ENHANCED KNOWLEDGE IN SCIENCES AND TECHNOLOGY
e-ISSN: 2773-6385

EKST

Vol. 6 No. 1 (2026) 150-159
https://publisher.uthm.edu.my/periodicals/index.php/ekst

University Student Enrolment Prediction in Malaysia with State Space Model

Fatin Illyana Mohd Zulhazmy[1], Kek Sie Long[1]*

[1] *Department of Mathematics and Statistics, Faculty of Applied Sciences and Technology, UTHM Kampus Cawangan Pagoh, Hab Pendidikan Tinggi Pagoh, KM 1, Jalan Panchor, 84600 Pagoh, Muar, Johor, MALAYSIA.*

*Corresponding Author: slkek.uthm.edu.my
DOI: https://doi.org/10.30880/ekst.2026.06.01.019

Article Info

Received: 4 January 2026
Accepted: 5 June 2026
Available online: 9 July 2026

Keywords

State Space Model, Enrolment, private Higher Education Institution, Mean Squared Error, Time Series Forecasting, public universities, Malaysia

Abstract

Accurate university student enrolment is crucial for informing decisions about academic infrastructure, personnel, and funding as it significantly impacts institutional capacity planning, resource allocation and national education policy. However, the accuracy of enrolment forecasting is influenced by demographic shifts, economic conditions and policy factors. This paper aimed to use a state space model to predict undergraduate student enrolment in Malaysian public and private universities. The enrolment dataset from 2002 to 2024, obtained from the Ministry of Higher Education (MOHE), was normalized to ensure numerical stability and comparability for modelling. Descriptive statistics, such as mean and standard deviation, were processed for analysis of standardization of the data. The performance of the model was evaluated using the mean squared error (MSE) to assess its effectiveness. Hence, the dynamic changes and long- term enrolment trends were detected by the state space model, which produced low prediction errors. Moreover, the state space model expressed the lowest MSE values compared to the simple moving average and exponential smoothing models for predicting enrolment in both public and private universities. These findings confirm that the study objectives had been fulfilled and the effectiveness of the state space model in predicting student enrolment in Malaysian universities is demonstrated.

1. Introduction

The Malaysian higher learning system was very important to the country's socio-economic development. Public and private institutions contributed immensely by providing skilled individuals and promoting the spirit of innovation towards the development of the country [4]. Enrolment forecasting for third-level institutions was imperative for the benefit of the government, institutional administration, parents and industries [3]. Student enrolment forecasting helped students make informed decisions about their academic future [3]. It assisted academic programs in remaining relevant to labour market trends. In the current modern world, the benefits from institutions of higher learning were significant. However, more research needed to be conducted to define the impact on various regions surrounding the country [4]. Enrolment numbers provided significantly to the budget formulation of higher learning institutions to align to the country's needs [5]. Models like time series often lacked accuracy in capturing the changing trends within the student enrolment numbers. Hence, this study recommended a state space model to predict university student enrolment in Malaysia using descriptive statistics, forecasting future enrolment through an appropriate state space model and evaluate the model's performance using the mean square error (MSE). This study proposed a model to forecast the number of bachelor's degree enrolment in higher education institutions. This category crucial for national workers planning and long-term development [8]. A multidimensional approach was required to make justifiable model comparison, as no single accuracy metric was suitable for all situations [9]. State space models were able to identify patterns within the changes observed over a period. State space models provided a strong model to interpret the complex times series given. This framework also allowed precise forecasting of these data [6]. The proposed state space model further improved the prediction accuracy compared to previous forecasting method. Flexible selection among the models could be adopted according to the characteristics given [5]. A smaller MSE indicated better prediction performance. This provided the ability to choose the model that closely matched the observed enrolment data [2].

