

VARX and GSTARX Models for Forecasting Currency Inflow and Outflow with Multiple Calendar Variations Effect

Suhartono*, Muhammad Munawir Gazali and Dedy Dwi Prastyo

Department of Statistics, Faculty of Mathematics, Computing, and Data Science
Institut Teknologi Sepuluh Nopember (ITS), 60111 Surabaya, Indonesia

*Corresponding author: suhartono@statistika.its.ac.id

Article history

Received: 22 August 2018

Received in revised form: 9 December 2018

Accepted: 17 December 2018

Published on line: 31 December 2018

Abstract VARX and GSTARX models are an extension of Vector Autoregressive (VAR) and Generalized Space-Time Autoregressive (GSTAR) models. These models include exogenous variable to increase the forecast accuracy. The objective of this research is to develop and compare the forecast accuracy of VARX and GSTARX models in predicting currency inflow and outflow in Bali, West Nusa Tenggara, and East Nusa Tenggara that contain multiple calendar variations effects. The exogenous variables that are used in this research are holidays in those three locations, i.e. EidFitr, Galungan, and Nyepi. The proposed VARX and GSTARX models are evaluated through simulation studies on the data that contain trend, seasonality, and multiple calendar variations representing the occurrence of EidFitr, Galungan, and Nyepi. The criteria for selecting the best forecasting model is Root Mean Square Error (RMSE). The results of a simulation study show that VARX and GSTARX models provide similar forecast accuracy. Furthermore, the results of currency inflow and outflow data in Bali, West Nusa Tenggara, and East Nusa Tenggara show that the best model for forecasting inflow and outflow in these three locations are VARX and GSTARX (with uniform weight) model, respectively. Both models show that currency inflow and outflow in Bali, West Nusa Tenggara, and East Nusa Tenggara have a relationship in space and time, and contain trends, seasonality and multiple calendar variations.

Keywords VARX, GSTARX, inflow, outflow, multiple calendar variations.

Mathematics Subject Classification 62M10, 62M30.

1 Introduction

In the case of inflows and outflows in Indonesia, local socio-cultural factors such as EidFitr, Chinese New Year, Nyepi, Galungan, Christmas, and Easter, are important influences in modeling it. Considering that Indonesia is a country with majority of Muslims, it causes a lot of economic data to have seasonal patterns which are influenced by two types of calendar, namely the Gregorian and the Hijri calendar. The effect of Gregorian calendar causes inflow and outflow become high or low in a certain month. Moreover, the Hijri calendar affects highly inflows

and outflows in the month around the EidFitr holidays. This effect is called the holiday effect [1,2,3]. This holiday effect cannot be identified as a seasonal effect because the EidFitr holidays occur at various dates and months in the Gregorian calendar from year to year.

The effects of local socio-cultural factors to the fluctuation of inflows and outflows also occur in both Bali and Nusa Tenggara island. These two islands consist of three provinces namely Bali, West Nusa Tenggara (NTB), and East Nusa Tenggara (NTT). Data from Bank Indonesia in Bali Province showed that the highest outflow during the period January 2003 to December 2016 was 2976.89 billion rupiah which occurred in July 2015. Whereas, the highest inflow in the same period was 2201.82 billion rupiah which occurred in July 2016. July 2015 is the month of the Galungan and EidFitr festivals, while July 2016 is the month of the EidFitr holiday [4]. The explanation about this high fluctuation of inflow and outflow at Bali and Nusa Tenggara island on EidFitr and Galungan holidays could be seen in [4,5,6].

NTB Province is located between Bali Province and NTT Province. The three provinces are in one corridor namely corridor V. The regions of Bali and Nusa Tenggara develop tourism products as one of the regional revenues through foreign exchange generated. Tourism prospects of Bali, NTB, and NTT as corridor V in the Masterplan for the Acceleration and Expansion of Indonesian Development (MP3EI) since 2011 recorded rapid growth [7]. The proximity of the locations of the three Provinces has caused many domestic and foreign tourists to continue their tour from Bali to NTB and NTT. Moreover, results of the May 2013 survey conducted by Bank Indonesia (BI) showed that 60% of tourists vacationing in Bali wanted to continue their tour to NTB and NTT [8]. Domestic and foreign tourists continue their tour from Bali to NTB and NTT due to the Nyepi holiday celebrated by Hindus in Bali.

Previous research on inflow and outflow in the Bali and NTT regions showed that modeling using calendar variations had better results [5]. This study used univariate time series methods such as ARIMA, ARIMA with exogenous variables (ARIMAX), time series regression (TSR), Transfer Function, and TBATS. The exogenous variables used in the study were EidFitr and Galungan. Furthermore, the method was developed into a more complex method using a hybrid model [4,6,9]. The purpose of using a more complex method is to get a model with better forecast accuracy. Nevertheless, these studies still have many weaknesses, particularly the lack of accurate forecasting results. The weakness is probably due to the method used in the study can only explain the relationship between of time and not involve the location relationship that supposedly occurred at these three provinces.

To overcome the relationship between location, this study proposes modeling inflow and outflow at BI in Bali and Nusa Tenggara regions that can handle the interrelationship between time and location. The methods are VAR and GSTAR models. Bowerman and O'Connel [10] stated there are three main components that are often used in time series regression modeling, i.e. trends, seasonal factors, and calendar variations. Calendar variations used in this study are EidFitr, Galungan, and Nyepi holidays or known as multiple calendar variations. Hence, the pattern components involved in this study are trend, seasonal, and multiple calendar variations. Moreover, VAR and GSTAR models that involving exogenous variables are known as VARX and GSTARX, respectively. Thus, the aim of this study is to propose VARX and GSTARX modeling to overcome trend, seasonal, and multiple calendar variations both using simulation study and real data about inflow and outflow at the BI in Bali and Nusa Tenggara regions.

2 Calendar Variation Model

The calendar variation model is a time series model used to forecast data based on seasonal patterns with varying periods [11]. Calendar variation model can be modeled using time series regression. In general, the calendar variation model is based on the regression method if there is a trend, seasonal, and dummy component of the effects of calendar variations with the effect on the week g (e.g EidFitr effect) on the data, then the inflow model can be written as:

$$Z_{i,t} = \delta_1 t + \delta_2 D_{1,t} + \delta_3 D_{2,t} + \delta_4 t D_{1,t} + \delta_5 t D_{2,t} + \sum_{m=1}^{12} \gamma_m S_{m,t} + \sum_{g=1}^4 v_g L_{g,t} + \sum_{g=1}^4 \varphi_g L_{g,t+1} + N_{i,t} \quad (1)$$

whereas, the calendar variation model for outflow data is written as follows:

$$Z_{i,t} = \delta_1 t + \delta_2 D_{1,t} + \delta_3 D_{2,t} + \delta_4 t D_{1,t} + \delta_5 t D_{2,t} + \sum_{m=1}^{12} \gamma_m S_{m,t} + \sum_{g=1}^4 v_g L_{g,t} + \sum_{g=1}^4 \nu_g L_{g,t-1} + N_{i,t}, \quad (2)$$

where δ is a linear trend parameter, $S_{1,t}, S_{2,t}, \dots, S_{12,t}$ is a seasonal dummy variable (in this study is the month), v, φ , and ν are parameters of calendar variation for EidFitr effect (in the study are the month when EidFitr occurs, one month after and before the EidFitr occurrence).

