Admission Test Modelling of State Islamic College in Indonesia Using Local Linear for Bivariate Longitudinal Data

Nidhomuddin¹, N Chamidah²*, A Kurniawan²

¹PhD Student, Faculty of Sciences and Technology, Airlangga University, Indonesia
²Department of Mathematics, Faculty of Sciences and Technology, Airlangga University, Indonesia

*Corresponding author’s e-mail: nur-c@fst.unair.ac.id

Abstract. The main challenge in the development of education is improving the quality of education and equitable education. Improving the quality of education is needed in order to create high-quality human resources. Equitable education carried out also still has problems with the uneven quality of education in each region. The aim of this research is to determine the effect of public senior high school status on the admission test scores, i.e., scholastic aptitude test score and Islamic test score of state Islamic colleges in Indonesia. In this research we use a local linear estimator to estimate nonparametric regression model for longitudinal data, and then apply it to the data of admission test scores of state islamic colleges in Indonesia based on the percentage of public senior high school. The result shows that the pattern of the scholastic aptitude test and Islamic test based on the percentage of public senior high school of state Islamic colleges in western Indonesia is different from that in eastern Indonesia. The result of this research can be used by the government to make policies so that the quality of state Islamic colleges will be better by increasing the admission test scores.

1. Introduction.

The main challenge in the development of education is improving the quality of education and equitable education [1]. Improving the quality of education is needed in order to create high-quality human resources [2]. Equitable education carried out also still has problems with the uneven quality of education in each region [3]. One of the important problems to improve the quality of education is the admission test because it can be used to determine the quality of college graduates [4].

Since 2014, the state Islamic colleges have carried out a joint admission test in Indonesian called “Ujian Masuk Perguruan Tinggi Keagamaan Islam Negeri (UM-PTKIN)”. There is still not many researches related to UM-PTKIN so that this research on UM-PTKIN is very important, because UM-PTKIN is a national standard to determine the quality of prospective students accepted by state Islamic colleges. Test materials of UM-PTKIN include scholastic aptitude test and Islamic test. Scholastic aptitude test gives a significantly contribution in predicting the success of the study [5]. Scholastic aptitude test as a selection tool is sometimes better than academic tests [6]. Educational Act No. 20, 2003 about the national education system confirms that education must be able to construct a religious personality or character.
The public senior high school status has an effect on students’ achievement when students were studying in college. Senior high school status has a significantly effect on the learning Introduction to Basic Mathematics of students of the Tarbiyah Faculty, the State Islamic College of Mataram [7]. Public senior high school status has an effect on students’ achievement in the Faculty of Science and Technology, the Sunan Kalijaga State Islamic College [8]. The relationship patterns between predictor variables (senior high school status) and response variables (scholastic aptitude test and Islamic test) on the admission test scores of state Islamic colleges in western Indonesia is different from that in eastern Indonesia. Therefore, a smoothing technique in nonparametric regression model especially locally modelling is more appropriate to analyse this case.

There are many researchers who have considered the smoothing techniques in nonparametric regression models. Splines smoothing for estimating nonparametric regression functions have been studied by [9-15]. In addition, [16] used local linear estimator in the bi-response nonparametric regression model for estimating median growth charts of children, [17] using local linear for estimating standard growth charts for the weight of children in east java, and [18] used local polynomial for improving of classification accuracy of cyst and tumor. The advantages of this estimator are able to estimate the function at each point such that the model closes to the real pattern, and also no need much data to estimate the model [19]. The aim of this research is to determine the effect of public senior high school status on the admission test scores (scholastic aptitude test and Islamic test) of state Islamic colleges in Indonesia. This research can be used by the government to make policies so that the quality of state Islamic colleges will be better by increasing the admission test scores.

