



Received : April 21, 2016  
Accepted : August 08, 2016  
Published Online : September 26, 2016

AJ ID: 2016.04.02.STAT.01  
DOI : 10.17093/aj.2016.4.2.5000186603

## Assessment of PISA 2012 Results With Quantile Regression Analysis Within The Context of Inequality In Educational Opportunity

Sevda Gürsakil | Econometrics Department, Uludağ University, Turkey, sdalgic@uludag.edu.tr  
Dilek Murat | Econometrics Department, Uludağ University, Turkey, dilekm@uludag.edu.tr  
Necmi Gürsakil | Econometrics Department, Uludağ University, Turkey, gursakal@uludag.edu.tr

### ABSTRACT

The importance of educational opportunity inequality has been increasing within the context of education systems during recent years. In addition to quality in education, opportunity equality is among the significant paradigms in countries of high educational performance. Thus, it is of utmost importance to research the relationship between socio-economic characteristics of the students and achievement based on opportunity equality. Especially to remove the gap observed in Turkish literature is among the objectives of the present study. The main objective of the study is to assess the socio-demographic characteristics that affect the achievement of students in mathematics within the context of educational opportunity equality for PISA 2012 Turkey sample. Data analysis was conducted with quantile regression (QR) and classical linear regression (OLS). As a result, it was determined that students' family background, familiarity with information and communication technology and school climate were affective on mathematics achievement. It was observed that as parentel education, educational resources at home, and index of family wealth increased, mathematics achievement increased as well. It was also observed that time of computer use had a negative effect on achievement in mathematics. Furthermore, study findings identified that the achievement of male students was higher than females.

### Keywords:

PISA, Quantile regression, Mathematics score, Inequality of educational opportunity

## Eğitimde Fırsat Eşitsizliği Bağlamında PISA 2012 Sonuçlarının Kantil Regresyon Analizi İle Değerlendirilmesi

### ÖZET

Eğitimde fırsat eşitsizliği eğitim sistemleri açısından son yıllarda gittikçe önemi artan bir kavram haline gelmiştir. Eğitimde kalite ile birlikte fırsat eşitliği olgusu yüksek eğitim performansına sahip olan ülkelerin önemli paradigmaları arasında yer alır. Bu yönü ile öğrencilerin sosyoekonomik durumu ile eğitim düzeyi ve başarı arasındaki ilişkinin fırsat eşitliği bağlamında araştırılması önem arz etmektedir. Özellikle Türkçe literatürde gözlenen açıklığın giderilmesi de bu çalışmanın amaçları arasında yer almaktadır. Çalışmada PISA 2012 Türkiye örneklemini için eğitimde fırsat eşitliği bağlamında matematik başarısını etkileyen sosyo-demografik özelliklerin değerlendirilmesi amaçlanmıştır. Verilerin analizinde kantil regresyon ve klasik doğrusal regresyon analizinden yararlanılmıştır. Sonuç olarak öğrencinin aile özgeçmişi, bilgi ve iletişim teknolojisi ile aşinalığı ve okul ortamının matematik başarısı üzerinde etkili olduğu tespit edilmiştir. Ailenin eğitim düzeyi, evdeki eğitim kaynakları ve ailenin refah düzeyi arttıkça matematik başarısının arttığı gözlenmiştir. Bilgisayar kullanım süresinin ise matematik başarısı üzerinde negatif etkiye sahip olduğu görülmüştür. Ayrıca erkek öğrencilerin kız öğrencilere göre daha başarılı olduğu da çalışmada ulaşılan bir diğer bulgudur.

### Anahtar Kelimeler:

PISA, Kantil regresyon, Matematik puanı, Eğitimde fırsat eşitsizliği



## 1. Introduction

The future of individuals and their future life standards are closely related to the education they receive. A higher level of education is significant in achieving higher standards of life. Thus, the main purpose of education with respect to personal development and benefit of the society is to elevate the development potential of individuals to the highest possible level. In other words, the education system should operate to provide equal opportunities for equal success and to shape their future for all independent of individual and sociocultural characteristics, socioeconomic standing and health conditions or other factors. Because, a failure in achievement of educational goals would influence the individual all through her or his life negatively. Therefore, it is necessary to scrutinize the problem of inequality in education comprehensively. In their report titled "Turkish Education System Equality and Educational Achievement Report and Analysis," Oral and McGivney (2014) stated that "highest performing education systems are those that integrate quality and equality and could provide quality education opportunities for all children."

The relationship between the socioeconomic status and the educational level and success of the student is a subject that has been often investigated in the literature. Hence, James S. Coleman during 1960's reported that familial characteristics of students in the United States had a significant impact on educational achievement in his study titled "Equality of Educational Opportunity". Under the light shed by that study, the subject of equality of opportunity in education was assessed from different perspectives since 1960's and the studies especially focus on reorganization of education systems in countries to provide better equality of opportunity. Heyneman and Loxley determined that the effect of school resources on educational achievement was more significant than individual traits in developed countries and as the income level of the country decreased, the impact of school resources on educational achievement increased in a study they conducted. Also in the PISA research, which is also the scope of our study, it was determined that, in countries with low performances in PISA, educational achievement was closely related to the socioeconomic status of the family.

