Journal of Data Science **6**(2008), 449-465

An Analysis of Mathematics and Science Achievements of American Youth with Nonparametric Quantile Regression

Maozai Tian, Xizhi Wu, Yuan Li and Pengpeng Zhou
Renmin University of China

Abstract: Considering the importance of science and mathematics achievements of young students, one of the most well known observed phenomenon is that the performance of U.S. students in mathematics and sciences is undesirable. In order to deal with the problem of declining mathematics and science scores of American high school students, many strategies have been implemented for several decades. In this paper, we give an in-depth longitudinal study of American youth using a double-kernel approach of nonparametric quantile regression. Two of the advantages of this approach are: (1) it guarantees that a Nadaraya-Watson estimator of the conditional function is a distribution function while, in some cases, this kind of estimator being neither monotone nor taking values only between 0 and 1; (2) it guarantees that quantile curves which are based on Nadaraya-Watson estimator not absurdly cross each other. Previous work has focused only on mean regression and parametric quantile regression. We obtained many interesting results in this study.

Key words: Double-kernel approach, nonparametric quantile regression, plug-in rule, science and mathematics achievements.

1. Introduction

Recently more and more attention has been paid to the importance of science and mathematics achievements of young students as the pace of change in our lives is becoming faster and faster. We need broad mathematics- and science-related knowledge and abilities for our everyday decision-making. Furthermore, intrinsic value of mathematical and scientific knowledge shape and define our common life, history, and culture. Mathematics and sciences are primary sources of lifelong learning and the progress of our civilization.

“We as a nation must take immediate action to improve the quality of mathematics and science teaching in every classroom in this country,” John Glenn, Former U.S. Senator and NASA astronaut, said when he presented his report.

"If we delay, we put at risk our continued economic growth and future scientific discovery...".

However, students' mathematics and science test scores have been going down in the United States for the past decades, and the dismal performance of U.S. students in mathematics and science have been stressed. American students in fourth grade are among the leaders on mathematics assessments worldwide, but by the time students graduate from high school, they are nearly last among 41 nations, according to the Third International Mathematics and Science Study (TIMSS). According to the U.S. National Assessment of Education Progress (NAEP), fewer than one-third of all U.S. students in grades four, eight, and twelve performed at or above proficient achievement levels. Nearly a third performed below basic performance levels.

Among the earlier controversial studies in this area, Coleman (1966) studies the effects of schooling inputs including class size on scholastic achievement and concludes that schooling inputs have a negligible effect on scholastic achievement. Hanushek (1986) provides a review of analyzing the effects of various inputs in the effects of class size of public schooling finding that the effect of class size reduction (and more generally increased expenditures on education) on achievement is ambiguous, wavering from positive to negative depending on the study. A number of studies have examined the effect of school quality on student achievement. (For example, Ehrenberg and Brewer, 1994, 1995, and Hanushek, 1996). The findings of these studies conclude that improving school resource can hardly improve students' performance on standardized achievement tests, which are indeed run counter to the conventional view point. Previous work has focused on the average effects using classical least squared methods.

Recently, quantile regression (QR), as introduced by Koenker and Bassett (1978), has been developed into a comprehensive approach to the statistical analysis of linear and non-linear response models and has been used in a broad range of application settings. Using quantile regression, several recent studies have modelled the performance of student on the standardized tests as a function of many factors such as the parents' socio-economic status, the number of parents and siblings, class size, teacher qualifications, *etc.* For example, Edide and Showalter (1998) uses quantile regression to estimate whether the relation between school quality and performance on standardized tests differs at different points in the conditional distribution of "tests score gains". Levin (2001) addresses the controversial topic of class size reduction, and controlling for a large number of observable characteristics and potential endogeneity in the class size variable, an educational production function, is estimated using a quantile regression technique while the main finding is that due to heterogeneity in the newly identified peer effect, class size reduction is a potentially regressive policy measure. Tian

(2006) investigates whether the family background factors alter performance on mathematical achievements of the stronger students in the same way that weaker students are affected by means of quantile regression approach. Tian's findings suggests that there may be differential family-background-factor effects at different points in the conditional distribution of mathematical achievements.

In this paper, we give an in-depth Longitudinal study of American youth using a double-kernel approach of nonparametric quantile regression (Yu and Jones, 1998). Two of the advantages of the approach are: (1) to insure that a Nadaraya-Watson estimator of the conditional function is a distribution function while, in some cases, this kind of estimator is neither monotone nor taking values only between 0 and 1; (2) to insure that quantile curves which are based on Nadaraya-Watson estimator not absurdly cross each other. Previous work has focused only on mean regression and parametric quantile regression.

This paper is organized as follows: Section 2 describes the data used for both science and mathematics achievement regression. Section 3 introduces the double kernel approach in the quantile regression. The resulting estimation and more detailed discussion about science and mathematics achievement are presented in Section 4 and 5 respectively. The relationship between science and mathematics achievements is discussed in Section 6. The conclusion is presented in the last section.

