Available online at http://www.iorajournal.org/index.php/ijgor/index

International Journal of Global Operations Research

e-ISSN: 2722-1016
p-ISSN: 2723-1739

Vol. 5, No. 4, pp. 288-294, 2024

Modeling of COVID-19 Growth Cases in Bandung Regency and Bandung City Using Vector Autoregression

Tubagus Robbi Megantara[1*], Rizki Apriva Hidayana[2], Abdul Gazir Syarifudin[3], Rika Amelia[4], Nenden Siti Nurkholipah[5]

[1,2,3,4,5]*Department of Mathematics, Faculty of Mathematics and Natural Sceinces, National University of the Republic of Indonesia, Bandung, Indonesia*

**Corresponding author email: tubagusrobbi@gmail.com*

Abstract

COVID-19 is a global health epidemic due to increasing infections and deaths. Indonesia has many confirmed cases with high daily case growth, including the Bandung City and Bandung Regency areas. High mobility between regions can impact the growth of COVID-19 cases. Strategies to prevent the growth of COVID-19 cases need to be carried out by considering the growth of COVID-19 cases in the nearest area. The Vector Autoregression (VAR) model is a forecasting model that can consider geographic impacts. This study aims to model the growth of cases in adjacent areas and have high mobility using the VAR method. The growth of COVID-19 cases in Bandung City and Bandung Regency is integrated into the VAR model to see the impact of each other. The VAR model also considers the impact of case growth in the past on its region's future. Transformation and differencing are carried out on the time series of case growth in each region to achieve time-series stationarity so that the VAR model can be carried out. First-order VAR becomes a model representing the growth of COVID-19 cases in Bandung City and Bandung Regency. The model shows that COVID-19 cases in each region will decrease over time and each region impacts each other. Decreasing cases growth can be caused because people who have been infected and vaccinated have sound immune systems to prevent re-infection. However, prevention still needs to be done to stop the pandemic. Therefore, restrictions on mobility between regions can be used as a strategy to prevent COVID-19 infection.

Keywords: Vector Autoregression, Region, COVID-19 growth cases, Forecast

1. Introduction

The COVID-19 outbreak has been considered a global health emergency as the growth of the number of confirmed cases and confirmed deaths (Sohrabi et al., 2020). Indonesia had 5,289,414 confirmed cases as of 22 February 2022, with a daily growth of 57,491 cases, according to the Indonesian COVID-19 Task Force. In addition, 10.4% of active cases, 86.8% of recovered cases and 2.8% of cases died. Unrestricted mobility between regions can lead to an accelerated COVID-19 infection. Bandung City and Regency have high mobility between regions.

The unknown peak conditions, the length of the outbreak duration and the infection rate from COVID-19 are the most worried by the public and the government (Zhang et al., 2020). COVID-19 puts people all over the world at risk of not only death from a viral infection and severe psychological stress (Duan & Zhu, 2020; Xiao, 2020). Symptoms of anxiety due to COVID-19 are also common among college students. However, good economic conditions can prevent these anxiety symptoms (W. Cao et al., 2020).

COVID-19 has impacted people, businesses and organizations worldwide, inadvertently changing global financial markets and economies (Nicola et al., 2020). Medium and long-term planning must be done to balance and revive the post-pandemic economy (Nicola et al., 2020). A mathematical and statistical model can be used to determine confirmed cases' estimates (Azhar et al., 2021). Forecasting models can be used to anticipate and determine the sustainable growth of COVID-19 cases, where the analysis results can be used in policy-making strategies (Ribeiro et al., 2020). The policies must ensure food security's economic and physical aspects (Hossain, 2020).

Vector Autoregression (VAR) is one of the forecasting models that can model the dynamic behaviour of the combination of many time series. Moran et al. (2022) uses VAR to calculate the effects of uncertainty shocks calibrated to match COVID-induced increases on the Canadian economy. Y. Cao & Francis (2021) uses VAR to forecast COVID-19 infections at the community level based on SARS-CoV-2 concentrations in wastewater. Brinca et al. (2021) uses bayesian structural VAR to estimate labour demand and supply shocks at the sector level surrounding the COVID-19 epidemic. Nakanishi et al. (2021) uses VAR to analyze Weekly mobility and infection data COVID-19 in Tokyo.