The objective of the study is to assess PISA 2012 Turkey sample with respect to equality of opportunity in education and socio-demographical features that affect achievement. Classical linear regression, as well as quantile regression were perused in this study. Classical linear regression analysis results provide us the effect of all explanatory variables (socio-demographic variables) at the conditional mean of the scores' distribution. However, we also want to know the effect of the explanatory variables at different parts of the score's distribution. Thus, we estimated quantile regressions (introduced by Koenker and Basset, 1978) at 5th, 25th, 50th, 75th and 95th quantiles. We followed Giambona and Porcu (2015), Santos (2007) and Fertig's (2003b) methods in this matter. Apart from estimating variables' effects at different parts of the distribution, quantile regressions have several other advantages compared to OLS (Ordinary Least Squares). First, they give less weight to outliers in the dependent variable when compared to the OLS. Second, estimation is a more robust method, because it allows the marginal effects of explanatory variables to differ across the quantiles of the dependent variable. Third, when error-terms are non-normal, quantile regression estimators may be more efficient than OLS

estimators. Finally, the semi-parametric nature of the approach relaxes the restrictions on the parameters to be constant across the entire distribution of the dependent variable (Rangvid, 2003: 12).

In the present study, a literature review of the studies that addressed PISA data with respect to equality of opportunity in education and quantile regression analysis is initially presented, and later on a detailed explanation of quantile regression analysis is given, followed by the actual quantile regression analysis to demonstrate the socio-demographic characteristics that affect the success of Turkey in PISA mathematics scores.

## 2. Literature Review

During recent years, several studies emerged that scrutinize inequality of opportunity in education with respect to PISA results along with the increasing importance of equality of opportunity in education. In one of these studies conducted by Natkhov and Kozina (2012), a measure of inequality of educational opportunity, which was estimated as a share of the variation in the 2009 PISA test scores explained by pre-determined family characteristics, was reported for 72 countries. The results of this study showed that there was a negative relationship between the inequality of educational opportunity and educational achievement. In countries where family background played a major role in determining individual progress, a lower mean educational achievement was found.

Tansel A. (2015), scrutinized the effects of inequality of opportunity in education on educational achievements in Turkey over time. Tansel used mathematics, science and reading achievement scores of 15 years old students in PISA between the years of 2003 and 2012. Study results demonstrated that inequality of opportunity in educational achievement declined marginally in Turkey in time and the most important determinants of the inequality in educational achievement were family background variables, which remained consistent over time.

In a study by Carvalho et al. (2013), a bi-dimensional (achievement-EOp and access-EOp) index and a Cobb-Douglas functional form were utilized. In that study recent PISA data for six Latin-American countries were used and rank reversals were observed with respect to orderings based on a single dimension. Authors of that study hypothesized two educational opportunity dimensions: (i) access, and (ii) achievement conditional on access, which relied on different perspective in EOp literature, and hence measured using a specific procedure before they were intergrated to form a composite index.

By simply using the variance or the standard deviation of test scores, Ferreira and Gignoux (2011) measured the inequality of achievement among all countries that participated in PISA 2006 survey. They utilized two alternative two-sample non-parametric procedures in order to assess the strength of the inequality measure on sample selection biases and utilized those on four countries with the smallest sample coverage in PISA (as a share of the total number of 15 years old students).

Martins and Veiga (2010) investigated mathematics achievements of 15 EU member students in PISA based on socioeconomic inequalities using 2003 data and a concentration index (CI) to analyze the differences between countries. The results of

the study demonstrated that the differences between these countries were significant. In Germany, Greece, Great Britain, Belgium, and Portugal, the inequality was greater, while in Sweden and Finland the results were reversed. According to the findings, socioeconomic factors explained 14.9% - 34.6% of the overall inequality in education.

Authors compared the most successful five OECD countries with the five most unsuccessful in PISA Mathematical literacy exams both in 2003 and 2006 based on social justice and equality of opportunity in education in a study by Aydın, Uysal and Sarier (2010). They scrutinized the data based on financial and human resources invested in education, the learning environment and organization in schools. Inadequacies in these dimensions were determined to have a negative impact on achievement of the students. As a result, the study reflected that the conditions and opportunities of the students were not equal. They have adequately emphasized that since the system could not provide equal opportunities and chances in education, it might not be possible to talk about success and failure.

Studies which assessed educational opportunity equality with quantile regression analysis provided a significant contribution to the literature during recent years. In one of these studies, Beblavy et al. (2014) investigated a subject which has been subject of extensive discussions previously. They analyzed the relationship between grouping based on ability and equality of educational opportunities in students with different performances in 4 OECD countries. Results demonstrated that in-classroom ability could have a positive or negative impact on educational equality based on the location of performance of the student on PISA achievement distribution. Based on quantile regression results, low performers were affected most by inequality effects in Belgium, while high performers were the most affected ones in Austria. The study was not able to find a significant relationship in Finnish students, while the results for Hungary were similar to those in Austria.