2. Longitudinal Data.
In longitudinal data, generally observations are made on n subjects which are mutually independent of each other, where each subject is collected repeatedly in different periods of time. If \( x_{ik} \) represents the predictor variable of the \( i \)-th subject at the \( k \)-th time, and \( y_{ik} \) represents the response variable measured of the \( i \)-th subject at the \( k \)-th time, then longitudinal data is given as \( (x_{ik}, y_{ik}); i = 1,2,...,n; k = 1,2,...,m \) where \( m_i \) is the number of replication for the \( i \)-th subject, and \( n \) is the number of subjects. The longitudinal nonparametric regression model for one predictor variable is given as follows:

\[
y_{ik} = f(x_{ik}) + \epsilon_{ik}, \quad i = 1,2,...,n; \quad k = 1,2,...,m_i
\]

3. Local Linear Estimator.
Suppose that paired data of observations \((y_i, x_i)\) follows nonparametric regression model as follows:

\[
y_i = f(x_i) + \epsilon_i, \quad i = 1,2,...,n
\]  

(1)

where \( \epsilon_i \) is a random error assumed to be independent with mean zero and variance \( \sigma^2 \), and \( f \) is a regression function to be estimated. The regression function (\( f \)) in (1) is estimated by a local linear estimator that gives as follows:

\[
\hat{f}(x) = \hat{\beta}(x_0)
\]  

(2)

where \( \hat{\beta}(x_0) \) is obtained by using weighted least square (WLS) method, i.e., by minimizing the following function:

\[
Q(x_0) = (y - X(x_0)\hat{\beta}(x_0))^T K_h(x_0)(y - X(x_0)\hat{\beta}(x_0))
\]  

(3)

So, we get:

\[
\hat{\beta}(x_0) = (X^T(x_0)K_h(x_0)X(x_0))^{-1}X^T(x_0)K_h(x_0)y
\]  

(4)

Based on (2) and (4), the form of local linear estimator for \( f(x) \) can be written as follows:

\[
\hat{f}(x) = \hat{\beta}(x_0)(X^T(x_0)K_h(x_0)X(x_0))^{-1}X^T(x_0)K_h(x_0)y
\]  

(5)

2
4. Generalized Cross Validation.
Bandwidth \((h)\) is a smoothing parameter that controls the smoothness of the curve. The small bandwidth \((h)\) value will give result in a very noisy estimation of function and the variance increases. In the contrary, if the bandwidth \((h)\) value is too large, the estimated function obtained will be oversmoother, so that consequently it gives a large bias. Therefore, an optimum bandwidth is needed [20].

A criterion for \(h\) will be limited to the linear estimator class. It means that for every \(h\) there is a \((n \times n)\) - smoother matrix \(A(h)\) such that:

\[
\hat{f}(x) = A(h)y
\]  

(6)

One of the optimum bandwidth selection methods used in nonparametric regression models is generalized cross validation (GCV) method which is given as follows:

\[
GCV(h) = \frac{\text{MSE}(h)}{\left(n^{-1} \text{tr}[I - A(h)]\right)^2}
\]

(7)

where mean square error (MSE) in (7) is given as follows:

\[
\text{MSE}(h) = n^{-1} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2
\]

(8)

According to [21] the optimum bandwidth \((h)\) is obtained by minimizing the GCV in (7).

5. Data and Steps of Analysis.
The secondary data used in this research is a set of admission test scores based on public senior high school status that was obtained from the secretariat of SPAN-PTKIN and UM-PTKIN from 2016 to 2018. The data was collected from 44 State Islamic Colleges. The data consists of 132 of scholastic aptitude tests, and 132 of Islamic tests. Next, to analyze the data, we conduct the following steps:

a. Calculate the correlation between scholastic aptitude test scores and Islamic test scores.
b. Determine the optimum bandwidth \((h)\) based on the minimum GCV value in (7).
c. Estimate the admission test scores by using nonparametric regression local linear estimator approach.
d. Calculate the value of MSE in (8).
e. Plot the admission test scores and the estimated admission test scores based on the percentage of public senior high school status.
f. Compare the admission test scores models of state Islamic colleges in western Indonesia with those in eastern Indonesia.

We create Open Source Software (OSS)-R code to analyze the data.

6. Results and Discussion
The variables used to model the admission test scores are scholastic aptitude test scores as the first response variable \((y^{(1)}_{ik})\), Islamic test scores as the second response variable \((y^{(2)}_{ik})\), and percentage of public senior high school status as the predictor variable \((x_{ik})\). Firstly, we calculate the correlation between scholastic aptitude test scores and Islamic test scores. Based on the output of correlation pearson test from OSS-R with \(\alpha = 0.05\), we obtain that the correlation value is 0.865 and P-value = 0.000. So, we conclude that there is correlation between the scholastic aptitude test scores and Islamic test scores.