2. Methodology

The methodology described the research framework of the study, including the data collection, data preprocessing, descriptive analysis, model development, model evaluation, and interpretation and conclusion. The research work was conducted with a dataset that includes the enrolment statistics for undergraduate students within Malaysia from 2002 to 2024 [11]. The datasets were obtained from the official reports published by the Ministry of Higher Education (MOHE) [10]. Data normalization was performed prior to analysis. These data were normalization to ensure numerical stability and comparability for modelling. It was followed by descriptive analysis, including the computation of mean, variance, and standard deviation. A state space model was formulated to predict trends in university student enrolment datasets. A mean squared error (MSE) was used to evaluate accuracy of the predictions made in this study [2]. The study concluded with a discussion on the future research directions. Table 1 shows the bachelor's degree enrolment of public and private universities from 2002 to 2024. Public universities have a higher enrolment compared to private universities. Both sectors exhibit an overall increasing trend, indicating a growing demand for higher education and a rising need for highly educated workers in the nation.

Table 1 Bachelor's Degree Enrolment Data, collected from MOHE

Year	Public Universities	Private Universities
2002	184,190	67,062
2003	192,288	90,631
2004	194,470	105,325
2005	209,148	101,395
2006	223,968	124,071
2007	247,881	140,699
2008	270,156	151,591
2009	272,012	198,760
2010	274,690	220,299
2011	299,179	180,065
2012	305,141	204,956
2013	331,410	198,653
2014	340,538	198,189
2015	324,894	266,622
2016	322,507	373,592
2017	332,023	320,561
2018	228,563	329,136
2019	350,102	313,214
2020	369,312	271,331
2021	378,806	258,775
2022	377,103	248,560
2023	375,904	263,555

2024	379,253	299,254

2.1 State Space Model Representation

Consider a general state space model [1] given by

$$\dot{x} = Ax + b, \tag{1}$$

$$y = Cx + d, \tag{2}$$

where x is the state vector and y is the output vector, while A is the state transition matrix, b is the model constant vector, C is the output coefficient matrix and d is the output constant vector. Here, $\dot{x}$ is the state rate of changes in time t. The initial condition is $x(0) = x_0$, where x_0 is a real-valued vector. It is known that (1) is the state equation and (2) is the output equation. The model is implemented in MATLAB to estimate parameters from the dataset.

2.2 Analytical Solution

Refer to (1), rewrite the state equation,

$$\dot{x} - Ax = b, \tag{3}$$

and multiply (3) by e^{-At},

$$e^{-At}\dot{x} - e^{-At}Ax = e^{-At}b, \tag{4}$$

Note that the left-hand side of (4) can be simplified by

$$\frac{d}{dt}\left(e^{-At}x\right) = e^{-At}b, \tag{5}$$

Integrating both sides of (5) from o to *t*,

$$e^{-At}x(t) - x(0) = -\left(e^{-At} - I\right)A^{-1}b, \tag{6}$$

where *I* is an identify matrix.

Then, the analytical solution of the state equation [12] is

$$x(t) = e^{At}x_0 - \left(I - e^{At}\right)A^{-1}b, \tag{7}$$

and the discrete sequence of the solution is given by

$$x_{k+1} = e^{A\Delta t}x_k - \left(I - e^{A\Delta t}\right)A^{-1}b, \tag{8}$$

with $\Delta t = t_{k+1} - t_k$ is a sampling time at *k* time step. The measured output becomes

$$y_k = Cx_k + d, \tag{9}$$

1.1 Least Squares Optimization Problem

For the prediction purpose, we introduce a least squares optimization problem [8] as follows,

$$\mathit{Minimize} \quad J(x) = \frac{1}{2}\left(y - y_s\right)^T\left(y - y_s\right) \tag{10}$$

subject to the state equation (1) and the output equation (2), where y_s is the observed dataset, which is the historical data, J is the objective function that minimizes the sum of prediction errors and ^{T} represents the transposition operator. The aim is to estimate the model parameters A, b, C and d to find the optimal state estimate   for predicting the observed output y_s so that the objective function J is minimized.

