Proceedings of the International Conference on Industrial Engineering and Operations Management

Publisher: IEOM Society International, USA DOI: 10.46254/AP06.20250190
Published: August 5, 2025

Comparative Study of VAR-IMMA and MICE PMM in Handling Missing Data

Embay Rohaeti
Mathematics Study Program
Pakuan University
Bogor, West Java, Indonesia
embay.rohaeti@unpak.ac.id

Muhammad Edy Rizal
Data Science Study Program
Tadulako University
Palu, Central Sulawesi, Indonesia
muhedyrizal@untad.ac.id

Abstract

This study aims to evaluate and compare the accuracy of two imputation methods, VAR-IMMA (Vector Autoregressive - Moving Average Imputation Method) and MICE PMM (Multiple Imputation by Chained Equations with Predictive Mean Matching), for handling missing data in multivariate time series. The evaluation was conducted using synthetic time series data generated from a stable VAR(1) model. Missing values were introduced under the Missing at Random (MAR) mechanism across five different missingness levels: 5%, 10%, 15%, 20%, and 25%. Each scenario was repeated 100 times. After imputation, each completed dataset was fitted using a VAR model, and the fitting accuracy was assessed by comparing the fitted values with the original data using RMSE and MAPE as performance metrics. The results showed that VAR-IMMA consistently outperformed MICE PMM across all scenarios, particularly at lower levels of missingness. These findings suggest that VAR-IMMA is more effective in preserving the temporal and inter-variable relationships in multivariate time series with MAR missing data.

Keywords
VAR-IMMA, MICE PPM, Data Imputation, Missing Data, Time Series

1. Introduction
One of the major challenges in multivariate time series forecasting is the occurrence of missing data, which can result from various causes such as recording errors, system failures, or external factors (Bansal et al., 2020; Jin et al., 2024; Yu et al., 2024). Incomplete data can significantly impact the reliability of statistical models, especially in applications requiring high precision in forecasting (Gong et al., 2023; Lee and Charles Huber, 2021; Nijman et al., 2022). Therefore, it is essential to implement appropriate techniques to handle missing data effectively to maintain the accuracy and validity of the forecasting results.

One commonly encountered missing data mechanism is Missing at Random (MAR), where the probability of missingness depends on the observed data but not on the missing data itself. This characteristic allows for statistical imputation methods to be applied, assuming that the information available is sufficient to estimate the missing values with reasonable accuracy (Sumertajaya et al., 2023).

This study focuses on two imputation methods designed to address missing data in multivariate time series: VAR-IMMA (Vector Autoregressive – Moving Average Imputation Method) and MICE PMM (Multiple Imputation by Chained Equations using Predictive Mean Matching). The VAR-IMMA method is a model-based approach that uses a combination of Vector Autoregressive (VAR) and Moving Average (MA) models to estimate missing values. This method is specifically tailored for multivariate time series data where the relationships between variables over time play a critical role in capturing the underlying data structure (Rohaeti et al., 2022). On the other hand, MICE PMM is a widely used iterative statistical method for handling missing data. It works by sequentially imputing each incomplete variable using regression models conditioned on the other variables in the dataset, and incorporates predictive mean matching to preserve the distributional characteristics of the data (Javadi et al., 2021; Mera-Gaona et al., 2021; Samad et al., 2022; Seu et al., 2022).

To compare the performance of these two methods, this study uses synthetically generated time series data with known structures. Missing values are introduced at varying proportions, simulating different levels of data loss under MAR assumptions. After the imputation is performed using both methods, the completed datasets are compared against the original data. The goal is to identify which method produces more accurate imputation results and thus provides a more reliable basis for subsequent forecasting.

Previous research by Bashir and Wei (2017) introduced the Vector Autoregressive Imputation Model (VAR-IM) and showed that it achieved significant improvements in imputation accuracy compared to other existing methods. The VAR-IM method was further developed into the VAR-IMMA approach, which demonstrated improved performance over its predecessor (Sumertajaya et al., 2023). In addition, Gaona et al. (2021) found that MICE PMM outperformed basic imputation methods across all levels of missingness in a simulation study on breast cancer data, suggesting the method's robustness under various conditions.

Based on the issues and findings, this study is titled "Evaluation of the Performance of VAR-IMMA and MICE Methods in Handling Missing Data in Multivariate Time Series."

1.1 Objectives

The objective of this study is to evaluate and compare the performance of two imputation methods: VAR-IMMA (Vector Autoregressive – Moving Average Imputation Method) and MICE PMM (Multiple Imputation by Chained Equations with Predictive Mean Matching), in handling missing data within multivariate time series. This study aims to determine which method is more effective in reconstructing missing values in multivariate time series data, based on model accuracy indicators.

2. Literature Review

2.1 Numerical Results

According to Rohaeti et al. (2022), time series data is divided into two types based on the number of variables used in the analysis: (1) Univariate time series analysis, involves only one variable or a single sequence of observations over time, using models such as Autoregressive (AR), Moving Average (MA), and Autoregressive Integrated Moving Average (ARIMA); (2) Multivariate Time Series, involves more than two variables in the sequence of observations over time, using models such as Vector Autoregressive (VAR).