Fertig (2003a) conducted a study with 15 – 16 years old German students where relationship between the individual-level reading test scores of students with individual and family background information and their characteristics of the school and class were scrutinized. Several quantile regression analyses were conducted in the study, which showed that the findings did not support several known explanations such as high level school regulations or sharing non-citizen students among the participants substantially. On the contrary, findings demonstrated that schools with a more homogenous body of students had better educational achievement.

A study by Schneeweis and Winter-Ebmer's study (2005) was focused on educational production in Austria and potential effect of peers on students' academic achievement where PISA 2000 data for 15 – 16 years old students were used to estimate peer effects. The estimates demonstrated that socioeconomic composition of peer groups had significant positive effects on student achievement. Also, quantile regressions suggested that peer effects favored low-ability students that is students with lower skills benefited more by exposure to clever peers, while high-skill students were not much affected. It was also reported that social heterogeneity did not have substantial adverse effects on academic achievement.

Using PISA 2000 reading test scores, Santos (2007) studied the factors and distribution of educational quality in Argentina. Survey weighted regressions and quantile regressions were used to estimate educational production functions at the mean of the distribution in the study. Study results showed that, since girls perform significantly better than boys, educational policies should address gender issues to improve mean reading scores. It was also determined that the classroom population should not be more than 40-45 students and quality resources such as libraries, laboratory equipment and multi-media technology should be provided for schools. Study findings suggested that autonomy of the teachers was also important at improving students' achievements, in addition to their relationships with students and their openness to institutional change.

Lounkaew (2013) contributed to the educational equality debate by comparing urban and rural areas using the extensive data variables available in Thai PISA data. Students' education production function estimates based on unconditional quantile regression showed that the student and family contributions, and school characteristics were not symmetrical across achievement distributions and gender. Oaxaca-Blinder decomposition results indicated the significance of nontangible school characteristics in explaining the differences between the achievements of urban and rural students. Achievement percentile decomposition exercises demonstrated the increasing role of these characteristics; as the achievement percentile increased, these characteristics explained the gap better.

Giambona and Porcu (2015) examined the individual background characteristics that influenced the achievement of Italian 15 years-old students in reading via the analysis of 2009 OECD-PISA survey data and using the quantile regression (QR) approach. Results demonstrated that the predictors had significant impact on reading achievement operating differently across quantiles. These findings suggested that different paths should be followed to improve the achievements of low and high performing readers. Specifically, certain predictors such as family background (parental education, computer availability at home, and availability of a desk for homework at home), the school program attended and, the region of student domicile all played a significant but a different role when low and high performing readers were compared. For instance, parental education demonstrated a positive impact on student reading and general programs performed better than occupational or technical programs, and Northern regions performed better than Central-Southern regions, and all had different effects on the the distribution of students' reading scores.

Conducted literature review revealed that a study which assessed PISA scores using quantile regression has not been conducted in Turkey. Thus, it was considered that the current study would significantly contribute to the literature.

### 3. Data and Method

PISA 2012 Turkey sample was used in the present study. PISA 2012 Turkey sample included 4848 data. Lost observations based on the variables included in this study were excluded and the data count was reduced to 2962.

The Programme for International Student Assessment (PISA) is an international educational research program where students' knowledge and skills on the fields of mathematics, science, and reading skills were assessed and the research is organized by Organisation for Economic Co-operation and Development (OECD) in member and non-member nations. The present study evaluates the level of basic knowledge and skills of students in 15 age group in OECD and other participating countries to participate in modern society (PISA 2009 National Preliminary Report, 2010: 1).

PISA project aims to measure the proficiencies of 15-year-old pupils that attend formal education at the end of compulsory education in using their knowledge and skills in situations they could encounter in today's information society, not the level of their learning achievements in subjects in the curriculum (mathematics, science and reading skills) (<http://earged.meb.gov.tr>).

PISA project focuses on only one of the fields of reading skills, mathematics and science in each period as the main field. However, the assessments in the other two fields are included in the context of the study as well. In a circle of nine years, each of these fields become the main field once. In 2000 main field was reading skills in PISA application, it was mathematics literacy in 2003, and it was science literacy in 2006. In 2009, a new nine years long cycle has started and the application again focused on reading skills (PISA 2009 National Preliminary Report, 2010: 2). Finally, in the most recent report published in 2012, PISA 2012, mathematics literacy was the primary field.

Since the primary field in PISA 2012 application was achievement in mathematics, this field was selected as the dependent variable in the present study. Explanatory variables used in the study were common variables used extensively in the literature and their definitions are presented in Table 1. Literature review was determinative in the selection of explanatory variables. In fact, in the literature the explanatory variables were categorized under the titles of i) students' characteristics, ii) students' family background, iii) familiarity with information and communication technologies, and iv) school climate. It was assessed that these explanatory variables would be an indicator of equality/inequality when considered with respect to inequality of opportunity.