To proceed, consider the following gradients with respect to the model parameters,

$$\frac{\partial J}{\partial A} = C^T\left(\hat{y} - y_s\right)\left(\hat{x}^T\right), \tag{11}$$

$$\frac{\partial J}{\partial b} = C^T(\hat{y} - y_s)(1^T), \tag{12}$$

$$\frac{\partial J}{\partial C} = (\hat{y} - y_s)(\hat{x}^T), \tag{13}$$

$$\frac{\partial J}{\partial d} = (\hat{y} - y_s)(1^T), \tag{14}$$

where $1 = (1,1, \ldots..,)^T$, $\hat{x}$ is the state estimate and $\hat{y}$ is the output estimate. With these gradients, the model parameters are estimated by using the following recursion equations [8],

$$A^{i+1} = A^i - \alpha_1\left(\frac{\partial J}{\partial A}\right)^i, \tag{15}$$

$$b^{i+1} = b^i - \alpha_2\left(\frac{\partial J}{\partial b}\right)^i, \tag{16}$$

$$C^{i+1} = C^i - \alpha_3\left(\frac{\partial J}{\partial C}\right)^i, \tag{17}$$

$$d^{i+1} = d^i - \alpha_4\left(\frac{\partial J}{\partial d}\right)^i, \tag{18}$$

Where $\alpha_1, \alpha_2, \alpha_3$ and α_4 are learning rates, and i is the iteration number. The optimal model parameter can be obtained once the first-order necessary conditions [8]

$$\frac{\partial J}{\partial A} = 0, \ \frac{\partial J}{\partial b} = 0, \ \frac{\partial J}{\partial C} = 0, \ \frac{\partial J}{\partial d} = 0, \tag{19}$$

are satisfied.

In addition, the gradient with respect to the state variables is given by

$$\frac{\partial J}{\partial x} = C^T(\hat{y} - y_s), \tag{20}$$

and the state estimate is obtained by the recursion equation below,

$$\hat{x}^{i+1} = x^i - \alpha_5\left(\frac{\partial J}{\partial x}\right)^i, \tag{21}$$

where α_5 is the learning rate and x is the solution to the state equation (1). Hence, the optimal state estimate is resulted when the first-order necessary condition

$$\frac{\partial J}{\partial x} = 0, \tag{22}$$

is presented.

2.3 Prediction Scheme

For implementing the prediction, we modify the solution sequence of the state equation (8) and the measured output (9) as follows,

$$\hat{x}_k = e^{A\Delta t} x_k - \left(I - e^{A\Delta t}\right) A^{-1} b, \tag{23}$$

$$\hat{y}_k = C\hat{x}_k + d, \tag{24}$$

where $\hat{x}_k$ and $\hat{y}_k$ are the state estimate and the output estimate at time step k, respectively. The calculation procedure is summarized as an iterative algorithm as follows.

Step 1 Solve the state equation (1) from (8) and measure the model output (2) from (9). Set the iteration Number $i = 0$ to begin the iteration.

Step 2 Compute the state estimate $\hat{x}^i$ and the output estimate $\hat{y}^i$ from (23) and (24), respectively.

Step 3 Evaluate the objective function J^{I+1} from (10).

Step 4 Calculate the gradients with respect to the model parameters from (11), (12), (13), and (14), and update the model parameters from (15), (16), (17), and (18).

Step 5 Calculate the gradient with respect to the state variable from (20) and update the state estimate $\hat{x}^{i+1}$ from (21).

Step 6 Test the convergence. If $\left\| J^{i+1} - J^i \right\| <$ tolerance, stop the iteration; otherwise, set $i = i + 1$, $x^i = \hat{x}^i$, go to Step 2 and repeat the iteration.

2.4 Accuracy of Prediction

A mean squared error (MSE) is applied to measure the accuracy of the prediction results [2]. The smallest value of MSE expresses the highest accuracy of the prediction results. The MSE is calculated from,

$$MSE = \frac{1}{N}\sum_{t=1}^{N}\left(y_t - \hat{y}_t\right)^2, \tag{25}$$

where y_t is the actual number of university enrolments at year t, $\hat{y}_t$ is the predicted number of university enrolments at year t, and N is the number of actual enrolments in universities.