2.2 Moving Average

According to Montgomery et al. (2015), Moving Average (MA) is a statistical technique commonly used in technical analysis to calculate the average of a set of data over a specific time period. Moving averages are divided into several types, including the Simple Moving Average (SMA) and the Exponential Moving Average (EMA). Both SMA and EMA are smoothing methods used to analyze patterns in time series data. The formula for the Simple Moving Average (SMA) is as follows.

$$\boldsymbol{EMA_t = \frac{(1-\alpha)^k y_{t-k} + \cdots + (1-\alpha)^1 y_{t-1} + (1-\alpha)^1 y_{t+1} + \cdots + (1-\alpha)^k y_{t+k}}{(1-\alpha)^1 + (1-\alpha)^2 + \cdots + (1-\alpha)^{k-1}}}$$

Where

M_t = Forecast for the upcoming period
y_{t-1} = Value from the previous period up to n periods
n = Number of periods averaged

Meanwhile, the formula for the Exponential Moving Average (EMA) is as follows.

$$EMA_t = \frac{(1-\alpha)^k y_{t-k} + \cdots + (1-\alpha)^1 y_{t-1} + (1-\alpha)^1 y_{t+1} + \cdots + (1-\alpha)^k y_{t+k}}{(1-\alpha)^1 + (1-\alpha)^2 + \cdots + (1-\alpha)^{k-1}}$$

Where

α = Smoothing factor
EMA_t = EMA values at time t
y_{t-k} = Actual value at previous and future times within a symmetric time window of size k
$(1-\alpha)^i$ = Exponential weight for each data point
k = time index

2.3 Missing Data

According to Sumertajaya et al. (2023), actual data often contains incompleteness. This incompleteness, commonly referred to as missing data, is one of the most prevalent data quality issues. It can be caused by various factors, such as respondent non-compliance, technical problems during data recording, incomplete information, or lack of knowledge. Moreover, missing data is a critical issue, especially in time series datasets, where data points are interdependent over time. In analyzing missing data, it is important to consider the types, patterns, and tests for missing data to ensure appropriate handling.

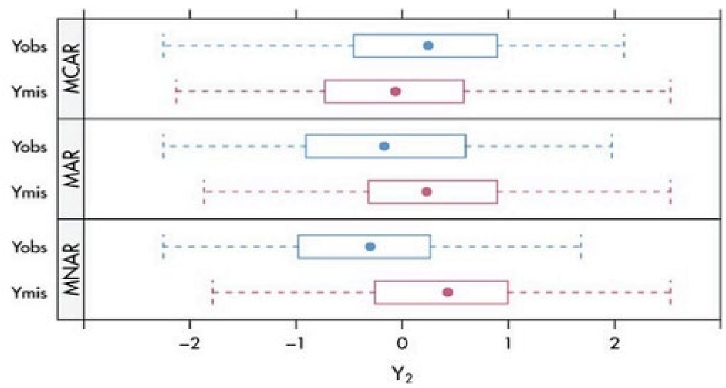


Figure 1. Pattern of Missing Data
Source: (Buuren, 2018)

Table 1. Description and Characteristics of Missing Data Patterns

Mechanism	Description	Visual Characteristics
MCAR (Missing at Completely Random)	Missing data is entirely random and not dependent on any variable, whether observed or missing.	Boxplots of Y_{obs} and Y_{miss} are nearly identical, indicating similar distributions.
MAR (Missing at Random)	Missing data is related only to observed data.	Boxplots of Y_{obs} and Y_{miss} show slight differences but still overlap, indicating moderate distribution differences.
MNAR (Missing Not at Random)	Missing data depends on the value of the missing variable itself or on unobserved data not present in the dataset.	Boxplots of Y_{obs} and Y_{miss} are significantly different with no overlap, indicating strong bias.

According to Buuren (2018), missing data can be categorized into three types: Missing Completely at Random (MCAR), Missing at Random (MAR), and Missing Not at Random (MNAR). Each type requires different handling strategies, so identifying the type and pattern of missing data is essential before applying any treatment. These patterns and definitions are illustrated in Figure 1.

The Figure 1 shows missing data patterns based on the three main mechanisms: MCAR, MAR, and MNAR. Each mechanism is illustrated using boxplots for observed data (Yobs) and missing data (Ymis) with respect to variable $\boldsymbol{Y_2}$. The descriptions and visual characteristics of each mechanism are summarized in Table 1.