Next, we determine the optimal bandwidth \((h)\) based on the minimum GCV criterion. The result of this process is given in Table 1. Also, plot of GCV values versus bandwidth \((h)\) is given in Figure 1.
Table 1. Values of bandwidth, GCV and MSE

| Bandwidth | GCV       | MSE      |
|-----------|-----------|----------|
| 3.88      | 3263.362  | 2899.872 |
| 3.89      | 3263.352  | 2900.683 |
| 3.90      | 3263.351  | 2901.498 |
| 3.91      | 3263.357  | 2902.316 |
| 3.92      | 3263.372  | 2903.138 |

Based on Table 1 and Figure 1, we get the optimum bandwidth of 3.90. Also, the average value of MSE is 2901.498. Plot of estimated scholastic aptitude test scores based on the percentage of public senior high school status is given in Figure 2.

Figure 1. Plot of GCV values versus bandwidth ($h$)

Figure 2. Plot of estimated scholastic aptitude test score versus percentage of public senior high school

For example, the estimation of admission test scores model for western Indonesia is represented by “UIN Sunan Ampel ($\hat{y}_{291}$)” in 2016, while the estimation of admission test scores model for eastern
Indonesia is represented by “UIN Alauddin Makassar ($\hat{y}_{421}$)” in 2016. The estimated model of scholastic aptitude tests scores ($\hat{y}^{(1)}$) is given as follows:

\[
\begin{align*}
\hat{y}^{(1)}_{291} &= 533.74 + 5.28 (x - 43.37); \quad 39.47 < x < 47.27 \\
\hat{y}^{(1)}_{421} &= 463.36 - 2.64 (x - 74.01); \quad 70.11 < x < 77.91 
\end{align*}
\]  

(9)

For local linear model of “UIN Sunan Ampel” is given as follows:

\[
\hat{y}^{(1)}_{291} = 533.74 + 5.28 (x - 43.37); \quad 39.47 < x < 47.27
\]

It means that if the percentage of public senior high school increases one percent, then scholastic aptitude test score of “UIN Sunan Ampel” increases of 5.28.

Further, for local linear model of “UIN Alauddin Makassar” is given as follows:

\[
\hat{y}^{(1)}_{421} = 463.36 - 2.64 (x - 74.01); \quad 70.11 < x < 77.91
\]

It means that if the percentage of public senior high school increases one percent, then scholastic aptitude test score of “UIN Alauddin Makassar” decreases of 2.64.

Plot of estimated Islamic test scores based on the percentage of public senior high school status is given in Figure 3.

![Figure 3. Plot of estimated Islamic test scores versus percentage of public senior high school](image)

The estimated model of Islamic test scores ($\hat{y}^{(2)}$) is given as follows:

\[
\begin{align*}
\hat{y}^{(2)}_{291} &= 550.34 + 9.13 (x - 43.37); \quad 39.47 < x < 47.27 \\
\hat{y}^{(2)}_{421} &= 473.72 - 2.17 (x - 74.01); \quad 70.11 < x < 77.91 
\end{align*}
\]  

(10)

For local linear model of “UIN Sunan Ampel” is given as follows:

\[
\hat{y}^{(2)}_{291} = 550.34 + 9.13 (x - 43.37); \quad 39.47 < x < 47.27
\]

It means that if the percentage of public senior high school increases one percent, then Islamic test score of “UIN Sunan Ampel” increases of 9.13.
For local linear model of “UIN Alauddin Makassar” is given as follows:
\[
y_{421}^{(2)} = 473.72 - 2.17 (x - 74.01); \quad 70.11 < x < 77.91
\]
It means that if the percentage of public senior high school increases one percent, then Islamic test scores of “UIN Alauddin Makassar” decreases of 2.17.

7. Conclusion
In average, the pattern of scholastic aptitude test scores and Islamic test scores based on the percentage of public senior high school of state Islamic colleges in western Indonesia is different from that in eastern Indonesia. In this case, of course, local linear estimator approach in nonparametric regression is more appropriate to model admission test scores based on public senior high school status. Based on the estimated models we get; the government can make policies that increase the admission test scores in order to the quality of education in state Islamic colleges would be better.

