Spatio-Temporal models with intervention effect for modelling the impact of Covid-19 on the tourism sector in Indonesia

M Prastuti¹, L Aridinanti¹ and W P Dwiningtyas¹
¹Department of Business Statistics, Institute of Technology Sepuluh Nopember, Surabaya, Indonesia

E-mail: mike_p@statistika.its.ac.id

Abstract. Coronavirus is a virus that attacks the respiratory system, this disease due to viral infection is called Covid-19. Since it was found at December 2019 in China, the Covid-19 outbreak has spread to several countries, one of them is Indonesia. This causes the tourism sector in Indonesia to decline drastically. Therefore, this study aims to analyze the impact of Covid-19 outbreak on the number of foreign tourist arrivals to Indonesia especially in Jakarta, Bali, and Surabaya. This data is ordered by times from several different locations and has a relationship with each other or is called spatio-temporal data. Furthermore, the data will be added the intervention variable is time where the covid-19 outbreak found in December 2019. The data will be analyzed using a spatio-temporal models. The result shows that seasonal GSTARX-GLS models tend to give more accurate forecast than VARX and seasonal GSTARX-OLS models.

1. Introduction

Coronavirus is a virus that attacks the respiratory system. The disease caused by viral infection is known Covid-19. Covid-19 cause disturbances the respiratory system, acute pneumonia, and also the death. This virus was first found in the city of Wuhan, China at the end of December 2019. This virus spread quickly and has spread to other regions in China and to several countries. In line with China’s increasing role in the global economy, the economic impact of the Covid-19 is predicted to be widespread and significant, especially for countries that have strong economic relations with China such as Indonesia. The Covid-19 outbreak will have a significant impact on the tourism sector in Indonesia.

Spatio-temporal models is a statistical model which is a combination of times and locations dependency elements in a multivariate time series data. Generalized Space Time Autoregressive (GSTAR) is a model developed from the spatio-temporal models [1]. The GSTAR models which use the parameter estimation method as use in the Seemingly Unrelated Regression (SUR) model, namely Generalized Least Square (GLS) method, so it is called GSTAR-GLS. The GSTAR model produces correlated residuals between variables, hence the GLS method is able to overcome this case. GSTAR-GLS produces more efficient parameter estimation than GSTAR with estimation method using Ordinary Least Square (OLS). This can be seen from the standard error value generated by parameter.
of GSTAR-GLS method is smaller than GSTAR-OLS method. So that the forecast results of GSTAR-GLS more accurate forecast than GSTAR-OLS [2].

In its application, the GSTAR-GLS model often influenced by other variables outside the model or called exogenous variables. The GSTAR model with exogenous variables is called GSTARX [3]. In this case X is a symbol or notation of exogenous variables or predictor variables which can be scaled metric (interval or ratio) or non-metric (nominal or ordinal) [4,5]. Forecasting models by adding many predictor variables are proven to be able to improve the accuracy of forecast on multivariate time series data. Several studies used exogenous variables or predictor variables are Andayani applied GSTARIMA and GSTARIMAX models where X is a transfer function for forecasting rice price data in six provinces in Java with X is grain prices. In this study GSTARIMAX models provides better forecast results than GSTARIMA models [6]. Novianto also applied GSTARIX models with trend effect, seasonal and intervention effects on tourist coming to Indonesia, the results show that GSTARIX-OLS model yields the smallest errors in two locations, and GSTARIX-GLS yields the smallest errors in one location [7].

Based on some of these studies, it shows that a model involving predictor variables will produces better forecast than individual models. Therefore, this study will use GSTARX-GLS. This method will be applied to indicator data in tourism sector issued by the Badan Pusat Statistik (BPS), namely the number of foreign tourist visiting Indonesia, especially in Jakarta, Bali and Surabaya. Furthermore the data will be added intervention variable is time when Coronavirus outbreak was found in December 2019 in China.