2.4 VAR-IMMA

According to Sumertajaya et al. (2023), the VAR-IMMA method is a missing data imputation method based on VAR and moving average. The VAR-IMMA algorithm is as follows:

1. VAR-IMMA begins with initial imputation using the moving average method. Next is the determination of the time window, which is the range of values to be averaged using the moving average. The time window is defined as $2k$, with k being an integer.
2. Initial analysis is carried out by averaging all values within the time window. If there is missing data, such as in the case of consecutive missing data, the missing data will be ignored. However, if all data within the time window are missing, then the time window must be expanded each time period until there is at least one data point. Once there is at least one data point in the time window, the average is calculated to obtain the initial imputation data.
3. The initial imputation values obtained are then modeled using VAR(p), resulting in the initial coefficient matrixβ_0.
4. The initial imputation stage is complete, and the process continues with the Estimation Maximization (EM) stage. The EM stages are as follows:
   4.1. The first step in EM is to build the best VAR(p) model based on the initial imputation data.
   4.2. Next, in the E-Step, the estimated values from the best VAR(p) model are used to update the estimation of the missing values.
   4.3. Then, in the M-Step, the VAR(p) model is updated based on the latest imputation results.
   4.4. The E-step and M-step process is repeated until convergence is achieved, which is when $\beta_{iter} - \beta_{iter-1} < \zeta$, where β_{iter} is the coefficient matrix at the current iteration, and βiter-1βiter-1 is the coefficient matrix at the previous iteration. $\beta_{iter} - \beta_{iter-1}$ is the Frobenius norm of the difference between the two matrices, while ζ is a constant.
5. After convergence is reached, rolling forecasting is performed for each missing data point. Rolling forecasting is carried out one period ahead, and if there are consecutive missing data points, the forecast result at period t will be used for forecasting at period t+1. Once the forecasting process is completed for all missing data, the final result is obtained, which is the final imputed data.

2.5 MICE PPM

According to Gaona et al. (Gaona et al., 2021), Multiple Imputation by Chained Equations (MICE) is a repeated imputation method based on fully conditional specification, where different models impute incomplete values.

$$\bar{P} = \frac{1}{m}\sum_{i=1}^{m}\widehat{P_\iota}$$

Where

$\bar{P}$ = Mean value

$\frac{1}{m}$ = Total number of measurements

$\sum_{i=1}^{m}\widehat{P_\iota}$ = Total estimate of value PP at the ii-th observation

After handling missing data using the VAR-IMMA method, the next step is handling missing data using the MICE method. The algorithmic steps of MICE are as follows:

a. Imputation process: At this stage, each variable is initially imputed using a simple method such as mean imputation. All missing values are replaced by the observed mean of the respective variable.
b. Reassignment of missing data in F_x: After imputation, the missing values are reassigned using F_x.
c. Each variable is linked to missing data: The observed values in F_x are linked to other variables in the imputation model, which may consist of a subset of variables in the dataset. At this stage, $F_x$becomes the dependent variable, while the other variables serve as independent variables.
d. Substitution of F_x values: The missing values for F_x are then substituted with imputations from each variable linked in the previous step. These are then used as independent variables in each variable linked to other variables, including both observed and imputed values.
e. Repeat steps b–d until the n-th iteration: At this stage, for each F_x with missing values, steps b–d are repeated. This process continues until all F_x values with missing data have been substituted. The repetition across these

variables is called one iteration or cycle. At the end of the iteration, all missing values have been replaced with imputations from F_x that reflect the relationships observed in the data.

f. Repeat steps b–e for n iterations: In the next stage, steps b–e are repeated for n iterations, and the data are filled or updated in each iteration with imputed values. The purpose of repeating these steps is to achieve stable imputations.

The MICE method used is the mean imputation method or MICE Predictive Mean Matching (PMM). The MICE method with the PMM (Predictive Mean Matching) approach works by forming relationship patterns for each variable with missing data, then matching the relationship values with observed values. This process preserves the original data distribution and prevents outliers due to extreme imputations. The missing values are then filled with actual values from observations that have the closest predictions.

3. Methods

This study employed a simulation-based experimental design to evaluate the performance of two imputation methods, VAR-IMMA (Vector Autoregressive – Moving Average Imputation Method) and MICE PMM (Multiple Imputation by Chained Equations with Predictive Mean Matching), in handling missing data in multivariate time series. The simulation approach allowed for a controlled comparison, as the true values of the data were known and could be used to assess imputation accuracy.

Synthetic multivariate time series data were generated using a stable Vector Autoregressive model of order one, VAR(1), with five variables and 200 time points. The data generation process used a coefficient matrix and white noise components that met the stationarity conditions, ensuring that the simulated series reflected realistic time-dependent behavior.

To simulate missing data, values were omitted under the Missing at Random (MAR) mechanism at five different proportions: 5%, 10%, 15%, 20%, and 25%. For each missingness level, the process was repeated 100 times to account for randomness and to obtain stable average performance estimates.

Two imputation methods were applied to the incomplete datasets. VAR-IMMA is a model-based imputation approach that utilizes both the temporal structure of the data through the VAR model and smoothing via moving average iteration. This method is designed specifically for multivariate time series data and aims to preserve both time-lagged and inter-variable dependencies. In contrast, MICE PMM is a general-purpose multiple imputation method that fills in missing values through iterative regression modeling, using predictive mean matching to ensure the imputed values are plausible within the distribution of the observed data.

Following imputation, each completed dataset was fitted using a VAR model. The imputed values were then compared to the original (complete) data. This study is motivated by practical applications in time series modeling, where imputation is often followed by model fitting. Therefore, the performance of an imputation method should be evaluated not only based on how well it recovers the original values, but also on how well it supports downstream modeling. By fitting a VAR model after imputation and comparing the resulting model output to the complete data, we assess the suitability of each imputation method for predictive modeling contexts, which in this case, is used for multivariate time series modeling.

