Re-Evaluating the Performance of MIMSA as a Criterion for Time-Series Model Selection

Maha Rani Pratama, Jonathan Hoseana, Agus Sukmana

***Abstract*—A novel time-series model selection criterion, known as MIMSA (Mutual Information Model Selection Algorithm), has demonstrated satisfactory performance in a recent study, but not so much in an earlier study. In this paper, we conduct modest numerical experiments to re-evaluate the performance of MIMSA in detecting time-series models and producing predictions, in comparison to standard information- and error-based criteria. The results confirm that MIMSA demonstrates less superior performance in detecting time-series models, and a tendency to select higher-order models. Furthermore, the prediction performance of MIMSA varies depending on the type of the model used to generate the actual data, with better performance observed in the cases of AR models with smaller orders than in the cases of MA models.**

***Index Terms*—time series, model selection, MIMSA, AR, MA.**

I. Introduction

THE theory of time series and its wide-ranging applications have been a fascinating research subject for decades. Over the last three years, in particular, time series theory has been applied in predictive studies of various real-world phenomena, including wind speed [29], traffic flow [42], stock prices [32], [45], and climate dynamics [37]. Complementarily, its prediction accuracy has also been assessed from the viewpoint of machine learning theory [38].

On the other hand, the problem of time-series model selection has also been a subject of extensive research [12], [13], with the existence of various criteria to evaluate the suitability of each of some pre-specified candidate models. The core of the problem lies in the identification of a single model which best represents a given dataset, possesses the most desirable properties, and can subsequently be employed for data predictions. To ensure the accuracy of such predictions, the criterion used for the model selection must be chosen judiciously. There are two classes of criteria from which one could choose a desired criterion: information-based criteria and error-based criteria. Information-based criteria are built upon likelihood functions, and are formulated with the aim of achieving an optimal balance between the model's fittingness for the dataset and the model's complexity. Popular information-based criteria include AIC (Akaike Information Criterion) [4], BIC (Bayesian Information Criterion) [39], AIC_c (Corrected Akaike Information Criterion) [16], and HQIC (Hannan-Quinn Information Criterion) [11]. On the other hand, error-based criteria assess the performance of a model by comparing actual dataset values with those predicted by the model. Common error-based criteria include RMSE (Root Mean Square Error) [33], MAPE (Mean Absolute Percentage Error) [24], and MASE (Mean Absolute Scaled Error) [19].

Information-based criteria for model selection have been applied in numerous studies. AIC alone, for instance, has been applied in disease mapping [23], business forecasting [30], and copula fitting [10], [7]. Moreover, as an alternative to BIC, AIC has been applied in distribution fitting [9], [44], portfolio risk estimation [21], non-life insurance capital computation [35], and carbon-dioxide emission analysis [31]. Likewise, AIC_c has been applied in studies on the spread of COVID-19 [1], the forecasting of value-added tax [34], and the distribution of passenger flow at rail stations [43], while HQIC has been applied in studies on the effect of climate risks on bond yields [14], [15] as an alternative to AIC. On the other hand, the error-based criterion RMSE has been applied in automatic building extraction [40], sea-level oscillation prediction [22], battery state-of-charge identification [5], and electric power system operation [26], while MAPE and MASE have been applied as alternatives in studies on short-term mortality forecasting [17], COVID-19 cases forecasting [41], residential building energy efficiency [20], and axial compression capacity of circular concrete-filled steel tubes [8].

In addition to the above criteria, the so-called LIC (Likelihood Information Criterion) and its derivatives are also known for use for time-series model selection [27]. Such criteria are information-based, and measure how well a model captures the relationship between past and future data in a time series, while also accounting for the model's complexity and sample size. Since the information generated by LIC encompasses various aspects of dependency within a time series, an algorithm has been developed to implement this criterion, known as MIMSA (Mutual Information Model Selection Algorithm) [2], [3].

In a recent study conducted by Akca and Yozgatligil [3], both MIMSA and BIC demonstrate the best performance in the selection of time-series models representing given time-series datasets. However, contrasting results are found in an earlier thesis by Akca [2], where the performance of MIMSA is observed to be inferior to that of BIC. These differing results on the performance of MIMSA constitute the motivation of the present study. In the present study, we thus aim to re-evaluate the performance of MIMSA, through numerical simulations involving two modest types of time-series models: AR and MA, and carry out performance comparisons using the aforementioned information-based

Manuscript received January 30, 2025; revised April 27, 2025.

Maha Rani Pratama is a graduate of the Mathematics Programme, Faculty of Science, Parahyangan Catholic University, Bandung, Indonesia (e-mail: maharanipratamaa@gmail.com).

Jonathan Hoseana is a researcher at the Center for Mathematics and Society, Faculty of Science, Parahyangan Catholic University, Bandung, Indonesia (corresponding author, e-mail: j.hoseana@unpar.ac.id).

Agus Sukmana is a researcher at the Center for Mathematics and Society, Faculty of Science, Parahyangan Catholic University, Bandung, Indonesia (e-mail: asukmana@unpar.ac.id).

and error-based criteria. At each stage of our simulations, the best model selected using MIMSA is evaluated to assess the accuracy of its predictions.

We organise our work as follows. In the upcoming section II, we begin by defining the theoretical concept of mutual information and subsequently MIMSA itself, and provide empirical methods to estimate both mutual information and MIMSA based on given datasets. In section III, we describe an algorithm utilised in our numerical simulations, and analyse the results it generates with the aim of re-evaluating the performance of MIMSA. In the final section IV, we state our conclusions and describe avenues for further studies.

II. Mutual information and MIMSA

In this section, we first review the theoretical concept of mutual information, and describe a method to estimate its value in an empirical setting (subsection II-A). Subsequently, we recall the theoretical definition of MIMSA, and also describe a method to estimate its value in an empirical setting (subsection II-B).

A. Mutual information

Informally, mutual information is a metric representing the degree at which two variables mutually depend or cooperate [2], [3]. For an illustration, let X be a random variable having range $R_X = \{1,2,3,4,5,6\}$ representing the outcome of rolling a fair die, and Y be a random variable having range $R_Y = \{0,1\}$ taking the value 0 if the outcome of the same roll is even and 1 otherwise. In this case, the value of Y provides information on the value of X, and vice versa. We therefore say that the two random variables share mutual information. On the other hand, if Z denotes a random variable having range $R_Z = \{1,2,3,4,5,6\}$ representing the outcome of rolling a different fair die, then the random variables X and Z do not share mutual information. Indeed, the outcome of the first roll provides no information whatsoever on the outcome of the second roll.

