Available online at http://journal.uad.ac.id/index.php/Konvergensi

Jurnal Ilmiah Matematika 7 (2), 2020, 81-87

Forecasting Tourist Visit Using the Vector Autoregressive Exogenous Method (VARX)

Erni Muschilati[1*], Nursyiva Irsalinda[2]

Mathematics Department, Ahmad Dahlan University of Yogyakarta, Ringroad Selatan Street, Kragilan, Tamanan, Banguntapan, Yogyakarta, 55191
E-mail: ernimuslichati50@gmail.com
* Corresponding Author

ARTICLE INFO

Keywords
Fuzzy
Time Series
Forecasting
Traveler
Calendar Effects
Vector Autoregressive Exogenous (VARX)
MAPE

ABSTRACT

Forecasting is an activity to predict what will happen in the future by paying attention to information from the past and the present. A regression model that explains the past movement of the variable itself and also all other variables without distinguishing which endogenous and exogenous variables are called Vector Autoregressive (VAR). But in practice, endogenous variables are supported by exogenous variables. The Vector Autoregressive Exogenous (VARX) model is a development of the VAR with the addition of exogenous variables. The purpose of this study is to form the best model in the VAR method with the addition of an exogenous variable in the form of an effect calendar for forecasting the number of tourists coming to the Special Region of Yogyakarta (DIY). The data used in this study are time series data for 10 years from January 2009 to December 2018 in the form of tourist visit data in the Special Region of Yogyakarta (DIY). The results obtained indicate that the effect calendar variable that affects tourist visitor data in DIY is at Christmas. After being analyzed using MAPE, the best model is the VARX (1.0) model which produces a smaller. So, it can be concluded that the VARX model with the addition of an effects calendar is suitable for predicting tourist visits.

INTRODUCTION

Forecasting is a technique of estimating value for the future by paying attention to data or information from the past or present [1]. To make forecasts, a method that studies time series is time series analysis. Time series analysis can be used for data that has only one variable (univariate) and data that has many variables (multivariate). One of the forecasting models used in univariate time series data is ARIMA (Autoregressive Integrated Moving Average), while the model that is often used for multivariate time series data is Vector Autoregressive (VAR) [2].

The VAR model is a combination of several Autoregressive (AR) models, where these models form a vector between the variables which influence each other [3]. The advantage of the VAR method is that the model is simple, there is no need to distinguish between exogenous and endogenous variables because in this method all variables are considered endogenous variables, and the estimation of the model is simple, that is, it can use the least squares method (OLS) [4]. However, in practice, in time series analysis, endogenous variables are often influenced by variables that are determined outside the model called exogenous variables. The development of the Vector Autoregressive (VAR) model with the addition of exogenous variables in the model is called the Vector Autoregressive Exogenous (VARX) Model. Exogenous variables in VARX are determined outside the model and have the character of influencing endogenous variables in a system of equations.

Yogyakarta is one of the provinces that has its own magnet for tourists, apart from being known as a struggle area, cultural center, and education center, it is also known for its rich natural and cultural charm. In the holiday seasons and new years, the number of domestic tourists to DIY will increase more than on normal days. The increase in the number of tourists in each district in DIY from time to time is caused by the effect of calendar variations.

The addition of elements of the effect calendar can improve the forecasting of the method used. On this basis, in this study the Vector Autoregressive Exogenous (VARX) method was used to predict tourist visits in DIY with the addition of an effect calendar element.

METHODS

Stationarity

A data is said to be stationary if the variance value and the average of each lag are constant at all times. To find out whether the data is stationary in the variant, it can be seen from the lambda value (λ) or the rounded value of the Cox Box transformation. If the value of $\lambda = 1$ then the data is stationary in variants. If the data is not stationary in variants, data transformation can be performed. The stationarity test in the mean can be used the Augmented Dickey Fuller (ADF) unit root test. If the data is not stationary in average, then differencing is performed [5].

Ordinary Least square Methods

Least square method is a method to calculate a and b in the regression equation as α and β approximation, such that the sum of error squares has the smallest value. The estimation of β_1, β_2 is obtained the following formula (1)

$$\hat{\beta} = (X^T X)^{-1} X^T Y \tag{1}$$

There are two testing parameter that is used for knowing whether the independent variables inside the regression model significantly influence the dependent variable, regression significance test (F test) and individual regression coefficient test (t test). The important things in classic assumption test are normality test, heteroscedasticity test, autocorrelation test, multicollinearity test, and error dependency test [3].