VAR is commonly used to model spatiotemporal schemes (Rajab et al., 2022). Monllor et al. (2020) uses VAR to forecast the number of infections, hospitalizations, and ICU bed numbers in Italy due to COVID-19. Rajab et al. (2022) uses VAR to forecast the number of new cases and deaths in the United Arab Emirates, Saudi Arabia, and Kuwait, respectively. De Souza Melo et al. (2021) uses VAR to model COVID-19 cases in Sao Paulo based on the daily isolated people, the daily number of infections and the daily number of deaths. Gianino et al. (2021) uses VAR and the Granger Causality test to analyze the possible effects that any control measure would have on the overall growth of the pandemic.

This study shows the effect of the impact of COVID-19 cases in Bandung City and Bandung Regency, each other. The causal effect and dynamic movement patterns regarding future COVID-19 growth cases using VAR. VAR modelling can be done using the Rstudio software (Pfaff, 2008).

2. Materials and Method

VAR is a commonly used model in multivariate time series modelling (Tsay, 2013). The VAR model consists of several univariate time series, where each time series has a relationship. VAR modelling can be done if each time series meets the stationary assumption. In general, the order p VAR model can be represented by order 1 VAR. The first order VAR model for bivariate is given by equation (1):

$$\boldsymbol{X}_t = \boldsymbol{\Phi}\boldsymbol{X}_{t-1} + \boldsymbol{a}_t, \tag{1}$$

Time series stationarity can be checked using the Augmented Dickey-Fuller Test (Azar & Nasr, 2015). After the time series stationarity is met, the order of the VAR model can be determined based on four selection criteria, namely AIC, HQ, SC, and FPE that measure the distance between observations and the model class. Order selection criteria that have a small value impact the quality of the VAR model (Siraj-Ud-Doulah, 2019).
Where $\boldsymbol{\Phi} \in M_{2\times 2}(\mathbb{R})$ is the coefficient matrix and a_t, is the residual white noise vector with an average of zero, it can be written $\boldsymbol{a}_t \in N(0, \boldsymbol{I}_2)$. The matrix $\boldsymbol{\Phi}$ can provide information about the impact dynamically on the time series from time to time. Equation (1) can be rewritten in matrix form, given by equation (2):

$$\begin{bmatrix} X_{1,t} \\ X_{2,t} \end{bmatrix} = \begin{bmatrix} \phi_{1,1} & \phi_{2,1} \\ \phi_{2,1} & \phi_{2,2} \end{bmatrix} \begin{bmatrix} X_{1,t-1} \\ X_{2,t-2} \end{bmatrix} + \begin{bmatrix} a_{1,t} \\ a_{2,t} \end{bmatrix}. \tag{2}$$

If each time series does not affect each other, then the non-diagonal entry of the coefficient matrix is zero.

VAR modelling can be done when the time series meets the assumptions of the VAR model, that the time series is stationary. The assumption of stationarity for bivariate order 1 VAR can be met when the absolute value of each eigenvalue of the coefficient matrix is less than one, where the eigenvalues are solutions of equation (3):

$$|\lambda \boldsymbol{I}_2 - \boldsymbol{\Phi}| = 0. \tag{3}$$

Parameters $\boldsymbol{\Phi}$ and $\boldsymbol{a_t}$ in equation (1) can be estimated using the Ordinary Least Square (OLS) method, as discussed by Nalita et al. (2021). Validation of the use of OLS can be achieved by fulfilling the assumptions of residual normality, homoscedasticity and independence (de Souza & Junqueira, 2005). The white noise of a residual can be tested through the Ljung Box statistic. The VAR model is evaluated based on the residual forecast results against the actual data using MAPE, MAE and RSME.

3. Results and Discussion

Daily confirmed cases of COVID-19 in Bandung City ($X_{1,t}$) and Bandung Regency ($X_{2,t}$) from 01 January 2022 to 21 February 2022, illustrated in Figure 1. Descriptive statistics for each region are presented in Table 1.

