shift along response variables. These findings also confirmed the assumption check previously that there was no violation of the homoscedasticity assumption. Given the results, it was considered unreasonable to proceed to test for scale and skewness shift.

CHAPTER 5

CONCLUSIONS AND SUGGESTIONS

Conclusions

Relationship between Educational Attainment and its Predictor

There were several predictors that had significant relationships with educational attainment on the conditional mean location and other location as well. There was no variable which had significant parameter values on conditional mean and all quantiles. Variables which had significant parameter values on the conditional mean and almost all quantiles were income per capita, managerial / professional and intelligence. Parameters for managerial / professional and intelligence were significant on higher quantiles and not significant on lower quantiles. Income per capita had significant parameters on middle and lower quantiles.

There was one variable where parameters were not significant on any conditional distributions, that was hours of child work per week. Hours of study had significant parameters only on very low quantiles, suggesting that hours of study at school might only benefit students who had low performance. It could be inferred that increasing hours for the remedial class would only help students who performed very poorly. Compared to intelligence, for very low performers, amount of hours of study at school was more important to determine student's attainment.

Skilled manual and non-skilled manual labor had significant parameter values on the conditional mean, but no significant effects on other important quantiles. A closer investigation of its parameter plots revealed interesting information, that these variables only had significant parameters around very limited quantiles.

Significant tests of parameter difference on certain quantiles were failed to provide evidence of significant difference between quantiles. This means that there would be no scale and skewness shift

Comparison between OLS regression and Quantile Regression

These results suggested that the relationships between education and its predictors might differ at different location across conditional distributions. This information could not be obtained if researchers only used the OLS regression method to analyze the data. Suggesting relationships as they were presented only by OLS regression could neglect important issues, for example using OLS regression it would be suggested that there was no relationship between educational attainment and hours of study at school, while actually there was significant relationship on only the very low quantiles.

Information regarding scale and skewness shift can only be gained using quantile regression. Unfortunately, in the current study, there was no evidence to suggest scale and skewness shift. But it was still information about the nature of relationship between educational attainment and its predictors.

Suggestion

It seems obvious that more studies need to be conducted to investigate relationship between educational attainment and its predictors by adding more relevant variables which had not yet been studied in current research such as type of school attended, hours of study at home, or other variables related to group characteristics. Studies should also be open to the possibility of interaction between predictors in the model. Using other categories for some predictors should also be considered in future research. For example using more categories for parents' education such as primary education, high school and higher education might provide more information

about the nature of the relationship. Parent occupation can also be represented by a continuous variable such as Socio-Economic Index of occupation (SEI) (Reiss, 1961).

Methods for obtaining more information from Quantile Regression should be developed. For example, methods for estimating scale and skewness shift for more than one variable and for continuous variables should be developed to obtain more information about the nature of relationships between variables. Research to investigate effect size of quantile regression and its comparison with *R*-squared should also be conducted to gain more insights about the similarities and differences between the two analyses.

Results from the current study should be interpreted carefully for it was only a correlational study. Causal relationships should be investigated more thoroughly using more appropriate design. Suggestions that implied influence of predictors to outcome variable should be treated as tentative suggestions needing more evidence to support them.

REFERENCE

Berry, W. D. (1993). *Understanding regression assumptions*. Thousand Oaks, CA: Sage Publications Inc.

Brown, B. B., and Steinberg, L. (1991). *Noninstructional influences on adolescent engagement and achievement. final report: project 2*. Madison, WI: National Center on Effetive Secondary Schools.

Buchinsky, M. (1994). Changes in the us wage structure 1963-1987: Application of quantile regression. *Econometrica, 62*, 405-405.

Cade, B. S., and Noon, B. R. (2003). A gentle introduction to quantile regression for ecologists. *Frontiers in Ecology and the Environment, 1*, 412-420.

Chen, C., Lee, S. Y., and Stevenson, H. W. (1996). Long-term prediction of academic achievement of American, Chinese, and Japanese adolescents. *Journal of Educational Psychology, 88*, 750-759.

Cooper, H., Valentine, J. C., Nye, B., and Lindsay, J. J. (1999). Relationships between five after-school activities and academic achievement. *Journal of Educational Psychology, 91*, 369-378.

D'Amico, R. (1984). Does employment during high school impair academic progress. *Sociology of Education, 57*(3), 152-164.

Departemen Pendidikan Nasional. (2008). Sistem Pendidikan Nasional (*National Education System*). Retrieved October 25, 2008, from http://www.depdiknas.go.id/

Diseth, Ã. G. (2002). The relationship between intelligence, approaches to learning and academic achievement. *Scandinavian Journal of Educational Research, 46*, 219-230.

Eide, E., and Showalter, M. (1999). Factors affecting the transmission of earnings across generations: A quantile regression approach. *Journal of Human Resources, 34*, 253-267.

Gettinger, M. (1985). Time allocated and time spent relative to time needed for learning as determinants of achievement. *Journal of Educational Psychology, 77*, 3-11.

Gettinger, M., and White, M. A. (1979). Which is the stronger correlate of school learning? Time to learn or measured intelligence? *Journal of Educational Psychology, 71*, 405-412.

Gilchrist, W. (2000). *Statistical modelling with quantile functions*. Boca Raton, FL: Chapman and Hall/CRC.

Hahn, J. (1995). Bootstrapping quantile regression estimators. *Econometric Theory, 11*, 105-105.

Haile, G. A., and Nguyen, A. N. (2008). Determinants of academic attainment in the United States: A quantile regression analysis of test scores. *Education Economics, 16*, 29-57.

Halle, T. G., Kurtz-Costes, B., and Mahoney, J. L. (1997). Family influences on school achievement in low-income, African American children. *Journal of Educational Psychology, 89*, 527-537.

Hanushek, E. A. (1995). Interpreting recent research on schooling in developing countries, *World Bank Research Observer, 10,* 227-246.

Hanushek, E. A. (1979). Conceptual and Empirical Issues in The Estimation of Educational Production Functions. *Journal of Human Resources, 14*, 351-388.

Hao, L., and Naiman, D. Q. (2007). *Quantile Regression*. Thousand Oaks: Sage Publications Inc.