3 VAR Model

One of the modeling in multivariate time series analysis is vector autoregressive (VAR). In time series modeling, before identifying the appropriate model for time series data, the data must be stationary first, both in the mean and variance. If time series data is not stationary in the variance, Box-Cox transformations are required. Whereas if the data is not stationary in mean then differencing is done on the data. In general, the equation of the VAR model with an order or vector AR(p) can be written as follows [12]

$$(\mathbf{I} - \varphi_1 B - \dots - \varphi_p B^p) \mathbf{Z}_t = \varphi_0 + \alpha_t. \quad (3)$$

If there is differencing on the seasonal VAR model, then the equation becomes [13]:

$$(\mathbf{I} - \varphi B)(\mathbf{I} - \Phi B^s)(1 - B)(1 - B)^s \dot{\mathbf{Z}}_t = \alpha_t, \quad (4)$$

where

φ_p = matrix $m \times m$ of the p -parameter

Φ = matrix $m \times m$ of the seasonal p -parameter

α_t = vector error $m \times 1$ of residual at time t

$\dot{\mathbf{Z}}_t$ = vector $m \times 1$ of data at time t .

After the time series data fulfills stationary condition, the next step in forming a VAR model is to identify the appropriate model. The identification of this model can be done by observing the pattern of the sample Matrix of Cross Correlation Function (MCCF) and the Matrix of Partial Cross Correlation (MPCCF).

4 GSTAR Model

The GSTAR model is a generalization of the Space-Time Autoregressive (STAR) model which is also a specification of the Vector Autoregressive (VAR) model. GSTAR is a time series model that reveals linear dependencies of space and time, expressed by spatial weight (\mathbf{W}). Let $\{\mathbf{Z}(t) : t = 0, \pm 1, \pm 2, \dots, T\}$ is a space-time data of N locations, then the GSTAR model with time order p and spatial order $\lambda_1, \lambda_2, \dots, \lambda_p$ or referred to as GSTAR ($p; \lambda_1, \lambda_2, \dots, \lambda_p$) can be written as follow [11, 14]:

$$\mathbf{Z}(t) = \sum_{k=1}^p \left(\Phi_{k0} + \sum_{l=1}^{\lambda_k} \Phi_{kl} \mathbf{W}^{(l)} \right) \mathbf{Z}(t-k) + \mathbf{a}(t), \quad (5)$$

where $\Phi_{k0} = \text{diag}(\phi_{k0}^1, \dots, \phi_{k0}^N)$, $\Phi_{kl} = \text{diag}(\phi_{kl}^1, \dots, \phi_{kl}^N)$ and $\mathbf{a}(t)$ is a residual model that satisfies identically, independently, distributed with mean $\mathbf{0}$ and variance Σ . For example, GSTAR model for time and spatial order one is as follows:

$$\mathbf{Z}(t) = \Phi_{10} \mathbf{Z}(t-1) + \Phi_{11} \mathbf{W}^{(1)} \mathbf{Z}(t-1) + \mathbf{a}(t). \quad (6)$$

There are several matrices of spatial weights (\mathbf{W}) in the GSTAR model. The weighted value selected meets the requirements $w_{ii}^{(k)} = 0$ and $\sum_{j \neq i}^N w_{ij}^{(k)} = 1$. In this research used four spatial weights, i.e. uniform weight, weight based on an inverse of the distance between locations, weight based on normalization of cross-correlation, and weight based on normalization of partial cross-correlation inference.

5 Research Methodology

This research was conducted in two studies, i.e. simulation study and application on monthly data inflow and outflow at BI area of Bali and Nusa Tenggara. Simulation studies were conducted by generating data that have trend, seasonal, and multiple calendar variations patterns as well as linear noise series. Multiple calendar variations consisted of three feast days in Bali and Nusa Tenggara regions, i.e. EidFitr, Galungan and Nyepi. The linear noise series followed GSTAR(1₁) or VAR(1) model with the parameters fulfilled the stationary conditions, i.e. the eigenvalue is less than 1, or $|\lambda \mathbf{I} - \Phi| = \mathbf{0}$, $|\Phi| = |\lambda \mathbf{I}|$, with $|\lambda_i| < 1$. The parameters of this model is

$$\Phi_1 = \begin{bmatrix} 0.45 & 0.25 & 0.25 \\ 0,15 & 0.40 & 0.15 \\ 0.20 & 0.20 & 0.35 \end{bmatrix}.$$

Simulation data was generated by assuming there were three locations with residuals satisfied multivariate normal distribution. Simulation studies were generated using six scenarios of residual covariance matrix. The residual covariance matrix that be used to generate six simulation data were as follows:

1. No correlation between residuals at three locations (1st scenario): the residual covariance matrix is

$$\boldsymbol{\mu} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 1.00 & 0.00 & 0.00 \\ 0.00 & 1.00 & 0.00 \\ 0.00 & 0.00 & 1.00 \end{bmatrix}.$$

2. No correlation between residuals at three locations with lower variance (2nd scenario): the residual covariance matrix is

$$\boldsymbol{\mu} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} 0.40 & 0.00 & 0.00 \\ 0.00 & 0.50 & 0.00 \\ 0.00 & 0.00 & 0.60 \end{bmatrix}.$$

3. All residuals at three locations were correlated (3rd scenario): the residual covariance matrix is

$$\boldsymbol{\mu} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} 1.00 & 0.50 & 0.45 \\ 0.50 & 1.00 & 0.55 \\ 0.45 & 0.55 & 1.00 \end{bmatrix}.$$

4. Not all residuals at three locations were correlated (4th scenario): the residual covariance matrix is

$$\boldsymbol{\mu} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} 1.00 & 0.50 & 0.45 \\ 0.50 & 1.00 & 0.00 \\ 0.45 & 0.00 & 1.00 \end{bmatrix}.$$

5. All residuals at three locations were correlated with lower variance (5th scenario): the residual covariance matrix is

$$\boldsymbol{\mu} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} 0.45 & 0.30 & 0.40 \\ 0.30 & 0.50 & 0.35 \\ 0.40 & 0.35 & 0.60 \end{bmatrix}.$$

6. Not all residuals at three locations were correlated with lower variance (6th scenario): the residual covariance matrix is

$$\boldsymbol{\mu} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} 0.45 & 0.30 & 0.40 \\ 0.30 & 0.50 & 0.00 \\ 0.40 & 0.00 & 0.60 \end{bmatrix}.$$

Furthermore, the inflow and outflow data that be used as case study is secondary data obtained from Bank Indonesia. The period of data from January 2003 to December 2016 is used as in-sample dataset and from January to October 2017 as out-of-sample dataset. The research variables in this study are shown in Table 1.