The accuracy of each imputation method was assessed using two standard metrics: Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE). These metrics were averaged over 100 simulation repetitions at each missingness level to evaluate the robustness and consistency of each method's performance.

4. Data Collection

The data used in this study were simulated to allow for controlled experimentation in evaluating the performance of imputation methods. The simulated dataset consisted of five variables, each with 5082 observations, generated from a stable Vector Autoregressive (VAR) model of order one, VAR(1). After we obtained 5082 observations, we performed burning process by removing the first 100 observations and collected the remaining 4982 observations. The simulation was conducted to resemble the structure of multivariate time series data assumed to be stationary.

The data generation process was repeated 100 times to create independent simulation runs. These replicated datasets were used for both introducing missingness and performing imputation evaluations at multiple levels of data loss. The missing proportions are 5%, 10%, 15%, 20%, and 25%, which was introduced randomly using missing at random (MAR) mechanism.

5. Results and Discussion

This section presents the comparative evaluation of the VAR-IMMA and MICE PMM methods across five levels of missing data: 5%, 10%, 15%, 20%, and 25%. The results are summarized numerically in Table 2 and illustrated graphically in Figure 4 to Figure 8. Figure 2 and Figure 3 shows illustrations of the generated and missing data that were used in this study.

5.1 Numerical Results

Table 2 shows the average RMSE and MAPE values along with their standard errors (SE) based on 100 simulation repetitions for each level of missingness. Across all levels, the VAR-IMMA method consistently achieved lower RMSE and MAPE values compared to MICE PMM.

Table 2. Evaluation Summary of 100 Iterations

Proportion	**Metric**	**VAR-IMMA**		**MICE PPM**	
		Mean	**SE**	**Mean**	**SE**
5%	MAPE	2.087022	1.030175	2.207897	0.919010
10%		2.030817	0.834653	2.521124	1.330119
15%		2.007304	0.852527	3.050228	1.686791
20%		1.984663	0.844531	3.301249	1.677551
25%		1.969538	0.837095	3.573409	1.494608
5%	RMSE	0.995341	0.004954	1.082012	0.255555
10%		0.991595	0.005535	1.256256	0.500937
15%		0.988061	0.006188	1.551210	0.621727
20%		0.983610	0.007574	1.676186	0.613335
25%		0.990658	0.121716	1.854589	0.537854

At the lowest missingness level of 5%, the average MAPE for VAR-IMMA was 2.09 with an RMSE of 0.995, while MICE PMM yielded a slightly higher MAPE of 2.21 and RMSE of 1.082. As the proportion of missing data increased, the performance gap widened.

At 25% missingness, VAR-IMMA's average RMSE remained relatively low (0.991), whereas MICE PMM's RMSE increased significantly to 1.855. A similar pattern is observed in the MAPE scores, where VAR-IMMA maintained stable performance (MAPE = 1.97), while MICE PMM's MAPE rose to 3.57. This suggests that VAR-IMMA is more robust to increasing levels of missingness and better preserves the temporal and multivariate structure of the data.

These findings support the hypothesis that a model-based imputation method like VAR-IMMA, which explicitly incorporates time-lagged dependencies and cross-variable relationships, is better suited for multivariate time series data compared to MICE PMM, which is primarily distribution-based and not inherently time-aware.

5.2 Graphical Results

Figure 2 presents the distribution of the generated data in the first simulation. As can be shown, all five variables are on different ranges. The data shown in Figure 2 is from the first iteration, but it represents the characteristics of all of the 100 generate data, where the values range is from around 20 to 100 with variety of ranges across variables.

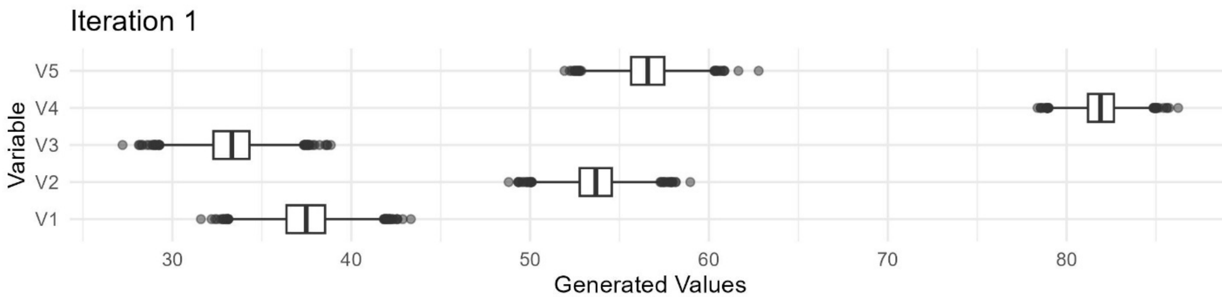


Figure 2. Distribution of the Generated Data, Illustrated by the First Iteration

Figure 3 illustrates how missing values were randomly introduced under the MAR mechanism with 5% missingness. Red line in the figures indicates the location of the missing data. The data shown in Figure 3 is from iteration 11, where the missing data were present in the first and second variables. This is not necessarily true in other generated data, where the missing data may present in variables other than the first and second variables.