2. Data

The data, which represent science and mathematics achievement samples of seventh-grade to twelfth-grade students from 1987 to 1992, are taken from Public Opinion Laboratory, Northern Illinois University, DeKalb, Illinois. (Reference to http://www. lsay.org). Beginning in the fall of 1987, the LSAY is a longitudinal panel study of public middle and high school students. About 60 seventh graders were randomly selected in each of the 52 schools and the total sample size was 3116 students. These students were followed for six years from grade 7 to grade 12, writing mathematics and science achievement tests and completing student questionnaires annually. With a focus on mathematics and science education, the information from students, parents, and teachers was also included in the study.

Outcome measurements are seven item scores: basic mathematics skills (BAS), algebra (ALG), geometry (GEO), quantitative literacy (QLT), biology (BIO), physics (PHY), and environmental sciences (ENV).

The BAS subscale measures the achievement in the basic mathematics skills using the items on the mathematics achievement test calibrated to measure an understanding of basic mathematics skills, and so on. The first four item scores are the measures of the latent true score for mathematics achievement with measurement errors respectively. The last three item scores are the measures of the corresponding latent true score for science achievement with measurement errors.

Item scores in each subject are imputed scores which include non-aberrant observed scores when available. These items are stored as continuous variables in the data file. These scores are comparable across grade levels within each school subject. There are missing data because some children were absent during testing. However, all available student achievement scores are used as dependent measures.

3. Methodology

3.1 Double kernel approach

The characteristic of a longitudinal study is that individuals are measured repeatedly through time. In the analysis of longitudinal data, we are usually interested in the estimation of the underlying curve which produces the observed measurements.

Recently, quantile regression methods have become increasingly popular in many applications in longitudinal studies because of its useful features: (1) given predictors, the models can show the character of the entire condition of a response variable; (2) both the recent advances in computing resources and the ready availability of linear programming algorithms make the estimation easy; (3) the resulting estimated coefficients are robust; (4) quantile regression estimators may be more efficient than those from least squared in the case that the error term is non-normal.

Quantile regression can also be studied through several aspects, such as the parametric, nonparametric and semi-parametric quantile regression model. It is well known that the main concern with parametric modeling is the search for a suitable parametric model with limited number parameters which gives a reasonable fit to the data. This could be a very difficult task since there is often little a priori knowledge of the underlying mechanisms that generate the data. Fitting an incorrect regression model can be misleading. In order to overcome this fundamental difficulties, an attractive alternative is nonparametric curve estimation approach.

Condition distribution is a vital ingredient for quantile regression. Yu and Jones(1998) and Hall *et al.*(1999) have recently considered several methods for estimating conditional distribution. In this article, we employ the *local linear double-kernel smoothing* method proposed by Yu and Jones (1998). Specifically, suppose that $\{(X_1, Y_1), \cdots, (X_n, Y_n)\}$ is a set of independent observations from some underlying distribution $F(x, y)$ with density $f(x, y)$, and the concerned centers of responses Y_i's are considered to be realizations from the condition

$F(y|x)$ or density $f(y|x)$ of Y given $X = x$. Define $\widehat{F}_{h_1,h_2}(y|x) = \widehat{a}$. where

$$(\widehat{a}, \widehat{b}) = \operatorname{argmin} \sum_i \left(\Omega\left(\frac{y - Y_i}{h_2}\right) - a - b(X_i - x) \right)^2 \times K\left(\frac{x - X_i}{h_1}\right),$$

where h_1 and h_2 are the bandwidth in the x and y directions respectively. The functions K and Ω are two kernel functions.

3.2 Bandwidth selection

The important issue with the kernel fitting approach is the bandwidth selection. There are several different ways to select the bandwidth in the x direction. Here one rule for it simply modifies the bandwidth h_{mean} that would be used for mean regression and can be implemented as follows:

(1) Employing the Ruppert, Sheater and Wand (1995) technique to obtain h_{mean}. The technique is based on the asymptotic mean square error (AMSE) together with the 'plug-in' rule to replace any unknown quantity in the AMSE by its estimator.

(2) Calculate $h_p = h_{mean}\left[\frac{p(1-p)}{\phi\{\Phi^{-1}(p)\}^2}\right]$, where ϕ and Φ are the standard normal density and distribution functions.

Similarly, from minimizing the AMSE of estimator over the bandwidth b_p in the y direction, the b_p can be chosen according to

$$\frac{b_p h_p^3}{b_{1/2} h_{1/2}^3} = \frac{\sqrt{2\pi}\phi\left(\Phi^{-1}(p)\right)}{2\left\{(1-p)I(p \geq 1/2) + pI(p < 1/2)\right\}},$$

where $b_{1/2}$ taken to be $h_{1/2}$ and $I(\cdot)$ is an ordinary indicator function. For further details see Yu and Jones (1998).

4. Science Achievement

4.1 Descriptive statistics

Descriptive Statistics of Science Achievement are presented in Table 1. Notice that the magnitude of average science scale score results increases monotonically from grades 7 to 11, but decreases in Grade 12. It is also notable that the average science score passed 60 score only in Grade 11, that is 61.44. Generally speaking, the students perform badly in science during high school (from grade 7 to 12).

