Analysis of Factors Affecting Tuition Fee in a Private University: A Data Mining Using VAR Model

Wahyuddin S1,2, Fauzi Insan Estiko3,4 and Estiko Rijanto5,1
1 Magister of information system, Universitas Komputer Indonesia
2 STMIK Dipanegara Makassar, Indonesia
3 Faculty of Economics and Business, Universitas Indonesia
4 Department of Economics and Finance, Hanyang University, Seoul, South Korea
5 Indonesian Institute of Sciences (Lembaga Ilmu Pengetahuan Indonesia), Indonesia

E-mail: wahyudipanegara@live.com, therealfauzi2@gmail.com, estiko.rijanto@lipi.go.id

Abstract. The purpose of this study is to deals with a data mining to extract knowledge regarding factors which affect tuition fee and predict future amount of the tuition fee. Specifically this paper aims to know what factors do affect tuition fee of a private university in eastern part of Indonesia. We used a vector autoregressive (VAR) model with variables including tuition fee, inflation rate, number of enrolled students and regional minimum wage. The data covers from January 2010 to December 2018 and were collected from the private university, the Bank Indonesia, the Central Bureau of Statistics (CBS), and the South Sulawesi office CBS. We carried out a step-by-step procedure that consists of stationarity test, optimal lag determination, and significance of parameters test as well as un-correlatedness of residuals and structure stability tests. We found that the tuition fee and the number of students are affected by inflation. This result gives impact on tuition fee prediction.

1. Introduction
Tuition fee in a university may change year by year due to changing inflation rate, the number of students, regional minimum wage, and other factors. Moreover, detail of cost structure and cost breakdown, which can be used to predict the amount of tuition fee, often can not be accessed. It is then necessary to extract knowledge from limited available data. Higher tuition fee means that fewer families can gain the education and training they need to grow and prosper in their communities. Underserved students as well as families from deeper into the middle class, find it difficult to afford a college education [1]. Economists and other analysts have long been interested in understanding the demand for higher education. Examples of such work include studies focused on quantifying price elasticities for various student populations, estimating student sensitivity to changes in financial aid packages, or constructing university-specific demand functions [2]. Despite colleges and universities increase tuition fees, the number of students who apply to attend colleges keeps getting larger each year. This has occurred in large part because, students and their parents instinctively understand that whether one goes to college may matter [3].

In Indonesia, we already have huge number of colleges and universities, but many parties have been engaging business in higher education by establishing new colleges or universities. For public universities in Indonesia, we have a government regulation which regulates unified study expense named Biaya Kuliah Tunggal (BKT) and unified tuition fee named Uang Kuliah Tunggal (UKT) [4]. The BKT includes all operational expenses which are directly related to study activities of students in one semester in a certain study program. The UKT is the cost that must be paid by every enrolled student according to individual economic capability. The BKT becomes the basis to calculate expenses which are shared between the society and the government. BKT and UKT are decided by the Minister
of Research, Technology, and Higher Education based on proposal from each university. On the other hand, for private universities in Indonesia, the amount of tuition fee that must be paid by each enrolled student may vary depending on many factors. Calculation of such cost may rely on different approaches i.e. investment based cost, activities based cost, or others. However, each private university must have its own supply and demand equilibrium. Published papers that address tuition fee analysis of universities or colleges in Indonesia are still limited [5-7]. Ahmad R. Budiman and B. Setyadin [5] estimated the real expenses that are paid by students in Malang State University, a state owned university, based on questioner using the Slovin sample size formula with number of samples of 392 students. From the statistical calculation they obtained that direct cost shares 21.45% and indirect cost shares 78.55% of the total education expenses. Juanda and Nikki V. Lestari [6] conducted study of unit cost of the educational process in faculty of medicine at University of Muhammadiyah Malang, a private university. Total education cost, excluding experiments, includes: general variable cost, new student specific variable cost, final year student specific variable cost, and fixed cost. They found that the fixed cost encompasses 92.5% of the total cost. Their results indicated that the cost paid by the student was more lesser than the real cost to be burdened, so that it could not be told being efficient. Yuni, Matrutty, S. Sipakoly [7] calculated the ideal amount of tuition fee for the Diploma IV study program on Applied Business Administration at the Ambon State Polytechnique. They used activity based costing approach and they compared the calculated tuition fee with the UKT. They found that the UKT is only around 50% of the calculated ideal tuition fee.