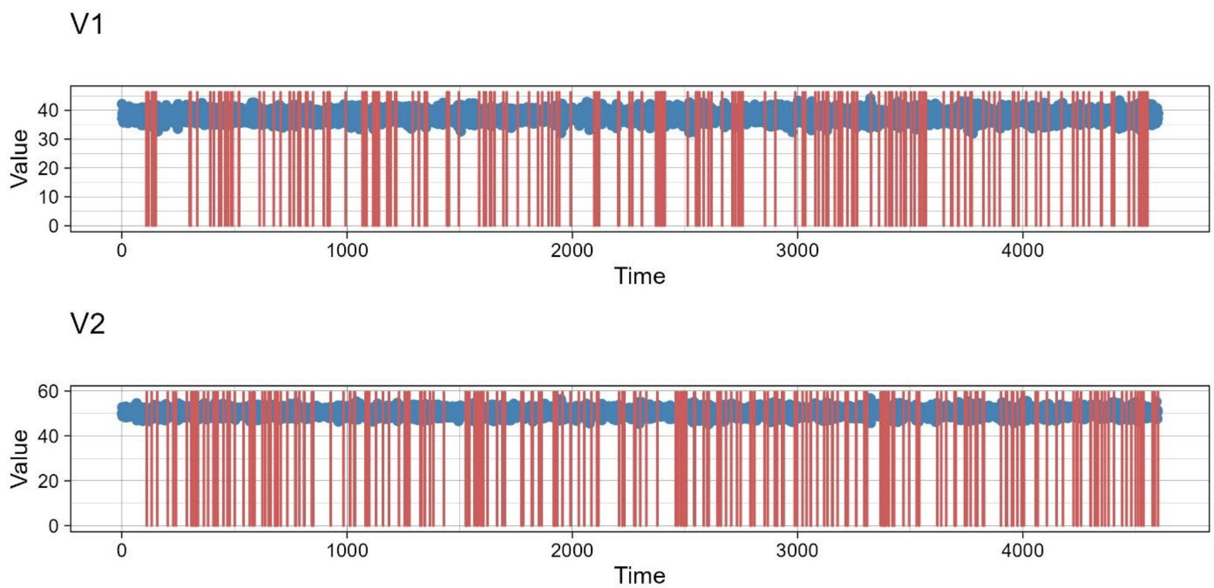


Figure 3. Illustration of Randomly Assigned Missing Data with a Proportion of 5% from Iteration 11

Figure 4 to Figure 8 provide visual comparisons of RMSE and MAPE scores at each missingness level. The left figures show the RMSE values, while the right figures show the MAPE scores in percent.

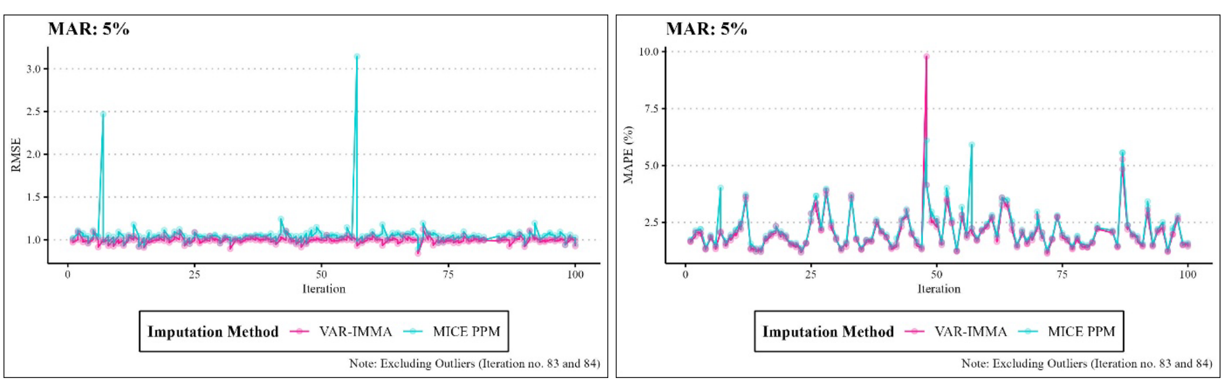


Figure 4. RMSE and MAPE Scores of VAR-IMMA's and MICE PPM's Imputation at 5% Missing Data

