Multi Proximity: Jurnal Statistika Universitas Jambi
Vol. 3 No. 2 - October 2024

E-ISSN: 2830-554X https://online-journal.unja.ac.id/multiproximity

The Best Moving Average Smoothing for GSTAR Model with Missing Value

Fadhlul Mubarak[1]*, Vinny Yuliani Sundara[2], Atilla Aslanargun[3]

[1]Department of Agribusiness, Universitas Jambi Jambi, Indonesia
[2]Department of Statistics, UIN Sulthan Thaha Saifuddin Jambi, Indonesia
[3]Department of Statistics, Eskişehir Technical University, Türkiye

e-mail: mubarakfadhlul@gmail.com
DOI: https://doi.org/10.22437/multiproximity.v3i2.32274

ABSTRACT

This research identified the best imputation technique for the price of gold in Turkey, Saudi Arabia, and Indonesia which has been applied to the GSTAR model based on the smallest RMSE value. The moving average smoothing technique with k = 2, 3, 4, and 5 have been used in this study. However, the moving average smoothing technique with k = 3 is the best technique for importing gold price data during weekends in Turkey, Saudi Arabia, and Indonesia.

Keywords: Gold prices, imputation techniques, RMSE, weekend.

INTRODUCTION

The Generalized Space Time Autoregressive (GSTAR) is one of the models used in statistical analysis for forecasting. The model is a multivariate model model formed from the influence of space and time. There have been several studies related to GSTAR [1], [2], [3], [4], [5], [6], [7]. One important element in the model is a spatial weighting matrix. The weighting matrices that have been widely used in many studies of various fields is the weighting distance inverse [6], [7], [8], [9], [10], [11], [12], [13], [14]. This spatial weighting has also been used in this study. The weighting assumes that the closer the distance between the object, the greater the effect on the dependent variable.

Missing value is one of the problems that is often faced in statistical analysis, especially time series topic [6], [15], [16], [17], [18]. Many things that cause incomplete data include holidays, system damage, etc. In this study it has implemented a solution to overcome missing value with moving average smoothing. The solution that has been applied is a moving average smoothing on k = 2, 3, 4, and 5. The GSTAR model has been applied to each data that has been carried out by the Imputation of Moving Average Smoothing. To choose the best moving average smoothing average root mean square error (RMSE). RMSE has been applied in various fields [19], [20]. Finally, based on the smallest RMSE value, the GSTAR model with the best missing value imputation technique.

METHOD

The data that has been used in this study is the price of gold in Indonesia, Turkey, and Saudi Arabia from January 1, 2018 to August 31, 2021 in the troy ounce unit. The price of gold has been displayed in Turkey at the TRY/troy ounce, in Saudi Arabia on the SAR/troy ounce, and in Indonesia at IDR/troy ounce. But the data that has been obtained from the World Gold Council has a missing data for weekends so that the imputation moving average smoothing technique has been used as equation 1 to complete it.

$$MA_t = \frac{T_{n-s+1} + T_{n-s+2} + \cdots + T_n}{t} \tag{1}$$

where MA_s is the mean over the last *t* data-points, T_{n-s+1} is the initial data, T_n is the last data, and t is the total data used for smoothing. Furthermore, the data that the imputation technique has been divided into 2 parts including data on training (January 1, 2018-December 31, 2020) and data testing (January 1, 2021-August 31, 2021). The generalized space time autoregressive (GSTAR) model that has been used in this study can be seen in the equation 2. This research has been limited to the GSTAR model (1).

$$\boldsymbol{Y}_a(t) = \sum_{k=1}^{p} \sum_{l=0}^{\lambda_k} \boldsymbol{\phi}_{kl}^{(a)} \boldsymbol{W}^{(l)} \boldsymbol{Y}_i(t-k) + \boldsymbol{\varepsilon}_a(t) \tag{2}$$

where $\boldsymbol{Y}_a(t)$ is independent variable, p is autoregressive order, q is moving average order, d is differencing order, λ_k is spatial order of the k-autoregressive condition, m_k is spatial order of k-th moving average condition, $\boldsymbol{\phi}_{kl}^{(a)}$ is parameter for autoregressive, $\boldsymbol{W}^{(l)}$ is spatial weighting matrix, $\boldsymbol{\theta}_{kl}^{(a)}$ is parameter for moving average, and $\boldsymbol{\varepsilon}_a(t)$ is error. The weighting that has been used in this study is the weighting of inverse distances such as equation 3.