Table 1. Descriptive Statistics

No	Region	Min	Mean	Max
1	Bandung City	0	366	1614
2	Bandung Regency	0	110.9	586

The growth rate of COVID-19 cases in Bandung City and Bandung Regency has a fairly high rate, as shown in Table 1. The case growth rate is shown by the distance from the average value to the biggest value that is so far away. The city of Bandung has an average confirmed case of COVID-19, which is three times larger than the Bandung Regency area.

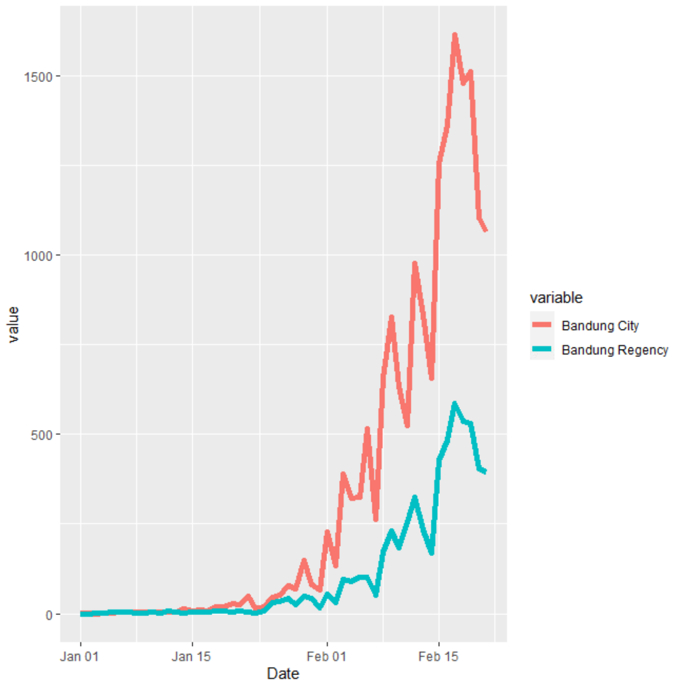

Figure 1. Daily data of COVID-19 cases

The daily COVID-19 cases in Bandung City and Bandung Regency do not have a stationary data pattern, as shown in Figure 1. Both time series have an exponential upward trend starting from early February 2022, so it is necessary to perform a logarithmic transformation to obtain a time series closer to the horizontal. However, time-series data with a zero value can make the logarithmic transformation process incorrect. Therefore, one confirmed case is added to each regional time series so that the logarithmic transformation process can be carried out, where the transformation results can be seen in Figure 2.

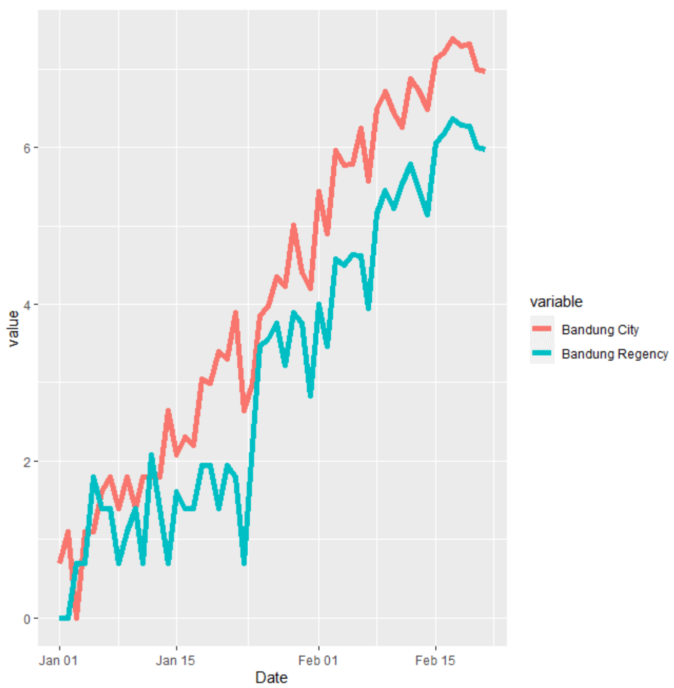

Figure 2. Logarithmic transformation

The transformed time series does not yet have a stationary pattern, as shown in Figure 2. However, the time series has a pattern linear uptrend, so it is necessary to do differencing to eliminate it. The stationarity can be achieved by differencing the time series, as shown in Figure 3.