More formally, the mutual information between two random variables X and Y with ranges R_X and R_Y is defined as

$$I(X,Y) = \int_{R_Y}\int_{R_X} f_{XY}(x,y)\,\ln\left[\frac{f_{XY}(x,y)}{f_X(x)f_Y(y)}\right]\mathrm{d}x\,\mathrm{d}y, \quad (1)$$

where f_X and f_Y denote the marginal probability density functions of X and Y, while f_{XY} denotes the joint probability density function of X and Y [2], [3]. The formula (1) constitutes a theoretical formula of the mutual information between two random variables. Let us next provide a method to estimate mutual information in an empirical setting.

Let $\mathcal{X} = (x_1,\ldots,x_N)$ and $\mathcal{Y} = (y_1,\ldots,y_N)$ be datasets of values of two random variables X and Y. Fix a positive integer n, denoting the number of subintervals into which the intervals $[\min(\mathcal{X}),\max(\mathcal{X})]$ and $[\min(\mathcal{Y}),\max(\mathcal{Y})]$ are to be partitioned. Defining the subintervals' widths

$$h_{\mathcal{X}} = \frac{\max(\mathcal{X})-\min(\mathcal{X})}{n} \quad (2)$$

and

$$h_{\mathcal{Y}} = \frac{\max(\mathcal{Y})-\min(\mathcal{Y})}{n}, \quad (3)$$

we partition the interval $[\min(\mathcal{X}),\max(\mathcal{X})]$ into n adjacent subintervals $I_1^{\mathcal{X}},\ldots,I_n^{\mathcal{X}}$, where

$$I_i^{\mathcal{X}} = [\min(\mathcal{X})+(i-1)h_{\mathcal{X}}, \min(\mathcal{X})+i\,h_{\mathcal{X}})$$

for every $i \in \{1,\ldots,n-1\}$, and

$$I_n^{\mathcal{X}} = [\min(\mathcal{X})+(n-1)h_{\mathcal{X}}, \max(\mathcal{X})],$$

and the interval $[\min(\mathcal{Y}),\max(\mathcal{Y})]$ into n adjacent subintervals $I_1^{\mathcal{Y}},\ldots,I_n^{\mathcal{Y}}$, where

$$I_i^{\mathcal{Y}} = [\min(\mathcal{Y})+(i-1)h_{\mathcal{Y}}, \min(\mathcal{Y})+i\,h_{\mathcal{Y}})$$

for every $i \in \{1,\ldots,n-1\}$, and

$$I_n^{\mathcal{Y}} = [\min(\mathcal{Y})+(n-1)h_{\mathcal{Y}}, \max(\mathcal{Y})].$$

Next, letting

$$\mathcal{S} = \{(x_1,y_1),\ldots,(x_N,y_N)\}, \quad (4)$$

we define for every $i,j \in \{1,\ldots,n\}$,

$$P_i^{\mathcal{X}} = \frac{\left|\{(x,y)\in\mathcal{S} : x \in I_i^{\mathcal{X}}\}\right|}{N}, \quad (5)$$

$$P_j^{\mathcal{Y}} = \frac{\left|\{(x,y)\in\mathcal{S} : y \in I_j^{\mathcal{Y}}\}\right|}{N}, \quad (6)$$

$$P_{i,j}^{\mathcal{X}\mathcal{Y}} = \frac{\left|\{(x,y)\in\mathcal{S} : x \in I_i^{\mathcal{X}} \text{ and } y \in I_j^{\mathcal{Y}}\}\right|}{N}. \quad (7)$$

Clearly, we have that

$$\sum_{i=1}^{n} P_i^{\mathcal{X}} = 1, \quad \sum_{j=1}^{n} P_j^{\mathcal{Y}} = 1, \quad \text{and} \quad \sum_{j=1}^{n}\sum_{i=1}^{n} P_{i,j}^{\mathcal{X}\mathcal{Y}} = 1.$$

Using the above constructs, as an estimate for the theoretical mutual information given by equation (1), we may calculate the empirical mutual information between $\mathcal{X}$ and $\mathcal{Y}$ as

$$\mathcal{I}_n(\mathcal{X},\mathcal{Y}) = \sum_{j=1}^{n}\sum_{i=1}^{n} P_{i,j}^{\mathcal{X}\mathcal{Y}}\,\ln\left(\frac{P_{i,j}^{\mathcal{X}\mathcal{Y}}}{P_i^{\mathcal{X}}\,P_j^{\mathcal{Y}}}\right). \quad (8)$$

Example. Suppose one wishes to compute the empirical mutual information between $\mathcal{X} = (-3,0,0,4,5)$ and $\mathcal{Y} = (2,3,4,3,6)$ using $n = 2$ subintervals: $\mathcal{I}_2(\mathcal{X},\mathcal{Y})$. First, the formulae (2) and (3) lead to $h_{\mathcal{X}} = 4$ and $h_{\mathcal{Y}} = 2$, so that the interval $[\min(\mathcal{X}),\max(\mathcal{X})] = [-3,5]$ is partitioned into the adjacent subintervals $I_1^{\mathcal{X}} = [-3,1)$ and $I_2^{\mathcal{X}} = [1,5]$, whereas the interval $[\min(\mathcal{Y}),\max(\mathcal{Y})] = [2,6]$ is partitioned into the adjacent subintervals $I_1^{\mathcal{Y}} = [2,4)$ and $I_2^{\mathcal{Y}} = [4,6]$. Next, the set $\mathcal{S}$ specified in (4) is given by

$$\mathcal{S} = \{(-3,2),(0,3),(0,4),(4,3),(5,6)\}.$$

The formulae (5) and (6) then lead to

$$P_1^{\mathcal{X}} = \frac{3}{5}, \quad P_2^{\mathcal{X}} = \frac{2}{5}, \quad P_1^{\mathcal{Y}} = \frac{3}{5}, \quad \text{and} \quad P_2^{\mathcal{Y}} = \frac{2}{5},$$

while the formula (7) gives

$$P_{1,1}^{\mathcal{X}\mathcal{Y}} = \frac{2}{5}, \quad P_{1,2}^{\mathcal{X}\mathcal{Y}} = \frac{1}{5}, \quad P_{2,1}^{\mathcal{X}\mathcal{Y}} = \frac{1}{5}, \quad \text{and} \quad P_{2,2}^{\mathcal{X}\mathcal{Y}} = \frac{1}{5}.$$

Substituting the above quantities into the formula (8) gives the following value of empirical mutual information between $\mathcal{X}$ and $\mathcal{Y}$:

$$\mathcal{I}_2(\mathcal{X},\mathcal{Y}) = \sum_{j=1}^{2}\sum_{i=1}^{2} P_{i,j}^{\mathcal{X}\mathcal{Y}}\,\ln\left(\frac{P_{i,j}^{\mathcal{X}\mathcal{Y}}}{P_i^{\mathcal{X}}\,P_j^{\mathcal{Y}}}\right) \approx 0.0138.$$

B. The MIMSA

Informally, a desirable time-series model selection criterion is that which selects models which provide predictions that are reasonably consistent with the corresponding values in the actual dataset [36]. More precisely, such a criterion is that for which the mutual information between the actual dataset and the prediction dataset is maximised. This motivates the introduction of the MIMSA (Mutual Information Model Selection Algorithm) for time-series model selection, which declares as the best time-series model one which minimises the criterion

$$\text{MIMSA} = -\ln\left[I\left(Y_t, \hat{Y}_t\right)\right] + \frac{2k\,(k+1)}{n-k-1}, \qquad (9)$$

where Y_t denotes the t-th term of the actual time-series dataset, $\hat{Y}_t$ denotes the t-th term of the prediction time-series dataset generated using the candidate time-series model, n denotes the dataset's size, and k denotes the number of parameters to be estimated in the candidate time-series model [2], [3].