$$W_{ij} = \begin{bmatrix} 0 & w_{12} & \cdots & w_{1k} \\ w_{21} & 0 & \cdots & w_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ w_{k1} & w_{k2} & \cdots & 0 \end{bmatrix} \tag{3}$$

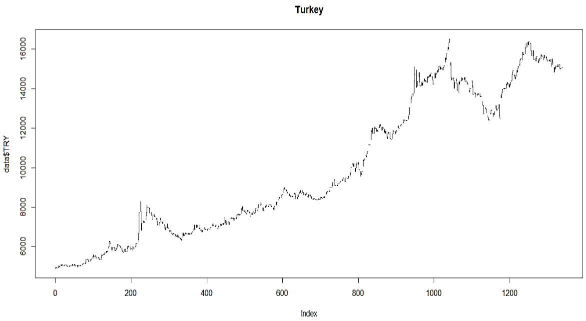
where $i \neq j$ for the distance between the country and $w_{ij} = w_{jk}$. Root mean square error (RMSE) which has been used to select the best model can be seen in equation 4.

$$RMSE = \sqrt{\frac{\sum_{a=1}^{n} \left(Y_a - \hat{Y}_a\right)^2}{n}} \tag{4}$$

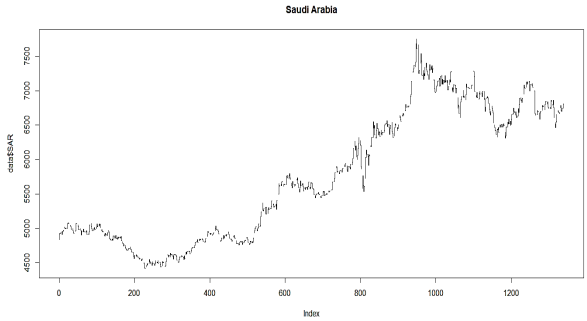
where Y_a is the actual value of the price of gold for each country, $\hat{Y}_a$ is the prediction value of the price of gold for each country, and n is the amount of data testing.

RESULTS AND DISCUSSION

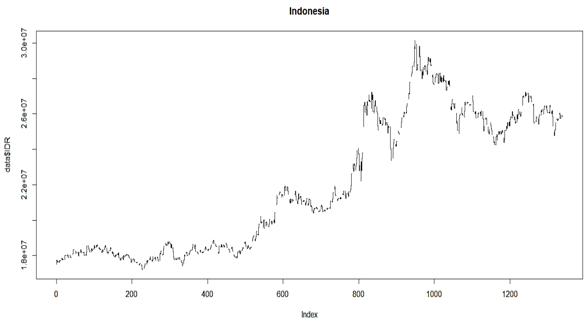
The data that has been obtained can be seen in Figure 1. Based on Figure 1, the dashed line indicated that the data had missing values. The index in the figure has been represented date (0 for January 1, 2018 and 957 for August 31, 2021. While the results of the moving average smoothing have been obtained can be seen in Figure 2. Based on the figure, on the x-axis is the date that has been represented 0 for January 1, 2018 and 12 for January 12, 2018.



(a)



(b)

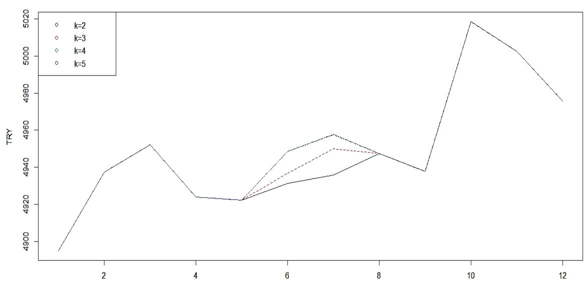


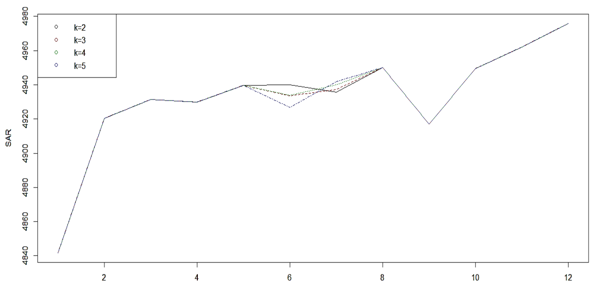
(c)

Figure 1. Graphs of gold price for each country.