2. Literature review

2.1 Generalized Space Time Autoregressive (GSTAR)

If know a series \( \{Z(t) : t = 0, \pm 1, \pm 2, \cdots, \pm T\} \) is a time series multivariate of \( N \) component, then GSTAR models of the Autoregressive order (time) and spatial order \( \lambda_1, \lambda_2, \lambda_3, \cdots, \lambda_p \), GSTAR \((p; \lambda_1, \lambda_2, \lambda_3, \cdots, \lambda_p)\) in matrix notation can written as follows [1].

\[
Z(t) = \sum_{s=1}^{p} \Phi_{s0} Z(t-s) + \sum_{k=1}^{\lambda_s} \Phi_{sk} W^{(k)} Z(t-s) + e(t)
\]

With

\[
\Phi_{s0} = \text{diag}(\phi_{s0}^{(1)}, \cdots, \phi_{s0}^{(N)}) \text{ is a time parameter matrix}
\]

\[
\Phi_{sk} = \text{diag}(\phi_{sk}^{(1)}, \cdots, \phi_{sk}^{(N)}) \text{ is a spatial parameter}
\]

e(t) are independent, identic, and multivariate normal distributed with mean 0 and variance-covariance matrix \( \sigma^2 I_N \)

the weighted value is chosen so that it fulfill the requirements \( w_{ij}^{(k)} = 0 \) and \( \sum_{ij} w_{ij}^{(k)} = 1 \).

2.2 Generalized Least Square (GLS)

Generalized Least Square (GLS) is a estimator of regression parameter that attention correlation of residuals between equations, where the residual values is obtained from Ordinary Least Square (OLS) estimation, which will be used in calculations to estimate the coefficient of regression in the SUR equation system. In general, the SUR model for N equations where each equation consist of K predictor variables can be written as in equation below.
\[ Z_{t1} = \beta_{10} + \beta_{11}X_{t1,1} + \beta_{12}X_{t1,2} + \ldots + \beta_{1K_1}X_{t1,K_1} + \epsilon_{t1} \]
\[ Z_{t2} = \beta_{20} + \beta_{21}X_{t2,1} + \beta_{22}X_{t2,2} + \ldots + \beta_{2K_2}X_{t2,K_2} + \epsilon_{t2} \]
\[ \vdots \]
\[ Z_{tN} = \beta_{N0} + \beta_{N1}X_{tN,1} + \beta_{N2}X_{tN,2} + \ldots + \beta_{NK_N}X_{tN,K_N} + \epsilon_{tN} \]

Where \( t = 1,2, \ldots, T \) with \( T \) many observations, \( K \) many predictor variables, and \( N \) many equations in the system and \( \epsilon_{t1}, \epsilon_{t2}, \ldots, \epsilon_{tN} \) have correlation each others \([8]\).

3. Methodology

In this study, the data used is secondary data from Badan Pusat Statistik (BPS). The data is the number of foreign tourist arrivals to Indonesia, especially in Jakarta, Bali and Surabaya. This data is monthly periods from January 2000 until July 2020. The intervention variables as a dummy variables which is time when the Covid-19 outbreak was found in December 2019 in China. Analysis steps in this study include the following.

1. Identify of data patterns on the number of foreign tourists at three locations, i.e Jakarta, Bali and Surabaya using time series plot.
2. Identify of time order of GSTARX-GLS model using MCCF plot and MPCCD plot as well as the minimum AIC values.
3. Identify the weight of the location in GSTARX-GLS model using normalized weights of partial inference cross-correlation.
4. Estimate parameters using OLS and GLS estimation methods.
5. Test the significance of the parameters of GSTARX models.
6. Check the assumption of white noise and multivariate normal distribution.
7. Choose the best model based on the Root Mean Square (RMSE) value.
8. Estimating the value of data in the future
9. Interpretation of the results.
10. Draw conclusions.

4. Analysis

4.1 Characteristics of the number of foreign tourist arrivals to Indonesia

Data of the foreign tourist arrivals to Indonesia especially in Jakarta, Bali and Surabaya shows indicates a monthly seasonal pattern. So this data is not stationary in means. Based on the graph know that from January 2000 until November 2019 there was an increase in the number of tourist arrival, but in December 2019 it started to decline due to the Covid-19 outbreak.

Figure 1. Time series plot of the number of foreign tourist arrivals to Indonesia.
4.2 Modelling of the number of foreign tourist arrivals use VARX

The data has seasonal patterns, hence do differencing 1 and 12 so that the data be stationary. After differencing, it was found that the data were stationary. This stationary data will be processed using VARX and GSTARX. The first step is to estimate the time order of the VARX model using Matrix Partial Cross Correlation Function (MPCCF) plot and minimum AIC value.