Empirically, consider an actual time-series dataset $(y_t)_{t=1}^N$ along with a prediction time-series dataset $(\hat{y}_t)_{t=1}^N$ generated by a time-series model involving k parameters. By replacing the mutual information $I\left(Y_t, \hat{Y}_t\right)$ in the formula (9) with the estimate (8) obtained using n subintervals, one obtains the following empirical formula for MIMSA:

$$\text{MIMSA}_n = -\ln\left[\mathcal{I}_n\left((y_t)_{t=1}^N, (\hat{y}_t)_{t=1}^N\right)\right] + \frac{2k(k+1)}{n-k-1}. \quad (10)$$

EXAMPLE. Consider an actual time-series dataset

$$(y_t)_{t=1}^5 = (1, 1.6165, 0.8376, 0.1949, -0.9822)$$

along with a prediction time-series dataset

$$(\hat{y}_t)_{t=1}^5 = (1, 0.5, 0.25, 0.125, 0.0625)$$

generated using the AR(1) recursion $\hat{y}_t = 0.5\hat{y}_{t-1}$ with $\hat{y}_0 = 1$. Let us employ the formula (10) with $n = 2$ to calculate the empirical MIMSA_2 associated to the AR(1). Using the method described in the previous subsection, one obtains the empirical mutual information $\mathcal{I}_2\left((y_t)_{t=1}^5, (\hat{y}_t)_{t=1}^5\right) \approx 0.1184$, so that the desired empirical MIMSA is given by

$$\text{MIMSA}_2 = -\ln\left[\mathcal{I}_2\left((y_t)_{t=1}^5, (\hat{y}_t)_{t=1}^5\right)\right] + \frac{2\cdot 2\cdot (2+1)}{5-2-1} \approx 8.1336.$$

III. RESULTS AND DISCUSSION

In this section, we describe an algorithm which we use to conduct our model-fitting tests, and subsequently the obtained results. Although we shall execute the algorithm using AR and MA models, we only describe the algorithm in the case of AR models, the case of MA models being similar. To evaluate the performance of MIMSA in comparison to AIC, BIC, AIC_c, and HQIC, we carry out a series of simulations involving various scenarios, designed by combining the values parameters ranging from 0.1 to 0.5 while ensuring the validity of the stationarity assumption. For each scenario, we generate two time-series datasets: one of size 300 as a training dataset and another of size 75 as a testing dataset. The training dataset is to be used to assess each criterion's ability to detect the correct order of the associated time-series model, whereas the testing dataset is to used to evaluate the predictive performance. The latter is conducted using the error criteria RMSE, MAPE, and MASE.

A. The AR model-fitting test algorithm

For every $p \in \{1, 2, 3\}$, we carry out the following steps to generate time-series datasets using the time-series model AR(p) to be used to test the fittingness of the models AR(1), ..., AR(6).

1) Choose values for the parameters $\varphi_{p,1}$, ..., $\varphi_{p,p}$ and the standard deviation σ, and generate the time-series dataset $(y_t)_{t=1}^{300}$ using the time-series model AR(p), namely,

$$y_t = \varphi_{p,1}y_{t-1} + \cdots + \varphi_{p,p}y_{t-p} + \varepsilon_t,$$

with $y_0 = 1$. The dataset $(y_t)_{t=1}^{300}$ is to be used as our actual training dataset. In addition, generate also the time-series dataset $(x_t)_{t=1}^{75}$ using the same values of parameters and standard deviation, initial value $x_0 = 1$, and the model

$$x_t = \varphi_{p,1}x_{t-1} + \cdots + \varphi_{p,p}x_{t-p} + \varepsilon_t,$$

as our actual testing dataset.

2) For every $i \in \{1, \ldots, 6\}$, we test the fittingness of the model AR(p) with respect to the actual dataset generated in the previous step, in the following way.
   (a) Fit the model AR(i), namely,

$$y_t = \varphi_{i,1}y_{t-1} + \cdots + \varphi_{i,i}y_{t-i} + \varepsilon_t,$$

   to the actual training dataset $(y_t)_{t=1}^{300}$, using the maximum likelihood method to estimate the values of the parameters $\varphi_{i,1}$, ..., $\varphi_{i,i}$.
   (b) Using the estimated values of the parameters $\varphi_{i,1}$, ..., $\varphi_{i,i}$, generate the time-series datasets $(\hat{y}_t)_{t=1}^{300}$ with $\hat{y}_0 = y_0$ using the model

$$\hat{y}_t = \varphi_{i,1}\hat{y}_{t-1} + \cdots + \varphi_{i,i}\hat{y}_{t-i}.$$

   (c) Calculate the value of $\text{MIMSA}_5\left((\hat{y}_t)_{t=1}^{300}, (y_t)_{t=1}^{300}\right)$. This quantifies the fittingness of the model AR(i) with respect to our actual training dataset $(y_t)_{t=1}^{300}$ generated using the AR(p) recursion on step (1). To provide comparisons, also calculate the corresponding values of AIC, BIC, AIC_c, and HQIC.
   (d) Recall our actual testing dataset $(x_t)_{t=1}^{75}$ generated on step (1). Compute the errors RMSE, MAPE, and MASE of $(\hat{y}_t)_{t=1}^{75}$ with respect to $(x_t)_{t=1}^{75}$.
3) From step (2), we obtain the values of MIMSA_5, AIC, BIC, AIC_c, and HQIC for every $p \in \{1, 2, 3\}$ and $i \in \{1, \ldots, 6\}$. For every $p \in \{1, 2, 3\}$, choose $i \in \{1, \ldots, 6\}$ for which the value of MIMSA_5 is the smallest over all values of MIMSA_5 computed for all $i \in \{1, \ldots, 6\}$. This allows us to conclude that according to MIMSA, the best model representing our actual training dataset $(y_t)_{t=1}^{300}$ generated for the associated value of p is AR(i).