3. Results and Discussion

There was a total of 23 data points on student enrolment in Malaysian public and private universities from 2002 and 2024. Table 2 shows that the range of the public university enrolment data was 195,063, with a standard deviation of 66,075.27. In contrast, the range of the private university enrolment data was 306,530, with a standard deviation of 85,160.18. The data were suitable for prediction because they represent a continuous annual time series across 23 years. Thus, the model can capture long-term trends and stable variability over time. Since these enrolment data values are quite large, these data were normalized through a standardization as follows,

$$\boldsymbol{normalized\ data} = \frac{\boldsymbol{original\ data - mean\ of\ original\ data}}{\boldsymbol{standard\ deviation\ of\ original\ data}}$$

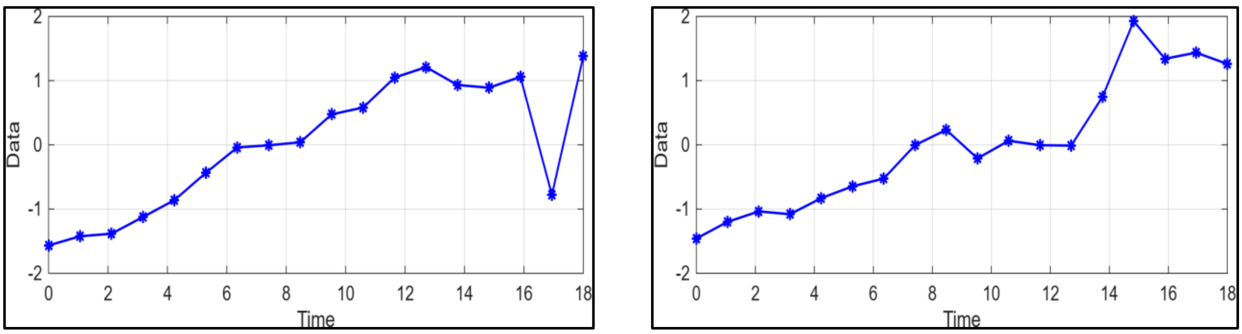
The ratio of 80:20 was considered to predict 18 data points and remained five data points was used for validation.

2

Table 2 Descriptive Statistics of the Public and Private Universities

Descriptive Statistics	Public Universities	Private Universities
Count	23	23
Mean	294,936.4348	214,186.7826
Median	305,141	204,956
Variance	4,365,941,870	7,252,255,734
Standard Deviation	66,075.27427	85,160.17692
Minimum	184,190	67,062
Maximum	379,253	373,592

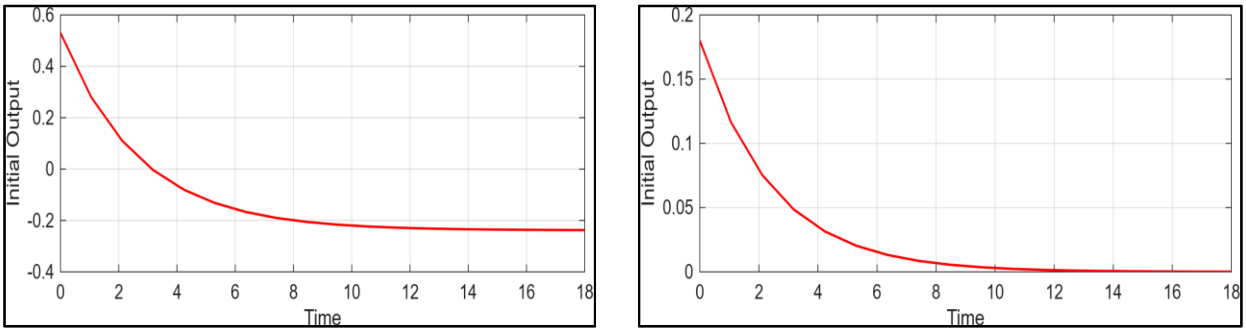
Fig. 1(a) and Fig. 1(b) showed the normalized enrolment data for both public and private universities from 2002 to 2019. The enrolment of public universities showed an overall increase over time with reasonable fluctuations, indicating a stable enrolment pattern. The presence of this long-term upward trend suggested that the data had a clear structure and was suitable for time series forecasting using a state space model. The stable trend and manageable variability supported the use of this data for enrolment prediction. The enrolment of private universities also demonstrated the enrolment trend generally increasing over time, where greater fluctuations than public universities indicating higher variability. Nevertheless, the clear upward trend suggested an underlying structure, supporting the use of the state space model as a flexible approach for forecasting the enrolment of public and private universities.