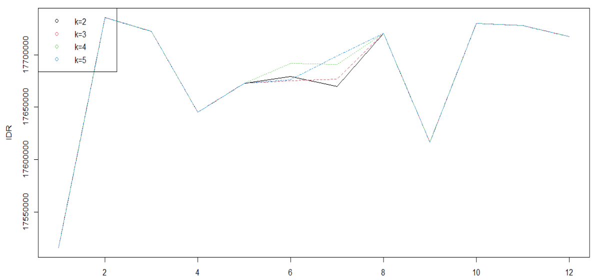
$$W_{ij} = \begin{bmatrix} 0 & \frac{1}{2128.51} & \frac{1}{9101.73} \\ \frac{1}{2128.51} & 0 & \frac{1}{7348.93} \\ \frac{1}{9101.73} & \frac{1}{7348.93} & 0 \end{bmatrix} \tag{5}$$

The weighting matrix that has been used in this study is like equation 5. The distance between Turkey and Saudi Arabia is 2128.51 km. Furthermore, the distance between Turkey and Indonesia is 9101.73 km and the distance between Saudi Arabia and Indonesia is 7348.93 km.





(a) (b)



(c)

Figure 2. Graphs of moving average smoothing for each country.

Table 1. Parameter estimates

k	Parameter	Estimate	k	Parameter	Estimate
2	$\boldsymbol{\phi}_{10}^{(1)}$	0.997279***	4	$\boldsymbol{\phi}_{10}^{(1)}$	0.997046***
2	$\boldsymbol{\phi}_{11}^{(1)}$	0.014271.	4	$\boldsymbol{\phi}_{11}^{(1)}$	0.015163.
2	$\boldsymbol{\phi}_{10}^{(2)}$	0.98839***	4	$\boldsymbol{\phi}_{10}^{(2)}$	0.98866
2	$\boldsymbol{\phi}_{11}^{(2)}$	0.02295.	4	$\boldsymbol{\phi}_{11}^{(2)}$	0.02243.
2	$\boldsymbol{\phi}_{10}^{(3)}$	1.003e+00***	4	$\boldsymbol{\phi}_{10}^{(3)}$	1.003e+00***
2	$\boldsymbol{\phi}_{11}^{(3)}$	-3.139e+04	4	$\boldsymbol{\phi}_{11}^{(3)}$	-2.921e+04
3	$\boldsymbol{\phi}_{10}^{(1)}$	0.997218***	5	$\boldsymbol{\phi}_{10}^{(1)}$	0.996912***
3	$\boldsymbol{\phi}_{11}^{(1)}$	0.014503.	5	$\boldsymbol{\phi}_{11}^{(1)}$	0.015671.
3	$\boldsymbol{\phi}_{10}^{(2)}$	0.988534***	5	$\boldsymbol{\phi}_{10}^{(2)}$	0.988493***
3	$\boldsymbol{\phi}_{11}^{(2)}$	0.022682.	5	$\boldsymbol{\phi}_{11}^{(2)}$	0.022751
3	$\boldsymbol{\phi}_{10}^{(3)}$	1.003e+00***	5	$\boldsymbol{\phi}_{10}^{(3)}$	1.003e+00***
3	$\boldsymbol{\phi}_{11}^{(3)}$	-3.025e+04	5	$\boldsymbol{\phi}_{11}^{(3)}$	-2.827e+04