Table 1 clearly reveals the tendency of the average scores from Grade 7 to Grade 12. The significant change in average score results between grades occurred at grade 8, where there were almost 6 points increase in students' average score.

That is from 52.28 to 57.81. The difference between the maximum and minimum values of all the six grades' average score is 61.44-50.41=11.03.

Table 1: Descriptive statistics of science achievement

Grade	Number	Minimum	Maximum	Mean	Std.
Grade 7	3077	26.62	89.99	50.41	10.22
Grade 8	2742	16.85	87.41	52.28	12.77
Grade 9	2440	24.80	96.62	57.81	12.50
Grade 10	2250	17.97	97.23	59.12	14.51
Grade 11	1838	18.09	99.84	61.44	15.60
Grade 12	1485	13.23	103.24	59.84	19.07

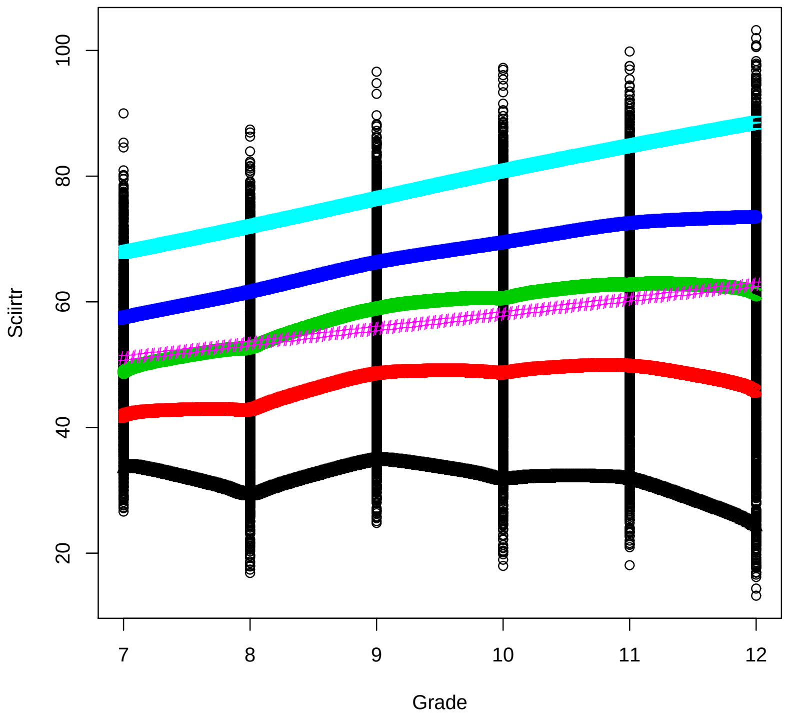

Figure 1: The science achievements of American youth of seventh- and twelfth-grade as a function of grades they were in. Five quantile curves with $p = 0.05$, 0.25, 0.5, 0.75, and 0.95.

4.2 Double-kernel regression quantile curve of science achievement

Recently, Quantile regression has been used as a standard analysis tool for modelling the performance of students on standardized exams as a function of socio-economic characteristic like their family background factors and policy variables like class size, school expenditure, and teacher qualifications. A simple reason for this is that the effect of the covariate can be very different for high- and low-level of covariates and that potentially different solutions at distinct quantiles may be interpreted as differences in the response of the dependent variable to changes in the regressors at various points in the conditional distribution of the dependent variable. Here the interesting covariate is time, i.e., the Grade.

Figure 1 depicts five quantile curves with $p = 0.05, 0.25, 0.5, 0.75$ and 0.95. Superimposed on the plot is a straight line with filled dots (marked by "#") representing the ordinary least squared estimate of the mean effect. From the figure, we see somewhat different trends over grade in the five quantiles. Note that there is nonlinearity and some heteroscedasticity in these data set. In fact, quantile plots can give a quick impression of the location, spread and shape of Y conditional on given $X = x$. Specifically, suppose that $Y = m(X) + \epsilon$. If the error term ϵ is with homoscedasticity, all regression quantile curves will be parallel. In another word, the imparalleled regression quantile curves indicate heteroscedasticity.

Furthermore, we can see that there is a much steeper increase in science achievement at higher quantiles, such as 0.75 and 0.95 quantiles. In fact, very little increases occurred at the intermediate part of the distribution, such as median. However, there is a marked decrease in the science achievement across all the lower part of the distribution, such as 0.05 quantiles. It is clear that between 1987 and 1992 there was an increase in the number of students of seventh- to twelfth-grade in the high science score as well as a decrease in the number of students in the low science score. Based on this there is a conclusion that " the good got better and the bad got worse". The result similar to that of Buchinsky (1998, 2001), Chaudhuir and Samarov (1997) and Bailar (1991) while analyzing economics phenomena with conclusion "the rich got richer and the poor got poorer."

Then turn to the OLS results. The OLS estimates suggest that the effect is linear with the intercept 34.628 and slope 2.345. The linear effect is an essentially negligible quadratic effects for some quantiles such as 0.25 and 0.50 quantiles. However, the quantile regression estimates give a very different picture. Generally speaking, ordinary least squares regression underestimates the magnitude of these effects at higher quantiles such as 0.75 and 0.95 quantiles, and overestimates the magnitude of these effects at lower quantiles such as 0.05 and 0.25.