8. References
[1] Badan Pusat Statistik 2017 Potret Pendidikan Indonesia Statistik Pendidikan 2017 Badan Pusat Statistik Press Jakarta.
[2] Mujahidun 2016 Pemerataan Pendidikan Anak Bangsa: Pendidikan Gratis Versus Kapitalisme Pendidikan Tarbiyatun 7 38-52.
[3] Idrus M 2012 Mutu Pendidikan dan Pemerataan Pendidikan di Daerah Psikopedagogia 1(2) ISSN: 2301-6167
[4] Harman G 1994 Student Selection and Admission to Higher Education: Policies and Practices in The Asian Region Higher Education 27(3) 313-339.
[5] Lohman D F 2004 Aptitude for College: The Importance of Reasoning Tests for Minority Admissions. In R. Zwick (Ed.) Rethinking the SAT: The Future of Standardized Testing in University Admissions RoutledgeFalmer New York pp. 41-55.
[6] Koljatic M, Silva M and Cofre R 2013 Achievement Versus Aptitude in College Admissions: A Cautionary Note Based on Evidence from Chile. International Journal of Educational Development 33(1) 106-115.
[7] Andriani P 2010 Pengaruh Asal Sekolah dan Jurusan Terhadap Hasil Belajar Pengantar Dasar Matematika Mahasiswa Fakultas Tarbiyah IAIN Mataram Beta 3(2) 118-133.
[8] Qudratullah M F 2014 Pengaruh Jalur Penerimaan Mahasiswa dan Asal Sekolah Terhadap Prestasi Mahasiswa di Fakultas Sains dan Teknologi UIN Sunan Kalijaga Fourier 3(1) 12-23.
[9] Lestari B, Fatmawati and Budiantara I N 2017 Estimasi Fungsi Regresi Nonparametrik Multirespon Menggunakan Reproducing Kernel Hilbert Space Berdasarkan Estimator Smoothing Spline Proceed. National Seminar on Math. and Its Appl.(SNMA) 2017 Fac. Sci. Tech. Airlangga University Surabaya pp. 243-250.
[10] Lestari B, Anggraeni D and Saifudin T 2018 Estimation of Covariance Matrix based on Spline estimator in Homoscedastic Multi-Responses Nonparametric Regression Model in Case of Unbalance Number of Observations Far East J. Mathematical Sciences (FJMS) 108(2) 341-355.
[11] Lestari B, Fatmawati, Budiantara I N and Chamidah N 2018 Estimation of regression function in multi-response nonparametric regression model using smoothing spline and kernel estimators Journal of Physics: Conference Series 1097 012091.
[12] Lestari B, Chamidah N and Saifudin T 2019 Estimasi fungsi regresi dalam model regresi nonparametrik birespon menggunakan estimator smoothing spline dan estimator kernel Jurnal Matematika, Statistika & Komputasi (JMSK) 15(2) 20-24.
[13] Lestari B, Fatmawati and Budiantara I N 2019 Spline Estimator and Its Asymptotic Properties in Multiresponse Nonparametric Regression Model Songklanakarin Journal of Science and Technology In Press.
[14] Chamidah, N., and Lestari, B. 2016. Spline Estimator in Homoscedastic Multi-Response Nonparametric Regression Model in Case of Unbalanced Number of Observations. Far East
Journal of Mathematical Sciences (JAMS), 100, 1433-1453.
[15] Islamiyati A, Fatmawati and Chamidah N 2018 Estimation of Covariance Matrix on Bi-Response Longitudinal Data Analysis with Penalized Spline Regression Journal of Physics: Conf. Series 979 012093.
[16] Chamidah N and Rifada M 2016 Estimation of median growth curves for children up twoyears old based on biresponse local linear estimator AIP Conference Proceedings 1718 110001.
[17] Chamidah N, Tjahjono E, Fadilah A R and Lestari B 2018 Standard Growth Charts for Weight of Children in East Java Using Local Linear Estimator Journal of Physics: Conf. Series 1097 012092.
[18] Chamidah N, Gusti K H, Tjahjono E and Lestari B 2019 Improving of Classification Accuracy of Cyst and Tumor Using Local Polynomial Estimator TELKOMNIKA 17(3) 1492-1500.
[19] Nottingham Q J and Cook D F 2011 Local Linear Regression for Estimating Time Series Data Journal of Computational Statistics and Data Analysis 37 209-217.
[20] Liang H and Cheng D Z 2005 Assesment of Esophageal Pressure in Gastroesophageal Reflux Disease by Local Regression Annals of Biomedical Engineering 33 847-853.
[21] Wu H and Zhang J T 2006 Nonparametric Regression Methods for Longitudinal Data Analysis Willey-Interscience New Jersey.