Label	Description	Mean	Std. dev
<b>MATH</b>	Mean Plausible values in math (Dependent Variable)	454.37	89.05
<b>Students' Characteristics</b>			
<b>AGE</b>	Age of student	15.81	0.27
<b>ANXMATH*</b>	Mathematics anxiety	0.25	1.03
<b>BELONG*</b>	Sense of belonging to school	0.15	1.08
<b>DICIPLINE*</b>	Disciplinary climate	-0.06	0.92
	The index is derived from the students' reports on how often they experienced the following		
	in their lessons: (i) students do not listen to what the teacher says;		
	(ii) there is noise; (iii) the teacher has to wait for a long time for the students to		
	quiet down; (iv) students cannot work well; and (v) students do not start		
	working for a long time when the lesson begins. Items are inverted for		
	scaling: higher values indicate better disciplinary climate		

Label	Description	Mean	Std. dev
<b>AGEATISCED 1</b>	Age at ISCED 1	6.87	0.53
<b>GENDER</b>	Gender of student 1=female n=1451 0=male n=1511		
<b>REPEAT</b>	If the student has grade repetition 1=Yes n=428 0=No n=2554		
<b>Students' Family Background</b>			
<b>HEDRES*</b>	The index is based on the items measuring the existence of educational resources at home, including a desk and a quiet place to study, a computer that students can use for school work, educational software, books to help with students' school work, technical reference books and a dictionary	-0.55	1.08
<b>PARED*</b>	Highest parent education in years	8.90	3.72
<b>WEALTH*</b>	The index is based on students' responses on whether they had the following at home: a room of their own; Internet access; a dishwasher; a DVD player; 3 other country specific items; and their responses on the number of cell phones, TVs, computers, cars and the rooms with a bath or shower. Higher values represent higher wealth	-1.51	0.96
	are available for them to use at home: (i) a desktop computer; (ii) a portable laptop or notebook; (iii) an Internet connection; (iv) a videogames console (v) a cell phone; (vi) MP3/MP4 or iPod or similar; (vii) a printer; and (viii) a USB stick. Higher values indicate greater ICT availability at home		
<b>Familiarity with Information and Communication Technology</b>			
<b>ICTSCHOOL*</b>	The index was derived from students' reports on whether any of the following are available for them to use at school: (i) a desktop computer; (ii) a portable laptop or notebook; (iii) an Internet connection; (iv) a printer; (v) a USB stick. Higher values indicate more ICT resources for learning at school	-0.39	1.14
<b>ICTHOME*</b>	The index was derived from students' reports on whether any of the following are available for them for use at home: (i) a desktop computer; (ii) a portable laptop or notebook; (iii) an Internet connection; (iv) a video games console, (v) a cell phone, (vi) MP3/MP4 or iPod or similar (vii) a printer; (viii) a USB stick. Higher values indicate more ICT resources for learning at school	-1.13	1.24
<b>TIMEINT*</b>		103.03	116.89
<b>School Climate</b>			
<b>STUDREL*</b>	Teacher student relation	0.19	1.07
<b>TCHBEHFA*</b>	Teacher behaviour: formative assessment	0.17	1.02
<b>TCHBEHSA*</b>	Teacher behaviour: student orientation	0.30	1.04
<b>TCHBEHTD*</b>	Teacher behaviour: teacher-directed instruction	0.39	1.16
<b>TEACHSUP*</b>	Teacher support	0.19	0.96

**Table 1.** Description of the Variables

\*: These indices were computed by OECD and available in the PISA database.

### 3.1. Quantile Regression Analysis

As is known, the objective of classical regression analysis is to establish a relationship between a dependent variable (response variable) and independent variables



(explanatory, predictor variables). In applications, response variable for any constant value of explanatory variable is accepted as a random variable and the value of response variable is generally summarized with the arithmetic mean in this framework. Classical regression analysis focuses on the mean and utilizes a function that determines the conditional mean of the response value for any constant explanatory variable (Hao and Naiman, 2013: 1).

In the following classical linear regression model:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

where  $y_i$  is a continuous response variable and  $x_i$  is an explanatory variable. It is assumed that  $\varepsilon_i$  random variable has a normal distribution with a zero mean and  $\sigma^2$  variance. After this model was applied to data,  $E[y|x]$ , which is the conditional mean of  $y$  for given  $x_i$  variable is calculated.

Parameters of this model is found by minimizing the residual total of squares:

$$\min \sum [y_i - (\beta_0 + \beta_1 x_i)]^2$$

There is no problem in classical regression analysis when the hypotheses of classical regression analysis are proven, in other words under ideal conditions. However, these hypotheses do not always conform to the real world. The corresponding response values to constant explanatory variable values could not always reflect a symmetrical distribution. Furthermore, the variances of these values could (heteroscedasticity). Heavy tailed distributions and outliers could be encountered (Jalali and Babanezhad, 2011: 1947). General approach within the framework of classical regression analysis is to exclude the outliers from the analysis. Which translates into losing valuable data for social sciences. Quantile regression facilitates the understanding of outliers at the end of the tails of the distribution instead of excluding them. Under these conditions, quantile regression becomes a tool for more productive analyses instead of using classical regression. Quantile regression, similar to other regression models, aims to explain the relationship between the variables.