Optimal Lag Examination

Optimal Lag Examination is used to determine optimal lag length which be used to determine the parameter estimation of VAR model. This can be caused by causal relationship and VAR model which are sensitive toward the lag length, so it is necessary to determine the proper optimal lag length [6]. To determine the optimal lag length in VAR model, it can be used information criteria as Akaike Information Criteria (AIC). The calculation of AIC is written as follows (2):

ISSN Print: 2087-8796, ISSN Online: 2774-3241

$$AIC = \ln\left(\det\left(\Sigma_p\right)\right) + \frac{2M^2P}{T} \quad (2)$$

With

Σp = covariance matrix of the residuals for the VAR(p) model
M = number of variables p = number of lags used T = amount of data
The optimal lag is on the smallest value obtained in calculation.

Dummy Regression

In regression analysis, it often happens that the dependent variable is influenced not only by the independent variable which is quantitative, but also by a qualitative character. However, this qualitative variable needs to be converted into quantitative form so that it can apply the regression method by forming dummy variables into the regression equation. The dummy variable has a value of 1 or 0. A value of 1 indicates the presence of an attribute while a value of 0 indicates the absence of an attribute. The general form of the dummy regression model is as follows:

$$Y_t = \beta_0 + \beta_1 D_{1t} + \beta_2 D_{2t} + \cdots + \beta_m D_{mt} + \varepsilon_t \quad (3)$$

where β_0 is the intercept and $\beta 1$, $\beta 2$, … , βm is the parameter coefficient associated with the dummy variables $D1t$, $D2t$, … , Dmt , and εt are the residuals of the dummy regression model [3].

Vector Autoregressive (VAR)

The vector autoregressive model is a development of the autoregressive (AR) model with more than one variable. In the VAR model, all variables are considered as endogenous and related variables. The general form of the VAR model with the order p (VAR (p)) is as follows [3]:

$$Y_t = \alpha_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_m Y_{t-p} + \varepsilon_t$$

Where:

Y_{t-i} = a vector of size $n \times n$ containing n variables that are included in the VAR model at times *t* and *t-i, i* = 1,2, ..., p.
α = intercept vector of size $n \times 1$
β_i = a coefficient matrix of size $n \times n$ for every *i* = 1,2, ..., p.
ε_t = error value vector of size $n \times 1$
p = lag VAR

Vector Autoregressive Exogenous (VARX)

The VARX model is a development of the VAR model with the addition of exogenous variables in the model. The general form of the VARX model (p) where p is the order (lag) of the endogenous variable and q is the lag of the exogenous variable can be written as follows [7]:

$$Y_t = \alpha_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_m Y_{t-p} + \theta X_t + \varepsilon_t$$

Where:

Y_{t-i} = a vector of size $n \times n$ containing n variables that are included in the VAR model at times *t-i, i = 1,2, ..., p.*
α = intercept vector of size $n \times 1$
β_i = a coefficient matrix of size $n \times n$ for every *i* = 1,2, ..., p.
X_t = vector of the exogenous variable at time *t-i, i = 1,2, ..., p.*
θ_j = matrix of exogenous variable sized parameters $n \times q$ for every *i = 1,2, ...,q.*
ε_t = error value vector of size $n \times 1$
p = lag VAR

Figures

The model's ability to forecast can be seen from the calculation of forecasting accuracy. One statistical measure that can be used to measure the accuracy of forecasting is the Mean Absolute Percentage Error (MAPE). The MAPE value is formulated as follows [8]:

$$MAPE = \frac{\sum_{t=1}^{n} \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right|}{n} \times 100\%$$

ISSN Print: 2087-8796, ISSN Online: 2774-3241

where Y_i is the actual observed value, $\hat{Y}_t$ is the forecast value, and n is the number of forecasts done. According to [9]the MAPE criteria are as follows:

Table 1. MAPE criteria

MAPE	Information
<10%	Very Good
10% - 20%	Well
20% - 50%	Enough
>50%	Bad

RESULTS AND DISCUSSION

Determination of Dummy Variables

The determination of the dummy variable in this study is based on the months that have big days which are thought to have an impact on fluctuations in tourist visits, such as the Christian New Year, Christmas Day, Education Day, Heroes' Day, Independence Day, and Youth Pledge Day.

$$D_{i,t} = \begin{cases} 1, \text{ the } t\text{-month with special events} \\ 0, \text{ another month} \end{cases}$$

With i = 1,2,3,…,6. With the dummy used, namely the Christian New Year ($D_{1,t}$), Christmas Day ($D_{2,t}$), Education Day ($D_{3,t}$), Heroes' Day ($D_{4,t}$), Independence Day ($D_{5,t}$), Young Pledge Day ($D_{6,t}$).