International Association for Evaluation of Educational Achievement. (2003). Trends in International Mathematics and Science Study (TIMSS). Retrieved August 28, 2008, from http://nces.ed.gov/timss/TIMSS03Tables.asp?Quest=3andFigure=5

Koenker, R. (2008). quantreg: Quantile Regression (Version R package version 4.22).

Koenker, R. (2005). *Quantile regression*. New York: Cambridge University Press.

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

Koenker, R., and F.Hallock, K. (2000). Quantile regression: An introduction. December 28,2000. from http://www.econ.uiuc.edu/~roger/research/intro/rq.pdf

Laidra, K., Pullmann, H., and Allik, J. (2007). Personality and intelligence as predictors of academic achievement: A cross-sectional study from elementary to secondary school. *Personality and Individual Differences, 42*, 441-451.

Moore, D. S. (2007). *The basic practice of statistics* (4th ed.). New York: W.H. Freeman and Co.

Mortimer, J. T., and Finch, M. D. (1996). The effects of work intensity on adolescent mental health, achievement, and behavioral adjustment: new evidence from a prospective study. *Child Development, 67*, 1243-1261.

Pedhazur, E. J. (1997). *Multiple regression in behavioral research : Explanation and prediction* (3rd ed.). Fort Worth, TX: Harcourt Brace College Publishers.

Programme for International Student Assessment. (2006). PISA 2006: Science Competencies for Tomorrow's World. 2008, from http://www.pisa.oecd.org/document/2/0,3343,en_32252351_32236191_39718850_1_1_1_1,00.html#tables_figures_dbase

R Development Core Team. (2008). R: A language and environment for statistical computing (Version 2.7.2): R Foundation for Statistical Computing.

Reiss, A. J. (1961). *Occupations and social status*. New York: The Free Press.

Rohde, T. E., and Thompson, L. A. (2007). Predicting academic achievement with cognitive ability. *Intelligence, 35*, 83-92.

Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research, 75*, 417.

Strauss, J., Beegle, K., Sikoki, B., Dwiyanto, A., Herawati, Y., and Witoelar, F. (2004). *The third wave of the indonesia family life survey: Overview and field report*. Santa Monica, CA: RAND corp.

Suara Pembaruan. (2008). Anak-anak, Sebuah Industri Baru (*Children, A New Industry*). *Suara Pembaruan Daily*. Retrieved September 17, 2008, from http://www.suarapembaruan.com/News/2008/06/01/Utama/ut01.htm

United Nations Statistics Division. (2008). Central government expenditure allocated to education (Publication.: http://data.un.org/Data.aspx?q=government+expenditure+educationandd=SOWCandf=inID%3a84

Velez, E., Schiefelbein, E., and Valenzuela, J. (1993). *Factors affecting achievement in primary education: A review of the literature for latin america and the caribbean*: Human Resources Development and Operations Policy, World Bank.

White, K. R. (1982). The relation between socioeconomic status and academic achievement. *Psychological Bulletin, 91*, 461-481.

Wiley, D. E., and Harnischfeger, A. (1974). Explosion of a myth: Quantity of schooling and exposure to instruction, major educational vehicles. *Educational Researcher, 3*, 7-12.

TABLES

Table 1
Descriptive Statistics

	Educational Attainment	Child Employment	Hours of Study	Managerial Professional	Skilled Manual	Skilled non Manual	Parent's Education	Income per Capita	Intelligence
mean	32.74	1.34	34.02	71*	150*	295*	242*	1.552099	10.35544
std dev	5.570	7.692	9.783	0.292	0.399	0.488353	0.467153	2.358701	2.356293
Q.01	19.53	0	3.53	0	0	0	0	0.045454	3
Q.05	24	0	6	0	0	0	0	0.102031	6
Q1	29	0	34	0	0	0	0	0.414881	9
median	32.46	0	36	0	0	0	0	0.888194	11
Q3	36.84	0	37	0	0	1	1	1.8	12
Q.95	41.79	0	44.35	1	1	1	1	5.0175	13
Q.99	43.97	49	49.47	1	1	1	1	9.282	13

Note: *= frequency of value of 1

Table 2.
Correlation Matrix of Predictors of Educational Attainment

Variables	1	2	3	4	5	6	7	8	9
1.Educational Attainment	-								
2.Children Employment	-0.03	-							
3.Hours of Study	0.113**	-0.194***	-						
4.Managerial/Professional	0.222***	-0.046	0.095**	-					
5.Skilled Manual	0.003	-0.001	-0.093*	-0.161**	-				
6.Skilled non Manual	0.038	0.012	0.078*	-0.258***	-0.400***	-			
7.Parents' Education	0.207***	-0.083*	0.149***	0.440***	-0.044	0.042601	-		
8.Income per Capita	0.248***	-0.01635	0.067	0.288***	-0.104**	0.119**	0.334***	-	
9.Intelligence	0.183***	-0.076*	0.203***	0.140***	-0.039	0.082*	0.165***	0.119**	-

Note: *=p<.05, **=p<.01, ***=p<.001

Table 3
ParameterEstimates of OLS Regression with Potential Outliers Included and Excluded

Variables	Included	Excluded
(Intercept)	27.037186***	27.03961***
Hours of child work	.001246	-.02603
Hours of study	.032736	.03308
Managerial / Professional	3.150483***	3.30325***
Skilled manual	1.266354*	1.36181*
Skilled non manual	.925007	.84783
Income per capita	.389984***	.46689***
Parent Education	.617959	.45562
Intelligence	.277017***	.27816***

Note: ***=p<.001, **=p<.01, *=p<.05

Table 4
R-square and F Values of OLS Regression with Potential Outliers Included and Excluded

	Included	Excluded
F	12.65***	13.69***
R-squared	.1196	.1291

Note: ***=p<.001

Table 5
Parameter Estimates of OLS Regression, OLS Regression Excluding Outliers and Median Regression

Parameter	OLS regression	OLS excluding outliers	Median Regression
(Intercept)	27.037186***	27.03961***	25.95442***
Hours of child work	.001246	-.02603	-.00831
Hours of study	.032736	.03308	0.03758
Managerial / Professional	3.150483***	3.30325***	3.66749**
Skilled manual	1.266354*	1.36181*	1.03746
Skilled non manual	.925007	.84783	1.02059
Income per capita	.389984***	.46689***	.34860*
Parent Education	.617959	.45562	1.22901
Intelligence	.277017***	.27816***	.34042*

