Longitudinal Study of the Positive Academic Emotion Effects on Children's Mathematics and Science Achievements-Using Nonparametric Quantile Regression

Yu Wang, Lingyu Kong, Dianjun Lu*

School of Mathematics and Statistics of Qinghai Normal University, Xining, China

*Corresponding author

Keywords: Quantile regression; academic emotion; mathematics and science achievements; nonparametric quantile regression; double-kernel approach

Abstract: This article attempts to address the interesting and important research topic and to give an in-depth longitudinal study of Xining high school students using a so called "double-kernel" nonparametric quantile regression approach. The conclusion is that the process-science-push's and process-math-push's points is 4, the result-science-push's point is 3, etc. All the findings are useful for parents understand students' learning status, especially for educators make educational plans.

1. Introduction

It is well known that quantile regression, as introduced by Koenker and Bassett[1], has found many applications in longitudinal studies because of its useful features: (1)Given predictors, it characterizes the entire condition distribution 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 squares in the case that the error term is non-normal. Details referred to Buchinsky[2].

Taking full advantage of the quantile regression, for example, Eide and Showalter[3] used quantile regression to estimate whether the relation between school and performance on the standardized tests differs at different points under the condition of 'test score gains'. Following this research method, this paper will study the impact of academic emotion on students' academic performance when it is at what value.

In the existing literature[4-9] on the correlation between academic emotion and academic achievement, academic emotion is divided into four dimensions: positive high arousal, positive low arousal, negative high arousal and negative low arousal. On the basis of reading literatures, after repeated discussions with several mathematics education experts and front-line teachers who have been teaching mathematics for many years, this paper finally determines that the dimension of mathematics academic emotion should be divided according to students' mathematics learning process and mathematics learning results, which is closer to the actual situation of students. Therefore, we propose to divide mathematics learning process and mathematics learning results according to the degree of pleasure. It includes four dimensions: positive emotion in the process of mathematics learning, negative emotion in the process of mathematics learning, positive emotion in the result of mathematics learning and negative emotion in the result of mathematics learning.

In fact, students' academic performance is greatly influenced by two conditions: subjective and objective conditions. Academic emotion plays a very important role under subjective conditions. This study examined the effects of positive academic emotions on students' math and science achievements.

In this paper we employ the local linear double-kernel smoothing method proposed by Yu and Jones[10]. The method is an attractive nonparametric quantile regression procedure. A nonparametric quantile regression model is used to analyze the data of three grades of senior high school students in Xining City from 2015 to 2018.

  DOI: 10.25236/icrtpe.2019.173

2. Real Data and Descriptive Analysis

The data, which represents science and mathematics achievement samples of senior 1 to senior 3 students from 2015 to 2018, are taken from Xining City, Qinghai Province. About 55 senior one students were randomly selected in each of the 9 schools and the total sample size was 495 students. These students were followed for three years from senior 1 to senior 3, writing mathematics and science achievement tests and completing student questionnaires annually. With a focus on mathematics and science education, the information from student's academic emotion was also included in the study. The descriptions for the variables used in the empirical analysis are as follows: MATHIRTR denotes mathematics achievement, SCIIRTR science achievement, PMHPH positive emotions in learning process's push on mathematics achievement, PSCPH positive emotions in learning process's push on science achievement, RMHPH positive emotions in learning result's push on mathematics achievement, RSCPH positive emotions in learning result's push on science achievement. The descriptive statistics for the variables used in the empirical analysis are listed in Figure 1.

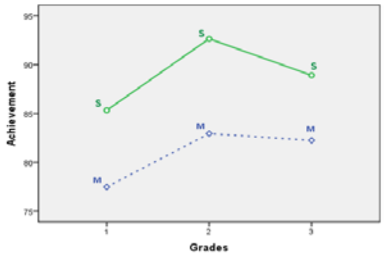

Fig. 1 Mathematics and science achievements for three grades

Figure 1 The solid line with filled dots marked by the capital letter "S" and the dashed line with filled dots marked by the capital letter "M" represent 3 average science achievements and mathematics achievements of senior 1 to senior 3 students from senior 1 to senior 3 i.e., from 2015 to 2018, respectively. Note that students seem to do better in science than in mathematics.

2.1 Mathematics and science achievements

Descriptive statistics of mathematics and science achievements and the corresponding plots are shown in Figure 1. The figure clearly reveals the tendency of the average scores from senior 1 to senior 3. The significant changes in average achievements occurred at senior 2 for both courses: mathematics and science. As can be seen from the slopes that change sharply. From this point of view, senior 2 is crucial to the high school students in their mathematics and science achievements.