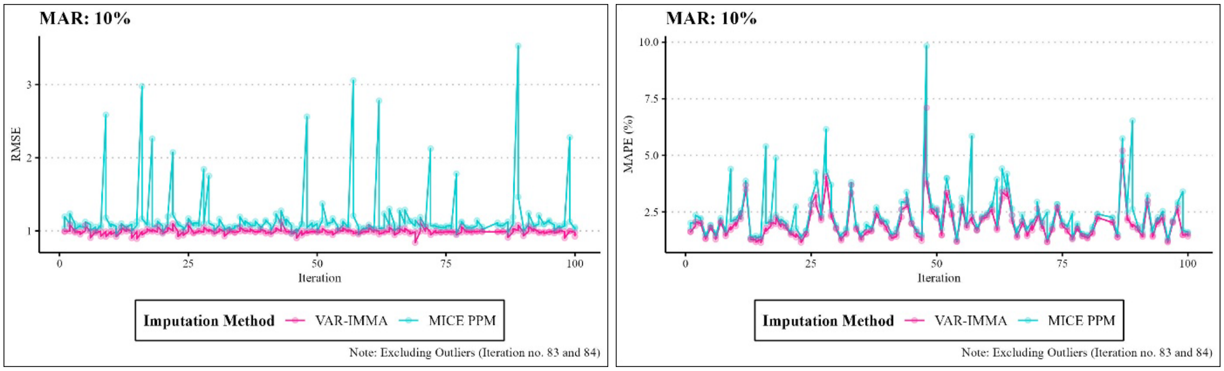


Figure 5. RMSE and MAPE Scores of VAR-IMMA's and MICE PPM's Imputation at 10% Missing Data

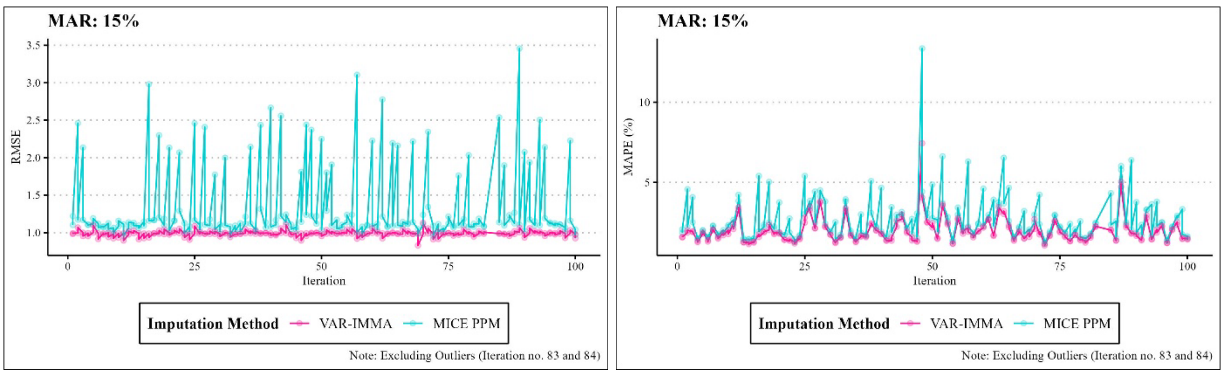


Figure 6. RMSE and MAPE Scores of VAR-IMMA's and MICE PPM's Imputation at 15% Missing Data

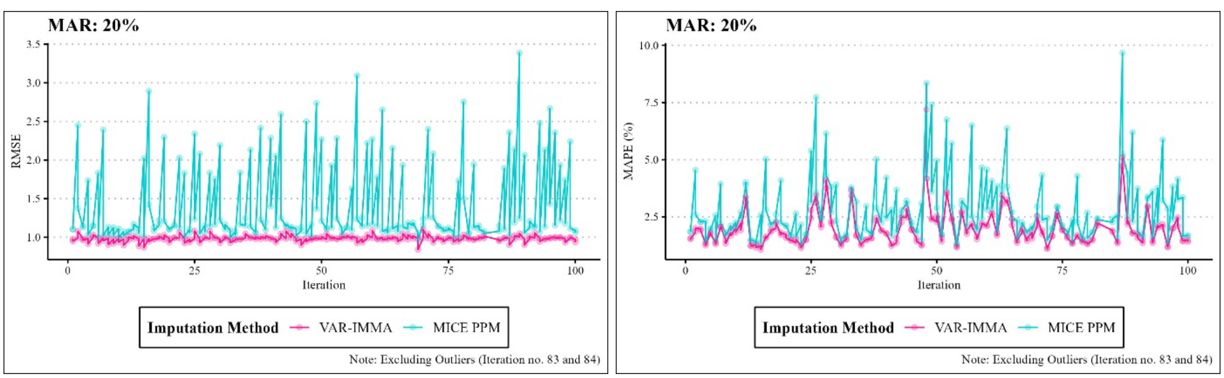
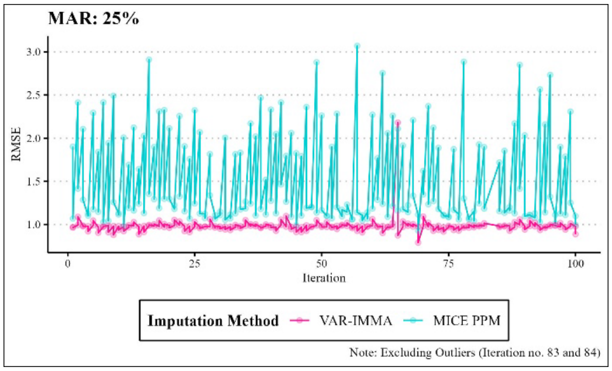