(a) (b)

Fig. 1 Normalized Historical Enrolment: (a) Public Universities; (b) Private Universities

Refer to state space model in (1) and (2), the model parameters for public universities were A = 0.373, b = -0.089, C = 1 and d = 0, while for private universities were A = -0.4120, b = 0.00001, C = 1 and d = 0. Fig. 2(a) and Fig. 2(b) presented the outputs of the state space model for both public and private universities. Both figures displayed a smooth model response over time, indicating that the model was stable. It validated the successful optimization of the model parameters. The reduction in prediction errors suggested that the state space model was able to capture enrolment dynamics over time.



(a) (b)

Fig. 2 Initial State Space Model Output: (a) Public Universities; (b) Private Universities

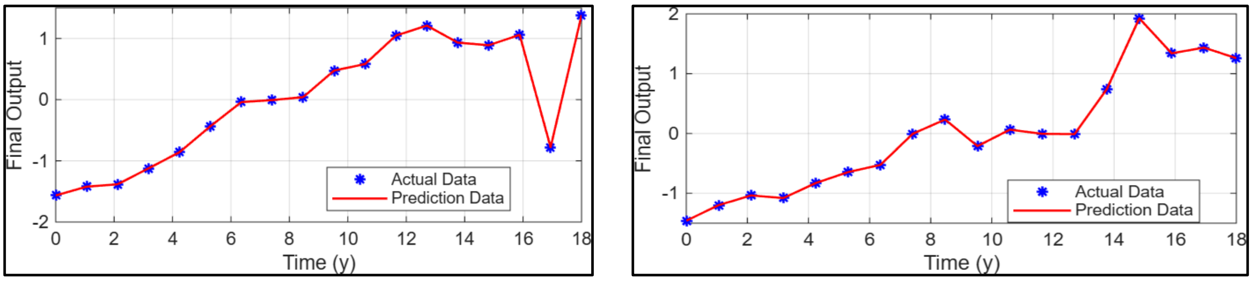
3.1 Final State Space Model Output

Table 3 summarized the model output for both public and private universities, including the number of iterations, elapsed time in seconds, and the mean squared error (MSE) values before and after optimization in MATLAB. For public universities, the model converged in 50 iterations within 0.007714 seconds. The MSE value dropped dramatically from 1.3129 before optimization to 2.1896×10^{-9}after optimization. The model showed a significant improvement in its fit. Similarly, for private universities, the model converged after 41 iterations with an elapsed time of 0.005273 seconds. The MSE was decreased from 1.0115 to 1.8397×10^{-9}. Prediction accuracy showed a significant improvement after the optimization process.

Table 3 Output of the State Space Model for Public and Private Universities

Institution Type	Iteration Number	Elapsed Time	Initial MSE	Final MSE
Public Universities	50	0.007714	1.3129	2.1896×10^{-9}
Private Universities	41	0.005273	1.0115	1.8397×10^{-9}

Fig. 3(a) and Fig. 3(b) showed the final state space model output for both public and private universities. The red line represented the predicted enrolment value. Meanwhile, the blue markers showed the actual student enrolment data. It was observed that the predicted data closely followed the actual data. The underlying trend and dynamic behaviour of student enrolment in Malaysia were successfully captured by the final state space model. Therefore, a small gap was still observed between the estimated and actual outputs upon closer inspection.