2.2 The positive emotions in learning process and result's push

Positive Academic Emotion refers to all kinds of positive emotional experiences related to students' academic activities in the process of teaching or learning, which can be divided into positive emotions in the learning process and positive emotions in the learning result. It is often said that children are easily influenced by their emotions. Here, the positive emotions in learning process and result's push are shown in Figure 2 respectively. Specifically, the solid line with filled dots marked by the capital letter "S" and the dashed line with filled dots marked by the capital letter "M" represent positive emotion's pushes to student's mathematics and science achievement, respectively.

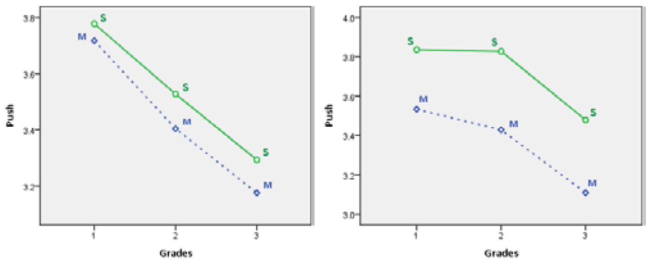

Fig. 2 Positive emotions in learning process and result's push to math and sci achievement

As can be seen from Figure 2, It is clear that PMHPH<PSCPH (left part) and RMHPH<RSCPH (right part) consistently, for the whole high school and that all the two push-curves decrease monotonically across all the three grades in high school. The findings imply that among the two kinds of the positive emotions in learning process and results' push, the science push is the most powerful, respectively. And the findings also show that the positive emotions in learning process and results' push on the high school student is becoming weaker and weaker as he grows up.

From the comparison of the left and right parts, we can see that the positive emotions in the learning process of each grade are higher than the positive emotions in the learning result.

3. Methodology

Quantile regression, as introduced by Koenker and Bassett[1], is gradually developing into a comprehensive approach to the statistical analysis of linear and non-linear response model. There are at least four equivalent mathematical definitions of quantile regression:

1)Definition based on the conditional quantile function

Let $q_p(x)$ be the *p-th* quantile of the dependent variable *Y* given *X=x*. In this case, $q_p(x)$ can be found by solving: $F\left(q_p(x)\middle|x\right)=P\left(Y\le q_p(x)\middle|X=x\right)=p$, where *F* is the cumulative distribution of *Y*;

2)Definition based on the quantile regression model [11]: $Y=x^T\beta+\varepsilon$, where the error term ε is assumed to satisfy $\text{Quantile}_p(\varepsilon)=0$. In standard linear regression model, the error term is assumed to be a Guassian error;

3)Definition based on a check function[1]: $\min_{\beta\in\Theta}\mathrm{E}\left\{\rho_p(\mathrm{Y}-\mathrm{X}'\beta)\middle|\mathrm{X}=x\right\}$, where $I_A(z)$ the usual indicator function of the set A, Θ is a parametric space for β, $\rho_p(z)=pzI_{[0,\infty)}(z)-(1-p)I_{(-\infty,0)}(z)$ is called check function;

4)Definition based on asymmetric Laplace density[12], Let $f(\varepsilon)$is the probability density of the model error ε ,we have $f(\varepsilon)\approx\exp\left\{-\sum_{i=1}^{n}\rho_p\left(y_i-x_i^T\beta\right)\right\}$.

In this paper, we use the definition based on Koenker and Bassett's definition, in which the dependent variable Y is the mathematics and science achievement and the independent variable X is the external expressure variables mentioned before.

3.1 Double kernel approach

We have seen that the condition distribution is a vital ingredient for quantile regression. However, there are some problems associated with the existing kernel-weighting estimation of the the condition distribution. For example, some estimators of the condition distribution are not a distribution function, see [13] for example, the others are dissatisfactory for the quantile curve based on these estimators may cross one another, which is of course absurd, see Hall *et al.*[14]for example. To avoid these problems, we employ the "double kernel" approach of Yu and Jones[10] in this paper.

We know that the condition distribution is a vital ingredient for quantile regression. Yu and Jones[10] and Hall *et al.*[14]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[10]. 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 interest centers on the responses Y_i considered to be realizations from the condition distribution $F(y|x)$ or density $f(y|x)$ of Y given $X=x$. Define $\hat{F}_{h_1,h_2}(y|x)=\hat{a}$, where $\left(\hat{\boldsymbol{a}},\hat{\boldsymbol{b}}\right)=\mathbf{arg\,min}\sum_i\left(\Omega\left(\frac{\boldsymbol{y}-\boldsymbol{Y}_i}{\boldsymbol{h}_2}\right)-\boldsymbol{a}-\boldsymbol{b}(\boldsymbol{X}_i-\boldsymbol{x})\right)^2\times\boldsymbol{K}\left(\frac{\boldsymbol{x}-\boldsymbol{X}_i}{\boldsymbol{h}_1}\right)$, 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.