Figure 7. RMSE and MAPE Scores of VAR-IMMA's and MICE PPM's Imputation at 20% Missing Data



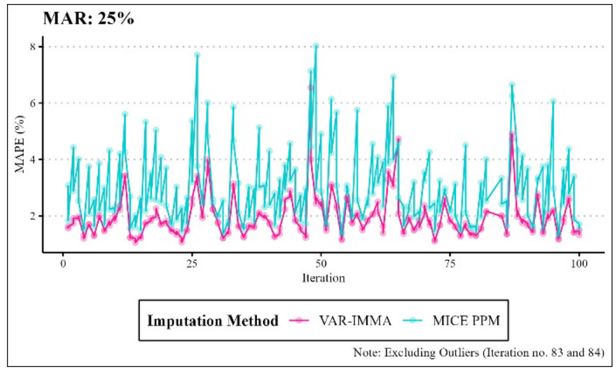


Figure 8. RMSE and MAPE Scores of VAR-IMMA's and MICE PPM's Imputation at 25% Missing Data

As can be shown in Figure 4 to Figure 8, VAR-IMMA consistently appears below MICE PMM in both RMSE and MAPE bars, consistent with the summary results from Table 2. The variance in error metrics for MICE PMM is also visibly higher, especially as missingness increases, which indicates less stable performance across simulations.

These visual comparisons help clarify the consistency and reliability of VAR-IMMA under different missing data conditions. The compact distribution of error metrics for VAR-IMMA suggests stronger adaptability in maintaining data integrity during the imputation process, even when substantial portions of the data are missing.

5.3 Proposed Improvements

This study proposes an improved approach to handling missing data in multivariate time series by demonstrating the effectiveness of the VAR-IMMA method compared to the widely used MICE PMM. Unlike MICE PMM, which does not explicitly model temporal dependencies, VAR-IMMA make use of the autoregressive and moving average structures inherent in time series data to impute missing values. The simulation results show that VAR-IMMA consistently yields lower RMSE and MAPE across various levels of missingness, indicating that it better preserves both the temporal and multivariate structure of the data.

This improvement is particularly relevant in applications where maintaining the dynamic relationships among variables is critical for reliable forecasting. By evaluating both methods under controlled simulation settings, this study offers empirical evidence supporting the adoption of VAR-IMMA as a more accurate and robust imputation method for time series data with MAR-type missingness.

5.4 Validation

Validation of the simulation process and the imputation methods was conducted through repeated experimentation. Each simulation was replicated 100 times for every level of missingness to ensure that the results were not influenced by random variation. The consistency of RMSE and MAPE values across repetitions, especially in the case of VAR-IMMA, supports the reliability of the method under MAR conditions.

In addition, the results show clear trends in error metrics: as the proportion of missing data increases, MICE PMM's performance declines more steeply compared to VAR-IMMA. This consistent pattern across all missingness levels validates the comparative advantage of VAR-IMMA in maintaining model accuracy despite data loss.

Although statistical hypothesis testing was not explicitly applied in this study, the standard errors of the RMSE and MAPE values across repetitions indicate that the observed differences in performance between methods are not due to random variation. Future studies may include formal hypothesis tests such as paired t-tests or Wilcoxon signed-rank tests to further confirm the significance of performance differences.

6. Conclusion

This study evaluated and compared the performance of two imputation methods, VAR-IMMA (Vector Autoregressive – Moving Average Imputation Method) and MICE PMM (Multiple Imputation by Chained Equations with Predictive Mean Matching) in handling missing data in multivariate time series. Using synthetically generated data under a

VAR(1) model, missing values were introduced at varying proportions under the Missing At Random (MAR) mechanism. Both methods were applied to impute the missing data, and the accuracy of the imputed datasets was assessed using RMSE and MAPE metrics after fitting a VAR model.

The results demonstrate that VAR-IMMA consistently outperforms MICE PMM in terms of lower RMSE and MAPE values across all levels of missingness. Moreover, VAR-IMMA exhibits more stable performance as the proportion of missing data increases, while MICE PMM's accuracy deteriorates more rapidly. These findings highlight the importance of using imputation methods that account for the temporal and cross-variable dependencies inherent in multivariate time series data.

The key contribution of this study is to provide empirical evidence that supports the adoption of VAR-IMMA as a more effective method for imputing missing values in time series settings, particularly when the goal is to preserve data integrity for subsequent modeling. This has practical implications in fields such as environmental monitoring, finance, and public health, where time-dependent data are often incomplete but critical for decision-making.

Future studies may extend this work by applying VAR-IMMA to real-world datasets, evaluating its performance in the presence of other missing data mechanisms (such as MNAR), or combining it with machine learning models to explore hybrid approaches.

References

Bansal. P.. Deshpande. P. and Sarawagi. S.. Missing value imputation on multidimensional time series. PVLDB. vol. 14. 2020.