Some dummy variables are used in this study, i.e. dummy for trend, seasonal, and multiple calendar variations. Multiple calendar variations represent about EidFitr, Galungan and Nyepi. Moreover, these dummy variables are shown in Table 2.

6 Simulation Study

The simulation study was conducted to find out the performance of the methods used in modeling data that have trend, seasonal, multiple calendar variations, and linear noise series. The data in the simulation study is obtained from the following equation:

$$Z_t^{(i)} = T_t^{(i)} + S_t^{(i)} + IF_t^{(i)} + Gl_t^{(i)} + Ny_t^{(i)} + N_t^{(i)}, \quad (7)$$

where $i = 1, 2, 3$ indicate the locations and

Table 1: Research Variables (Billion IDR)

Inflow		Outflow	
Variable	Description	Variable	Description
$Z_{1,t}$	Bank Indonesia Bali	$Z_{1,t}$	Bank Indonesia Bali
$Z_{2,t}$	Bank Indonesia NTB	$Z_{2,t}$	Bank Indonesia NTB
$Z_{3,t}$	Bank Indonesia NTT	$Z_{3,t}$	Bank Indonesia NTT

- $T_t^{(i)}$ is a component for trend, where $T_t^{(i)} = \delta^{(i)}t$. The coefficients for trends used in all scenarios are the same, i.e. $\delta^{(1)} = 0.45$, $\delta^{(2)} = 0.35$, $\delta^{(3)} = 0.25$.
- $S_t^{(i)}$ are the seasonal components obtained from the following equation

$$S_t^{(i)} = \beta_1^{(i)} S_{1,t} + \beta_2^{(i)} S_{2,t} + \dots + \beta_{12}^{(i)} S_{12,t}, \tag{8}$$

where:

$$\begin{aligned} S_t^{(1)} &= 25S_{1,t} + 28S_{2,t} + 30S_{3,t} + 28S_{4,t} + 25S_{5,t} + 21S_{6,t} + 16S_{7,t} + 13S_{8,t} \\ &\quad + 9S_{9,t} + 13S_{10,t} + 16S_{11,t} + 21S_{12,t}. \\ S_t^{(2)} &= 22S_{1,t} + 24S_{2,t} + 27S_{3,t} + 24S_{4,t} + 22S_{5,t} + 17S_{6,t} + 13S_{7,t} + 10S_{8,t} \\ &\quad + 6S_{9,t} + 10S_{10,t} + 13S_{11,t} + 17S_{12,t}. \\ S_t^{(3)} &= 17S_{1,t} + 21S_{2,t} + 24S_{3,t} + 21S_{4,t} + 17S_{5,t} + 12S_{6,t} + 9S_{7,t} + 6S_{8,t} \\ &\quad + 4S_{9,t} + 6S_{10,t} + 9S_{11,t} + 12S_{12,t}. \end{aligned}$$

The equation of simulation study for multiple calendar variations consisting of EidFitr, Galungan, and Nyepi in three locations is as follows:

- $IF_t^{(i)}$ is the calendar variation for EidFitr that following

$$IF_t^{(i)} = \nu_1^{(i)} IF_{1,t} + \dots + \nu_4^{(i)} IF_{4,t} + \nu_1^{(i)} IF_{1,t-1} + \dots + \nu_4^{(i)} IF_{4,t-1}. \tag{9}$$

where:

$$\begin{aligned} IF_t^{(1)} &= 51IF_{1,t} + 55IF_{2,t} + 59IF_{3,t} + 63IF_{4,t} + 61IF_{1,t-1} + 57IF_{2,t-1} + 54IF_{3,t-1} + 49IF_{4,t-1}. \\ IF_t^{(2)} &= 49IF_{1,t} + 53IF_{2,t} + 57IF_{3,t} + 61IF_{4,t} + 60IF_{1,t-1} + 55IF_{2,t-1} + 51IF_{3,t-1} + 47IF_{4,t-1}. \\ IF_t^{(3)} &= 48IF_{1,t} + 52IF_{2,t} + 56IF_{3,t} + 60IF_{4,t} + 59IF_{1,t-1} + 54IF_{2,t-1} + 50IF_{3,t-1} + 46IF_{4,t-1}. \end{aligned}$$

- $Gl_t^{(i)}$ is the calendar variation for Galungan that following

$$Gl_t^{(i)} = \varsigma_1^{(i)} Gl_{1,t} + \dots + \varsigma_4^{(i)} Gl_{4,t} + \varsigma_1^{(i)} Gl_{1,t-1} + \dots + \varsigma_4^{(i)} Gl_{4,t-1}. \tag{10}$$

where:

$$Gl_t^{(1)} = 34Gl_{1,t} + 41Gl_{2,t} + 47Gl_{3,t} + 53Gl_{4,t} + 52Gl_{1,t-1} + 46Gl_{2,t-1} + 40Gl_{3,t-1} + 33Gl_{4,t-1}.$$

Table 2: Dummy Variables

Dummy Variables	Description
Trend	t , with $t = 1, 2, \dots, n$
Seasonal	$S_{1,t} = \begin{cases} 1, & \text{January in the } t\text{-th period} \\ 0, & \text{otherwise} \end{cases}$
	\vdots
Eid Fitr	$S_{12,t} = \begin{cases} 1, & \text{December in the } t\text{-th period} \\ 0, & \text{otherwise} \end{cases}$
	$IF_{i,t} = \begin{cases} 1, & \text{month when Eid Fitr (period } t) \text{ occurs in week-}i, \text{ with } i=1,2,3 \\ 0, & \text{otherwise} \end{cases}$
	$IF_{i,t-1} = \begin{cases} 1, & \text{month before Eid Fitr (period } t) \text{ in week-}i, \text{ with } i=1,2,3 \\ 0, & \text{otherwise} \end{cases}$
Galungan	$IF_{i,t+1} = \begin{cases} 1, & \text{month after Eid Fitr (period } t) \text{ in week-}i, \text{ with } i=1,2,3 \\ 0, & \text{otherwise} \end{cases}$
	$GL_{i,t} = \begin{cases} 1, & \text{month when Galungan (period } t) \text{ occurs in week-}i, \text{ with } i=1,2,3 \\ 0, & \text{otherwise} \end{cases}$
	$GL_{i,t-1} = \begin{cases} 1, & \text{month before Galungan (period } t) \text{ in week-}i, \text{ with } i=1,2,3 \\ 0, & \text{otherwise} \end{cases}$
Nyepi	$GL_{i,t+1} = \begin{cases} 1, & \text{month after Galungan (period } t) \text{ in week-}i, \text{ with } i=1,2,3 \\ 0, & \text{otherwise} \end{cases}$
	$Ny_{i,t} = \begin{cases} 1, & \text{month when Nyepi (period } t) \text{ occurs in week-}i, \text{ with } i=1,2,3 \\ 0, & \text{otherwise} \end{cases}$
	$Ny_{i,t-1} = \begin{cases} 1, & \text{month before Nyepi (period } t) \text{ in week-}i, \text{ with } i=1,2,3 \\ 0, & \text{otherwise} \end{cases}$
	$Ny_{i,t+1} = \begin{cases} 1, & \text{month after Nyepi (period } t) \text{ in week-}i, \text{ with } i=1,2,3 \\ 0, & \text{otherwise} \end{cases}$