4.3 With a first order auto-regressive point of view: today and yesterday

It is important and interesting for us to have a true understanding of the students' past. What is the tendency for the student's science achievement over the six grades? Why is the history of the past score of student important?

Figure 2 is a scatter plot of 6 years of science achievement data; it is plotted against last year's science achievement with a simple autoregressive point of view. The phenomenon here is that there is a strong tendency for data to cluster almost along the 45 degree line, which can be interpreted that this year's science achievement is more or less close to that of the last year at higher quantiles.

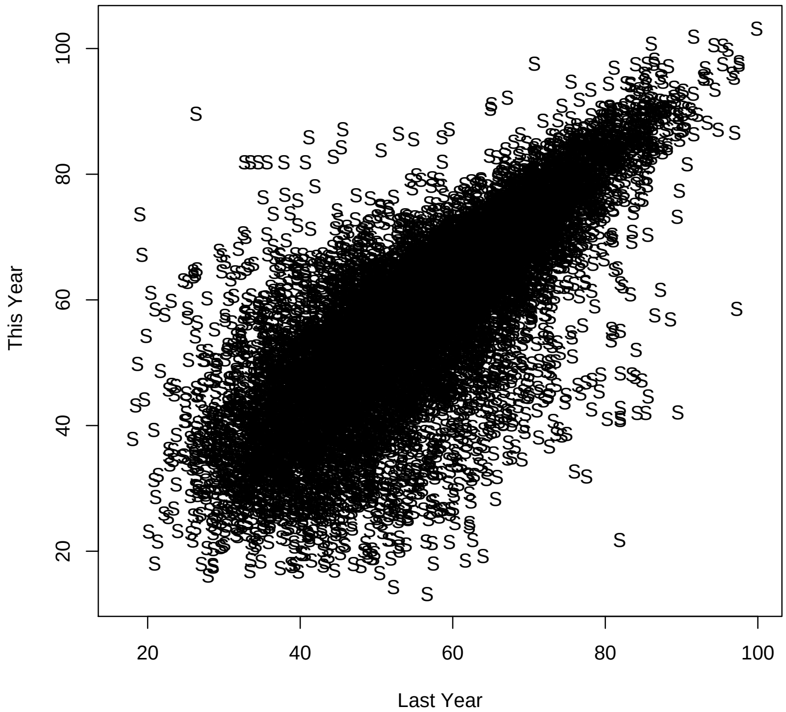

Figure 2: A scatter plot illustrates the data of 6 years of sciences achievements of seventh- and twelfth-grade American students from 1987 to 1992. The data is almost scattered around the 45 degree lines implying that this year's science achievement is roughly close to that of the last year at higher quantiles.

Figure 3 presents several estimated quantile regression curves with $p = 0.025$, 0.075, 0.125, 0.175, 0.225, 0.275, 0.325, 0.375, 0.425, 0.475, 0.525, 0.575, 0.625, 0.675, 0.725, 0.775, 0.825, 0.875, 0.925 and 0.975.

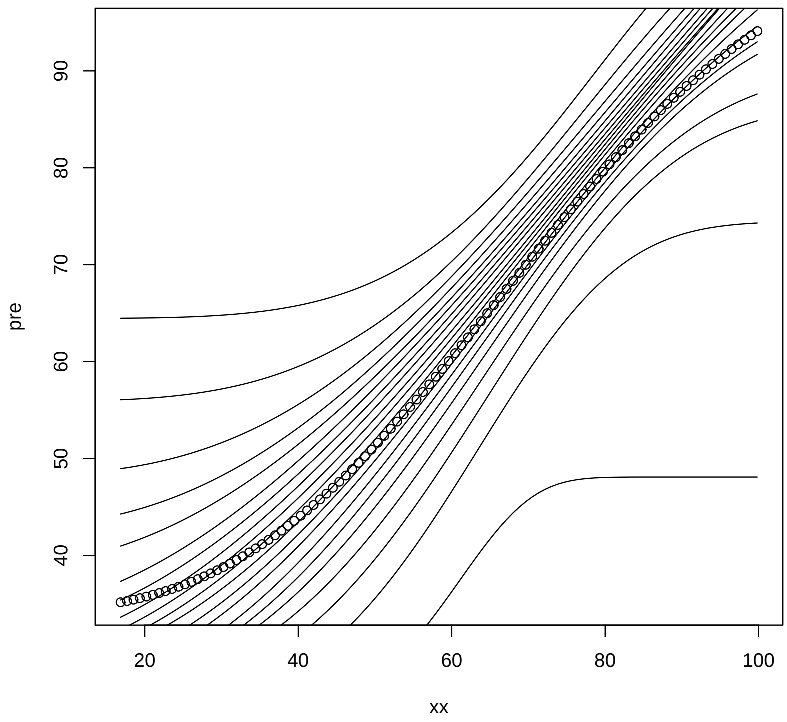

Figure 3: The 'xx' indicates last year's science achievement and 'pre' indicates the predicted quantile values. The solid lines on the scatter plot are estimated conditional quantile functions with quantiles $p = 0.025$, 0.075, 0.125, 0.175, 0.225, 0.275, 0.325, 0.375, 0.425, 0.475, 0.525, 0.575, 0.625, 0.675, 0.725, 0.775, 0.825, 0.875, 0.925 and 0.975. The dotted line represents a classical least squared regression.