Estimating and Testing the Significance of the Calendar Variation Model Parameters with Linear Regression Analysis

The process of estimating model parameters with the dummy variable effect of calendar variations in equation (3) uses the least squares method. Meanwhile, to test the significance of the model parameters is done by partial test using the test.

The test results of the variation model parameter of the Tourist Visit in DIY

Table 2. Testing Results of Calendar Variation Model Parameters in DIY

Variable	t_{count}	Test Result
Constant	17.570	Significant
$D_{1,t}$	1.285	Not Significant
$D_{2,t}$	4.4220	Not Significant
$D_{3,t}$	2.145	Not Significant
$D_{4,t}$	-0.096	Not Significant
$D_{5,t}$	-0.768	Not Significant
$D_{6,t}$	-.0753	Not Significant

From Table 2, the effect of calendar variations that affect the number of tourist visits in Yogyakarta is at Christmas.

Data Stationarity Test

The data stationarity test in variants uses the Box Cox transformation. From the Box Cox transformation of each tourist visit data in DIY, a rounded value (λ) of 0.00 is obtained, therefore it is transformed into $\ln(x)$, so that λ = 1 is obtained. After the data is stationary in variance, then the stationary test is carried out on the average using the Augmented Dickey Fuller (ADF) test.

ISSN Print: 2087-8796, ISSN Online: 2774-3241

Table 4. ADF data transformation

Variable	t_{count}	t_{table}	**Result**
Gunung Kidul	-8.703166	-2.893508	Stationary
Yogyakarta City	-3.144635	-2.897490	Stationary
Kulon Progo	-3.805406	-2.897048	Stationary
Bantul	-7.534915	-2.2894607	Stationary
Sleman	-5.436033	-2.896616	Stationary

Optimum Lag Determination

Before determining the VAR method, it is necessary to select the appropriate lag. Identification of the appropriate lag is seen from the Akaike Information Criterion (AIC) value on the lag. In the end, the selected lag value is a significant lag value and has the smallest AIC value. Then given the AIC value to reinforce the lag to be selected until lag 7 as follows:

Table 6. AIC Value

Lag	**AIC**
1	-7.82549*
2	-7.76592
3	-7.78605
4	-7.25021
5	-7.60006
6	-7.67081
7	-7.54416

Based on Table 6, it can be seen that the smallest AIC value lies in lag 1 with a value of -7.82549, so the VAR model is chosen 1.

Estimation of the VARX model parameters

The parameter estimation for VARX modeling in this study used the least squares method. The results of VARX modeling (3.0) for variable tourist visits in Gunung Kidul ($Y_{1,t}$), Yogyakarta City ($Y_{2,t}$), Kulon Progo ($Y_{3,t}$), Bantul ($Y_{4,t}$), and Sleman ($Y_{5,t}$) as endogenous variables and calendar effects at Christmas (C) as exogenous variables are as follows:

$$\hat{Y}_{1,t} = 0.415524 + 0.39912Y_{1,t-1} - 0.0262Y_{2,t-1} + 0.49656Y_{3,t-1} + 0.02656Y_{4,t-1} + 0.10285Y_{5,t-1} + 0.4103C$$

$$\hat{Y}_{2,t} = 9.14216 + 0.10976Y_{1,t-1} + 0.36383Y_{2,t-1} - 0.0816Y_{3,t-1} + 0.07487Y_{4,t-1} - 0.1914Y_{5,t-1} + 0.592363C$$

$$\hat{Y}_{3,t} = 7.34658 + 0.12332Y_{1,t-1} - 0.1286Y_{2,t-1} + 0.75707Y_{3,t-1} - 0.1662Y_{4,t-1} - 0.1914Y_{5,t-1} + 0.612827C$$

$$\hat{Y}_{4,t} = 6.33439 + 0.08555Y_{1,t-1} + 0.19511Y_{2,t-1} + 0.00248Y_{3,t-1} + 0.20892Y_{4,t-1} - 0.0116Y_{5,t-1} + 0.2744C$$

$$\hat{Y}_{5,t} = 8.08815 + 0.12689Y_{1,t-1} - 0.0977Y_{2,t-1} - 0.1498Y_{3,t-1} + 0.02718Y_{4,t-1} + 0.4452Y_{5,t-1} + 0.25102C$$

Evaluation of the VARX Model Results

Evaluation of the results of the VARX model is carried out by calculating the MAPE value for the data forecast for the sample out of tourist visits in Yogyakarta.