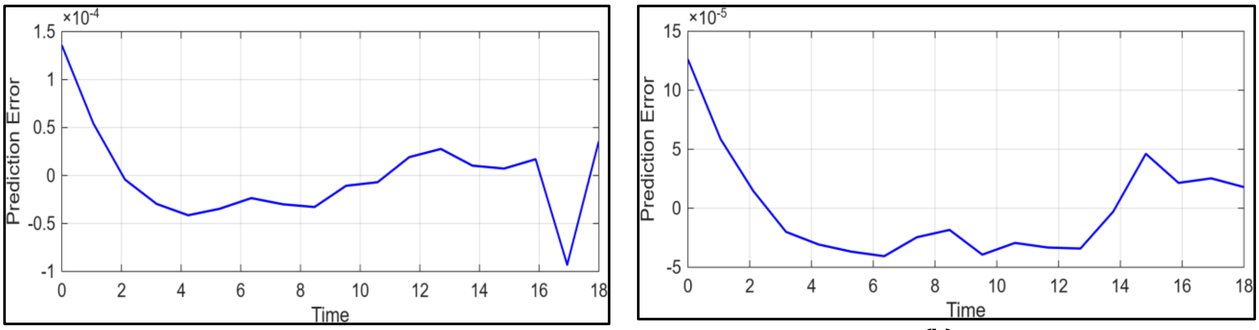


(a) (b)

Fig. 3 Optimized State Space Model Outputs: (a) Public Universities; (b) Private Universities

3.2 Prediction Error

Fig. 4(a) and Fig. 4(b) presented the prediction error between the actual student enrolment and the estimated output from the state space model. Small prediction errors close to zero were observed for both datasets. Although slight changes were observed in certain years. The overall magnitude of the error was low. This situation indicated that the model successfully captured the underlying enrolment trends. This observation aligned with the low MSE value as given in Table 3. It confirmed that the high accuracy of the state space model for both datasets.



(a) (b)

Fig. 4 Prediction Error Analysis (a) Public Universities; (b) Private Universities

3.3 Comparison of Forecasting Methods

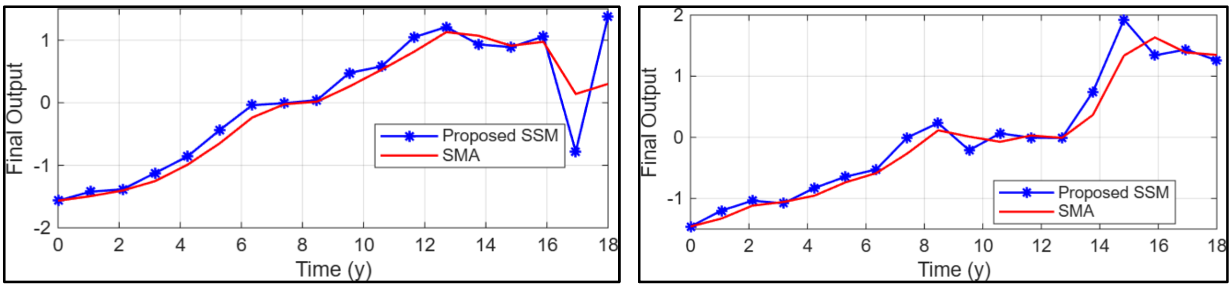
A comparison of the prediction performance of the state space model with the simple moving average (SMA) and the exponential smoothing method (ESM) was made for benchmarking purposes. The SMA was easy to use but it responded slowly to trends and could not capture dynamic pattern. In contrast, the ESM reacted more quickly to changes but it was limited in handling complex dynamics. This comparison was conducted using the same datasets of the enrolment for both public and private universities.

Table 4 compared the MSE values for the three prediction methods. The state space model yielded the lowest MSE value of 1.7642×10^{-5} for the enrolment of public universities, compared to the SMA of 0.1261, and the ESM of 0.0645. This result indicated that the state space model provided higher prediction accuracy for public universities enrolment. Similarly, the state space model outperformed the SMA value of 0.0439 and the ESM value of 0.0344 for the enrolment of private universities, displaying the lowest MSE value of 4.3198×10^{-8}. Overall, the results demonstrated that the state space model predicted enrolment more accurately than the SMA and ESM methods for both public and private universities.