The above algorithm yields:

1) the best model chosen among AR(1), ..., AR(6) for each $p \in \{1, 2, 3\}$ and $i \in \{1, \ldots, 6\}$ according to

TABLE I
MODEL IDENTIFICATION FREQUENCIES IN THE CASE WHERE OUR ACTUAL TRAINING DATASET IS GENERATED USING THE MODEL AR(1).

Parameter	Criterion	AR(1)	AR(2)	AR(3)	AR(4)	AR(5)	AR(6)
$\varphi_{1,1} = 0.8$	MIMSA_5	5	5	4	9	9	18
	AIC	36	5	4	0	3	2
	BIC	48	2	0	0	0	0
	AIC_c	38	5	3	0	2	2
	HQIC	46	3	0	0	0	1

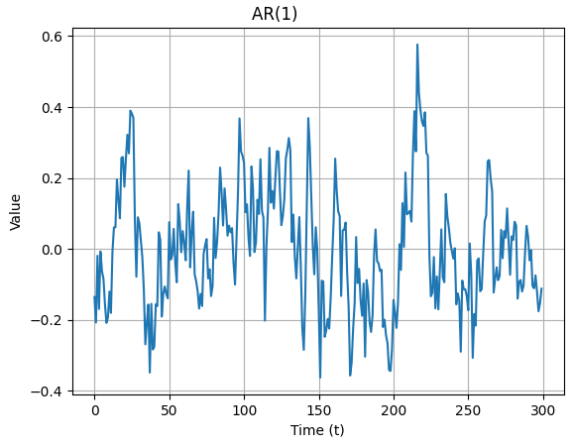

Fig. 1. A training dataset generated using AR(1) with $\varphi_{1,1} = 0.8$.

each of the criteria: MIMSA_5, AIC, BIC, AIC_c, and HQIC;
2) the values of the errors RMSE, MAPE, and MASE for each $p \in \{1, 2, 3\}$ and $i \in \{1, \ldots, 6\}$.

B. Results of AR model-fitting test

We execute the algorithm described in the previous subsection 50 times. Out of these 50 executions, let $n_i^{(1)}$, $\ldots$, $n_i^{(5)}$ be the number of times AR(i) is selected as the best model according to the criteria MIMSA_5, AIC, BIC, AIC_c, and HQIC, respectively.

In Table I, we present a performance comparison of MIMSA against AIC, BIC, AIC_c, and HQIC in the case where our actual training and testing datasets are generated using AR(1) with $\varphi_{1,1} = 0.8$. A training dataset generated using this model is visualised in Figure 1. Table I shows that BIC has the highest frequency in correctly identifying the order of the utilised AR model. Indeed, out of 50 algorithm executions, BIC correctly identifies AR(1) as many as 48 times and misclassifies it as AR(2) twice. On the other hand, the frequencies of correctly identifying the model in the same case for criteria other than BIC are lower, namely, 5 for MIMSA, 36 for AIC, 38 for AIC_c, and 46 for HQIC.

Computing the prediction errors, we find that the model AR(1) exhibits the smallest average RMSE value of 0.2775, compared to the models AR(2), $\ldots$, AR(6) whose average RMSE values increase with the models' order. This shows that the model AR(1) provides predictions with the highest accuracy with respect to our actual testing dataset, compared to the other models. Conversely, the model AR(6) exhibits the largest average RMSE value of 0.3743, demonstrating the worst prediction performance. The model AR(1) also exhibits the smallest average MAPE value of 1097.94%,

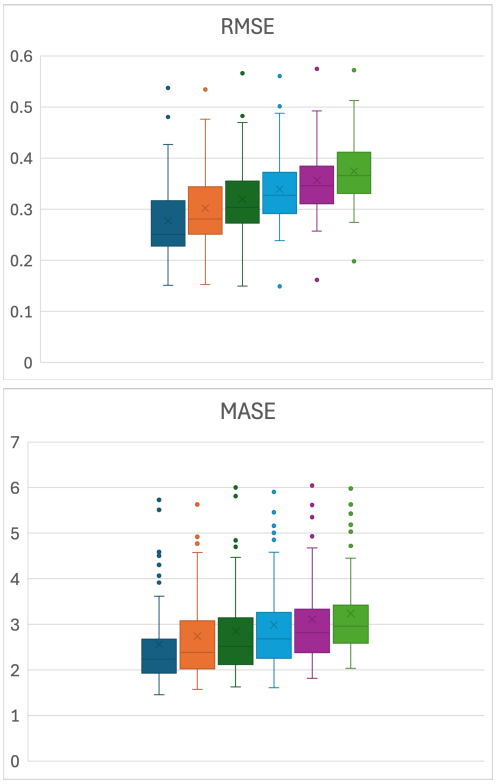

Fig. 2. Boxplots representing the values of RMSE and MASE associated to our testing datasets generated using the model AR(1).

TABLE II
MODEL IDENTIFICATION FREQUENCIES IN THE CASE WHERE OUR ACTUAL TRAINING DATASET IS GENERATED USING THE MODEL AR(2).

Parameter	Criterion	AR(1)	AR(2)	AR(3)	AR(4)	AR(5)	AR(6)
$\varphi_{2,1} = 0.05$ $\varphi_{2,2} = 0.5$	MIMSA_5	0	8	15	7	6	14
	AIC	0	34	9	3	2	2
	BIC	0	48	1	1	0	0
	AIC_c	0	36	8	3	1	2
	HQIC	0	45	3	2	0	0

indicating the best prediction performance. The average MAPE values also increase with the models' order, showing that higher-order AR models, despite being more complex, are less effective in capturing our actual testing dataset generated using AR(1). Consistently, the model AR(1) also exhibits the smallest average MASE value of 2.5471, while the model AR(6) exhibiting the highest average MASE value of 3.2371. This once again indicates that the prediction performance of the model AR(1) is better than that of the model AR(6).

Figure 2 shows boxplots visualising the values of RMSE and MASE associated to our testing datasets generated using the model AR(1). We observe that the model AR(1) itself exhibits the best performance compared to the other AR models. This is apparent from the lower mean and median values of both error criteria, which indicate a lower prediction error level. Additionally, the error distribution does not exhibit significant variation, with a narrower range, although some outliers are still present in MASE.