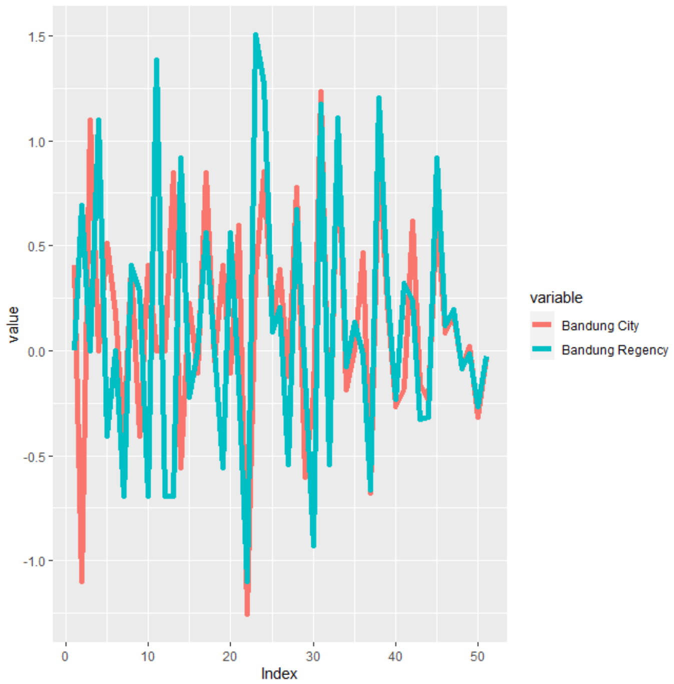

Figure 3. Differencing of logarithmic transformation

The stationarity pattern of the time series can be seen in Figure 3, where the data is spread between -1.5 to 1.5. Stationarity was further checked using the Augmented Dickey-Fuller test, shown in Table 2. The time series has stationarity for a significant level of 1% to carry out VAR modelling.

Table 2. Time series stationarity

No	Region	Actual	Logarithmic Transformation	Differencing
1	Bandung City	Not stationary	Not stationary	Stationary
2	Bandung Regency	Not stationary	Not stationary	Stationary

The smallest value for each criterion is used for selecting the model order, where lag represents the order of the VAR model. The criteria for selecting the smallest AIC, HQ, SC, and FPE models are at lag 1, as shown in Table 3. Therefore, order 1 VAR was chosen as a time series model approach resulting from differencing of logarithmic transformations.

Table 3. Model selection criteria

No	AIC(n)	HQ(n)	SC(n)	FPE(n)
1	-2.68021	-2.61933	-2.51303	0.068559
2	-2.65811	-2.53635	-2.32375	0.070168
3	-2.53543	-2.3528	-2.0339	0.079564
4	-2.42318	-2.17968	-1.75447	0.089542
5	-2.35799	-2.0536	-1.5221	0.096528
6	-2.23323	-1.86797	-1.23016	0.111029
7	-2.25829	-1.83216	-1.08805	0.110675
8	-2.46391	-1.9769	-1.12649	0.092872
9	-2.33383	-1.78594	-0.82923	0.110158
10	-2.16335	-1.55458	-0.49157	0.137812

The first-order VAR model in the form of equation (1) which corresponds to the time series from the differencing results to the logarithmic transformation, is given by equation (4):

$$\nabla \boldsymbol{X}_t = \boldsymbol{\Phi} \nabla \boldsymbol{X}_{t-1} + \boldsymbol{a}_t, \tag{4}$$

$\nabla \boldsymbol{X}_t$ is a time series differencing vector from the logarithmic transformation of daily COVID-19 cases with one additional case. The time series differencing vector of the logarithmic transformation is expressed as equation (5):

$$\nabla \boldsymbol{X}_t = \begin{bmatrix} \log(X_{1,t}+1) - \log(X_{1,t-1}+1) \\ \log(X_{2,t}+1) - \log(X_{2,t-1}+1) \end{bmatrix}, \quad (5)$$

$\boldsymbol{X_t}$ is a two-dimensional vector with elements $X_{1,t}$ and $X_{2,t}$, respectively representing COVID-19 cases in Bandung City and Bandung Regency at time t.