Define $\hat{q}_p(x)$ to satisfy $\hat{F}_{h_1,h_2}\left(\hat{q}_p(x)\middle|x\right)=p$ so that $\hat{q}_p(x)=\hat{F}^{-1}_{h_1,h_2}(p|x)$.

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's[15] technique to obtain h_{mean}. The technique is based on the asymptotic mean square error (AMSE) together with the "plugin" rule to replace any unknown quantity in the AMSE by its estimator.

(2) Calculate $h_p = h_{mean}\left\lfloor p(1-p)/\phi\left\{\Phi^{-1}(p)\right\}^2\right\rfloor$, 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 $\frac{\boldsymbol{b_p h_p^3}}{\boldsymbol{b_{1/2} h_{1/2}^3}} = \frac{\sqrt{2\pi}\varphi\left(\Phi^{-1}(\boldsymbol{p})\right)}{2\left\{(1-\boldsymbol{p})\boldsymbol{I}(\boldsymbol{p}\ge 1/2)+\boldsymbol{pI}(\boldsymbol{p}<1/2)\right\}}$, where $b_{1/2}$ is taken to be $h_{1/2}$ and $I(\cdot)$ is a ordinary indicator function. For further details see Yu and Jones[10] .

4. Statistical Analysis

In this section, we analyze the longitudinal data set Science. The interesting covariates are PSCPH, RSCPH and PMHPH、RMHPH. It is important to investigate the tendency of student's achievements of different quantiles in the conditional distribution in the case of various positive emotions.

4.1 Double-Kernel Quantile Regression on Science Achievements

The template is designed so that author affiliations are not repeated each time for multiple authors of the same affiliation. Please keep your affiliations as succinct as possible (for example, do not differentiate among departments of the same organization). This template was designed for two affiliations.

1) The positive emotions in the learning process's push on children's science achievements

First, we consider the relationship between the science achievement and the positive emotions in the learning process's push on science achievement. The dependent variable is the science achievements and covariate is the positive emotions in the learning process's push on science. The double-kernel regression results which were estimated at five different quantiles (5%, 25%, 50%, 75% and 95%) are reported in Table 1.

Table 1 Double-kernel regression results for SCIIRTR-PSCPH

Quantiles	**Learning process's push on science achievement**				
	1	***2***	***3***	***4***	***5***
5%	37.84	39.46	40.26	42.32	40.39
25%	49.25	51.81	53.23	55.88	54.81
50%	63.22	64.15	66.46	67.41	66.29
75%	76.29	78.99	80.99	83.24	81.15
95%	95.26	96.34	98.38	102.90	100.15

Table 1 depicts the double-kernel quantile regression results which were made at five different quantiles (5%, 25%, 50%, 75% and 95%). We can see that the magnitude of the estimates for all quantiles increases monotonically when moving from a lower value of process-science-push to a higher value of process-science-push, i.e., from 1 to 4, but decreases in 5, the highest push from process on science.

2) The positive emotions in the learning result's push on children's science achievements

Secondly, we consider the relationship between the science achievement and the positive emotions in the learning result's push on science push. The dependent variable is the science achievements and covariate is now the positive emotions in the learning result's push on science. The double-kernel regression results are reported in Table 2.

Table 2 Double-kernel regression results for SCIIRTR-RSCPH

Quantiles	**Learning result's push on science achievement**				
	1	***2***	***3***	***4***	***5***
5%	37.83	38.73	40.09	39.88	38.64
25%	51.68	53.19	55.95	54.70	53.13
50%	64.57	65.07	67.49	66.85	64.91
75%	87.45	88.79	90.47	89.56	88.75
95%	102.33	103.84	106.94	102.83	102.54

Table 2 depicts the double-kernel quantile regression results which were made at five different quantiles (5%, 25%, 50%, 75% and 95%). It is clear that the magnitude of the estimates for all quantiles increases consistently when moving from the starting point of the result-science-push, i.e., 1 to a higher value of 3, but decreases monotonically when moving from 3 to the highest push of result on science, i.e., 5. That is to say, the value 3 is a maximum point at which all effects of result-science-push of different quantiles are maximum.

4.2 Double-Kernel Quantile Regression on Mathematics Achievements

1) The positive emotions in the learning process's push on children's mathematics achievements

First, we consider the relationship between the mathematics achievement and the positive emotions in the learning process's push on mathematics achievement. The dependent variable is the mathematics achievements and covariate is the positive emotions in the learning process's push on mathematics. The double-kernel regression results are reported in Table 3.

Table 3 Double-kernel regression results for MATHIRTR-PSCPH

Quantiles	**Learning process's push on mathematics achievement**				
	1	***2***	***3***	***4***	***5***
5%	31.32	32.41	34.53	37.61	35.79
25%	46.98	47.98	49.98	51.49	50.47
50%	61.49	62.48	64.97	67.98	65.96
75%	75.99	77.00	79.98	82.99	80.00
95%	92.99	94.98	95.99	97.98	96.96

Table 3 depicts the double-kernel quantile regression results which were made at five different quantiles. It is clearly that the process-math-push value 4 is the turning point.