Bashir. F. and Wei. H. L.. Handling missing data in multivariate time series using a vector autoregressive model-imputation (VAR-IM) algorithm. Neurocomputing. vol. 276. pp. 23–30. 2017.

Buuren. S. van.. Flexible Imputation of Missing Data. CRC Press. 2018. Available: https://doi.org/10.18637/jss.v085.b04

Gaona. M.. Neumann. U.. Vargas-Canas. R. and López. D. M.. Erratum: Evaluating the impact of multivariate imputation by MICE in feature selection. PLoS One. vol. 16. 2021. Available: https://doi.org/10.1371/journal.pone.0261739

Gong. Y.. Liu. G.. Xue. Y.. Li. R.. and Meng. L.. A survey on dataset quality in machine learning. Information and Software Technology. vol. 162. 107268. 2023. Available: https://doi.org/10.1016/J.INFSOF.2023.107268

Javadi. S.. Bahrampour. A.. Saber. M. M.. Garrusi. B. and Baneshi. M. R.. Evaluation of four multiple imputation methods for handling missing binary outcome data in the presence of an interaction between a dummy and a continuous variable. Journal of Probability and Statistics. vol. 2021. pp. 1–14. 2021. Available: https://doi.org/10.1155/2021/6668822

Jin. M.. Koh. Y.. Wen. Q.. Zambon. D.. Alippi. C.. Webb. G. I.. King. I. and Pan. S.. A survey on graph neural networks for time series: forecasting, classification, imputation, and anomaly detection. IEEE Transactions on Pattern Analysis and Machine Intelligence. vol. 46. pp. 10466–10485. 2024.

Lee. J. H. and Huber. J. C.. Evaluation of multiple imputation with large proportions of missing data: how much is too much? Iranian Journal of Public Health. vol. 50. pp. 1372–1380. 2021.

Mera-Gaona. M.. Neumann. U.. Vargas-Canas. R. and López. D. M.. Evaluating the impact of multivariate imputation by MICE in feature selection. PLoS One. vol. 16. 2021. Available: https://doi.org/10.1371/journal.pone.0254720

Montgomery. D. C.. Jennings. C. L. and Kulahci. M.. Introduction to Time Series Analysis and Forecasting. Wiley. 2015.

Nijman. S. W. J.. Leeuwenberg. A. M.. Beekers. I.. Verkouter. I.. Jacobs. J. J. L.. Bots. M. L.. Asselbergs. F. W.. Moons. K. G. M. and Debray. T. P. A.. Missing data is poorly handled and reported in prediction model studies using machine learning: a literature review. Journal of Clinical Epidemiology. vol. 142. pp. 218–229. 2022. Available: https://doi.org/10.1016/J.JCLINEPI.2021.11.023

Rohaeti. E.. Sumertajaya. I. M.. Wigena. A. H. and Sadik. K.. The prominence of vector autoregressive model in multivariate time series forecasting models with stationary problems. BAREKENG: Jurnal Ilmu Matematika dan Terapan. vol. 16. pp. 1313–1324. 2022. Available: https://doi.org/10.30598/barekengvol16iss4pp1313-1324

Samad. M. D.. Abrar. S. and Diawara. N.. Missing value estimation using clustering and deep learning within multiple imputation framework. Knowledge-Based Systems. vol. 249. 2022. Available: https://doi.org/10.1016/j.knosys.2022.108968

Seu. K.. Kang. M.-S. and Lee. H.. An intelligent missing data imputation techniques: a review. International Journal on Informatics Visualization. vol. 278. 2022.

Sumertajaya. I. M.. Rohaeti. E.. Wigena. A. H. and Sadik. K.. Vector autoregressive-moving average imputation algorithm for handling missing data in multivariate time series. IAENG International Journal of Computer Science. vol. 50. 2023.

Yu. C.. Wang. F.. Shao. Z.. Qian. T.. Zhang. Z.. Wei. W. and Xu. Y.. GinAR: an end-to-end multivariate time series forecasting model suitable for variable missing. Proceedings of the ACM. pp. 3989–4000. 2024. Available: https://doi.org/10.1145/3637528.3672055

Biographies

Embay Rohaeti is a lecturer in the Mathematics Study Program at Universitas Pakuan, with research expertise in mathematical modeling. Her scholarly work spans various applied mathematics topics, including time series analysis, statistical modeling, and numerical methods. She has co-authored numerous publications in national and international journals, contributing to fields such as tuberculosis modeling, financial forecasting, and environmental statistics. Her research has been cited multiple times, reflecting her academic influence and collaboration with fellow researchers.

Muhammad Edy Rizal is a lecturer in the Data Science Study Program at the Faculty of Mathematics and Natural Sciences, Universitas Tadulako. He earned his master's degree in 2023 from the Department of Statistics and Data Science at IPB University, where he developed a strong foundation in statistical modeling and computational techniques. His research interests lie primarily in time series analysis and machine learning, with a particular focus on their applications in disaster prevention and mitigation. He has authored several studies exploring predictive models and data-driven approaches to enhance early warning systems and improve resilience against natural hazards. Through his academic work, he aims to contribute to the development of intelligent solutions for disaster risk management in Indonesia and beyond.