ISSN Print: 2087-8796, ISSN Online: 2774-3241

Table 7. Evaluation of the Results of the VARX Model with MAPE

Variable	MAPE
Gunung Kidul	8.6580%
Yogyakarta City	8.1308%
Kulon Progo	11.1852%
Bantul	6.5978%
Sleman	7.2862%

From Table 7 it can be seen that each MAPE value for all variables is small, so based on Table 1 it can be concluded that the VARX (1.0) model has very good forecasting capabilities, so it can be used for forecasting tourist visits in Yogyakarta Province for the next periods.

Forecasting

The results of the forecast for tourist visits in Yogyakarta for the period January 2019 to December 2019 in Table 6 show that the highest increase occurred in December due to the calendar effect at Christmas.

Table 6. Forecasting tourist visits in Yogyakarta Province

Date	Gunung Kidul	Yogyakarta City	Kulon Progo	Bantul	Sleman
Jan 2019	67162	267338	42523	181811	326967
Feb 2019	71950	277083	46006	185898	315900
Mar 2019	76605	283352	49164	189297	309249
Apr 2019	80992	287637	52005	191951	305207
Mei 2019	85038	290682	54544	194053	302686
June 2019	88711	292925	56800	195752	301062
July 2019	92007	294631	58793	197147	299976
Aug 2019	94937	295966	60545	198306	299221
Sep 2019	97523	297036	62080	199279	298675
Oct 2019	99792	297908	63419	200099	298266
Nop 2019	101774	298630	64583	200794	297949
Dec 2019	156001	541236	121722	264924	382643

CONCLUSION

The forecasting results of the VARX model (1.0) experienced the highest increase in December due to the effect of the Christmas calendar and the VARX model (1.0) resulted in a MAPE value of 8.6580% for the tourist visit variable in Gunung Kidul, 8.1308% for the tourist visit variable in Yogyakarta City, 11.1852% for the tourist visit variable in Kulon Progo, 6.5978% for the tourist visit variable in Bantul, and 7.2862% for the tourist visit variable in Sleman. It can be concluded that the VARX (1.0) model has very good forecasting capabilities, so it can be used for forecasting tourist visits in Yogyakarta Province in the coming period.

ISSN Print: 2087-8796, ISSN Online: 2774-3241

REFERENCES

[1] Chang, P.C., Wang, Y.W., dan Liu, C.H. 2007. *The Development of a Weighted Evolving Fuzzy Neural Network*. Expert System With Application.
[2] Gujarati, D., dan Portner, D.N. 2003. *Basic Econometrics: Dasar-dasar Ekonometrika Edisi 5. Alih Bahasa Raden Carlos M*. Jakarta: Salemba Empat.
[3] Gujarati, Damodar. 2004. *Basic Econometrics*. New York: McGraw-Hill.
[4] Makridakis, McGee, dan Wheelright. 1999. *Introduction to Time Series Analysis and Forecasting*. Hoboken, New Jersey: Willey.
[5] Nachrowi, D.N., dan Usman, H. 2004. *Teknik Pengambilan Keputusan*. Grasindo: Jakarta.
[6] Ocampo, S., dan Rodriguez, N. 2011. *An Introductory Review of a Structural VAR-X Estimation and Applications*. Borradores de Economia.
[7] Wei, W.W.S. 1990. *Time Series Analysis*. Addison Wesley: New York.
[8] Wei, W.W.S. 2006. *Time Series Analysis Univariate and Multivariate Method 2ed*. New York: Pearson Education.
Durbin, J., & Koopman, S. J. (2012). *Time series analysis by state space methods*. Oxford university press.
[9] Widoarjono, A. 2007. *Ekonometrika Teori dan Aplikasi untuk Ekonomi dan Bisnis*. Edisi Kedua Yogyakarta: Fakultas Ekonomi UII.

ISSN Print: 2087-8796, ISSN Online: 2774-3241