The above papers contributed to the literature on tuition fee in Indonesia from the viewpoint of accountancy. Whereas, this paper deals with data mining to extract knowledge regarding factors which affect tuition fee from available data. We use vector autoregressive (VAR) method. Specifically, this paper aims to know what factors affect tuition fee of a private university in Makassar, the capital of South Sulawesi province, Indonesia. The knowledge obtained from this data mining result can be used to predict future tuition fee based on the significant factors.

2. Methods

In this paper, we set a hypothesis that tuition fee at STMIK Dipanegara Makassar is affected by inflation rate, the number of enrolled students, and the regional minimum wage. Since theoretically it is not clear yet what factors do affect the tuition fee, we use an analysis method known as vector autoregressive (VAR). VAR was first proposed by Sims (1980) [8] as an alternative model besides autoregressive distributed lag (ADL) models by minimizing theoretical approach. It can be used to analyze dynamic relationships between time-series variables based on available data without knowing underlying theory in advance. Each variable is explained by its own lagged values and the lagged values of all other endogenous variables in the model observed.

The basic form of a VAR model consisting of a set of K variables \( Y_t = [y_{1(t)}, y_{2(t)}, \ldots, y_{K(t)}]^T \) that is observed at time \( t = 0, 1, 2, \ldots, T \) and defined with lag order \( p \) is given in equation (1) [9][10].

\[
Y_t = \theta + A_1 Y_{t-1} + A_2 Y_{t-2} + \ldots + A_p Y_{t-p} + e_t \quad (1.a)
\]

\[
Y_t = \theta + \sum_{i=1}^{p} (A_i Y_{t-i}) + e_t \quad (1.b)
\]

where \( A_i \) is \((K \times K)\) coefficient matrix for each \( i = 1, 2, \ldots, p \). \( e_t \) is a \(K\)-dimensional white noise process with \( E(e_t) = 0\), and \( \theta_t = [\theta_{1(t)}, \theta_{2(t)}, \ldots, \theta_{K(t)}]^T \) is a fixed \((K \times I)\) vector of intercept terms allowing for the possibility of a non-zero mean, \( E(Y_t) \). \( e_t \) is time invariant with positive definite covariance matrix.

A VAR model assumes that each variable is stationer. An appropriate variable transformation is necessary when the original time-series data is not stationer. A common method to evaluate stationarity of a time-series data is the Augmented Dickey Fuller (ADF) test. The ADF test assesses existence of any unit root in an autoregressive process model with lag order \( p \) as given in equation (2) [11].
\[
\Delta Y_{k(t)} = \alpha + \delta t + \beta Y_{k(t-1)} + \sum_{i=1}^{p} (\varphi_i \Delta Y_{k(t-1)}) + \epsilon(t)
\]

(2)

where \( \Delta Y_{k(t)} = Y_{k(t)} - Y_{k(t-1)} \). The time-series data is stationary if \( \beta < 0 \).

Optimal lag order \( p \) for the VAR model in equation (1) can be determined using Final Prediction Error (FPE) [12] or Schwarz Information Criterion (SIC) [13].

\[
FPE(p) = \left[ \frac{T+K+p+1}{T-K_p-1} \right]^K \det \Sigma_{\epsilon}(p)
\]

(3)

\[
SIC(p) = \ln(\Sigma_{\epsilon}(p)) + \frac{\ln(T)}{T} p K^2
\]

(4)

where \( T \) denotes the sample size, \( K \) the dimension of the time series, \( p \) the lag order of the VAR fitted to the data, and \( \Sigma_{\epsilon}(p) \) is the maximum likelihood (ML) estimator of \( \Sigma_{\epsilon}(p) \) obtained by fitting a VAR(p) model to the data. \( \Sigma_{\epsilon}(p) \) is the white noise covariance matrix.

To evaluate fitness of the VAR model, diagnostic tests are carried out including un-correlatedness of the residuals and stability of the model.