3. $Ny_t^{(i)}$ is the calendar variation for Nyepi that following

$$Ny_t^{(i)} = \omega_1^{(i)} Ny_{1,t} + \dots + \omega_4^{(i)} Ny_{4,t} + \psi_1^{(i)} Ny_{1,t-1} + \dots + \psi_4^{(i)} Ny_{4,t-1}. \quad (11)$$

where:

$$Ny_t^{(1)} = 27Ny_{1,t} + 32Ny_{2,t} + 36Ny_{3,t} + 40Ny_{4,t} + 39Ny_{1,t-1} + 35Ny_{2,t-1} + 30Ny_{3,t-1} + 26Ny_{4,t-1}.$$

$$Ny_t^{(2)} = 22Ny_{1,t} + 26Ny_{2,t} + 31Ny_{3,t} + 36Ny_{4,t} + 35Ny_{1,t-1} + 29Ny_{2,t-1} + 24Ny_{3,t-1} + 21Ny_{4,t-1}.$$

4. Components for noise series used in simulation studies are linear noise series. The linear noise series equation used is

$$N_t^{(i)} = \phi_1^{(i)} N_{1,t-1} + \phi_2^{(i)} N_{2,t-1} + \phi_3^{(i)} N_{3,t-1} + a_t^{(i)}, \quad (12)$$

where:

$$N_t^{(1)} = 0, 45N_{t-1}^{(1)} + 0, 25N_{t-1}^{(2)} + 0, 25N_{t-1}^{(3)} + a_t^{(1)},$$

$$N_t^{(2)} = 0, 15N_{t-1}^{(1)} + 0, 40N_{t-1}^{(2)} + 0, 15N_{t-1}^{(3)} + a_t^{(2)},$$

$$N_t^{(3)} = 0, 20N_{t-1}^{(1)} + 0, 20N_{t-1}^{(2)} + 0, 35N_{t-1}^{(3)} + a_t^{(3)}.$$

The time series plot of the trend, seasonal, and multiple calendar variations are shown in Figure 1. Furthermore, identification of noise series is shown through the matrix plot between noise at the time t (N_t) and noise at the time $t - 1$ (N_{t-1}) for each location as shown in Figure 2. Based on the matrix plot of the residuals at scenario 1 to 6 in Figure 2, it can be seen that the relationship between N_t and N_{t-1} is a linear pattern. Each scenario is replicated 10 times, and all simulation data is analyzed using VARX and GSTARX.

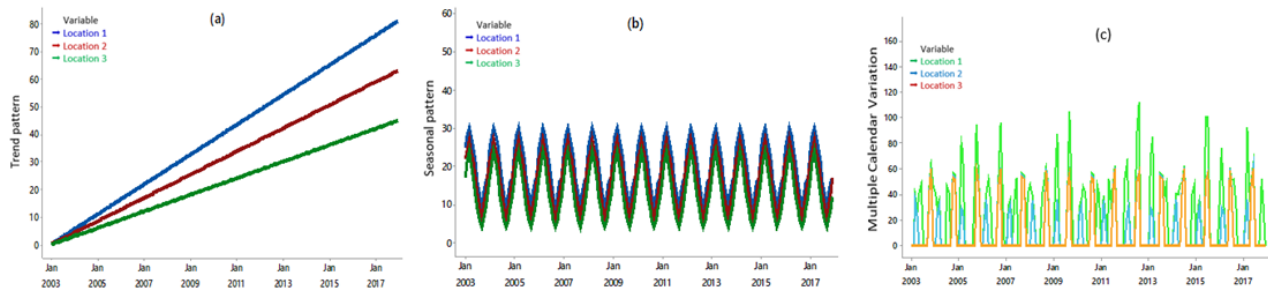


Figure 1: Time Series Plot of (a) Trend, (b) Seasonal, and (c) Multiple Calendar Variations

After implementing the VARX and GSTARX models to each simulation data scenario, comparison of forecast accuracy was carried out. The best model was determined based on the smallest RMSE value at the out-of-sample dataset. The RMSE value of simulation data at scenario 1 to 6 were replicated 10 times and can be seen in Figure 3 to 8.

Figure 3 to 8 show that the VARX and GSTARX models give the same forecast results. It can be seen from the mean and median of RMSE at out-of-sample dataset in all three locations for the six simulation studies are relatively the same. These results prove that GSTAR model, particularly GSTAR(1₁) is a special form of VAR(1) model, and imply the results of these both models are the same. The theoretical explanation about the relationship between GSTAR(1₁) and VAR(1) could be found in Borovkova, Lopuhaa, and Ruchjana [15].