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

The results of the estimation of parameters have been displayed in Table 1. For each k, it has been seen 6 parameters that have consisted of $\boldsymbol{\phi}_{10}^{(1)}, \boldsymbol{\phi}_{11}^{(1)}, \boldsymbol{\phi}_{10}^{(2)}, \boldsymbol{\phi}_{11}^{(2)}, \boldsymbol{\phi}_{10}^{(3)},$ dan $\boldsymbol{\phi}_{11}^{(3)}$. For the influence of lag, $\boldsymbol{\phi}_{10}^{(1)}$ parameter represented the influence of the gold price lag in Turkey, $\boldsymbol{\phi}_{10}^{(2)}$ parameter represented the influence of the gold price lag in Saudi Arabia, and $\boldsymbol{\phi}_{10}^{(3)}$ parameter represented the influence of the gold price lag in Indonesia. In addition, for the effect of neighbors, the $\boldsymbol{\phi}_{11}^{(1)}$ parameter represented the effect of neighboring Saudi Arabia and Indonesia on Turkey, $\boldsymbol{\phi}_{11}^{(2)}$ parameter represented the effect of neighboring Turkey and Indonesia on Saudi Arabia, and $\boldsymbol{\phi}_{11}^{(3)}$ parameter represented the effect of neighboring Turkey and Saudi Arabia on Indonesia.

In the table it is also relatively clear that all the coefficients of the lag effect parameters have been significant at α =1%. In addition, the $\boldsymbol{\phi}_{11}^{(1)}$ and $\boldsymbol{\phi}_{11}^{(2)}$ parameter coefficients were significant at =10%, relatively. While the parameter coefficient $\boldsymbol{\phi}_{11}^{(3)}$ was insignificant, relatively.

Table 2. RMSE

k	**TRY**	**SAR**	**IDR**	**ALL**
2	1003.67	416.0062	949151.5	547993.3
3	1015.727	416.7726	948294.9	547498.7
4	1044.406	422.877	973881.6	562271.2
5	1065.534	424.8906	997608.4	575969.8

Based on the smallest RMSE value in Table 2, gold prices in Turkey and Saudi Arabia are best if forecasting is done that combines the GSTAR model and smoothing moving average k=2. Meanwhile, for forecasting gold prices in Indonesia and overalls it is best if you have combined the GSTAR model and the smoothing moving average k=3. For the imputation technique that has been applied to the GSTAR model which has produced the worst forecast value is the smoothing moving average with k=5.

CONCLUSION

Incomplete data had become a problem for conducting time series analysis, especially in the case of forecasting. Smoothing average imputation technique has been a solution to this problem. There have been several missing value imputation techniques that have been used in this study including smoothing moving averages with k=2, 3, 4, and 5. Based on the smallest RMSE value, the smoothing moving average technique with k=3 that has been applied to the GSTAR model is the best missing value imputation in the case of gold prices in Turkey, Saudi Arabia, and Indonesia.

Research related to the GSTAR model with this missing value still needs to be explored. This research limited to the GSTAR model which had an autoregressive order of 1 and a spatial order of 1. For further research, it is also necessary to try out the GSTAR model with an autoregressive order of 2, 3, and so on. In Figure 1, there are relatively indications that the data is not stationary, so further identification is still needed. If these indications are proven, it is also necessary to model the gold price data into generalized space time autoregressive integrated (GSTARI). The high-order spatial weighting matrix can also be used for further research exploration with the aim of looking at the spatial order of how much neighboring influence is significant at $\alpha = 5\%$ even at $\alpha = 1\%$.

REFERENCES

[1] S. Suhartono, D. D. Prastyo, H. Kuswanto, and M. H. Lee, "Comparison between VAR, GSTAR, FFNN-VAR and FFNN-GSTAR Models for Forecasting Oil Production," *MATEMATIKA*, vol. 34, no. 1, pp. 103–111, 2018.

[2] A. S. Abdullah *et al.*, "Implementation of Generalized Space Time Autoregressive (GSTAR)-Kriging model for predicting rainfall data at unobserved locations in West Java," *Applied Mathematics and Information Sciences*, vol. 12, no. 3, pp. 607–615, 2018.

[3] U. S. Pasaribu, U. Mukhaiyar, and M. N. Heriawan, "Spatial weight determination of GSTAR (1; 1) model by using kernel function," in *Journal of Physics: Conference Series*, IOP Publishing, 2018, p. 012223.

[4] N. M. Huda, U. Mukhaiyar, and U. S. Pasaribu, "The approximation of GSTAR model for discrete cases through INAR model," in *Journal of Physics: Conference Series*, IOP Publishing, 2021, p. 012100.