Note: ***=p<.001, **=p<.01, *=p<.05

Table 6
Parameters of OLS Regression and Quantile Regression

Variables	OLS	0.05	0.1	0.25	Median	0.75	0.9	0.95
(Intercept)	27.037186***	18.82697***	19.82032***	23.20550***	25.95442***	31.64282***	34.62273***	36.92061***
Hours of child work	0.001246	0.02118	0.05004	0.00665	-0.00831	-0.00588	0.02733	-0.00657
Hours of study	0.032736	0.10463*	0.10461*	0.05764	0.03758	-0.01031	0.02234	-0.01270
Managerial / Professional	3.150483***	0.69092	1.30782	2.71200*	3.66749**	3.13424*	2.52195*	2.07020
Skilled manual	1.266354*	-0.06556	0.92404	0.32168	1.03746	1.56734	0.90628	0.28624
Skilled non manual	0.925007	-0.17741	1.17619	1.09563*	1.02059	0.27095	0.86085	0.42696
Income per capita (per 1,000,000 rupiahs)	0.389984***	0.58732*	0.53431***	0.42645***	0.34860*	0.39898	0.15775	0.17006
Parent Education	0.617959	0.94689	1.54754	0.84232	1.22901	-0.19817	0.75589	0.70167
Intelligence	0.277017***	0.08142	0.05178	0.23488*	0.34042*	0.36119*	0.31145*	0.35003***

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

Table 7
Significance Test for Parameter Difference Between Quantile .25, .5, and .75

Quantile	***df* between**	***df* residual**	***F***	***p***
$Q^{.75}$- $Q^{.25}$	8	1500	1.9227	0.05298
$Q^{.75}$,$Q^{.5}$,$Q^{.25}$	16	2246	1.559	0.0719

FIGURES

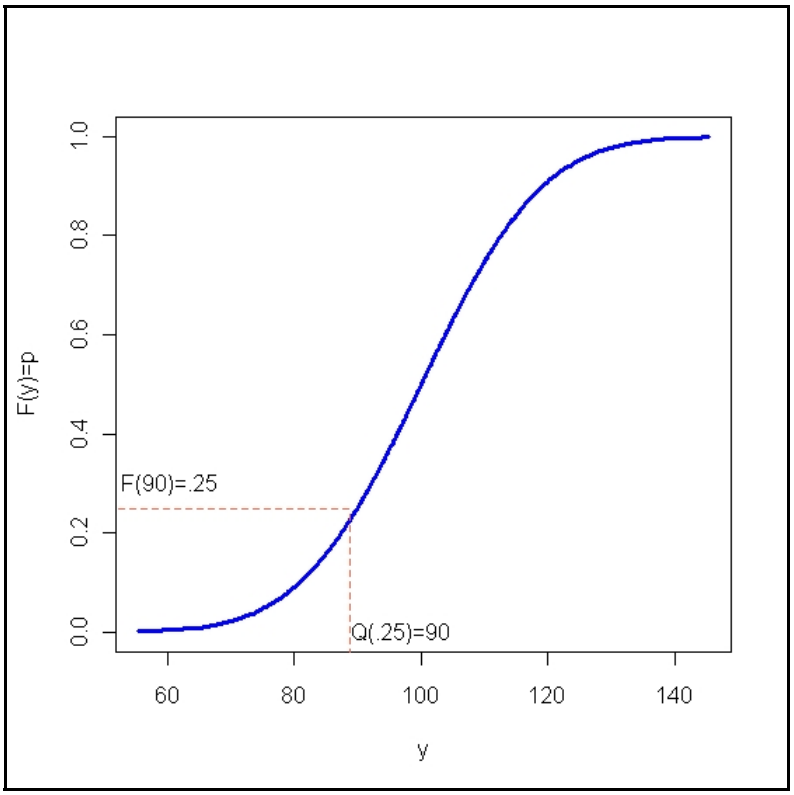

Figure 1.
Relationship between cumulative distribution function and quantile function.

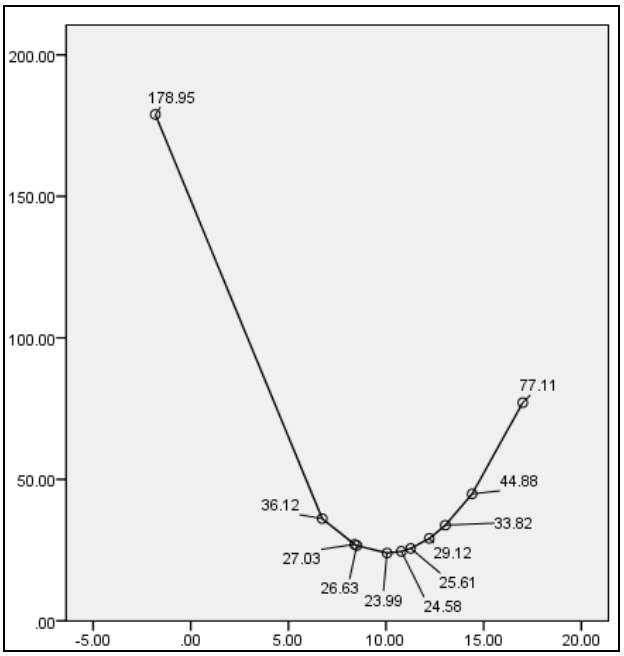

Figure 2.
Plots of *y* against mean of squared deviation.

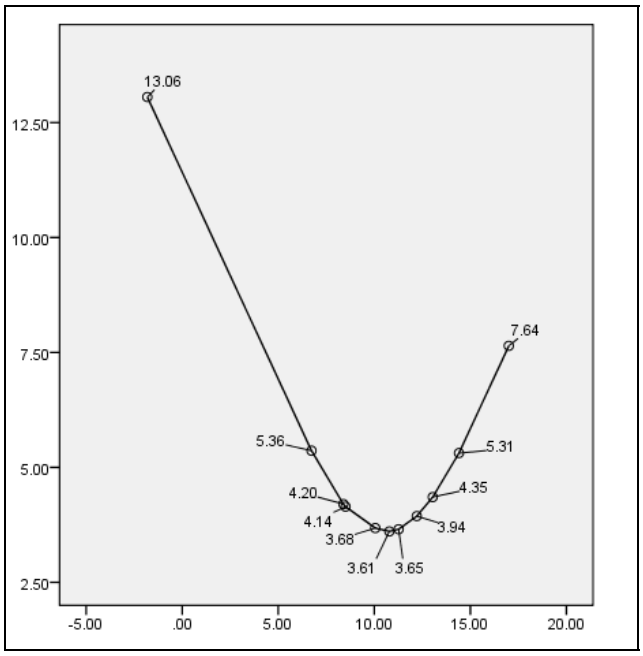

Figure 3.
Plots of *y* against mean of absolute deviation.