In order to analyze factors that affect tuition fee in a private university, we used data of tuition fee and quantity of income from STMIK Makassar, South Sulawesi province, Indonesia. We used monthly data of year-on-year inflation from January 2010 to December 2018. This data was collected from Bank Indonesia (BI) (BI, 2018) [14] and the Indonesian Central Bureau of Statistics (CBS) (CBS, 2018) [15]. The corresponding data of the regional minimum wage was collected from CBS South Sulawesi (CBS South Sulawesi, 2018) [16].

Let \( y_1(t), y_2(t), y_3(t), \) and \( y_4(t) \) denote the tuition fee, the inflation, the number of students, and the regional minimum wage at time \( t \), respectively. From the basic form of the VAR model in equation (1) we obtain a VAR model expressed in equation (5).

\[
\begin{bmatrix}
\Delta y_1(t) \\
\Delta y_2(t) \\
\Delta y_3(t) \\
\Delta y_4(t)
\end{bmatrix} = \begin{bmatrix}
\theta_1 \\
\theta_2 \\
\theta_3 \\
\theta_4
\end{bmatrix} + \begin{bmatrix}
\Delta y_1(t-1) \\
\Delta y_2(t-1) \\
\Delta y_3(t-1) \\
\Delta y_4(t-1)
\end{bmatrix} + \begin{bmatrix}
\varphi_1 y_1(t-2) \\
\varphi_2 y_2(t-2) \\
\varphi_3 y_3(t-2) \\
\varphi_4 y_4(t-2)
\end{bmatrix} + \cdots + \begin{bmatrix}
\varphi_p y_1(t-p) \\
\varphi_p y_2(t-p) \\
\varphi_p y_3(t-p) \\
\varphi_p y_4(t-p)
\end{bmatrix} + \begin{bmatrix}
\epsilon_1(t) \\
\epsilon_2(t) \\
\epsilon_3(t) \\
\epsilon_4(t)
\end{bmatrix}
\]

(5)

We first evaluate the time-series in terms of stationary using the ADF test method. Following the first step, then optimal lag order is calculated using the FPE-IC. The next step is selecting the coefficients according to significance of coefficient tests. After that, a reduced VAR model is derived that has only the significant coefficients. And then, the coefficients that are not significant are omitted from the VAR model. Subsequently, to evaluate fitness of the model, diagnostic tests are carried out including un-correlatedness of the residuals and structural stability of the model. And finally, impulse function test is carried out to verify the validity of the reduced VAR model.

If the original variable is not stationary, we calculate the first difference of the variable. If the first difference of the variable is not stationary, we further calculate the second difference of the variable. We will move to the second step only if the variable is stationary.

3. Results and Discussion

From 2010 to 2018 the number of academic staffs and supporting staffs at STMIK Dipanegara Makassar has almost been constant, but number of students varies. Figure 1 plots number of students and quantity of income from the students that are collected each year by the STMIK, in normalized unit. Variations of students number and income quantity imply variation of tuition fee paid by the students, like following this Figure 1:
Figure 1. Quantity of the students and income from the students

Figure 2 plot autocorrelation functions of each endogenous variable. Up to lag order 18 was included in stationarity test of each variable. The ADF test results are listed in table 1. We obtained stationarity conditions for all variable at the second-difference [11]. Therefore, a VAR model was estimated at the second-difference.

Table below explain the ADF test result, there are have 4 variable, like following this Table 1:

| Variable | ADF Statistic | p-value |
|----------|---------------|---------|
| $\Delta^2 y_1$ | -3.158 | 0.0259 |
| $\Delta^2 y_2$ | -3.453 | 0.0124 |
| $\Delta^2 y_3$ | -3.054 | 0.0353 |
| $\Delta^2 y_4$ | -2.926 | 0.0467 |
The FPE information criterion (IC) was applied to a tentative initial VAR model with maximum lag of 18, and the optimal lag number was found to be 12 as shown in Table 2 [12].