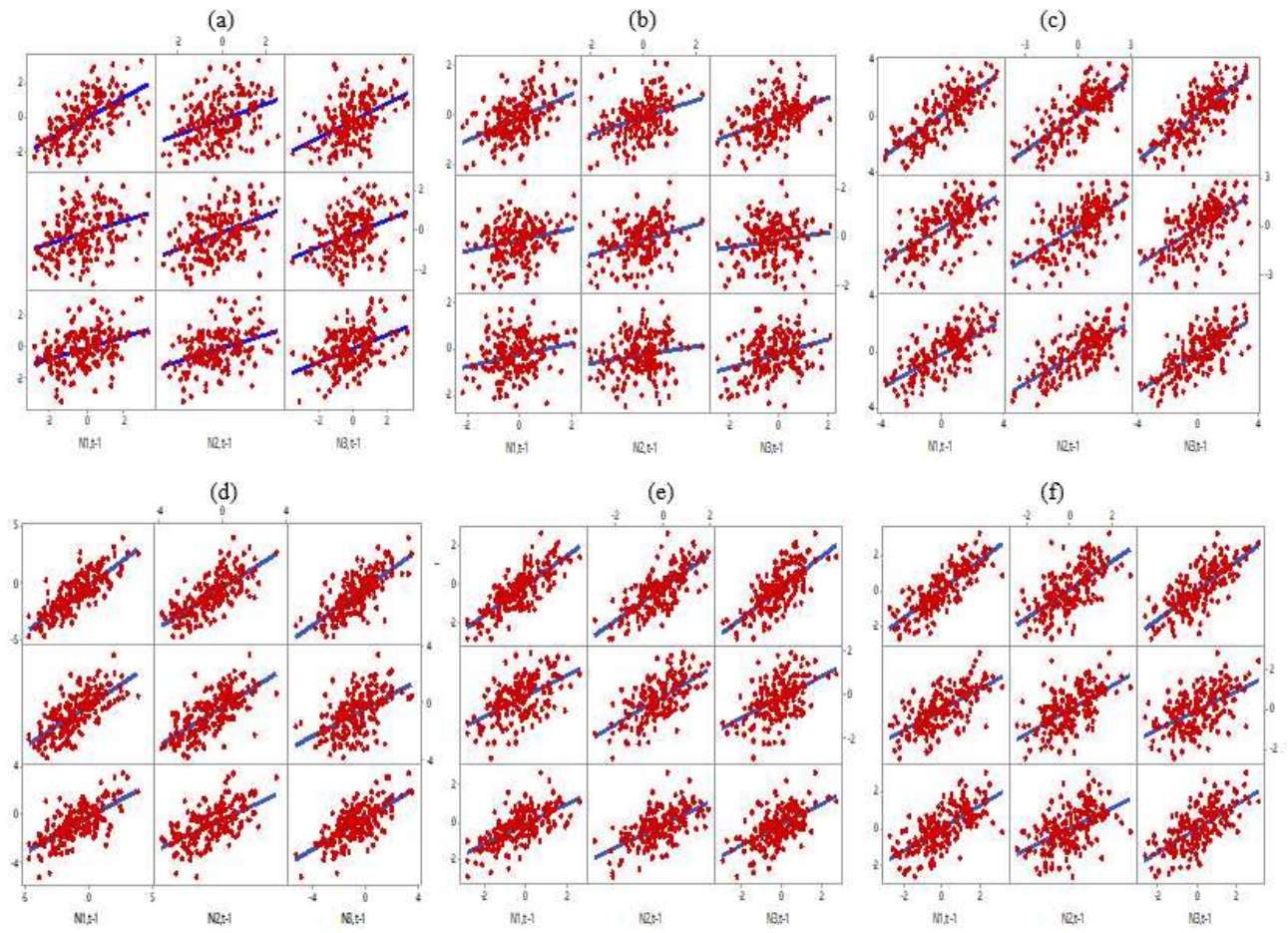


Figure 2: Scatter Plot between N_t and N_{t-1} from Linear Noise at (a) Scenario 1, (b) Scenario 2, (c) Scenario 3, (d) Scenario 4, (e) Scenario 5, (f) Scenario 6