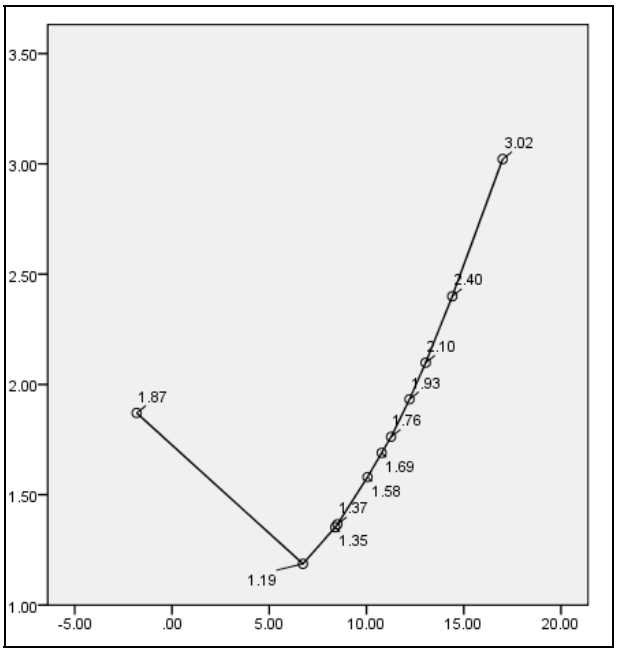

Figure 4.
Plots of *y* against weighted mean of absolute deviation for quantile .1.

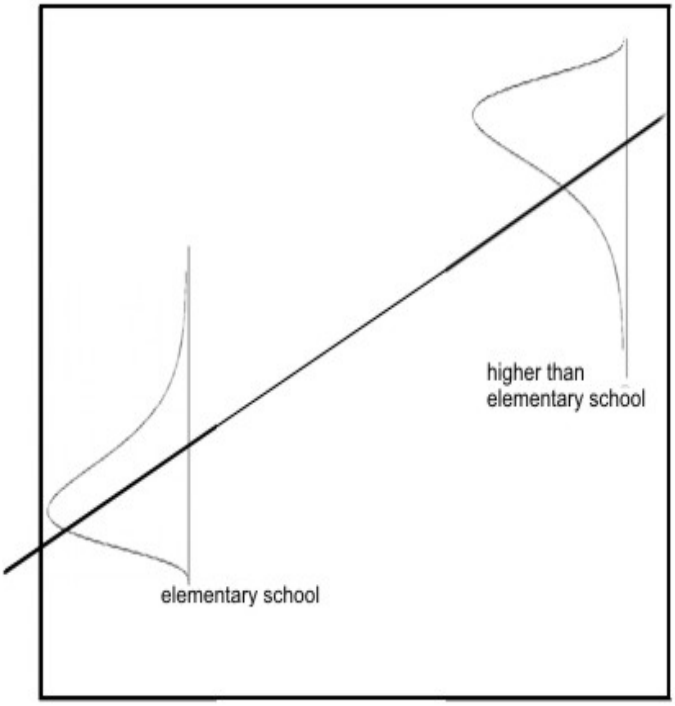

Figure 5.
Illustration of skewness shift.

(taken from RAND website : http://www.rand.org/labor/FLS/IFLS/)

Figure 6

Distribution of IFLS3 data.