Table 4 Comparison of MSE Values for Different Forecasting Methods

Institution Type	State Space Model	Simple Moving Average (SMA)	Exponential Smoothing Method (ESM)
Public Universities	1.7642×10^{-5}	0.1261	0.0645
Private Universities	4.3198×10^{-8}	0.0439	0.0344

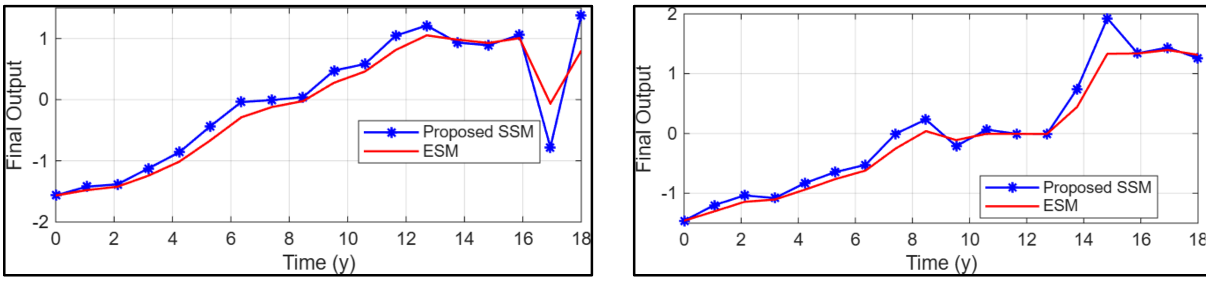
Fig. 5(a) and Fig. 5(b) showed a comparison of trends between the original data and the SMA for public and private universities. The blue line represented the original data, while the red line represented the SMA. The SMA responded more slowly and failed to capture significant fluctuations. The mean squared error results tested this difference. Overall, the SMA predicted the enrolment data less accurate compared to the state space model.



(a) (b)

Fig. 5 Prediction with SMA: (a) Public Universities; (b) Private Universities

Fig. 6(a) and Fig. 6(b) compared the prediction performance of the ESM for public and private universities. The blue line represented the original data, while the red line represented the ESM output. The ESM still produced smoother prediction and tends to miss sharp fluctuations. Overall, the ESM remained less accurate than the state-space model since the state space model could capture both long-term trends and sudden changes in the enrolment.



(a) (b)

Fig. 6 Prediction with ESM: (a) Public Universities; (b) Private Universities

4. Conclusion

This study employed a state space model to predict university student enrolment in Malaysia, using enrolment data from the Ministry of Higher Education (MOHE) spanning the period from 2002 to 2024 for both public and private universities. A prediction ratio of 80:20 was applied, with 80% of the data used for training and another 20% for validation. The datasets were analysed using descriptive statistics to identify trends and assess their adaptability for modelling. The enrolment data were predicted using the proposed state space model. The model parameters were estimated by minimizing the value of mean squared error (MSE). During the parameter estimation process in MATLAB, the MSE value was monitored. Parameter estimation was crucial for minimizing the prediction error of the proposed state space model. Based on the MSE, the model output was stabilized. It achieved the smallest gap between predicted and actual enrolment values.

The MSE values obtained from the proposed state space model were compared with the simple moving average (SMA) and exponential smoothing model (ESM) for both datasets. The proposed state space model produced the lowest MSE compared to the SMA and the ESM. Indirectly, the state space model showed a superior prediction performance for the student enrolment in both public and private universities. It was confirmed through the findings of this study that the proposed state space model was suitable for predicting student enrolment in Malaysian universities. Therefore, the objectives of this study were successfully achieved.

Acknowledgement

The authors would like to thank the Faculty of Applied Sciences and Technology, Universiti Tun Hussein Onn Malaysia for the support.

Conflict of Interest

The authors declare that there is no conflict of interest regarding the publication of the paper.