Table 2. FPE information criterion (IC) value

| IC       | Lag 1 | Lag 2 | Lag 3 | Lag 4 | Lag 5 | Lag 6 | Lag 7 | Lag 8 | Lag 9 |
|----------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| FPE (10^-8) | 6.512 | 5.606 | 6.034 | 7.259 | 8.229 | 10.769 | 14.964 | 21.328 | 30.136 |

Since the optimum order lag is 12, our VAR model can be expressed as in equations (6) and (7).

\[ \Delta^2 Y(t) = \sum_{i=1}^{12} (A_i \Delta^2 Y(t-i)) + e(t) \]  

\[ A_i = \begin{bmatrix} a_{i1} & a_{i2} & a_{i3} & a_{i4} \\ a_{i1} & a_{i2} & a_{i3} & a_{i4} \\ a_{i3} & a_{i4} & a_{i2} & a_{i1} \\ a_{i3} & a_{i4} & a_{i2} & a_{i1} \end{bmatrix} \]

Coefficient matrix \( A_i \) for every lag \( i \) was calculated by the use of ordinary least square (OLS) algorithm. Notice that for every lag \( i \) we have 16 coefficients, therefore total number of coefficients is 192. However, by evaluating significance of coefficient using t-statistic, only coefficients which satisfy the hypothesis are used, and the others are neglected.

Our concern is to know what variable affects the tuition fee. Therefore, in the following the coefficient which significantly affects the tuition fee is listed (See Table 3).

Table 3. Coefficient which significantly affect the tuition fee (lag order 12, significance code 0.01)

| Lag i | \( a_{i1} \) | \( a_{i2} \) | \( a_{i3} \) | \( a_{i4} \) |
|-------|------------|------------|------------|------------|
| Lag 1 | -0.590390  | 0          | 0          | 0          |
| Lag 12| 0          | 0.163848   | 0          | 0          |

From Table 3 we know that the second-difference of tuition fee is affected by the second-difference of tuition fee itself of lag 1 as well as by the second-difference of inflation of lag 12. The second-differences of both student quantity and regional minimum wage do not affect the second-difference of tuition fee. We can express tuition fee as in equation (8).

\[ \Delta^2 y_1(t) = -0.59\Delta^2 y_1(t-1) + 0.164\Delta^2 y_2(t-12) + e_1(t) \]  

It is also interesting to know what variable affects the number of enrolled students. The coefficient that significantly affects the quantity of the students is listed in Table 4.

Table 4. Coefficient which significantly affect the quantity of the students (lag order 12, significance code 0.01)

| Lag i | \( a_{i1} \) | \( a_{i2} \) | \( a_{i3} \) | \( a_{i4} \) |
|-------|------------|------------|------------|------------|
| Lag 1 | 0          | 0          | -0.630354  | 0          |
| Lag 2 | 0          | 0          | -0.601191  | 0          |
| Lag 12| 0          | -0.176517  | 0          | 0          |
From Table 4 we know that the second-difference of student quantity is affected by the second-difference of student quantity itself of lag 1 and 2 as well as by the second-difference of inflation of lag 12. The second-differences of both tuition fee and regional minimum wage do not affect the second-difference of student quantity.

To evaluate un-correlatedness of residuals, a Portmanteau-test was carried out which included 72 lag order. Figures 3 and 4 plots ACF and PACF of residual of the second-difference of tuition fee.

![Figure 3. Plots of serial correlation (Δ²y₁)](image)

![Figure 4. Plots of serial correlation (Δ²y₃)](image)

Structural stability was evaluated using cumulative summation (CUSUM) method. Figures 5 and 6 plots empirical fluctuation processes of tuition fee equation and student quantity equation, respectively.

![Figure 5. CUSUM plot of equation Δ²y₁](image)

![Figure 6. CUSUM plot of equation Δ23](image)
From the above results, it can be said that serial correlation of residuals does not exist and that the VAR model structure is stable. Next, response of tuition fee to an impulse shock of inflation is plotted in Figures 7 and 8 plots impulse response of student quantity to impulse shock of inflation.

Figure 7. Impulse response of tuition fee to inflation shock

![Image of Figure 7]

Figure 8. Impulse response of student quantity to inflation shock

Figure 7 is consistent with the result in Table 3 whereas Figure 8 supports the results in Table 4. From equation (8), we can predict that next year tuition fee is obtained from this year tuition fee plus a positive proportional term of the difference of inflations between this year and last year.