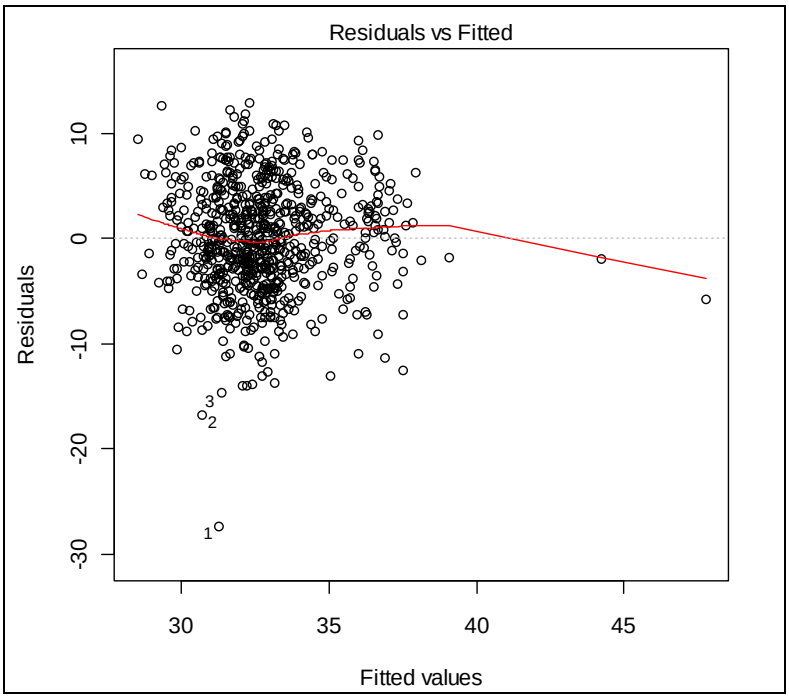

Figure 7.
Plots of residuals against predicted values

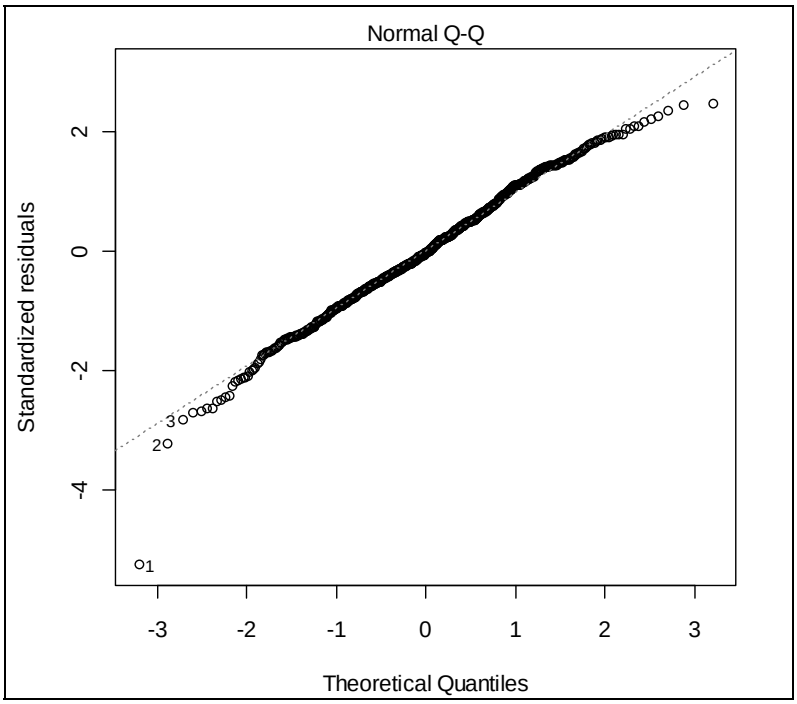

Figure 8.
Q-Q plot for normality of error term

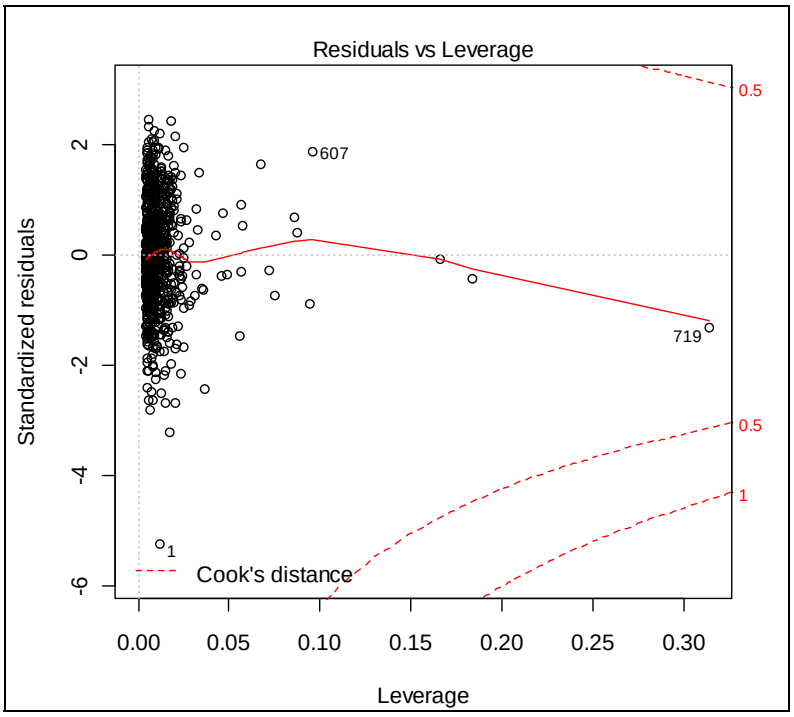

Figure 9

Plots of standardized residuals against Leverage

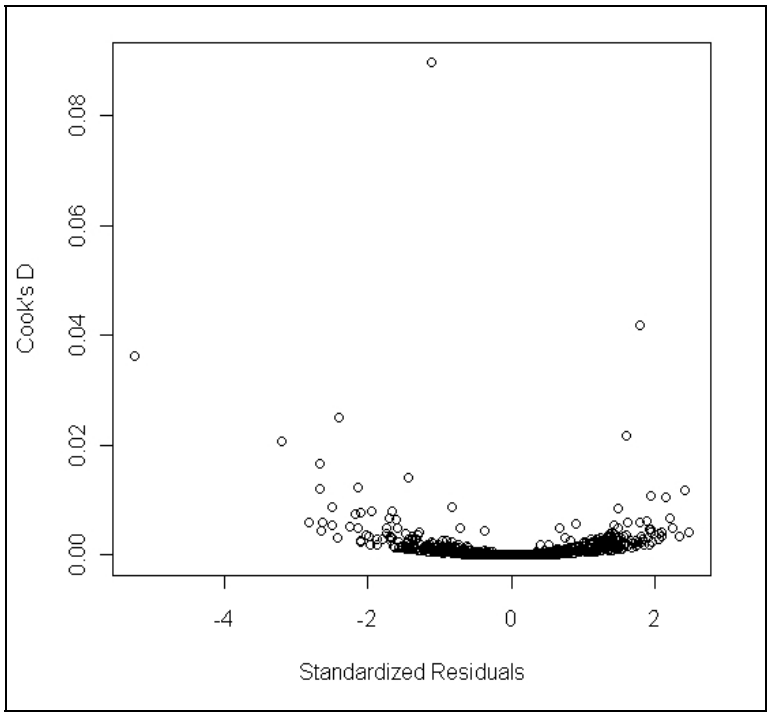

Figure 10
Plots of standardized residuals and Cook's *D*

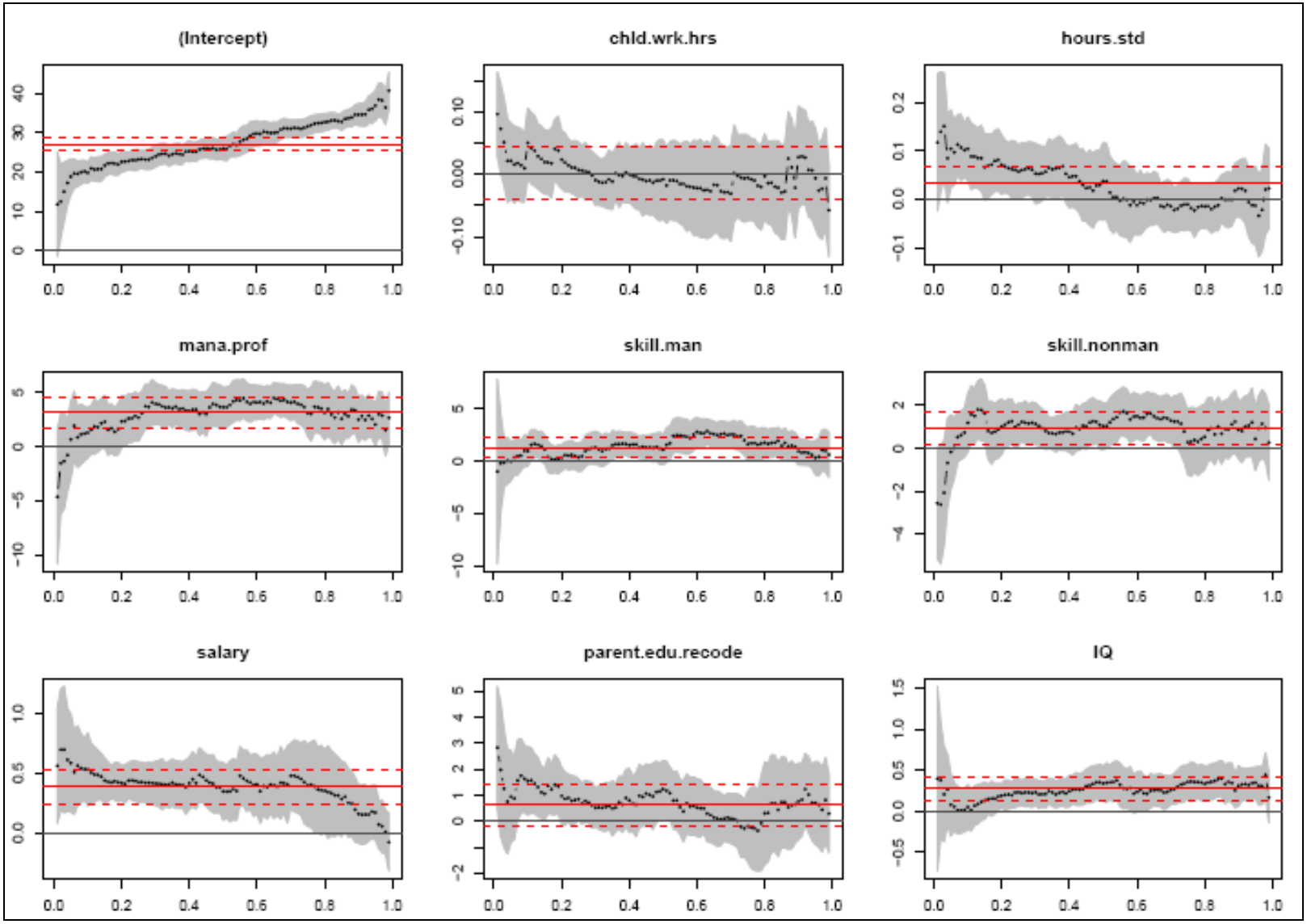

Figure 11

Parameters of OLS regression and quantile regression with 95% confidential intervals

APPENDIX A

SUMMARY TABLES FOR OLS REGRESSION AND QUANTILE REGRESSION ON EACH QUANTILE

Table A.1
Parameters, Standard Error, t-Value and p-Value for OLS Regression

	Estimate	**Std. Error**	***t* value**	**Pr(>|*t*|)**
(Intercept)	27.037186	1.043508	25.910	< 2e-16 ***
chld.wrk.hrs	0.001246	0.025443	0.049	0.960965
hours.std	0.032736	0.020546	1.593	0.111526
mana.prof	3.150483	0.832368	3.785	0.000166 ***
skill.man	1.266354	0.553743	2.287	0.022481 *
skill.nonman	0.925007	0.478217	1.934	0.053457 .
salary	0.389984	0.088777	4.393	1.28e-05 ***
parent.edu.recode	0.617959	0.482778	1.280	0.200941
IQ	0.277017	0.084584	3.275	0.001105 ***

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

Residual standard error: 5.255 on 745 degrees of freedom
Multiple *R*-squared: 0.1196, Adjusted *R*-squared: 0.1101
F-statistic= 12.65 on 8 and 745 DF, p-value= < 2.2e-16

Table A.2