Table 1. Minimum AIC value.

| Lag | MA (0) |
|-----|--------|
| AR (0) | 56,710 |
| AR (1) | 56,447 |
| AR (2) | **56,337** |
| AR (3) | 56,372 |
| AR (4) | 56,338 |
| AR (5) | 56,360 |

Based on the MPCCF plot and the smallest value of AIC, it can be seen that the order of VARX model is VARX (4)_{1,12}. The results of parameter estimation shows that there are 17 significant parameters of 39 parameters. The diagnostic check of residual shows that the model of VARX(4)_{1,12} fulfill white noise assumption and also has a multivariate normal distribution.

4.3 Modelling of the number of foreign tourist arrivals use GSTARX-OLS and GSTARX-GLS

The first step in the GSTAR model is to determine the time and spatial orders. Where the time order is estimated based on the MPCCF plot and minimum AIC value. For spatial orders in the GSTAR model it is limited to first order. So that order GSTARX models is (1,2,3,4,12)_{1,12}. The data has seasonal patterns so namely seasonal GSTAR for this models. Seasonal GSTARX with parameter estimated method is OLS have 21 significant parameters from 31 all parameters in GSTAR. Meanwhile, for seasonal GSTARX-GLS have 23 significant parameters from 31 all parameters. The intervention variable when time of the Covid-19 outbreak is found has a significant effect on the number of foreign tourist arrivals to Bali.

Furthermore is check diagnostic of residuals, GSTARX-OLS and GSTARX-GLS are satisfied of white noise assumption and multivariate normal distribution.

4.4 The results of forecast the number of foreign tourist arrivals

Based of RMSE value from three methods, it is found that seasonal GSTARX-GLS produces the smallest RMSE among other methods. Beside that, seasonal GSTARX-GLS yields the smallest RMSE in three locations, i.e Jakarta, Bali and Surabaya. the RMSE value from three methods in three locations shown in the following table 2.
Table 2. RMSE value from three methods.

| Methods       | Jakarta | Bali | Surabaya |
|---------------|---------|------|----------|
| VARX          | 21498   | 36206| 2403     |
| GSTARX-OLS    | 21198   | 33086| 2248     |
| GSTARX-GLS    | *21039  | *32687| *2228    |

The graph for time series plot of actual and forecast data by seasonal GSTARX-GLS methods shown on the following Figure 3.

![Figure 3](image)

**Figure 3.** Time series plot of actual and forecast data from the number of foreign tourist arrivals in (a) Jakarta, (b) Bali, and (c) Surabaya

Based on the graph, it can be seen that the forecast results are close to the actual value, even the forecast results at the time of the covid-19 outbreak also approach the actual value.

5. **Conclusion**

The results of study shows that seasonal GSTARX-GLS tend to give more accurate forecast than VARX and Seasonal GSTARX-OLS for forecasting the number of foreign tourist arrivals to Indonesia, i.e. Jakarta, Bali and Surabaya. Moreover, the results also shows that the Covid-19 outbreak which is happened in December 2019 is affect the number foreign tourist arrivals to Indonesia especially in Bali.
Acknowledgements
This research was supported by DRPM ITS under scheme of “Penelitian Dana Departemen”, project No. 1620/PKS/ITS/2020. The authors gratefully acknowledge the support from ITS and Business Statistic Department for funding this research.

References

[1] Borovkova S, Lopuhaä H P and Ruchjana B N 2008 Statistica Neerlandica 482-508
[2] Setiawan, Suhartono, and Prastuti M 2016 Malaysian Journal of Mathematical Sciences 53-65
[3] Suhartono, Wahyuningrum, S R, Setiawan and Akbar M S 2016 Malaysian Journal of Mathematical Sciences 91-103
[4] Box G E P, Jenkins G M and Reinsel G C 2008 Time Series Analysis: Forecasting and Control Fourth Edition (Hoboken, New Jersey: John Wiley & Sons, Inc)
[5] Bowerman B L and O’Connel 1993 Forecasting and Time Series: An Applied Approach Third Edition (Belmont, California: Duxbury Press)
[6] Andayani N et al 2018 IOP Conference Series : Earth and Environmental Science 187 012052
[7] Navianto M A 2018 IOP Conf. Series: Journal of Physics: Conf. Series 1097 012076
[8] Tsay R S 2005 Analysis of Financial Time Series : Financial Econometrics (University of Chicago : John Wiley & Sons, Inc)