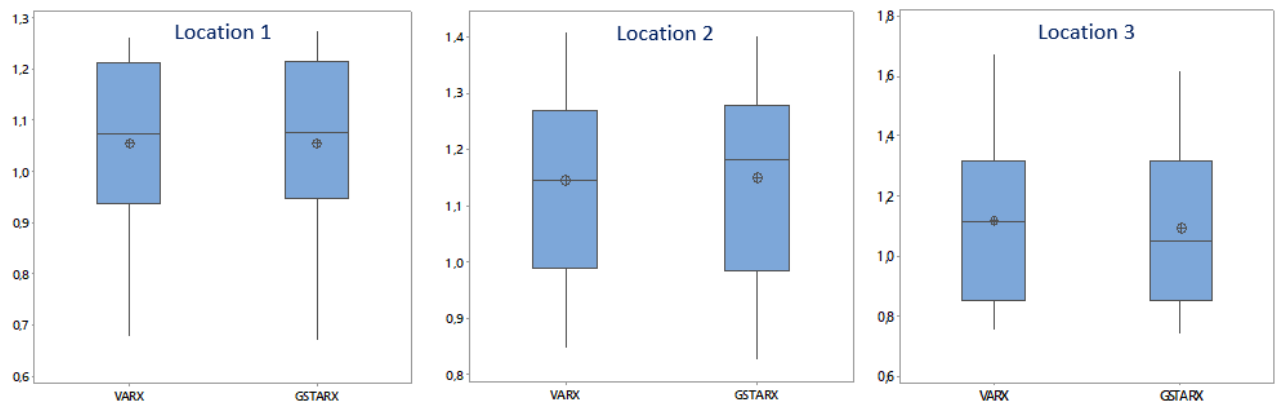


Figure 3: Box-Plot of RMSE in Three Locations at Scenario 1

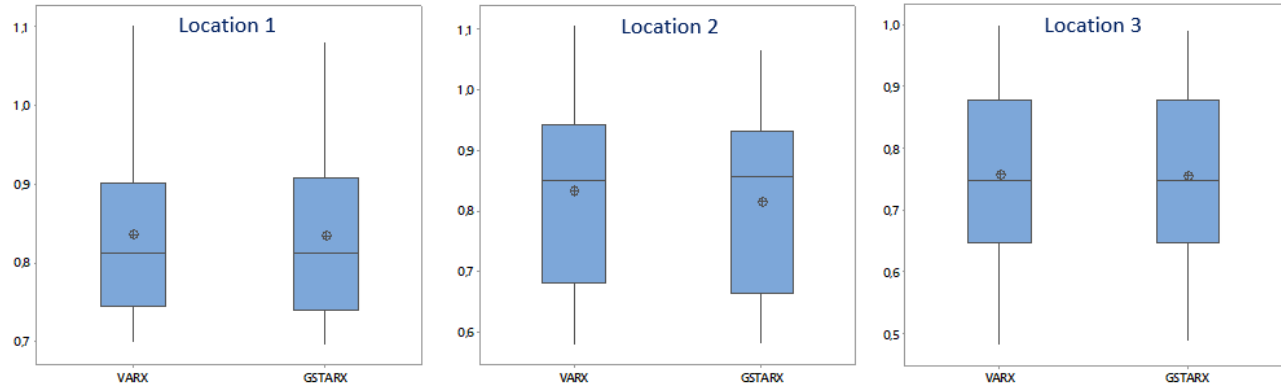


Figure 4: Box-Plot of RMSE in Three Locations at Scenario 2

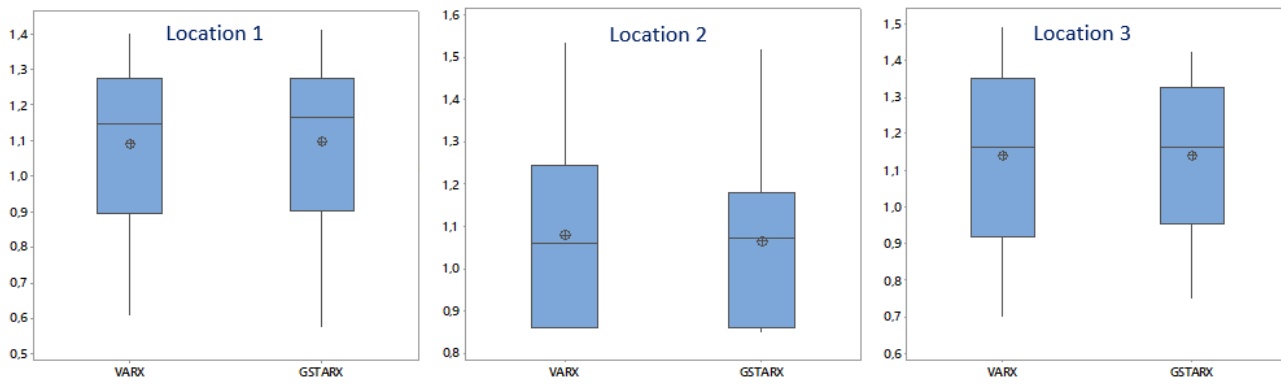


Figure 5: Box-Plot of RMSE in Three Locations at Scenario 3

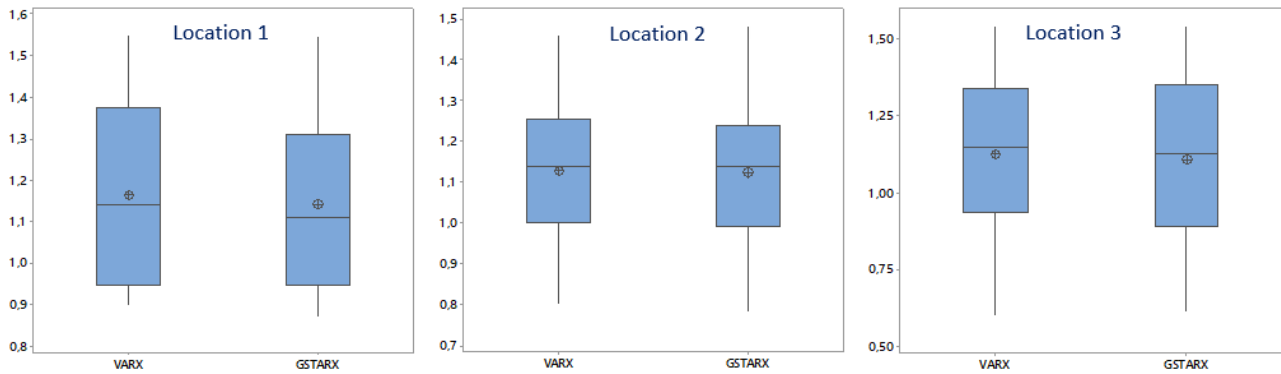


Figure 6: Box-Plot of RMSE in Three Locations at Scenario 4

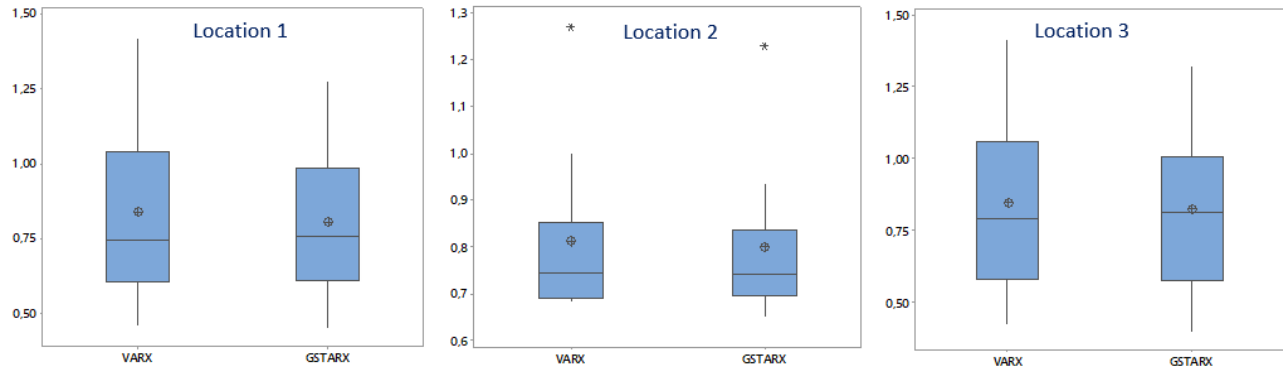


Figure 7: Box-Plot of RMSE in Three Locations at Scenario 5

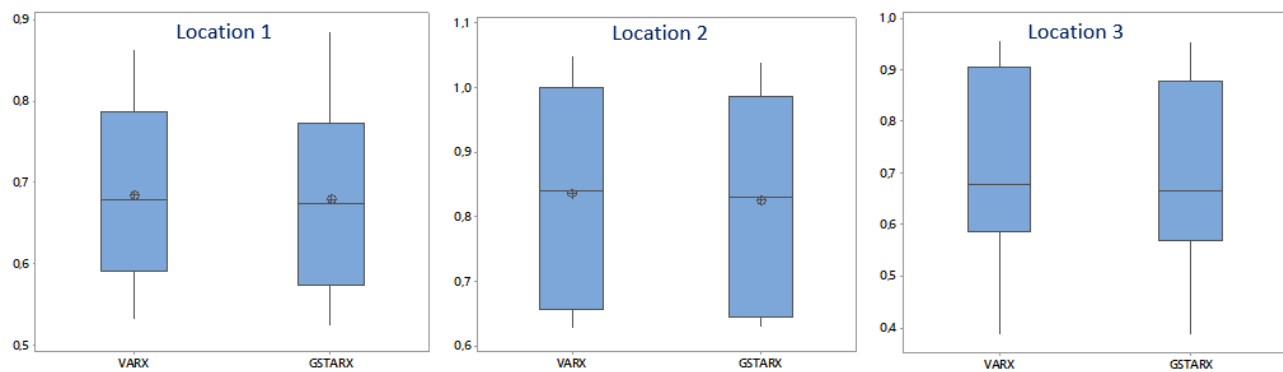


Figure 8: Box-Plot of RMSE in Three Locations at Scenario 6

7 Forecasting Inflow and Outflow with VARX and GSTARX

Outflow and outflow have high values around EidFitr, Galungan, and Nyepi events. EidFitr, Galungan, and Nyepi which occur in different weeks will affect differently on increasing inflows and outflows. The fluctuation of currency inflow and outflow in the Bali and Nusa Tenggara regions from 2003 to 2017 at each BI in Bali and Nusa Tenggara are presented in Figure 9.