Parameters, Standard Error, t-Value And p-Value for OLS Regression Excluding Outliers

	Estimate	**Std. Error**	***t* value**	**Pr(>|*t*|)**
(Intercept)	27.03961	1.02892	26.280	< 2e-16 ***
chld.wrk.hrs	-0.02603	0.02631	-0.989	0.322882
hours.std	0.03308	0.02062	1.605	0.109007
mana.prof	3.30325	0.81009	4.078	5.04e-05 ***
skill.man	1.36181	0.53992	2.522	0.011871 *
skill.nonman	0.84783	0.46604	1.819	0.069280 .
salary	0.46689	0.10196	4.579	5.48e-06 ***
parent.edu.recode	0.45562	0.47070	0.968	0.333376
IQ	0.27816	0.08278	3.360	0.000819 ***

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

Residual standard error: 5.089 on 739 degrees of freedom
Multiple R-squared: 0.1291, Adjusted R-squared: 0.1197
F-statistic= 13.69 on 8 and 739 DF, p-value= < 2.2e-16

Table A.3
Parameters, standard error, t-value and p-value for Median Regression

	Value	**Std. Error**	***t* value**	**Pr(>|*t*|)**
(Intercept)	25.95442	1.82181	14.24651	0.00000***
chld.wrk.hrs	-0.00831	0.03901	-0.21290	0.83146
hours.std	0.03758	0.04633	0.81108	0.41758
mana.prof	3.66749	1.17906	3.11052	0.00194**
skill.man	1.03746	0.93298	1.11199	0.26650
skill.nonman	1.02059	0.67414	1.51391	0.13047
salary	0.34860	0.15337	2.27295	0.02331*
parent.edu.recode	1.22901	0.70993	1.73118	0.08383
IQ	0.34042	0.13716	2.48192	0.01329*

Note: ***= $p<0.001$, **=$p<0.01$, *=$p<0.05$

Table A.4
Parameters, standard error, t-value and p-value for Quantile Regression Q.05