From the plot, we may conclude that under the low-science-score conditions, for example, failed to pass examinations (less than 60), these curves disperse sharply. It is also interesting to mention that if a student failed to pass the examination last year, he could hardly pass the examination this year.

However under the high-science-score conditions (greater than 60), all these quantile curves bunch tightly around the 45 degree line with a number of exceptions.

In short, as for the science achievement, if a student performs well in the last year, his score appears to have two tendencies: one is roughly around that of last year with very large quantile ($> 95\%$), and the other drops out with a very small quantile ($< 5\%$).

In addition, the tendency of the classical least squared regression curve presented by the dotted line reveals that there are several unusual points with high science achievements of last year and low science achievements this year. These points have strong effects exerted on the least squared fit. One consequence of this non-robustness is that the classical least squares regression provides a rather poor estimate of the conditional mean for the worst students in the sample.

5. Mathematics Achievement

5.1 Descriptive statistics

Descriptive statistics of mathematics achievements are given in Table 2. Note that the magnitude of average mathematics scale score results increases in a straight line from grades 7 to 11, but decreases in Grade 12. Obviously, the senior high school students do better than the junior high school students in mathematics score. And all the average mathematics scores for 3 years passed 60 score in the senior high school but failed to pass 60 score in the junior high school. Table 2 clearly reveals the tendency of the average scores from Grade 7 to Grade 12. The maximum and minimum values of all the six years' average mathematics achievements are 64.76 of grade 11 and 50.40 of grade 7.

Table 2: Descriptive statistics of mathematics achievement

Grade	Number	Minimum	Maximum	Mean	Std.
Grade 7	3065	27.56	86.92	50.40	10.22
Grade 8	2749	23.49	92.58	52.92	11.75
Grade 9	2435	22.87	98.59	57.36	14.01
Grade 10	2264	23.18	101.35	61.59	16.01
Grade 11	1832	25.48	104.70	64.76	17.26
Grade 12	1467	23.05	106.90	64.25	19.07

5.2 Double-kernel regression quantile curves

Figure 4 presents five quantile curves with $p = 0.05, 0.25, 0.5, 0.75$ and 0.95. The only straight line (marked by "*") represents the ordinary least squares estimated (OLS) of the mean effects. This plot seems support the saying "the good got better and the bad got worse".

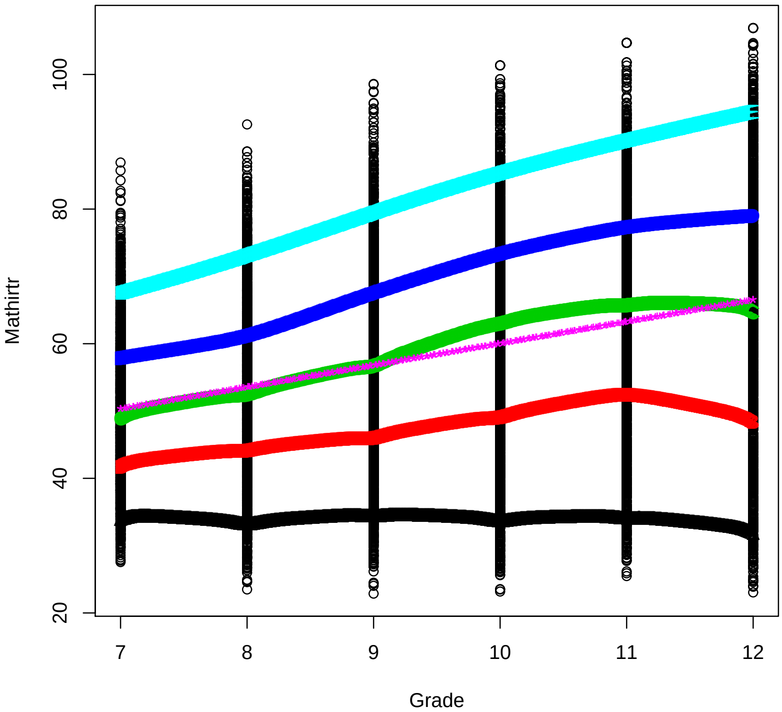

Figure 4: Five quantile curves with $p = 0.05$, 0.25, 0.5, 0.75, and 0.95 for the mathematics achievements of American youth of seventh- and twelfth-grade as a function of their grades.

Specifically, notice that for 0.05-quantile the magnitude of the estimates decreases monotonically when moving from grade 7 to grade 12. For the intermediate parts, such as 0.25 and 0.5 quantiles, the magnitude of the estimates increases monotonically from grade 7 to grade 11 but decreases monotonically in the last year grade 12. Clearly, for higher quantiles, such as 0.75 and 0.95, the magnitude of the estimates increases monotonically when moving from grade 7 to grade 12.