Quantile regression models are used to estimate conditional mean functions and conditional quantile functions. Quantile regression is the generalized form of median regression for the determined quantiles. These regression models are less sensitive to marginal values and skewness when compared to OLS. Quantile regression initially emerged as a robust regression technique that ignores the normal distribution of error terms, one of the classical hypotheses and it is a method that was designed to present a more comprehensive regression view (Koenker, 2005:112).

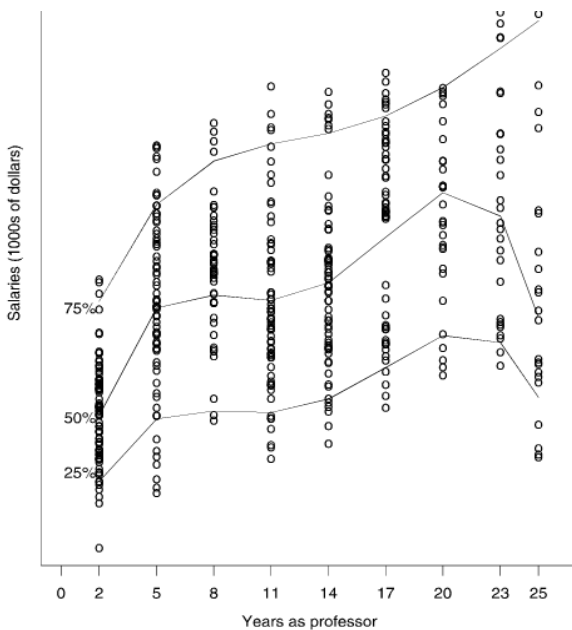
OLS method models the relationship between one or more explanatory variables and the conditional mean of  $Y$  dependent variable when  $X$  is given. On the other hand, quantile regression proposed by Koenker and Bassett (1978) provides a suitable method for modeling conditional quantile functions (Koenker and Hallock, 2001: 145). Quantile regression is especially useful in cases where conditional quantiles reflect variations. The method identifies the regression coefficients based on quantiles (Kurtoğlu, 2011: 33).

As stated by Mosteller and Tukey (1977, p. 266); "What the regression curve does is give a grand summary for the averages of the distributions corresponding to the set



of  $x$ 's. We could go further and compute several different regression curves corresponding to the various percentage points of the distributions and thus get a more complete picture of the set. Ordinarily this is not done, and so regression often gives a rather incomplete picture. Just as the mean gives an incomplete picture of a single distribution, so the regression curve gives a corresponding incomplete picture for a set of distributions."

Since numerous quantiles are modeled in quantile regression, it is possible to understand how the response variable is affected by the predictors in detail. Especially when the distribution of predictors changes, in conditional quantiles such as 5%, 25%, 50%, 75%, 90% and 95%, the distribution could be better determined. For instance, the relationship between the years as a professor and their salaries of 459 professors in the USA is presented in Figure 1. The figure is plotted for 0.25th, 0.50th, and 0.75th quantiles. The most interesting aspect of the figure is the fact that despite the quantile regression obtained for 0.25th and 0.50th quartiles gravitated downwards after working as a professor for 20 years, the regression for the 0,75th quartile gravitated upwards after 20 years. In short, these data, figure and results demonstrate that focusing on the mean in regression could be misleading in certain situations.



**Figure 1.** Salaries of 459 US statistics professors as a function of years as professor: three quantile regression curves with Q: 0.25, 0.50, 0.75 are shown

(Source: Yu Keming, Lu Zudi, Stander Julian, "Quantile regression: applications and current research areas", *The Statistician* (2003), 52, Part 3, 333-334)

Median is a special quantile. A median divides a data cluster into two, quartiles into four, deciles into ten, and percentiles into hundred. Generally all divisors are called "quantiles." Quantile regression models could be fitted to data with the minimization of a distance measurement that is generalized using algorithms based on linear programming. Provided that the quantile is shown as  $p$  ( $0 < p < 1$ ), weighted total of the distances to the theoretical line (without squaring) is minimized in quantile regression:

$$\min \sum |y_i - \beta_0 - \beta_1 x_i|$$

$p$  is accepted as the weight for the points above the matching line and  $1-p$  is accepted as the weight for the points below the matching line. For instance, quantile functions with different conditions could be matched for quantile values such as  $p=0.25$ ;  $0.50$ ;  $0.75$  (Hao and Naiman, 2013:33).

Conditional quantile is specified by the basic quantile regression model as a linear function of explanatory variables. The usual way to write the quantile regression model for the  $\tau$ th quantile is:

$$Q_y(\tau / X = x) = x_i' \beta(\tau) \quad \text{with} \quad \hat{\beta}(\tau) = \arg \min_{\beta \in \mathbb{R}^k} \sum \rho_\tau(y_i - x_i' \beta)$$

The minimization of the sum of absolute deviations residuals could be solved with the  $\tau$ th quantile ( $0 < \tau < 1$ ) of  $y$ :

$$\min_{\beta^\tau} \sum_{\sigma_k < 0} \tau |y_k - X_k \beta^\tau| + \sum_{\sigma_k > 0} (1-\tau) |y_k - X_k \beta^\tau|$$

Where  $\tau$  is the conditional quantile of interest and all positive residuals receive a weight of  $\tau$ , while the negative ones receive a weight of  $(1-\tau)$ . Thus, any component of the Quantile regression coefficients  $\beta(\tau)$  provides an estimate for the marginal effect of the associated explanatory variable on the response for the  $\tau$ th quantile, controlling for the remaining variables.