	Value	**Std. Error**	***t* value**	**Pr(>|*t*|)**
(Intercept)	18.82697	2.73160	6.89229	0.00000***
chld.wrk.hrs	0.02118	0.03724	0.56874	0.56970
hours.std	0.10463	0.04790	2.18419	0.02926*
mana.prof	0.69092	2.00274	0.34499	0.73020
skill.man	-0.06556	1.11938	-0.05857	0.95331
skill.nonman	-0.17741	1.04725	-0.16941	0.86552
salary	0.58732	0.25647	2.28999	0.02230*
parent.edu.recode	0.94689	1.01971	0.92858	0.35341
IQ	0.08142	0.20767	0.39209	0.69511

Note: ***= $p<0.001$, **=$p<0.01$, *=$p<0.05$

Table A.5
Parameters, standard error, t-value and p-value for Quantile Regression Q.10

	Value	**Std. Error**	***t* value**	**Pr(>|*t*|)**
(Intercept)	19.82032	1.69136	11.71855	0.00000***
chld.wrk.hrs	0.05004	0.03510	1.42579	0.15435
hours.std	0.10461	0.04104	2.54908	0.01100*
mana.prof	1.30782	1.60040	0.81718	0.41408
skill.man	0.92404	0.86186	1.07215	0.28400
skill.nonman	1.17619	0.81271	1.44725	0.14825
salary	0.53431	0.16158	3.30686	0.00099***
parent.edu.recode	1.54754	0.79325	1.95088	0.05145
IQ	0.05178	0.15176	0.34123	0.73302

Note: ***= $p<0.001$, **=$p<0.01$, *=$p<0.05$

Table A.6
Parameters, standard error, t-value and p-value for Quantile Regression Q.25

	Value	**Std. Error**	***t* value**	**Pr(>|*t*|)**
(Intercept)	23.20550	1.61054	14.40848	0.00000***
chld.wrk.hrs	0.00665	0.02593	0.25658	0.79757
hours.std	0.05764	0.03684	1.56485	0.11804
mana.prof	2.71200	1.33831	2.02644	0.04308*
skill.man	0.32168	0.83596	0.38481	0.70049
skill.nonman	1.09563	0.54397	2.01413	0.04436*
salary	0.42645	0.09399	4.53735	0.00001***
parent.edu.recode	0.84232	0.52013	1.61945	0.10577
IQ	0.23488	0.09719	2.41678	0.01590*

Note: ***= $p<0.001$, **=$p<0.01$, *=$p<0.05$

Table A.7
Parameters, standard error, t-value and p-value for Quantile Regression Q.75

	Value	**Std. Error**	***t* value**	**Pr(>|*t*|)**
(Intercept)	31.64282	1.51840	20.83958	0.00000***
chld.wrk.hrs	-0.00588	0.04046	-0.14542	0.88442
hours.std	-0.01031	0.03406	-0.30263	0.76226
mana.prof	3.13424	1.31613	2.38141	0.01750*
skill.man	1.56734	0.94443	1.65955	0.09743
skill.nonman	0.27095	0.91690	0.29551	0.76769
salary	0.39898	0.21908	1.82119	0.06898
parent.edu.recode	-0.19817	0.86857	-0.22815	0.81959
IQ	0.36119	0.14859	2.43081	0.01530*

Note: ***= $p<0.001$, **=$p<0.01$, *=$p<0.05$

Table A.8
Parameters, standard error, t-value and p-value for Quantile Regression Q.90

	Value	**Std. Error**	***t* value**	**Pr(>|*t*|)**
(Intercept)	34.62273	1.62002	21.37175	0.00000***
chld.wrk.hrs	0.02733	0.04683	0.58351	0.55973
hours.std	0.02234	0.03670	0.60881	0.54284
mana.prof	2.52195	1.25132	2.01544	0.04422*
skill.man	0.90628	0.81907	1.10647	0.26888
skill.nonman	0.86085	0.95098	0.90522	0.36564
salary	0.15775	0.17292	0.91224	0.36194
parent.edu.recode	0.75589	0.89915	0.84066	0.40081
IQ	0.31145	0.13457	2.31442	0.02092*

Note: ***= $p<0.001$, **=$p<0.01$, *=$p<0.05$

Table A.9
Parameters, standard error, t-value and p-value for Quantile Regression Q.95

	Value	**Std. Error**	***t* value**	**Pr(>|*t*|)**
(Intercept)	36.92061	2.36443	15.61504	0.00000***
chld.wrk.hrs	-0.00657	0.04836	-0.13590	0.89194
hours.std	-0.01270	0.04733	-0.26830	0.78854
mana.prof	2.07020	1.11305	1.85994	0.06329
skill.man	0.28624	1.01579	0.28179	0.77818
skill.nonman	0.42696	0.91153	0.46841	0.63963
salary	0.17006	0.12252	1.38802	0.16555
parent.edu.recode	0.70167	0.95588	0.73406	0.46314
IQ	0.35003	0.11923	2.93585	0.00343***

Note: ***= $p<0.001$, **=$p<0.01$, *=$p<0.05$

APPENDIX B

R SCRIPT FOR DESCRIPTIVE STATISTICS

```
#upload data file
data<-read.csv(file.choose(),header=T)
names(data)
attach(data)
salary<-salary.ave/1000000
data2=cbind(data,(salary.ave/1000000))
indeks=c(16,7,9,20,21,22,18,36,35)
data2=data2[indeks]

#summary
summari=matrix(1:81,nrow=9)
for (i in 1:9){
summari[1,i]=mean(data2[,i])
summari[2,i]=sd(data2[,i])
summari[3,i]=quantile(data2[,i],probs=0.01)
summari[4,i]=quantile(data2[,i],probs=0.05)
summari[5,i]=quantile(data2[,i],probs=0.25)
summari[6,i]=median(data2[,i])
summari[7,i]=quantile(data2[,i],probs=0.75)
summari[8,i]=quantile(data2[,i],probs=0.95)
summari[9,i]=quantile(data2[,i],probs=0.99)
rownames(summari)=c("mean","std
dev","Q.01","Q.05","Q1","median","Q3","Q.95","Q.99")
colnames(summari)=c("Ebtanas.tot","chld.wrk.hrs","hours.std","ma
na.prof",
"skill.man","skill.nonman","parent.edu.recode","salary","IQ")
}
summari=as.matrix(summari)
write.csv(summari,"summary_table.csv")
```

```
#chld.wrk.hrs more info
chld.wrk.recode=ifelse(chld.wrk.hrs>0,1,0)
chld.wrk.recode2=ifelse(chld.wrk.hrs>20,1,0)
chld.wrk.hrs2=chld.wrk.hrs[which(chld.wrk.hrs>0)]
jml0=sum(chld.wrk.recode)#amount of students who worked
jml20=sum(chld.wrk.recode2)#amount of students who worked more
than 20 hours
mean0=mean(chld.wrk.hrs2)#mean of working hours for students who
worked
mean20=mean(chld.wrk.hrs3)#mean of working hours for students
who worked more than 20 hours
```