The OLS estimates suggest an essentially negligible quadratic effects for some quantiles such as 0.25 and 0.50 quantiles. However, the quantile regression estimates give a very different picture. Generally speaking, ordinary least squares underestimates the magnitude of these effects at higher quantiles such as 0.75 and 0.95 quantiles, and overestimates the magnitude of these effects at lower quantiles such as 0.05 and 0.25.

5.3 Analysis of the relationship between today and yesterday

For the mathematics achievements data of 6 years, a scatter plot of this year's mathematics achievements against last year's shows that there is a tendency for data points to cluster almost along the 45 degree line, which implies that this year's mathematics achievement is close to that of last year.

Figure 5 presents several estimated quantile regression curves with $p = 0.025$, 0.075, 0.125, 0.175, 0.225, 0.275, 0.325, 0.375, 0.425, 0.475, 0.525, 0.575, 0.625, 0.675, 0.725, 0.775, 0.825, 0.875, 0.925 and 0.975. Dotted line is the least squared fit. The plot reveals that almost all these quantile curves are bunched tightly around the 45 degree line at higher quantiles with a few exceptions.

The conditional median and mean fits are almost the same in this data. This may be explained by the symmetry of the conditional density and a lack of extreme points in the data.

The shape of Figure 5 implies that this year's mathematics achievements are similar to that of last year. The consistence in learning mathematics may be explained by the nature of mathematics.

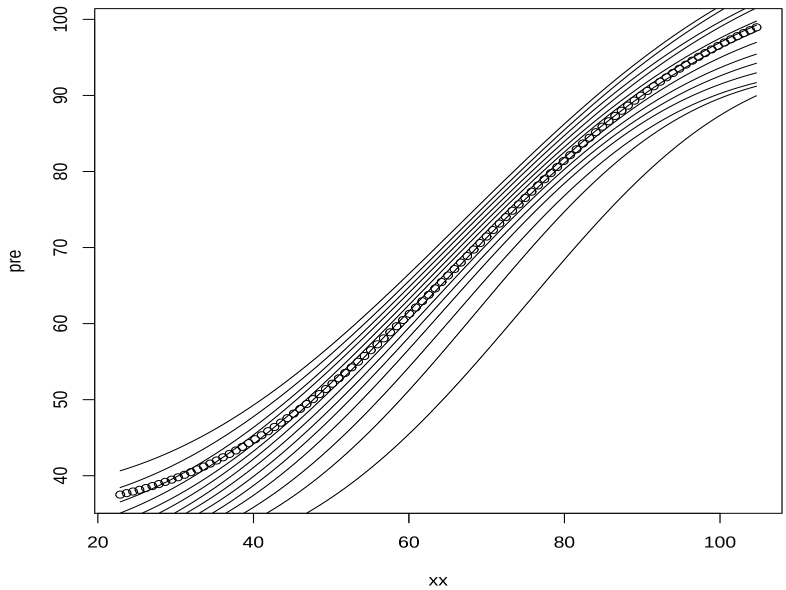

Figure 5: The 'xx' indicates last year's mathematics achievement and 'pre' indicates the predicted quantile values. The solid lines on the scatter plot are estimated conditional quantile functions with quantiles $p = 0.025$, 0.075, 0.125, 0.175, 0.225, 0.275, 0.325, 0.375, 0.425, 0.475, 0.525, 0.575, 0.625, 0.675, 0.725, 0.775, 0.825, 0.875, 0.925 and 0.975.

6. The Relationship between Science Achievement and Mathematics Achievement

It is well known that mathematics is an instrumental subject for other subjects, especially sciences. Hence it is necessary to study the effect of mathematics achievement on science achievement. In this section, we present linear quantile regression estimates of the effect of mathematics inputs on changes in science achievements.

Table 3 shows the comparison of quantile regression and OLS results conditional on mathematics achievement. The estimated standard errors are reported in parentheses. We first note that the effects of mathematics achievement on science achievement are significantly positive values and only the intercept at quantile 5% is a insignificantly negative value, -1.19, which can be interpreted as the estimated conditional quantile function of the science achievement of a student whose mathematics achievement is zero. That is to say, without mathematics background, the effect on learning science is nearly zero.

From the quantile regression results, we notice the following differences from OLS regression. The effects for the factor *Mathematics* are all positive significant across all quantiles of the conditional distribution of science achievement changes. The estimated effects are 0.74, 0.80, 0.77, 0.72 and 0.61 at 5%, 25%, 50%75% and 95% quantiles respectively. Obviously, the maximum marginal effect of *Mathematics* on science is at 25% quantile. Notice that the magnitude of the estimates decreases monotonically from a lower quantile to a higher quantile except for the 5% quantile(i.e., from 25% to 95% quantiles). Ordinary least-squares underestimates the magnitude of these effects at lower quantiles from 5% to median 50% and overestimates the magnitude of these estimates at higher quantiles from 75% to 95%.