As in the case of AR(1), Table II shows that BIC has the highest frequency in correctly identifying the order of the AR model, with 48 correct identifications. On the other hand, MIMSA has the lowest frequency in detecting the

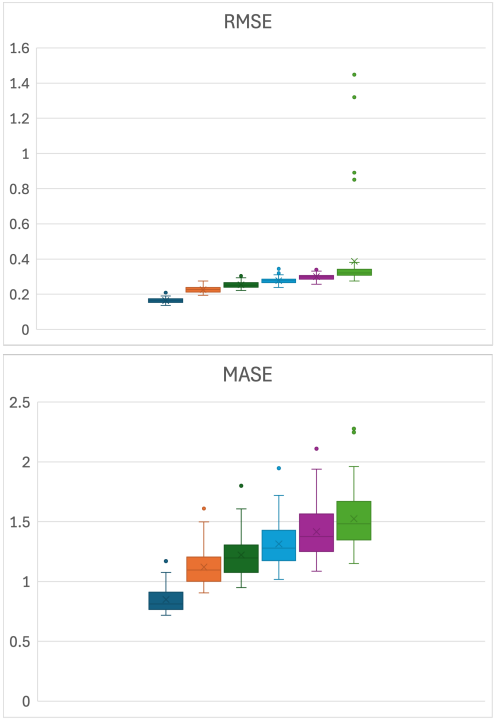

Fig. 3. Boxplots representing the values of RMSE and MASE associated to our testing datasets generated using the model AR(2).

TABLE III
MODEL IDENTIFICATION FREQUENCIES IN THE CASE WHERE OUR ACTUAL TRAINING DATASET IS GENERATED USING THE MODEL AR(3).

Parameter	Criterion	AR(1)	AR(2)	AR(3)	AR(4)	AR(5)	AR(6)
$\varphi_{3,1} = 0.1$ $\varphi_{3,2} = 0.1$ $\varphi_{3,3} = 0.7$	MIMSA_5	0	2	16	7	8	17
	AIC	0	0	41	4	2	3
	BIC	0	0	49	1	0	0
	AIC_c	0	0	41	4	2	3
	HQIC	0	0	45	5	0	0

correct AR model among the other criteria, with only 8 correct identifications. Computing the errors of the associated predictions, we conclude that the model AR(1) provides the best prediction performance, due to its lower RMSE, MAPE, and MASE values compared to those of the higher-order AR models. The increase of orders of the models from AR(2) to AR(6) leads to increases of the errors.

Let us next examine the RMSE and MASE boxplots shown in Figure 3. In the RMSE boxplot, the AR(1) model exhibits the lowest and most stable RMSE value compared to other models, with a relatively small error range. On the other hand, for the AR(2), ..., AR(6) models, the RMSE values tend to gradually increase, with greater variation and some significant outliers. This indicates that model performance deteriorates as the AR order increases, both in terms of the mean and the stability of prediction results. In the MASE boxplot, we observe a similar pattern, with an increase from AR(1) to AR(6).

From Table III we see the better performance of BIC compared to MIMSA and the other criteria, in identifying the correct order of the AR model. In particular, MIMSA incorrectly detects AR(6) as our actual training dataset's time-series model with the highest frequency of 17 times.

Evaluating the RMSE, MASE, and MAPE for our actual testing dataset generated using the model AR(3), we infer

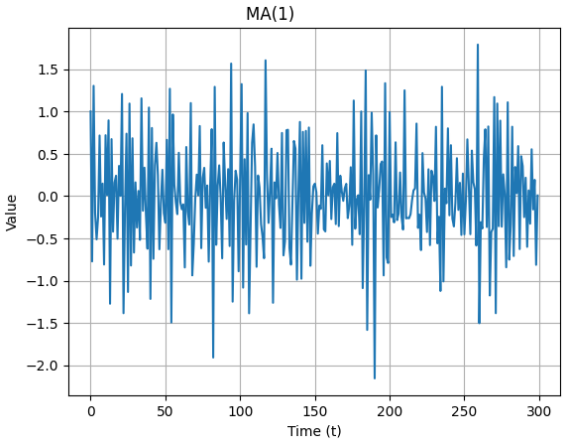

Fig. 4. A training dataset generated using MA(1) with $\theta_{1,1} = 0.8$.

that the model AR(1) demonstrates the best performance in providing predictions, with average RMSE and MASE values of 0.1980 and 1.0617. On the other hand, the model AR(2) exhibits an average MAPE value of 571.71%, while the models AR(3) to AR(6) exhibiting a significant decline of prediction performance. This indicates that as the models' order increases, the prediction errors become larger and the model tends to be less stable. Extreme values are also observed in higher-order AR models, both in MASE and MAPE.

C. The MA model-fitting tests

The MA model-fitting tests are carried out using an algorithm similar to that in subsection III-A, with MA(q),

$$y_t = \varepsilon_t - \theta_{q,1}\varepsilon_{t-1} - \theta_{q,2}\varepsilon_{t-2} - \cdots - \theta_{q,q}\varepsilon_{t-q},$$

replacing AR(p). The algorithm yields:

1) the best model chosen among MA(1), ..., MA(6) for each $q \in \{1, 2, 3\}$ and $i \in \{1, \ldots, 6\}$ according to each of the criteria: MIMSA_5, AIC, BIC, AIC_c, and HQIC;
2) the values of the errors RMSE, MAPE, and MASE for each $q \in \{1, 2, 3\}$ and $i \in \{1, \ldots, 6\}$.

In Table IV we present a comparison the performance of MIMSA with those of the AIC, BIC, AIC_c, dan HQIC, in the case where the our actual training and testing datasets are generated using MA(1), with $\theta_{1,1} = 0.8$. A training dataset generated using this model is visualised in Figure 4. Table IV shows that BIC possesses the highest frequency in correctly determining the order of the employed MA model.

We observe that BIC correctly identifies MA(1) as many as 49 times, and misclassifies it as MA(2) once. On the other hand, the frequencies of correctly identifying the model for criteria other than BIC are lower, namely, 10 for MIMSA, 38 for AIC, 39 for AIC_c, and 45 for HQIC.

Our error computations reveal that the model MA(3) exhibits the best prediction performance compared to the other models, with the average RMSE and MASE values of 0.7557 and 0.6808, respectively. By contrast, the MA(6) model exhibits the worst performance, with the highest average RMSE of 0.7879 and the highest average MASE of 0.7070, indicating the largest prediction error. Meanwhile,

TABLE IV
MODEL IDENTIFICATION FREQUENCIES IN THE CASE WHERE OUR ACTUAL TRAINING DATASET IS GENERATED USING THE MODEL MA(1).

Parameter	Criterion	MA(1)	MA(2)	MA(3)	MA(4)	MA(5)	MA(6)
$\theta_{1,1} = 0.8$	MIMSA_5	10	9	6	6	10	9
	AIC	38	5	5	1	0	1
	BIC	49	1	0	0	0	0
	AICC_c	39	4	5	1	0	1
	HQIC	45	2	3	0	0	0

the MA(2) and MA(4) models are almost comparable in performance with MA(3), but with slightly higher average RMSE values of 0.7592 and 0.7682, respectively. Additionally, the MA(6) model, which has the highest MASE value in most experiments, indicates that increasing the MA order does not always improve model performance and could lead to larger prediction errors.