APPENDIX C

R SCRIPT FOR CORRELATION MATRIX

```
#upload data file
data<-read.csv(file.choose(),header=T)
names(data)
attach(data)
data2=cbind(data,(salary.ave/1000000))
indeks=c(16,7,9,20,21,22,18,36,35)
data2=data2[indeks]
indeks1=c(1,2,3,4)
indeks2=c(1,5,6,7)
indeks3=c(1,8,9)

#make a correlation plot matrices
pairs(data2[indeks1])
pairs(data2[indeks2])
pairs(data2[indeks3])
pairs(data2)

#make a correlation matrix
cor.matrx=cor(data2)
cor.matrx=as.matrix(cor.matrx)
write.csv(cor.matrx,"correl_matrix.csv")

#make a correlation matrix with its significance
cor.sig=matrix(1:162,nrow=18)
for (i in 0:8){
j=(i*2)+1
l=(i+1)*2
cor.sig[j,]=cor.matrx[i+1,]
wek=c(cor.test(data2[,i+1],data2[,1])[3],
cor.test(data2[,i+1],data2[,2])[3],
cor.test(data2[,i+1],data2[,3])[3],
cor.test(data2[,i+1],data2[,4])[3],
cor.test(data2[,i+1],data2[,5])[3],
cor.test(data2[,i+1],data2[,6])[3],
cor.test(data2[,i+1],data2[,7])[3],
cor.test(data2[,i+1],data2[,8])[3],
cor.test(data2[,i+1],data2[,9])[3])
wek=as.matrix(wek)
rownames(wek)=NULL
```

```
cor.sig[l,]=t(wek)
cor.sig=matrix(cor.sig,nrow=18)
colnames(cor.sig)=colnames(cor.matrx)
}
a=1:18
rownames(cor.sig)=a
for (u in 0:8){
c=(u*2)+1
d=(u+1)*2
rownames(cor.sig)[c]=rownames(cor.matrx)[u+1]
rownames(cor.sig)[d]="p"
}

#make a csv file
write.csv(cor.sig,"correl_sig.csv")
```

APPENDIX D

R SCRIPT FOR OLS REGRESSION AND QUANTILE REGRESSION ANALYSIS

```
#upload quantreg library and data file
library(quantreg)
data<-read.csv(file.choose(),header=T)
names(data)
attach(data)
salary<-salary.ave/1000000

#OLS-Regression
ols<-lm(Ebtanas.tot~chld.wrk.hrs+hours.std+mana.prof+skill.man+
skill.nonman+salary+parent.edu.recode+IQ)
summary(ols)

#diagnostics
cook=cooks.distance(ols)
resid=resid(ols)
std.resid=(resid-mean(resid))/sd(resid)
plot(std.resid,cook,xlab="Standardized Residuals", ylab="Cook's
D")

#deletion of data with big cook's D
data.cook=data[-(which(cook > 0.02)),]
data.cook=as.data.frame(data.cook)
detach(data)

#regression on cleaned data
attach(data.cook)
salary=salary.ave/1000000
ols.cook<-
lm(Ebtanas.tot~chld.wrk.hrs+hours.std+mana.prof+skill.man+
skill.nonman+salary+parent.edu.recode+IQ, data=data.cook)
summary(ols)
detach(data.cook)
attach(data)
salary<-salary.ave/1000000

#Quantile Regression
median<-
rq(Ebtanas.tot~chld.wrk.hrs+hours.std+mana.prof+skill.man+
skill.nonman+salary+parent.edu.recode+IQ,tau=0.5)
```

```
summary(median,se="boot",R=200)

q05<-rq(Ebtanas.tot~chld.wrk.hrs+hours.std+mana.prof+skill.man+
skill.nonman+salary+parent.edu.recode+IQ,tau=0.05)
summary(q05,se="boot",R=200)

q10<-rq(Ebtanas.tot~chld.wrk.hrs+hours.std+mana.prof+skill.man+
skill.nonman+salary+parent.edu.recode+IQ,tau=0.1)
summary(q10,se="boot",R=200)

q25<-rq(Ebtanas.tot~chld.wrk.hrs+hours.std+mana.prof+skill.man+
skill.nonman+salary+parent.edu.recode+IQ,tau=0.25)
summary(q25,se="boot",R=200)

q75<-rq(Ebtanas.tot~chld.wrk.hrs+hours.std+mana.prof+skill.man+
skill.nonman+salary+parent.edu.recode+IQ,tau=0.75)
summary(q75,se="boot",R=200)

q90<-rq(Ebtanas.tot~chld.wrk.hrs+hours.std+mana.prof+skill.man+
skill.nonman+salary+parent.edu.recode+IQ,tau=0.9)
summary(q90,se="boot",R=200)

q95<-rq(Ebtanas.tot~chld.wrk.hrs+hours.std+mana.prof+skill.man+
skill.nonman+salary+parent.edu.recode+IQ,tau=0.95)
summary(q95,se="boot",R=200)

#make plot for parameters and 95%CI
plotCI<-
rq(Ebtanas.tot~chld.wrk.hrs+hours.std+mana.prof+skill.man+
skill.nonman+salary+parent.edu.recode+IQ,tau=1:99/100)
plotCI<-summary(plotCI,se="boot",R=200)
plot(plotCI)

#make a postscript file
postscript("QR_Plot.ps",horizontal=T, width=10,
height=7,paper="A4",colormodel="rgb")
plot(plotCI)
dev.off()

#Test Between Regression Line of Different Quantiles
anova(q25,q75)
anova(q25,median,q75)
```