4. Conclusion

We obtained stationarity conditions for all the variables at the second-difference. A VAR model was estimated at the second-difference of the variables with the optimal lag of 12. From the VAR model we know that tuition fee in STMIK Dipanegara Makassar is affected by inflation and the number of enrolled students is also affected by inflation. Moreover, the tuition fee is not affected by both the number of students and the regional minimum wage. This result gives impact on tuition fee prediction.

Acknowledgment

The authors thank to STMIK Dipanegara Makassar – South Sulawesi province for providing data regarding the number of students and staffs as well as the data of tuition fee from 2010 to 2018. This research was conducted as part of study by the first author at Magister of Information System, Universitas Komputer Indonesia (UNIKOM), Bandung – West Java province, Indonesia.
References

[1] C. Mulhern, R. Spies, D. D. Wu, and M. P. Staiger, “The Effects of Rising Student Costs in Higher Education: Evidence from Public Institutions in Virginia,” Ithaka S+R, New York, Aug. 2015.

[2] S. W. Hemelt and D. E. Marcotte, “Rising Tuition and Enrollment in Public Higher Education,” no. IZA DP No. 3827, p. 34, 2008.

[3] R. G. Ehrenberg, “The Economics of Tuition and Fees in American Higher Education,” *ILR Collect.*, p. 21, Oct. 2007.

[4] “Peraturan Menteri Riset, Teknologi, dan Pendidikan Tinggi Republik Indonesia Nomor 39 Tahun 2017.” Kemenkumham, No 779-2017.

[5] A. R. Budiman and B. Setyadin, “Analisis Satuan Biaya Pendidikan Mahasiswa Universitas Negeri Malang,” *Manaj. Pendidik.*, vol. 23, no. 5, pp. 467–478, Mar. 2012.

[6] A. Juanda and V. L. Nikki, “Analisis Perhitungan Biaya Satuan (Unit Cost) Penyelenggaraan Pendidikan Kedokteran (Studi Kasus Pada Fakultas Kedokteran Universitas Muhammadiyah Malang),” *J. Reviu Akutansi Dan Keuang.*, vol. 2, no. 1, pp. 227–241, 2012.

[7] Yuni, M. T.J.F.K, and S. Selly, “Penerapan Metode Activity Based Costing Dalam Menentukan Tarif SPP Program Studi Diploma IV Administrasi Bisnis Terapan Politeknik Negeri Ambon,” *Ejournal Polnam*, pp. 35–55, 2018.

[8] L. J. Christiano, “Christopher A. Sims and Vector Autoregressions*: Christopher A. Sims and Vector Autoregressions,” *Scand. J. Econ.*, vol. 114, no. 4, pp. 1082–1104, Dec. 2012.

[9] Dr. Bernhard Pfaff and Kronberg im Taunus, “Using The Vars Package,” *ResearchGate*, p. 34, Sep. 2007.

[10] B. Pfaff, “VAR, SVAR and SVEC Models: Implementation Within R Package vars,” *J. Stat. Softw.*, vol. 27, no. 4, 2008.

[11] *Stata Time-Series Reference Manual Release 15*, vol. 15. College Station, Texas: A Stata Press Publication, StataCorp LLC, 2017.

[12] H. Lütkepohl, *New introduction to multiple time series analysis*. Berlin: New York : Springer, 2005.

[13] A. A. Neath and J. E. cavanaugh, “Regression and time series model selection using variants of the schwarz information criterion,” *Commun. Stat. - Theory Methods*, vol. 26, no. 3, pp. 559–580, Jan. 1997.

[14] “Data Inflasi - Bank Sentral Republik Indonesia.” [Online]. Available: https://www.bi.go.id/id/moneter/inflasi/data/Default.aspx. [Accessed: 21-Feb-2019].

[15] “Badan Pusat Statistik.” [Online]. Available: https://www.bps.go.id/statisticable/2014/09/26/915/tingkat-inflasi-tahun-ke-tahun-gabungan-82-kota-sup-1-sup-2012-100.html. [Accessed: 21-Feb-2019].

[16] “Badan Pusat Statistik Provinsi Sulawesi Selatan.” [Online]. Available: https://sulsel.bps.go.id/brs.html. [Accessed: 21-Feb-2019].