Thus, the quantile regression model to be used in the present study could be written as follows:

$$Y_i = X_i' \beta_\tau + u_{\tau i}, \quad \text{Quant}(A_i / X_i) = X_i' \beta_\tau \quad \text{with} \quad \tau \in (0,1)$$

Where  $\text{Quant}(Y_i / X_i)$  denotes the  $\tau^{\text{th}}$  quantile of the math score  $Y_i$  conditional on the vector of explanatory  $X_i$  variables.

#### 4. Results

Table 2 presents the results obtained with OLS and QR from the 5th to 95th percentile of the distribution of the students' math scores. The first column reports the selected covariates, columns 2–11 reflect the coefficients for the three main quantiles 0.05, 0.25, 0.50, 0.75 and 0.95, reported with their level of significance, finally the last two columns report the OLS coefficients and the related levels of significance.

According to Giambona and Porcu (2015), if  $\beta$  coefficients of the quantile regression are significant and different from the OLS  $\beta$  regressions, then the use of quantile regression is more efficient than the regression on the average. Thus, when OLS and quantile regression results are compared, it is possible to argue that there were several differences between the results obtained with two models generally. Comparison of the model results based on gender demonstrates that both results show gender variable significantly affected mathematics achievement and male students were more successful than female students. Quantile regression findings showed that the gender variable coefficient increased from 0.05th quantile towards the 0.95th quantile. This finding was consistent with the main experimental results found in the literature, despite the fact that the gender gap decreased at the upper quantile of the distribution as observed in quantile regression coefficients. The mathematics achievement scores of students that never failed a year at school were higher than those who did and the impact of REPEAT increased across the quantiles.

It was observed that mathematics anxiety (ANXMATH) significantly affected achievement both in OLS and quantile regression, and quantile regression findings demonstrated that anxiety increased at a great extent at high achievement levels.

While the impact of BELONG variable on mathematics performance was significant based on OLS regression results ( $p = 0.0068$ ), the parameter of this variable was not significant based on quantile regression. Similarly, AGE AT ISCED1 variable was significant in OLS regression ( $p = 0.0034$ ), it was insignificant based on quantile regression results. Based on OLS findings, there was a negative correlation between schooling age and mathematics achievement. This could be interpreted as the mathematics achievement would decrease as the the schooling age would increase.

Disciplinary climate (DICIPLINE) was positively correlated with mathematics performance and the effect was increasing across the quantiles based on OLS and quantile regression results. The highest parental education (PARED) had a significant effect on mathematics scores and there was a positive correlation between the two. That is, mathematics achievement score increased with the parent's education level. Home educational resources (HEDRES) had significantly positive effects on mathematics score, but at the 95th quantile the effect was not significant. Between the index of family wealth (WEALTH) and mathematics score there was a significant and positive correlation based on OLS results, however, quantile regression scores varied across the quantiles. Wealth was not significant and negatively correlated with mathematics score in the 0.05th quantile, but there was a significantly positive correlation between the coefficient and mathematics score in the remaining four quantiles and the correlation increased from 0.25th to 0.95th quantiles.

There was no correlation between availability of computer resources at school (ICTSCHOLL) and mathematics score, whereas availability of computer resources at home (ICTHOME) has differing effects across different quantiles and based on the OLS. Based on OLS findings, there was a significant and positive correlation between ICTHOME and mathematics score, as well as 0.05th, 0.25th and 0.75th quantiles based on quantile regression findings. However, ICTHOME was not significantly correlated with mathematics score based on the median and the 0.95th quantile regression results. There was a negative correlation between the time of computer use (TIMEINT) and mathematics score for all results except for the quantile regression at the 0.05th quantile ( $p = 0.8209$ ). This finding demonstrates that mathematics score decreased with an increased time on computers.

An analysis of the variables included in the study to evaluate mathematics achievement based on school climate for the results of both regression models showed that student-teacher relationship (STUDREL) did not affect mathematics achievement, teacher behavior: formative assessment (TCHBEHFA) created a positive and significant effect only in OLS and median regression results, TCHBEHSO had a negative and significant effect in all regression results, and the coefficient for this variable demonstrated an increase from the 0.05th quantile towards the 0.95th quantile.