Author Contribution

The authors confirm contribution to the paper as follows: ***study conception and design****: Fatin Illyana, Kek Sie Long,* ***data collection****: Fatin Illyana;* ***analysis and interpretation of results****: Fatin Illyana, Kek Sie Long;* ***draft manuscript preparation****: Fatin Illyana. All authors reviewed the results and approved the final version of the manuscript.*

References

[1] A. F. Adedotun, T. Latunde, and O. A. Odusanya, "Modelling and forecasting climate time series with state-space model," *Journal of the Nigerian Society of Physical Sciences*, vol. 2, no. 3, pp. 149–159, 2020, doi: 10.46481/jnsps.2020.94.

[2] T. Chai and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)? Arguments against avoiding RMSE in the literature," *Geosci. Model Dev.*, vol. 7, pp. 1247–1250, 2014, doi: 10.5194/gmd- 7-1247-2014.

[3] C. Chhor, V. Sek, R. Norng, S. Hoeurn, and R. Sam, "Factors influencing students' enrollment decisions in higher education institutions: A case study of the National University of Battambang," *Eur. J. Contemp. Educ. E-Learn.*, vol. 2, no. 6, pp. 169–186, 2024, doi: 10.59324/ejceel.2024.2(6).10.

[4] E. V. Fakhrutdinova, J. S. Kolesnikova, N. I. Larionova, and L. G. Chumarova, “Impact of higher education on socio-economic development,” *Rev. On Line Política Gest. Educ.*, vol. 25, no. 6, pp. 3472–3482, 2021, doi: 10.22633/rpge.v25iesp.6.16102.

[5] S. Grapragasem, A. Krishnan, and A. N. Mansor, “Current trends in Malaysian higher education and the effect on education policy and practice: An overview,” *Int. J. Higher Educ.*, vol. 3, no. 1, pp. 85–93, 2014, doi: 10.5430/ijhe.v3n1p85.

[6] R. J. Hyndman, A. B. Koehler, R. D. Snyder, and S. Grose, “A state space framework for automatic forecasting using exponential smoothing methods,” *Int. J. Forecast.*, vol. 18, no. 3, pp. 439–454, 2002, doi: 10.1016/S0169-2070(01)00110-8.

[7] M. A. Islam, N. A. A. Rahim, T. C. Liang, and H. Momtaz, “Effect of demographic factors on e-learning effectiveness in a higher learning institution in Malaysia,” *Int. Educ. Stud.*, vol. 4, no. 1, pp. 112–121, 2011, doi: 10.5539/ies.v4n1p112

[8] X. Ma and F. Ding, “Gradient-based parameter identification algorithms for observer canonical state space systems using state estimates,” *Circuits, Syst., Signal Process.*, vol. 34, no. 5, pp. 1697–1709, 2015, doi: 10.1007/s00034-014-9911-5.

[9] N. Mehdiyev, D. Enke, P. Fettke, and P. Loos, “Evaluating forecasting methods by considering differentaccuracy measures,” *Procedia Comput. Sci.*, vol. 95, pp. 264–271, 2016, doi: 10.1016/j.procs.2016.09.332

[10] Ministry of Higher Education Malaysia, *Higher Education Statistics 2023*. Malaysia: Ministry of Higher Education Malaysia, 2023. [Online]. Available: https://mohe.gov.my/muat-turun/statistik/2023-3/1473-bab-1-makro-2023-pdf/file

[11] Ministry of Higher Education Malaysia, *Higher Education Statistics 2024*. Malaysia: Ministry of Higher Education Malaysia, 2024. [Online]. Available: https://www.mohe.gov.my/en/download/statistics/2024-4/1702-bab-1-makro-2024-update-pdf/file

[12] C. Venkateswarlu and R. R. Karri, “Linear filtering and observation techniques,” in *Optimal State Estimation for Process Monitoring, Fault Diagnosis and Control*, ch. 3, pp. 21–57. Amsterdam, The Netherlands: Elsevier, 2022, doi: 10.1016/B978-0-323-85878-6.00011-7.