The parameter $\boldsymbol{\Phi}$ in equation (4) is estimated using the OLS method, where the parameter estimation results are given by equation (6):

$$\boldsymbol{\Phi} = \begin{bmatrix} \phi_{1,1} & \phi_{2,1} \\ \phi_{2,1} & \phi_{2,2} \end{bmatrix} = \begin{bmatrix} -0.5275749 & 0.1437697 \\ 0.06852804 & 0.6668706 \end{bmatrix}. \quad (6)$$

Substituting equation (6) into equation (3), the eigenvalues are -0.5692928 and -0.2914114. The absolute value of each eigenvalue is less than one, so the assumption of stationarity for bivariate order 1 VAR has been met.

The residuals of the parameter estimation using the OLS method were examined using the Ljung-Box test, where the residuals of the two-time series were normally distributed for a significant level of 1%. Therefore, the parameter estimation results by the OLS method can be used for forecasting.

The causal relationship between each time series, both with the past time series itself or with other time series, can be seen in equation (6), where each time series influences the decline in future data. In addition, one time series influences the increase in data in the future of another time series. Therefore, both time series can be forecasted using a multivariate model.

The forecasting model for COVID-19 cases in Bandung City and Bandung Regency for time t is obtained by substituting equation (5) into equation (4), which can be written as

$$\begin{bmatrix} \log(\hat{X}_{1,t}+1) \\ \log(\hat{X}_{2,t}+1) \end{bmatrix} = \begin{bmatrix} \log(X_{1,t-1}+1) \\ \log(X_{2,t-1}+1) \end{bmatrix} + \boldsymbol{\Phi} \begin{bmatrix} \log(X_{1,t-1}+1) - \log(X_{1,t-2}+1) \\ \log(X_{2,t-1}+1) - \log(X_{2,t-2}+1) \end{bmatrix}, \quad (7)$$

where $\boldsymbol{\Phi}$ is given by equation (6). $\hat{X}_{1,t}$ and $\hat{X}_{2,t}$ are forecasts of COVID-19 cases in Bandung City and Bandung Regency at time t, respectively. The results of forecasting COVID-19 cases using order 1 VAR can be seen in Figure 4, where the forecast for Bandung City is on the left and Bandung Regency is on the right.