	Q=0.05	Prob.	Q=0.25	Prob.	Q=0.50	Prob.	Q=0.75	Prob.	Q=0.95	Prob.	OLS	Prob.
C	362,12	0,000	460,29	0,000	506,59	0,043	581,71	0,000	720,25	0,022	527,13	0,000
ANXMATH	-8,45	0,000	-12,79	0,000	-15,21	0,000	18,57	0,000	-19,81	0,000	-16,03	0,000
BELONG	-0,06	0,978	-2,65	0,105	-3,00	0,069	-3,38	0,054	-2,40	0,569	-3,69	0,007
DICIPLINE	8,10	0,003	5,28	0,001	12,26	0,000	15,18	0,000	11,38	0,000	10,74	0,000
GENDER	12,74	0,003	20,94	0,000	24,13	0,000	-37,49	0,000	45,30	0,000	28,03	0,000
HEDRES	7,83	0,000	8,74	0,000	10,01	0,000	8,08	0,001	2,92	0,480	9,32	0,000
ICTHOME	6,41	0,000	4,77	0,006	3,54	0,104	5,50	0,048	6,96	0,205	4,70	0,003
ICTSCHOOL	-2,20	0,108	-0,27	0,819	2,44	0,136	3,08	0,119	4,43	0,322	1,33	0,277
AGEATISCED1	-1,83	0,453	-5,21	0,041	-4,42	0,188	-6,16	0,098	-14,49	0,084	-7,10	0,003
PARED	3,20	0,000	2,78	0,000	3,65	0,000	4,31	0,000	3,94	0,000	4,06	0,000
STUDREL	-1,98	0,362	2,03	0,174	-0,34	0,852	-2,27	0,340	0,75	0,830	-0,01	0,989
TCHBEHFA	-1,20	0,702	2,82	0,253	6,42	0,002	4,35	0,000	2,23	0,630	4,28	0,025
TCHBEHFO	-13,98	0,000	-23,36	0,000	-24,96	0,000	-25,79	0,000	-26,63	0,000	-24,11	0,000
TCHBEHTD	5,61	0,020	4,31	0,039	0,94	0,633	-0,11	0,958	-3,60	0,508	2,52	0,147
TEACHSUP	2,42	0,340	5,27	0,008	4,14	0,090	7,93	0,090	11,74	0,010	4,49	0,013
TIMEINT	0,00	0,820	-0,03	0,007	-0,04	0,004	-0,07	0,000	-0,06	0,036	-0,05	0,000
WEALTH	-0,08	0,976	4,69	0,058	8,86	0,004	10,42	0,000	11,50	0,010	7,24	0,000
REPEAT	-37,69	0,000	-46,84	0,000	-58,40	0,000	-72,77	0,000	-78,03	0,00	-61,71	0,000
R <sup>2</sup>	0.137		0.176		0.215		0.260		0.260		0.380	

**Table 2.** Estimated Coefficients for OLS and QR Models for Q=0.05, Q=0.25, Q=0.50, Q=0.75 and Q=0.95

## 5. Conclusion

In this study we analyzed Turkish students' mathematics achievement determinants using the last PISA 2012 survey. We applied Quantile Regression (QR) to assess the impact of selected variables at more than one level of mathematics achievement distribution. We selected 0.05th, 0.25th, 0.50th, 0.75th and 0.95th quantiles for Quantile regression analysis. And also, we used the Ordinary Least Square (OLS) regression to compare with the quantile regression results. We found that students' characteristics, students' family background, familiarity with information and communication technology and school climate were effective factors for math achievement in both OLS and QR models. Furthermore, an aspect of QR results were superior to OLS findings, and that was its capacity to analyze the effect of variables on different distribution points.

Based on gender variable, it was observed that this variable affected mathematics achievement significantly and male students had a higher level of achievement when compared to females. However, QR results showed that gender gap decreased as the analysis moved to upper quantiles of the distribution.

OLS and QR model findings demonstrated that mathematics anxiety (ANXMATH) had a significant impact on achievement and anxiety had an increasing effect in higher levels of achievement.

OLS and QR results showed that disciplinary climate (DICIPLINE) was positively correlated with mathematics achievement and this effect increased across the quantiles. There was a positive and significant correlation between the highest

parental education (PARED) and mathematics score. That is, mathematics achievement score increased with the parents' education level. Home educational resources (HEDRES) estimation coefficients had significant and positive effects on mathematics achievement with the only exception of the 95th quantile. OLS results showed that there was a positive and significant correlation between index of family wealth (WEALTH) and mathematics score, but quantile regression findings differed across the quantiles. Wealth did not have a significant effect in the 0.05th quantile and was correlated negatively with mathematics score, however, in the other four quantiles, there was a positive and significant correlation between this coefficient and mathematics score, which increased through 0.25th to 0.95th quantiles.

There was no correlation between availability of computer resources at school (ICTSCHOLL) and mathematics score, while the effects of availability of computer resources at home (ICTHOME) differed across the OLS and different quantiles. OLS results suggested that there was a positive and significant correlation between ICTHOME and mathematics score in the 0.05th, 0.25th and 0.75th quantiles of quantile regression results. However, there was no correlation based on median and 0.95th quantile regression results. There was a negative and significant correlation between time of computer use (TIMEINT) and mathematics score for all regression results except the 0.05th quantile ( $p = 0.8209$ ). It could be deduced that mathematics score decreased with the time spent on computer.