On the other hand, the MA(2) model performs the best, with the lowest average MAPE of 363.41%. However, the MA(6) model again has the highest average MAPE of 734.23%, indicating the largest prediction error and therefore unreliability. The high MAPE values across all models, particularly in one of our experiments where the MA(6) model achieves an exceptionally high MAPE value of 20,214.22%, show that the MA(6) model is not suitable as a predictive model for the actual dataset generated using MA(1).

From the boxplots shown in Figure 5, we can see that the model MA(3) exhibits lower medians in the cases of RMSE and MASE, with a narrower data range. This indicates that the MA(3) model performs notably well in prediction. By contrast, the other MA models are associated to wider data ranges and outliers in both RMSE and MASE, suggesting that their performance tends to be less stable and has a higher potential for large errors. The significant outliers indicate that models with higher orders may produce highly inaccurate predictions in some cases.

In Table V, we present a performance comparison of MIMSA, AIC, BIC, AICc, and HQIC in identifying the correct model in the case where our actual training and testing datasets are generated using MA(2). It is apparent that BIC has the highest number of times in identifying the correct order, namely 49. On the other hand, MIMSA records a suboptimal performance, with a notable difference in the frequency of correctly identifying the MA(2) model, achieving correctness only 13 times.

In terms of prediction errors, our computation shows that the model MA(2) exhibits the best performance based on its lowest average RMSE value of 0.6426, thereby demonstrating its lowest level of prediction errors compared to the other models. The model MA(3) stands out in prediction accuracy, with the lowest average MAPE of 260.54%, making it the most reliable model in minimising relative error. Furthermore, both models perform almost equally in terms of MASE, the average MASE values for MA(2) and MA(3) being 0.7466 and 0.7465, respectively. By contrast, MA(6) and MA(5) both show the notably poor performance, the former exhibiting the highest average RMSE of 0.6721 and MASE of 0.7746, the latter exhibiting a MAPE value of 425.08%, clearly suggesting unreliability.

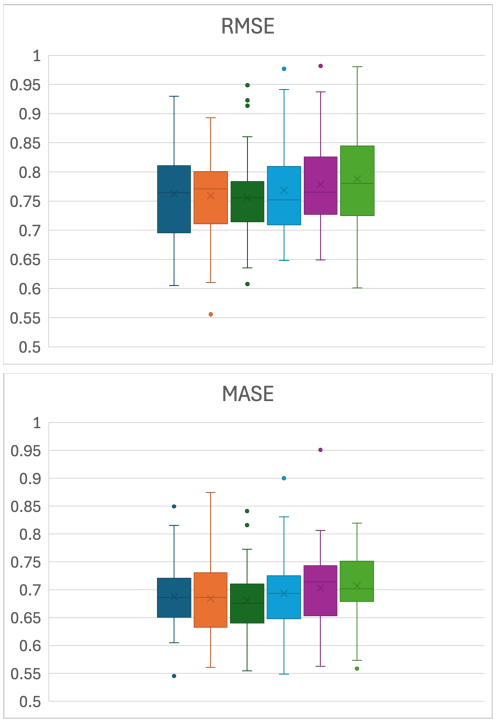

Fig. 5. Boxplots representing the values of RMSE and MASE associated to our testing datasets generated using the model MA(1).

TABLE V
MODEL IDENTIFICATION FREQUENCIES IN THE CASE WHERE OUR ACTUAL TRAINING DATASET IS GENERATED USING THE MODEL MA(2).

Parameter	Criterion	MA(1)	MA(2)	MA(3)	MA(4)	MA(5)	MA(6)
$\theta_{2,1} = 0.3$ $\theta_{2,2} = 0.5$	MIMSA_5	4	13	9	10	9	5
	AIC	0	38	5	4	2	1
	BIC	0	49	1	0	0	0
	AIC_c	0	39	5	4	2	0
	HQIC	0	45	4	1	0	0

A similar conclusion can be drawn from the boxplots shown in Figure 6. The model MA(2) demonstrates the best performance in prediction. In terms of RMSE, the MA(2) model has a low median value, with no outliers. On the other hand, despite the presence of outliers, MA(2) still excels in MASE, as indicated by a smaller median value and a narrower data range compared to the other models. By contrast, the model MA(6) consistently shows the worst performance, with higher median and average RMSE and MASE values, along with a wider data range.

Finally, the results in the case where the actual training dataset is generated using the model MA(3) shows that MIMSA performs worse than AIC, BIC, AIC_c, and HQIC in identifying the order of the utilised time-series model. This is demonstrated in Table VI.

In the case where the actual training dataset is generated using the model MA(3), the model MA(1) exhibits the lowest average RMSE, MAPE, and MASE values of 0.5895, 122.96%, and 0.7432, respectively. By contrast, the model MA(6) exhibits the highest average RMSE, MAPE, dan MASE values of 0.6491, 325.01%, and 0.8055, respectively.

Based on the RMSE and MASE boxplots in Figure 7, the model with the best overall performance tends to have lower average and median values, accompanied by a narrow interquartile range. For both RMSE and MASE, the model

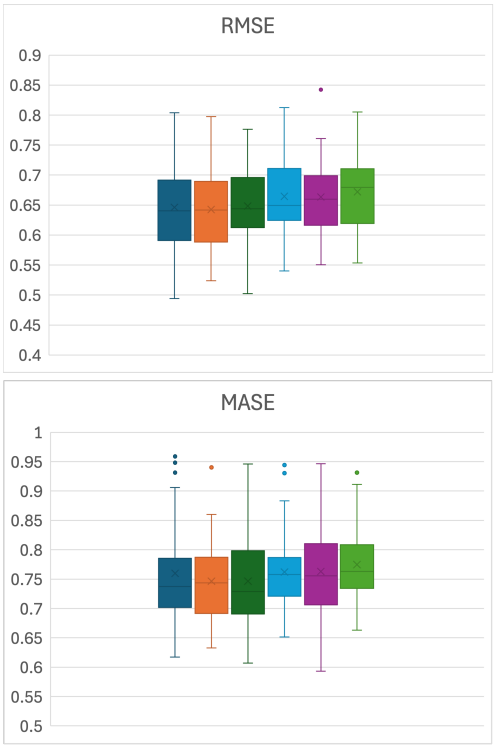

Fig. 6. Boxplots representing the values of RMSE and MASE associated to our testing datasets generated using the model MA(2).