Inflow and outflow modeling in the regions of Bali and Nusa Tenggara are carried out in two stages. The first stage is modeling trend patterns, seasonal, and multiple calendar variations using Time Series Regression (TSR) with the GLS estimation method. Multiple calendar variations in applied studies are the same as in simulation studies namely EidFitr, Galungan, and Nyepi. The use of calendar variations in all three locations or Provinces is adjusted to the third celebration of holidays in each Province. The second stage is residual modeling from the first stage using the VAR and GSTAR methods. Modeling in the second stage will be carried out like a simulation study, i.e the inflow and outflow data are modeled using the full model and restricted model.

Modeling of inflow and outflow at BI in Bali and Nusa Tenggara regions are carried out in two stages. The first stage is modeling trend, seasonal, and multiple calendar variations patterns using Time Series Regression (TSR) by implementing Generalized Least Square or GLS method for estimating the parameters due to the correlation between residuals. Moreover, multiple calendar variations in this case study are the same as in simulation study, i.e. EidFitr, Galungan, and Nyepi. The use of calendar variations in all these three locations is justified by the fact that there is celebration periodically for these three holidays in each province. The second stage is modeling of residuals from the first stage using VAR and GSTAR methods. Modeling process in this second stage was carried out as at a simulation study, i.e. the inflow and outflow data are modeled by using the full and restricted (by eliminating insignificant parameters) of VAR and GSTAR model.

8 Comparison of VARX and GSTARX

The comparison of forecast accuracy between VARX and GSTARX is done after modeling the residual of inflow and outflow data at the second stage. The results of forecast accuracy comparison between VARX and GSTARX (with certain spatial weight) for forecasting inflow at outflow data at BI in Bali and Nusa Tenggara regions are shown at Table 3 and 4, respectively.

Based on the RMSE criteria at out-of-sample dataset for selecting the best model, the results show that VARX([1,12]) model on full parameters (without considering significance) yield the smallest RMSE compared to the other methods for forecasting inflow data at BI in Bali and Nusa Tenggara regions. Otherwise, the GSTARX([2,12]₁) model on restricted parameters (only involving significance parameters) with uniform spatial weight give the most accurate forecast at out-of-sample dataset for forecasting outflow data at BI in Bali and Nusa Tenggara regions. These results in line with the conclusion from two previous forecasting competitions, i.e. M3 and M4 competition, stated that the complex models do not necessary yield better forecast than the simpler one [16,17].

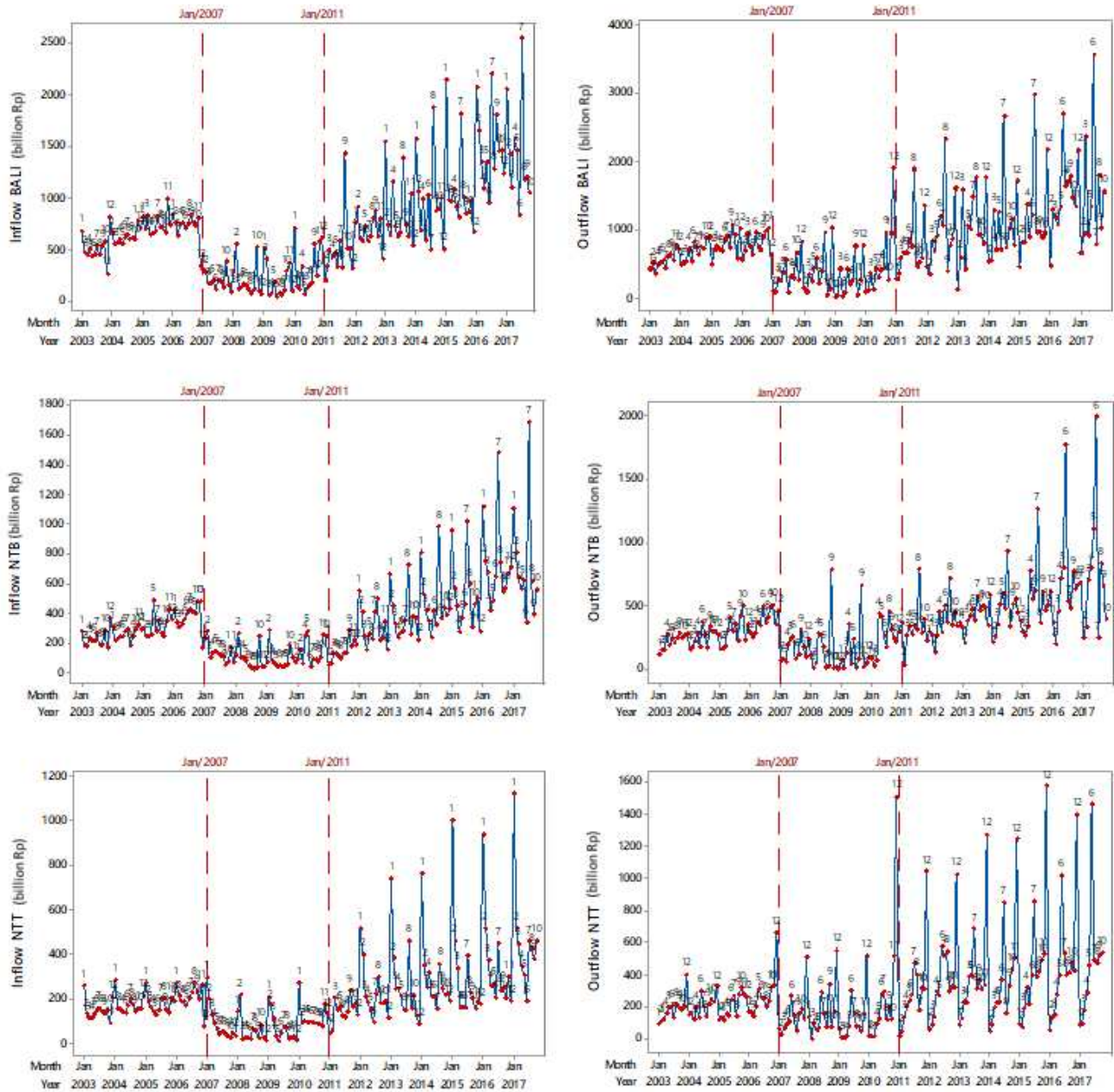


Figure 9: Time Series Plot of Inflow (Left) and Outflow (Right) at BI in Bali and Nusa Tenggara Regions

Table 3: The Results of Forecast Accuracy Comparison between VARX and GSTARX Models for Inflow Data at BI in Bali and Nusa Tenggara Regions