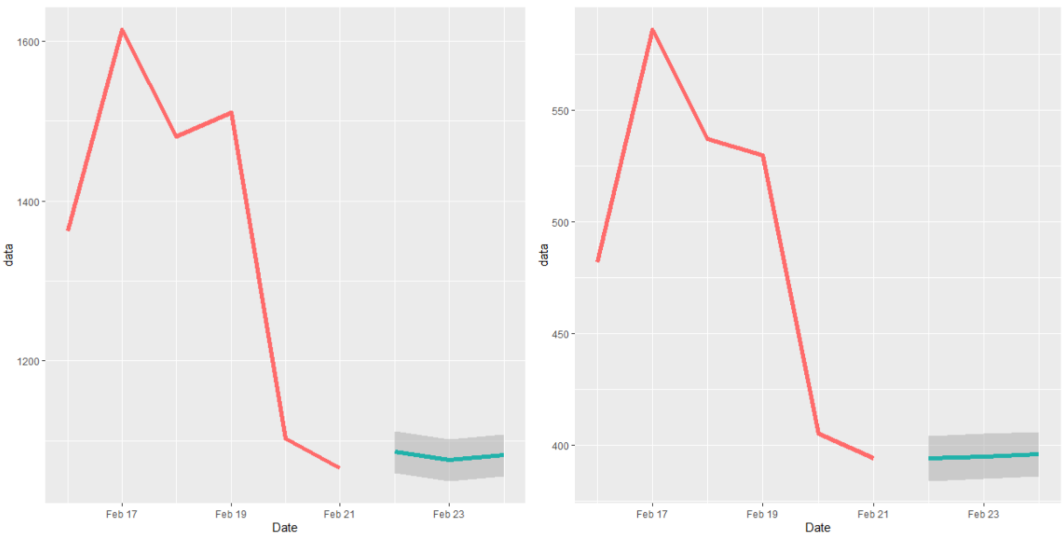

Figure 4. Forecasting COVID-19 cases

The absolute value of the mean error percentage (MAPE) is 11.69% and 13.47%, respectively. The root means squared error (RSME) are 62.13 and 23.08, respectively. The absolute mean error value (MAE) is 25.94 and 9.37, respectively. Therefore, the order 1 VAR model can be used for integrated forecasting of COVID-19 cases in the City of Bandung and Bandung Regency, where the forecast results show an increase of around 1060 to 1112 cases and an addition of about 384 to 404 cases for each region for February 22, 2022.

4. Conclusion

Logarithmic and differencing transformations are carried out to obtain a stationarity time series. The order of the VAR model is selected through four selection criteria, and the model is evaluated based on the residual forecasting results against the actual data. The first-order VAR model represents COVID-19 cases in Bandung City and Bandung Regency.

The causal relationship is shown only applies to differencing from the logarithmic transformation of the time series but still represents the cause and effect of the initial time series. The model shows that past daily cases of COVID-19 affect future daily cases. The endurance of humans who have been infected with COVID-19 can minimize the occurrence of cases of being re-infected with COVID-19 so that the daily cases are getting smaller as time goes by. However, this is also influenced by the vaccination rate and population.

Two adjacent areas have a causal relationship to make COVID-19 cases increase. If one region has a confirmed case of COVID-19 on a certain day, then other regions can be affected by a confirmed case of COVID-19 as well. Therefore, restrictions on mobility between regions can be used as a strategy to prevent COVID-19 infection.

References

Azar, S. A., & Nasr, N. I. (2015). The Power of the Augmented Dickey-Fuller Test to Detect Non-stationarity When There Is a Mix with a Stationary Variable. Empirical Economics Letters, 14(9), 885–890.

Azhar, M. A. R., Adi Nugroho, H., & Wibirama, S. (2021). The Study of Multivariable Autoregression Methods to Forecast Infectious Diseases. Proceedings - 2021 IEEE 5th International Conference on Information Technology, Information Systems and Electrical Engineering: Applying Data Science and Artificial Intelligence Technologies for Global Challenges During Pandemic Era, ICITISEE 2021, 83–88. https://doi.org/10.1109/ICITISEE53823.2021.9655881

Brinca, P., Duarte, J. B., & Faria-e-Castro, M. (2021). Measuring labor supply and demand shocks during COVID-19. European Economic Review, 139. https://doi.org/10.1016/j.euroecorev.2021.103901

Cao, W., Fang, Z., Hou, G., Han, M., Xu, X., Dong, J., & Zheng, J. (2020). The psychological impact of the COVID-19 epidemic on college students in China. Psychiatry Research, 287. https://doi.org/10.1016/j.psychres.2020.112934

Cao, Y., & Francis, R. (2021). On forecasting the community-level COVID-19 cases from the concentration of SARS-CoV-2 in wastewater. Science of the Total Environment, 786. https://doi.org/10.1016/j.scitotenv.2021.147451

de Souza Melo, A., da Penha Sobral, A. I. G., Marinho, M. L. M., Duarte, G. B., Vieira, A. A., & Sobral, M. F. F. (2021). The impact of social distancing on COVID-19 infections and deaths. Tropical Diseases, Travel Medicine and Vaccines, 7(1). https://doi.org/10.1186/s40794-021-00137-3

de Souza, S. V. C., & Junqueira, R. G. (2005). A procedure to assess linearity by ordinary least squares method. Analytica Chimica Acta, 552(1–2), 25–35.