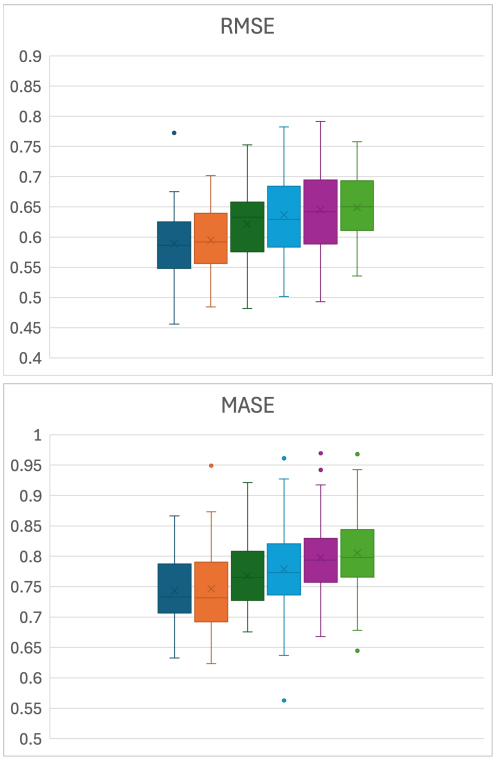

Fig. 7. Boxplots representing the values of RMSE and MASE associated to our testing datasets generated using the model MA(3).

TABLE VI
MODEL IDENTIFICATION FREQUENCIES IN THE CASE WHERE OUR ACTUAL TRAINING DATASET IS GENERATED USING THE MODEL MA(3).

Parameter	Criterion	MA(1)	MA(2)	MA(3)	MA(4)	MA(5)	MA(6)
$\theta_{3,1} = 0.1$ $\theta_{3,2} = 0.2$ $\theta_{3,3} = 0.5$	MIMSA_5	0	1	18	16	6	9
	AIC	0	0	35	10	2	3
	BIC	0	0	48	2	0	0
	AIC_c	0	0	37	8	2	3
	HQIC	0	0	44	5	0	1

MA(1) stands out, as it has lower average and median values compared to the other models, indicating a smaller prediction error. Additionally, the model MA(1) also shows a more stable distribution with fewer outliers. By contrast, the MA(6) model proves to be the worst-performing model based on the analysis of both error criteria.

IV. Conclusions and Future Research

The numerical experiments conducted in the present study lead to the following conclusions. First, MIMSA shows inferior performance compared to other criteria such as AIC, BIC, AIC_c, and HQIC in detecting the correct time series model. This is evident across AR(1), AR(2), and AR(3), as well as MA(1), MA(2), and MA(3), where MIMSA tends to select more complex models with higher orders. Moreover, based on the values of RMSE, MAPE, and MASE, MIMSA shows a range of different prediction performances, depending on the time-series model used to generate our actual dataset. In the case of AR models, MIMSA tends to be more consistent in providing predictions with high accuracy, especially for models with smaller orders, with better RMSE, MAPE, and MASE values compared to other models. However, in the case of MA models, the prediction performance of MIMSA is inconsistent. For example, in the case where our actual dataset is generated using MA(1), the most accurate RMSE and MAPE values are obtained from the MA(3) model, while in the case where our actual dataset is generated using MA(3), the MA(1) model provides better prediction results.

There are various recommendations for further studies. First, we recommend applying the method used in the present study to datasets with a larger size, to see whether the results are consistent or show any significant changes. One could also employ other classes of time-series models, such as ARMA and ARIMA. Additionally, since the present study consistently estimates mutual information using five subintervals, we also recommend the use of different numbers of subintervals. In fact, one could consider using more sophisticated methods to estimate mutual information, such as the non-parametric kernel density estimation [3] as well as the so-called k-nearest neighbours method [25], which may also necessitate numerical methods to compute double integrals, for which one could employ, for instance, the two-dimensional trapezoidal rule [28, sec. 6.7].

Acknowledgment

The authors would like to thank Taufik Limansyah and Melania Eva Wulanningtyas for their useful inputs.

References

[1] S. B. Abdullahi, A. H. Ibrahim, A. B. Abubakar, and A. Kambheera, "Optimizing Hammerstein-Wiener model for forecasting confirmed cases of COVID-19," IAENG International Journal of Applied Mathematics, vol. 52, no. 1, pp155-164, 2022.

[2] E. Akca, "A novel algorithm for time series model selection: mutual information model selection algorithm (MIMSA)," Masters thesis, Middle East Technical University, Turkey, 2017.

[3] E. Akca and C. Yozgatligil, "Mutual information model selection algorithm for time series," Journal of Applied Statistics, vol. 47, no. 12, pp2192-2207, 2020.

[4] H. Akaike, "A new look at the statistical model identification," IEEE Transactions on Automatic Control, vol. 19, no. 6, pp716-723, 1974.

[5] L. K. Amifia, "Model parameter identification of state of charge based on three battery modelling using Kalman filter," Engineering Letters, vol. 30, no. 3, pp1128-1137, 2022.

[6] G. E. Box, "Science and statistics," Journal of the American Statistical Association, vol. 71, no. 356, pp791-799, 1976.

[7] K. H. Chen and K. Khashanah, "Analysis of systemic risk: A vine copula-based ARMA-GARCH model," Engineering Letters, vol. 24, no. 3, pp268-273, 2016.

[8] Y. Cheng, H. Tao, X. Lu, Z. Xu, H. Fu, and C. Zhu, "Prediction method of axial compression capacity of CCFST columns based on deep learning," IAENG International Journal of Computer Science, vol. 51, no. 3, pp243-251, 2024.

[9] W. Cui, Z. Yan, and X. Peng, "A new Marshall Olkin Weibull distribution," Engineering Letters, vol. 28, no. 1, pp63-68, 2020.

[10] Y. Fang, L. Madsen, and L. Liu, "Comparison of two methods to check copula fitting," IAENG International Journal of Applied Mathematics, vol. 44, no. 1, pp53-61, 2014.

[11] E. J. Hannan and B. G. Quinn, "The determination of the order of an autoregression," Journal of the Royal Statistical Society: Series B (Methodological), vol. 41, no. 2, pp190-195, 1979.

[12] P. Hendikawati, Subanar, Abdurakhman, and Tarno, "ANFIS performance evaluation for predicting time series with calendar effects," IAENG International Journal of Applied Mathematics, vol. 51, no. 3, pp587-598, 2021.

[13] P. Hingley and G. Dikta, "Use of information criteria for finding Box-Jenkins time series models for patent filings counts forecasts," IAENG International Journal of Applied Mathematics, vol. 52, no. 1, pp7-17, 2022.

[14] J. Huang, "Does climate risk affect corporate bond yield? Take Shanghai as an example," IAENG International Journal of Applied Mathematics, vol. 53, no. 2, pp704-711, 2023.

[15] J. Huang, "Does climate risk have the same impact on corporate bond yields across credit ratings?," IAENG International Journal of Applied Mathematics, vol. 53, no. 2, pp566-572, 2023.