Finally, overall application results demonstrated that quantile regression approach was a more robust when compared to OLS since it could analyze the effects on different ends of the distribution. This fact shows that QR approach is significantly superior in demonstrating the differences in educational opportunity inequalities.

## References

- Aydın, A., Uysal, Ş. and Sarier, Y. (2010). Analysing the Results of Pisa Maths Literacy in Terms of Social Justice and Equality in Educational Opportunities. *Procedia Social and Behavioral Sciences*, 2(2), 3537–3544.
- Beblavy, M., Thum, A.-E., Potjagailo, G. & Werder, M. V. (2014). A Closer Look at The Effects of Within-School Ability Grouping in Secondary Schools: How Are Different Performers Affected?. IAAE 2014 Annual Conference, Queen Mary, University of London.
- Carvalho, M., Gamboa, L. F. & Waltenberg F. D. (2013). Equality of Educational Opportunity Employing PISA Data: Taking Both Achievement and Access Into Account, Paper Prepared for the IARIW-IBGE Conference on Income, Wealth and Well-Being in Latin America, Rio de Janeiro, Brazil.
- Coleman, J. S. (1966). *Equality of Educational Opportunity*, U.S. Government Printing Office, Washington.
- Ferreira F. H. G. & Gignoux J. (2011). The Measurement of Educational Inequality: Achievement and Opportunity. PSE Working Papers n2011-38. 2011.
- Fertig, M. (2003a). Who's to Blame? The Determinants of German Students' Achievement in the PISA 2000 Study. IZA Discussion Paper No. 739, RWI Essen and IZA Bonn.
- Fertig, M. (2003b). Educational Production, Endogenous Peer Group Formation and Class Composition – Evidence from the PISA 2000 Study. IZA DP No. 714.
- Giambona, F. & Porcu, M. (2015). Student Background Determinants of Reading Achievement in Italy. A Quantile Regression Analysis. *International Journal of Educational Development*, 44, 95–107.
- Hao, L. & Naiman, Q. D. (2013). *Quantile Regression*. Sage Publications, Inc, California.

- Heyneman, S. P. and Loxley, W. A. (1983). The Effect of Primary School Quality on Academic Achievement Across Twenty-Nine High and Low Income Countries. *American Journal of Sociology*, 88 (6), 1162-1194.
- <http://earged.meb.gov.tr>, erişim tarihi: 18.03.2016
- Jalali, N. & Babanezhad, M. (2011). Quantile Regression due to Skewness and Outliers. *Applied Mathematical Sciences*, 5(39), 1947 – 1951.
- Koenker, R. & Bassett, G. (1978). Regression Quantiles. *Econometrica*, 46(1), 33–50.
- Koenker, R. & Hallock, K. F. (2001). Quantile Regression. *Journal of Economic Perspectives*, 15(4), 143–156.
- Kurtoğlu, F. (2011). Quantile Regresyon: Teorisi ve Uygulamaları. Çukurova Üniversitesi Sosyal Bilimler Enstitüsü İstatistik Ana Bilim Dalı Ekonometri, Basılmamış Yüksek Lisans Tezi, Adana.
- Lounkaew, K. (2013). Explaining Urban–Rural Differences in Educational Achievement in Thailand: Evidence From PISA Literacy Data. *Economics of Education Review*, 37, 213–225.
- Martins, L. & Veiga, P. (2010). Do Inequalities in Parents' Education Play An Important Role in PISA Students' Mathematics Achievement Test Score Disparities?. *Economics of Education Review*, 29, 1016–1033.
- Mosteller, F. & Tukey, J. (1977). *Data Analysis and Regression: A Second Course in Statistics*. Reading, Mass., Addison-Wesley.
- Natkhov, T. & Kozina, N. (2012). Inequality of Educational Opportunity in a Cross-Section of Countries: Empirical Analysis of 2009 Pisa Data. National Research University Higher School of Economics, Moscow, Russia. Working paper Series: Education WP BRP 07/EDU/2012.
- Oral, I. & McGivney, E. J. (2014). Türkiye Eğitim Sisteminde Eşitlik ve Akademik Başarı Araştırma Raporu ve Analiz. Eğitim Reformu Girişimi Raporları.
- PISA 2009 Ulusal Ön Rapor, (2010). Milli Eğitim Bakanlığı, Eğitimi Araştırma ve Geliştirme Dairesi Başkanlığı, Ankara.
- Rangvid, B.S. (2003). Evaluating Private School Quality in Denmark. Working Paper No. 03-2, Aarhus School of Business, Denmark.
- Santos, M. E. (2007). Quality of Education in Argentina: Determinants and Distribution Using PISA 2000 Test Scores. *Wellbeing and Social Policy* 3(1), 93-119.
- Schneeweis, N. & Winter-Ebmer, R. (2005). Peer Effects in Austrian Schools. Institute for Advanced Studies, Vienna.
- Tansel, A. (2015). Inequality of Opportunities of Educational Achievement in Turkey over Time. IZA Discussion Paper No. 9005 April 2015
- Yu, K., Lu, Z. & Stander, J. (2003). Quantile Regression: Applications and Current Research Areas. *The Statistician* 52(3), 333-334.