[5] U. S. Pasaribu, U. Mukhaiyar, N. M. Huda, K. N. Sari, and S. W. Indratno, "Modelling COVID-19 growth cases of provinces in java Island by modified spatial weight matrix GSTAR through railroad passenger's mobility," *Heliyon*, vol. 7, no. 2, p. e06025, 2021.

[6] F. Mubarak and A. Aslanargun, "GSTARIMA Model with Missing Value for Forecasting Gold Price," *Indonesian Journal of Statistics and Its Applications*, vol. 6, no. 1, pp. 90–100, 2022.

[7] F. Mubarak, I. Sumertajaya, and M. N. Aidi, "The best spatial weight matrix order of radius for GSTARIMA," *International Journal of Engineering and Management Research (IJEMR)*, vol. 7, no. 2, pp. 494–497, 2017.

[8] H. Zhou, Q. Ding, and R. Jia, "Inverse of the distance matrix of a weighted cactoid digraph," *Appl Math Comput*, vol. 362, p. 124552, 2019.

[9] T. J. Souza, J. A. C. C. Medeiros, A. C. Gonçalves, and A. S. Martinez, "A method to identify an accidental control rod drop with inverse distance weighted interpolation," *Ann Nucl Energy*, vol. 145, p. 107462, 2020.

[10] H.-K. Wang and C.-F. Chien, "An inverse-distance weighting genetic algorithm for optimizing the wafer exposure pattern for enhancing OWE for smart manufacturing," *Appl Soft Comput*, vol. 94, p. 106430, 2020.

[11] Y. Shi *et al.*, "Expected output calculation based on inverse distance weighting and its application in anomaly detection of distributed photovoltaic power stations," *J Clean Prod*, vol. 253, p. 119965, 2020.

[12] K. Shukla, P. Kumar, G. S. Mann, and M. Khare, "Mapping spatial distribution of particulate matter using Kriging and Inverse Distance Weighting at supersites of megacity Delhi," *Sustain Cities Soc*, vol. 54, p. 101997, 2020.

[13] R. Qu *et al.*, "Predicting the hormesis and toxicological interaction of mixtures by an improved inverse distance weighted interpolation," *Environ Int*, vol. 130, p. 104892, 2019.

[14] F. Mubarak, A. Aslanargun, and İ. Sıklar, "Higher-Order Spatial Classification Using Google Trends Data during Covid-19," *Adv Appl Stat*, vol. 78, pp. 93–103, 2022.

[15] Y. Zhang, B. Zhou, X. Cai, W. Guo, X. Ding, and X. Yuan, "Missing value imputation in multivariate time series with end-to-end generative adversarial networks," *Inf Sci (N Y)*, vol. 551, pp. 67–82, 2021.

[16] J. Ma *et al.*, "Transfer learning for long-interval consecutive missing values imputation without external features in air pollution time series," *Advanced Engineering Informatics*, vol. 44, p. 101092, 2020.

[17] D. Li, L. Li, X. Li, Z. Ke, and Q. Hu, "Smoothed LSTM-AE: A spatio-temporal deep model for multiple time-series missing imputation," *Neurocomputing*, vol. 411, pp. 351–363, 2020.

[18] F. Mubarak, A. Aslanargun, and I. Siklar, "The Best K-Exponential Moving Average with Missing Values: Gold Prices in Indonesia, Saudi Arabia, and Turkey during COVID-19," in *Proceedings of The International Conference on Data Science and Official Statistics*, 2021, pp. 334–342.

[19] M. Ćalasan, S. H. E. A. Aleem, and A. F. Zobaa, "On the root mean square error (RMSE) calculation for parameter estimation of photovoltaic models: A novel exact analytical solution based on Lambert W function," *Energy Convers Manag*, vol. 210, p. 112716, 2020.

[20] M. W. Liemohn, A. D. Shane, A. R. Azari, A. K. Petersen, B. M. Swiger, and A. Mukhopadhyay, "RMSE is not enough: Guidelines to robust data-model comparisons for magnetospheric physics," *J Atmos Sol Terr Phys*, vol. 218, p. 105624, 2021.