[16] C. M. Hurvich and C. L. Tsai, "Regression and time series model selection in small samples," Biometrika, vol. 76, no. 2, pp297-307, 1989.

[17] W. Z. W. Husin, M. S. Zainol, and N. M. Ramli, "Common factor model with multiple trends for forecasting short term mortality," Engineering Letters, vol. 24, no. 1, pp98-105, 2016.

[18] R. J. Hyndman and G. Athanasopoulos, Forecasting: Principles and Practice, 2nd ed., OTexts, Australia, 2018.

[19] R. J. Hyndman and A. B. Koehler, "Another look at measures of forecast accuracy," International Journal of Forecasting, vol. 22, no. 4, pp679-688, 2006.

[20] M. Irfan, F. Ramlie, Widianto, M. Lestandy, and A. Faruq, "Prediction of residential building energy efficiency performance using deep neural network," IAENG International Journal of Computer Science, vol. 48, no. 3, pp731-737, 2021.

[21] I. Ismail, G. Y. Yee, A. Zhang, and H. Zhou, "Currency exchange portfolio risk estimation using copula-based value at risk and conditional value at risk," IAENG International Journal of Applied Mathematics, vol. 53, no. 3, pp892-898, 2023.

[22] A. F. Jamali, A. Mustapha, and S. A. Mostafa, "Prediction of sea level oscillations - Comparison of regression-based approach," Engineering Letters, vol. 29, no. 3, pp990-995, 2021.

[23] I. G. N. M. Jaya, Y. Andriyana, and B. Tantular, "Spatial prediction of malaria risk with application to Bandung City, Indonesia," IAENG International Journal of Applied Mathematics, vol. 51, no. 1, pp199-206, 2021.

[24] U. Khair, H. Fahmi, S. Al Hakim, and R. Rahim, "Forecasting error calculation with mean absolute deviation and mean absolute percentage error," Journal of Physics: Conference Series, vol. 930, 012002, 2017.

[25] A. Kraskov, H. Stögbauer, and P. Grassberger, "Estimating mutual information," Physical Review EStatistical, Nonlinear, and Soft Matter Physics, vol. 69, no. 6, 066138, 2004.

[26] Y. P. Li, J. S. Wang, and M. W. Wang, "Prediction model of short-term load in power system based on interval type-2 fuzzy logic," IAENG International Journal of Computer Science, vol. 48, no. 3, pp646-652, 2021.

[27] L. Li and Z. Xie, "Model selection and order determination for time series by information between the past and the future," Journal of Time Series Analysis, vol. 17, no. 1, pp65-84, 1996.

[28] P. Linz and R. Wang, Exploring Numerical Methods: An Introduction to Scientific Computing Using MATLAB, Jones and Bartlett Publishers, Massachusetts, 2003.

[29] H. Long, Y. He, W. Xiang, Z. Guan, H. Tan, and J. Yu, "Research on short-term wind speed prediction based on adaptive hybrid neural network with error correction," IAENG International Journal of Computer Science, vol. 50, no. 4, pp1290-1304, 2023.

[30] A. R. Lubis, N. M. B. Ginting, S. F. Ayu, R. E. Caraka, Y. Kim, P. U. Gio, and B. Pardamean, "Business predictive analytics of smallholder Indonesian maize using vector error correction," Engineering Letters, vol. 32, no. 2, pp429-435, 2024.

[31] Z. Luo, "Cultural diversity effect onto CO_2 emissions: Evidence from developed and emerging markets," IAENG International Journal of Applied Mathematics, vol. 54, no. 6, pp1198-1209, 2024.

[32] C. Ma, J. Wu, H. Hu, Y. N. Chen, and J. Y. Li, "Predicting stock prices using hybrid LSTM and ARIMA model," IAENG International Journal of Applied Mathematics, vol. 54, no. 3, pp424-432, 2024.

[33] A. Mendyk, R. Jachowicz, K. Fijorek, P. Dorozyński, P. Kulinowski, and S. Polak, "Kinetds: An open source software for dissolution test data analysis," Dissolution Technologies, vol. 19, no. 1, pp6-11, 2012.

[34] N. K. Micheni, E. B. Atitwa, and P. M. Kimani, "Prediction of domestic value-added tax in Kenya using SARIMA and Holt-Winters methods," IAENG International Journal of Applied Mathematics, vol. 55, no. 3, pp594-602, 2025.

[35] F. N. N. A. Mutalip, I. Ismail, and K. S. Long, "Capital requirement for non-life insurance industry using D-vine copula: An empirical evidence from Malaysia," IAENG International Journal of Applied Mathematics, vol. 54, no. 8, pp1694-1704, 2024.

[36] M. A. Pitt and I. J. Myung, "When a good fit can be bad," Trends in Cognitive Sciences, vol. 6, no. 10, pp421-425, 2002.

[37] A. B. Salsabila, S. Sutisna, and S. Purwani, "Using VARI model to forecast climate phenomena in big data era," IAENG International Journal of Applied Mathematics, vol. 54, no. 12, pp2711-2718, 2024.

[38] H. Sen, "Time series prediction based on improved deep learning," IAENG International Journal of Computer Science, vol. 49, no. 4, pp1133-1138, 2022.

[39] G. Schwarz, "Estimating the dimension of a model," The Annals of Statistics, vol. 6, no. 2, pp461-464, 1978.

[40] L. B. Theng, "Automatic building extraction from satellite imagery," Engineering Letters, vol. 13, no. 3, pp255-259, 2006.

[41] T. Toharudin, R. S. Pontoh, R. E. Caraka, S. Zahroh, A. Akbar, B. Pardamean, and R. C. Chen, "Indonesia in facing new normal: An evidence hybrid forecasting of COVID-19 cases using MLP, NNAR, and ELM," Engineering Letters, vol. 29, no. 3, pp749-758, 2021.

[42] X. Yang and L. Liu, "Research on traffic flow prediction based on chaotic time series," IAENG International Journal of Applied Mathematics, vol. 53, no. 3, pp1007-1011, 2023.

[43] J. Zhao, C. Zhu, and K. Zhao, "Spatial and temporal characteristics of the impact of TOD built environment on rail transit riders' travels," IAENG International Journal of Applied Mathematics, vol. 55, no. 1, pp65-73, 2025.

[44] T. Zhu, "Application of a new model for complex systems," Engineering Letters, vol. 29, no. 2, pp471-476, 2021.

[45] C. Zhu, P. Lu, W. Feng, and Y. Wang, "Bimodal stock price prediction based on Holt-Winters exponential smoothing and PCA whitening transformation," IAENG International Journal of Computer Science, vol. 52, no. 1, pp187-200, 2025.