2) The positive emotions in the learning result's push on children's mathematics achievements

Secondly, we consider the relationship between the mathematics achievement and the positive emotions in the learning result's push on mathematics push. The double-kernel regression results are reported in Table 4.

Table 4 Double-kernel regression results for MATHIRTR-RSCPH

Quantiles	**Learning result's push on mathematics achievement**				
	1	***2***	***3***	***4***	***5***
5%	35.48	36.99	37.98	39.95	38.98
25%	47.49	48.50	51.99	50.45	49.49
50%	62.98	63.99	66.96	65.99	64.00
75%	86.99	87.99	88.97	88.93	87.96
95%	98.59	99.99	102.99	100.49	99.29

Table 4 depicts the double-kernel quantile regression results which were made at five different quantiles (5%, 25%, 50%, 75% and 95%). As for the estimated curve with quantile 0.05, the lowest one in our analysis, the optimal value of the result-math-push is 4, but for the rest four quantile regression curves, the optimal value is 3.

5. Conclusion

The value 4 is a maximum point at which all effects of process-science-push of different quantiles are maximum. We may call the maximum point the optimal process-science-push point. Parents should pay attention to the limit value of process-science-push, which means that parents should always care about children's emotions. Excessive positive emotions are not suitable. The value of process-science-push should be no more than 4. The value 3 is a maximum point at which all effects of result-science-push of different quantiles are maximum. In fact, the result’s science push is moderate, the students' performance is better. May be that is the so called *modesty makes progress*. The value 4 is a maximum point at which all effects of process-math-push of different quantiles are maximum. As for the estimated curve with quantile 0.05, the lowest one in our analysis, the optimal value of the result-math-push is 4, but for the rest four quantile regression curves, the optimal value is 3. The finding implies that the student whose mathematics achievement is at the bottom of his class needs more result’s math-push if he wants to get the maximum mathematical achievement than those who do better in their mathematics.

Acknowledgement

This research was financially supported by Applied Basic Research Project of Qinghai province (2019-ZJ-7099). And it was partly supported by National Social Science Foundation Project (19BYY223).

References

[1] Koenker R, Bassett G. Regression quantiles[J]. Econometrica, 1978, 46: 33–50.

[2] Buchinsky M. Recent advances in quantile regression model[J]. J. Human Res., 1998, 38: 88–126.

[3] Eide E, Showalter M. The effect of school quality on student performance: A quantile regression approach[J]. Economics Letters, 1998, 58: 345–350.

[4] Pekrun R, Goetz T, Frenzel A, et al. Measuring emotions in students’ learning and performance: The Achievement Emotions Questionnaire (AEQ)[J]. Contemporary Educational Psychology, 2011, 36(1): 40-48.

[5] Pekrun R, Cusack A, Murayama K, et al. The power of anticipated feedback: Effects on students\" achievement goals and achievement emotions[J]. Learning and Instruction, 2014, 29:115-124.

[6] Goetz T, Sticca F, Pekrun R, Murayama R, Andrew J. Intraindividual relations between achievement goals and discrete achievement emotions: An experience sampling approach[J]. Learning and Instruction, 2016, 41:115-125.

[7] Pekrun R, Lichtenfeld S, Marsh H, et al. Achievement Emotions and Academic Performance: Longitudinal Models of Reciprocal Effects[J]. Child Development, 2017, 88(5): 1653-1670.

[8] Pekrun R. Emotion and Achievement During Adolescence[J]. Child Development Perspectives, 2017, 11(3):417-418.

[9] Putwain D, Becker S, Symes W, Pekrun R. Reciprocal relations between students’ academic enjoyment, boredom, and achievement over time[J]. Learning and Instruction, 2018, 54: 73-81.

[10] Yu K, Jones M. Local linear regression quantile estimation[J]. J. Am. Statist. Ass., 1998, 93: 228–238.

[11] Bailar B. Salary survey of U.S. colleges and universities offering degrees in statistics[J]. Amstat News, 1991, 182: 3–10.

[12] Yu K, Moyeed R. Bayian quantile regression[J]. Statist. Probab. Lett., 2001, 54: 437–447.

[13] Stone C. Consistent nonparametric regression[J]. The Annals of Statistics, 1977, 4: 595-645.

[14] Hall P, Wolff R, Yao Q. Methods for estimating a conditional distribution[J]. J. Am. Statist. Assoc., 1999, 94: 154–163.

[15] Ruppert D, Sheather S, Wand, M. An effective bandwidth selector for local least squares regression[J]. J. Am. Statist. Assoc., 1995, 90: 1257–1270.