Significant Parameter	Model (Type of Spatial Weight)	RMSE at Out-of-Sample		
		Bali	NTB	NTT
Full Model	VARX	298.77	223.777	119.965
	GSTARX (uniform)	318.40	228.071	134.742
	GSTARX (Inverse of the distance)	336.43	231.352	137.111
	GSTARX (Normalization of Cross Correlation)	309.144	227.767	123.611
	GSTARX (Normalization Partial Cross Correlation Inference)	319.458	231.685	134.219
Restrict	VARX	300.796	225.665	127.316
	GSTARX (uniform)	316.658	231.644	133.837
	GSTARX (Inverse of the distance)	332.888	234.809	135.971
	GSTARX (Normalization of Cross Correlation)	315.386	230.689	126.468
	GSTARX (Normalization Partial Cross Correlation Inference)	317.352	232.356	134.319

Table 4: The Results of Forecast Accuracy Comparison between VARX and GSTARX Models for Outflow Data at BI in Bali and Nusa Tenggara Regions

Significant Parameter	Model (Type of Spatial Weight)	RMSE at Out-of-Sample		
		Bali	NTB	NTT
Full Model	VARX	508.627	328.876	263.096
	GSTARX (uniform)	488.374	323.089	239.677
	GSTARX (Inverse of the distance)	492.392	320.010	239.522
	GSTARX (Normalization of Cross Correlation)	505.284	327.615	241.576
	GSTARX (Normalization Partial Cross Correlation Inference)	514.767	331.516	244.393
Restrict	VARX	517.823	334.614	271.494
	GSTARX (uniform)	484.667	323.766	238.337
	GSTARX (Inverse of the distance)	489.910	323.721	247.288
	GSTARX (Normalization of Cross Correlation)	503.532	329.576	237.093
	GSTARX (Normalization Partial Cross Correlation Inference)	516.823	334.251	253.799

9 Conclusion

Based on the results of simulation study, it could be concluded that VARX and GSTARX models at two stage modeling could capture well the data patterns, i.e. trend, seasonal, multiple calendar variations, and linear noise patterns, and produce relatively similar forecast errors. Moreover, the results of case study about inflow and outflow at BI in Bali and Nusa Tenggara regions showed that the best model for forecasting inflow and outflow are VARX([1,12]) model on full parameters and GSTARX([2,12]₁) model on restricted parameters with uniform spatial weight, respectively. In general, these results in line with two previous competition forecasting results, i.e. M3 and M4 competition, that showed complex models did not necessary produce more accurate forecast than the simpler one [16,17]. Furthermore, due to the simulation study was replicated 10 times as preliminary study, further research by running more replication is needed to validate these results. Additionally, the criteria evaluation for selecting the best model in this research is only using RMSE. Hence, further research could be done by implementing more forecasting accuracy criteria to select the best model for evaluating the robustness of the results.

Acknowledgement

This research was funded by a LPPM-ITS under scheme of “Penelitian Pascasarjana”, project No. 1337/PKS/ITS/2018. The authors thank to the ITS and General Director of DIKTI for funding and to anonymous referees for their useful suggestions.

References

- [1] Liu, L. M. Identification of time series model in the presence of calendar variation. *International Journal of Forecasting* 1986. 2: 357-372.
- [2] Sullivan, R, Timmermann, A., White, H. Dangers of data mining: The case of calendar effects in stock returns. *Journal of Econometrics*. 2001. 105: 249-286.
- [3] Seyyed, F. J., Abraham, A., Al-Hajji, M. Seasonality in stock returns and volatility: The Ramadhan effect. *Research in International Business and Finance*. 2005. 19: 374-383.
- [4] Reganata, G. P. *Forecasting Currency Inflow and Outflow using Multi Input Transfer Function and Hybrid ARIMA-Artificial Neural Network in Bali Province*, Unpublished MSc Thesis, Department of Statistics, Institut Teknologi Sepuluh Nopember, Surabaya: ITS. 2015.
- [5] Dana, I. M. G. M. *Forecasting Inflow and Outflow as Projection of Currency Demand in Bali Province using Transfer Function, TBATS, ARIMA, and ARIMAX*, Unpublished Final Project, Department of Statistics, Institut Teknologi Sepuluh Nopember, Surabaya: ITS. 2016.
- [6] Wijayanti, D. G. P. *Forecasting Currency Inflow and Outflow in Bali Province using Hybrid ARIMAX-Neural Networks-GARCH*, Unpublished Final Project, Department of Statistics, Institut Teknologi Sepuluh Nopember, Surabaya: ITS. 2017.
- [7] <http://kip.sloka.or.id/wp-content/uploads/2011/05/Bali.pdf> retrieved on 19 August 2017

- [8] <https://beritabali.com/read/2013/08/27/201308270007/60-Persen-Wisman-di-Bali-Ingin-Lanjutkan-Wisata-ke-NTB-dan-NTT.html> retrieved on 19 August 2017
- [9] Puka, A. O. B. *Hybrid ARIMAX-QR and QRNN Model for Forecasting Currency Inflow and Outflow at Bank Indonesia in NTT Province and National*, Unpublished MSc Thesis, Department of Statistics, Institut Teknologi Sepuluh Nopember, Surabaya: ITS. 2017.
- [10] Bowerman, B. L., O'Connell, R. T. *Forecasting and Time Series: An Applied Approach* (3rd ed.). Belmont, CA: Duxbury Press. 1993.
- [11] Lee, M. H., Suhartono, Hamzah, N. A. Calendar variation model based on ARIMAX for forecasting sales data with Ramadhan effect. In *Proceedings of the Regional Conference on Statistical Science*. 2010. 349-361.
- [12] Wei, W. W. S. *Time Series Analysis: Univariate and Multivariate Methods* (2nd ed.). New York: Addison Wesley. 2006.
- [13] Tsay, R. S. *Multivariate Time Series Analysis*. University of Chicago: John Wiley & Sons, Inc. 2014.
- [14] Wutsqa, D. U., Suhartono, Ulama, B. S. *Generalized Space-Time Autoregressive Modeling*. The 6th IMT-GT Conference on Mathematics, Statistics and its Application (ICMSA), Kuala Lumpur, Malaysia. 2010. 752-761.
- [15] Borovkova, S., Lopuhaa, H. P., dan Ruchjana, B. N. Consistency and asymptotic normality of least squares estimators in generalized STAR models. *Statistica Neerlandica*. 2008. 62: 482-508.
- [16] Makridakis S, Hibon M. The M3 competition: result, conclusions and implications. *International Journal of Forecasting*. 2000. 16: 451-476.
- [17] Makridakis, S., Spiliotis, E., Assimakopoulos, V. The M4 competition: results, findings, conclusions and way forward. *International Journal of Forecasting*. 2018. 34: 802-808.