For comparison, the differentials in science achievements between with-mathematics (MATHEMATICS=1) and without-mathematics (MATHEMATICS=0) are 0.74, 0.80, 0.77, 0.72 and 0.61 when from the lowest quantile to the highest quantile.

Table 3: Comparison of quantile regression and OLS results conditional on the controlling model

	Quantile regression results					O. L. S.
Description	5%	25%	50%	75%	95%	
Intercept	-1.19	4.54*	12.08*	20.53*	34.97*	14.01*
	(0.83)	(0.38)	(0.33)	(0.39)	(0.60)	(0.30)
Mathematics	0.74*	0.80*	0.77*	0.72*	0.61*	0.73*
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)

Selected the instrumental subject MATHEMATICS as the predictor variable, and the responsible variable is SCIENCE.

Figures 6 and 7 give a concise visual summary of the linear quantile regression results for the data. Figure 10 describes Intercept of the model and Figure 11 describes slope of Mathematics in the model. The solid line with filled dots (marked by the capital letter "I" and "E" in respective figures) represents the 5 point estimates of the effects of 'MATHEMATICS' for τ ranging from 0.05 to 0.95. In both figures, two dashed lines with filled dots, marked by the capital letter "U" and "L"respectively, represent the lower and upper confidence bounds. The area between the lower and upper confidence bound is a 90% pointwise confidence band. The horizontal dotted line with filled dots marked star "*" presents the ordinary least squared estimates of the mean effects.

The intercept of the model may be interpreted as the estimated conditional quantile function of the science distribution for a student whose mathematics achievement is null. From Figure 6, we can find that the estimated conditional quantile function is monotonically increasing.

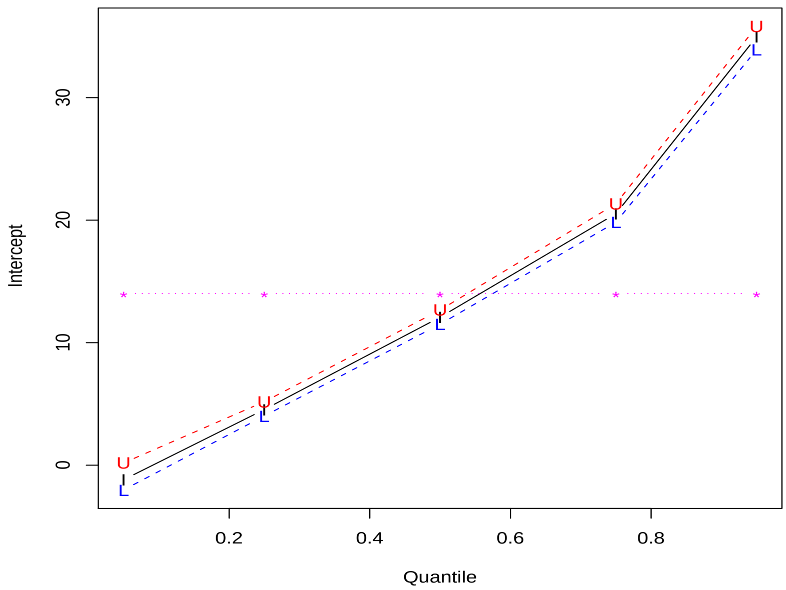

Figure 6: Intercept of the model. The solid line with filled dots (marked by the capital letter "I") represents the 5 point estimates of the effects of 'MATHEMATICS' for τ ranging from 0.05 to 0.95. Two dashed lines with filled dots marked by the capital letter "U" and "L" respectively are the lower and upper confidence bounds. The horizontal dotted line with filled dots marked star "*" presents the ordinary least squared estimates of the mean effects.

The differential between the maximum and the minimum values is 34.97-(-1.19)= 36.16 at 95% and 5% quantiles respectively. The slopes of MATHEMATICS are in fact marginal effects. Figure 7 reveals the marginal effects decrease monotonically from lower quantiles to higher quantiles (except for 5% quantile). The differential between the maximum and the minimum values is 0.80-0.61= 0.19 at 25% and 95% quantiles respectively.