Duan, L., & Zhu, G. (2020). Psychological interventions for people affected by the COVID-19 epidemic. The Lancet Psychiatry, 7(4), 300–302. https://doi.org/10.1016/S2215-0366(20)30073-0

Gianino, M. M., Nurchis, M. C., Politano, G., Rousset, S., & Damiani, G. (2021). Evaluation of the Strategies to Control COVID-19 Pandemic in Four European Countries. Frontiers in Public Health, 9. https://doi.org/10.3389/fpubh.2021.700811

Hossain, S. T. (2020). Impacts of COVID-19 on the agri-food sector: Food security policies of Asian productivity organization members. Journal of Agricultural Sciences - Sri Lanka, 15(2), 116–132. https://doi.org/10.4038/jas.v15i2.8794

Monllor, P., Su, Z., Gabrielli, L., & de La Paz, P. (2020). COVID-19 Infection Process in Italy and Spain: Are Data Talking? Evidence From ARMA and Vector Autoregression Models. Frontiers in Public Health, 8. https://doi.org/10.3389/fpubh.2020.550602

Moran, K., Stevanovic, D., & Touré, A. K. (2022). Macroeconomic uncertainty and the COVID-19 pandemic: Measure and impacts on the Canadian economy. Canadian Journal of Economics. https://doi.org/10.1111/caje.12551

Nakanishi, M., Shibasaki, R., Yamasaki, S., Miyazawa, S., Usami, S., Nishiura, H., & Nishida, A. (2021). On-site dining in Tokyo during the COVID-19 pandemic: Time series analysis using mobile phone location data. JMIR MHealth and UHealth, 9(5). https://doi.org/10.2196/27342

Nalita, Y., Rahani, R., Tirayo, E. R., Toharudin, T., & Ruchjana, B. N. (2021). Ordinary least square and maximum likelihood estimation of VAR (1) model's parameters and it's application on covid-19 in China 2020. Journal of Physics: Conference Series, 1722(1), 12082.

Nicola, M., Alsafi, Z., Sohrabi, C., Kerwan, A., Al-Jabir, A., Iosifidis, C., Agha, M., & Agha, R. (2020). The socio-economic implications of the coronavirus pandemic (COVID-19): A review. International Journal of Surgery (London, England), 78, 185–193. https://doi.org/10.1016/j.ijsu.2020.04.018

Pfaff, B. (2008). VAR, SVAR and SVEC models: Implementation within R package vars. Journal of Statistical Software, 27, 1–32.

Rajab, K., Kamalov, F., & Cherukuri, A. K. (2022). Forecasting COVID-19: Vector Autoregression-Based Model. Arabian Journal for Science and Engineering. https://doi.org/10.1007/s13369-021-06526-2

Ribeiro, M. H. D. M., da Silva, R. G., Mariani, V. C., & Coelho, L. D. S. (2020). Short-term forecasting COVID-19 cumulative confirmed cases: Perspectives for Brazil. Chaos, Solitons and Fractals, 135. https://doi.org/10.1016/j.chaos.2020.109853

Siraj-Ud-Doulah, M. (2019). Time Series Forecasting: A Comparative Study of VAR ANN and SVM Models. Journal of Statistical and Econometric Methods, 8(3), 21–34.

Sohrabi, C., Alsafi, Z., O'Neill, N., Khan, M., Kerwan, A., Al-Jabir, A., Iosifidis, C., & Agha, R. (2020). World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19). International Journal of Surgery (London, England), 76, 71–76. https://doi.org/10.1016/j.ijsu.2020.02.034

Tsay, R. S. (2013). Multivariate time series analysis: with R and financial applications. John Wiley & Sons.

Xiao, C. (2020). A Novel Approach of Consultation on 2019 Novel Coronavirus (COVID-19)-Related Psychological and Mental Problems: Structured Letter Therapy. Psychiatry Investigation, 17(2), 175–176. https://doi.org/10.30773/pi.2020.0047

Zhang, X., Ma, R., & Wang, L. (2020). Predicting turning point, duration and attack rate of COVID-19 outbreaks in major Western countries. Chaos, Solitons, and Fractals, 135, 109829. https://doi.org/10.1016/j.chaos.2020.109829