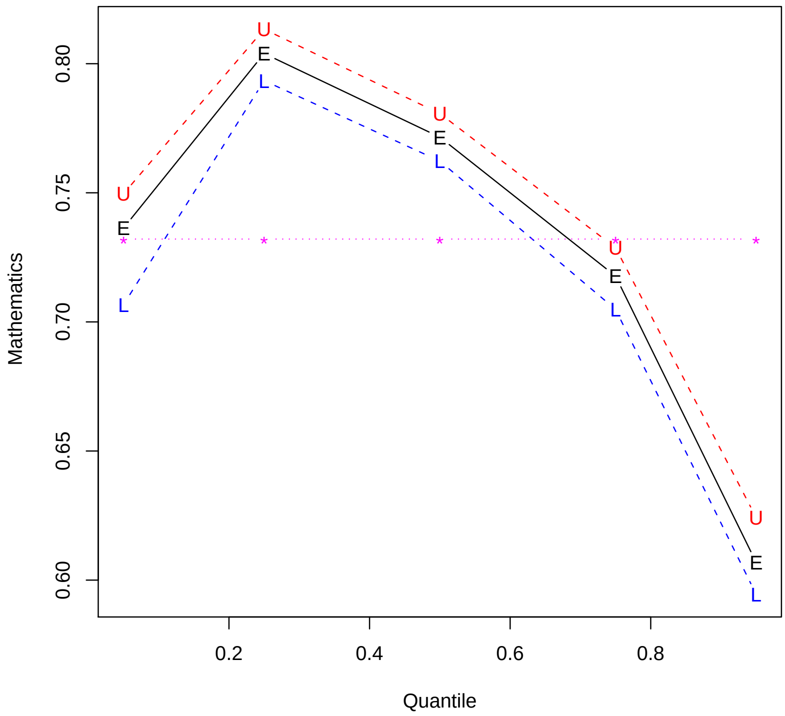

Figure 7: Slope of Mathematics in the model. The solid line with filled dots (marked by the capital letter "E") represents the 5 point estimates of the effects of 'MATHEMATICS' for τ ranging from 0.05 to 0.95. Two dashed lines with filled dots, marked by the capital letter "U" and "L" respectively are the lower and upper confidence bounds. The area between the lower and upper confidence bound is a 90% pointwise confidence band. The horizontal dotted line with filled dots marked star "*" presents the ordinary least squared estimates of the mean effects.

Acknowledgement

The authors gratefully acknowledge Canadian research chair Xin Ma for his real data, to Linda M. McDonald, Professor and Chair, for her help. Part of the work was performed while Maozai Tian was postdoctoral fellow at Canadian Center for Advanced Studies of National Databases, University of Alberta, Canada. This research is partially supported by National Social Sciences grant of China, No. 07BTJ002 (2007-2009), NSFC grant, No. 10431010 (2005-2008), Education Ministry Key Project, No. 05JJD910001(2005-2008), and Funds of Renmin University of China, No. 2006031611.

References

Coleman, J. (1966). Equality of educational opportunity. Report for US Department of Health, Education, and Welfare, Office of Education, Washington D. C.

Ehrenberg, R. G., Brewer, D. J. (1994). Do school and teacher characteristics matter? Evidence from high school and beyond. *Economics of Education Review* **13**, 1-17.

Ehrenberg, R. G., Brewer, D. J. (1995). Did teachers' verbal ability and race matter in the 1960s? Coleman revisited. *Economics of Education Review* **14**, 1-21.

Eide, E. and SHowalter, M. H. (1998). The effect of school quality on student performance: A quantile regression approach. *Economics Letters* **58**, 345-350.

Gallagher, Shelagh A. (1994). Middle school predictors of science achievement. *Journal for Research in Science Teaching* **31**, 721-734.

Hall,P., Wolff, R. C. L. and Yao, Q. (1999) Methods for estimating a conditional distribution. *J. Am. Statist. Assoc.* **94**, 154-163.

Hanushek, E. (1986). The economics of schooling: Production and efficiency in public schools. *Journal of Economic Literature* **24**, 1141-1177.

Hanushek, E. (1996). School resources and students performance. In *Does money matter? The effect of school resources on student achievement and adult success* (Edited by Burtless, G.), 43-73. Brookings Institution, Washington, D.C.

Hoffer, T. B. (1992). Middle School Ability Grouping and Student Achievement in Science and Mathematics. *Educational Evaluation and Policy Analysis* **14**, 205-227.

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

Levin, J. (2001). For whom the reductions count: A quantile regression analysis of class size and peer effects on scholastic achievement. *Empirical Economics* **26**, 221-246.

Reynolds, Arthur J. (1991). Note on Adolescents' Time-Use and Scientific Literacy. *Psychological Reports* **68**, 63-70.

Reynolds, Arthur Herbert J. Walberg. (1992). A Structural Model of Science Outcomes: An Extension to High School. *Journal of Educational Psychology* **84**, 371-82.

Reynolds, Arthur H. and Walberg J. (1992). A Structural Model of Science Achievement. *Journal of Educational Psychology.* **83** 97-107.

Ruppert,D., Sheather, S. J. and Wand, M. P . (1995). An effective bandwidth selector for local least squared regression. *J. Am. Statist. Assoc.* **90**, 1257-1270.

Tian, M. Z. (2006) A Quantile Regression Analysis of Family Background Factor Effects on Mathematical Achievement. *Journal of Data Science* **4**, 461-478.

Yu, K. and Jones, M. C. (1998) Local linear regression quantile estimation. *J. Americaan Statist. Ass.* **93**, 228-238.

Received December 9, 2006; accepted August 14, 2007.

Maozai Tian
School of Statistics
Renmin University of China
Beijing, 100872, China
mztian@ruc.edu.cn

Xizhi Wu
School of Statistics
Renmin University of China
Beijing, 100872, China
xzwu2003@ruc.edu.cn

Yuan Li
School of Statistics
Renmin University of China
Beijing, 100872, China
i222@ruc.edu.cn

Pengpeng Zhou
School of Statistics
Renmin University of China
Beijing, 100872, China
peng.zhou@ruc.edu.cn