

# Nonlinearity Test on Time Series Data

## Case Study: The Number of Foreign Tourists

Rahma Dwi Khoirunnisa<sup>1</sup>, Wahyu Wibowo<sup>2</sup>, Agus Suharsono<sup>3</sup>

<sup>1</sup> Department of Statistics , Institut Teknologi Sepuluh Nopember, ITS

<sup>2</sup> Department of Statistics , Institut Teknologi Sepuluh Nopember, ITS

<sup>3</sup> Department of Statistics , Institut Teknologi Sepuluh Nopember, ITS  
r4hm4dwi25@gmail.com

**Abstract**— *Time series data analysis is a method for modeling a data pattern. Forecasting is one of the main points in a time series analysis. difficult to choose the method of parametric models that are not linear. Before forecasting time series data nonlinearity testing should be done in order to explain the nonlinear relationships in the variable and testing procedures to detect the presence of nonlinear relationships. Some alternative methods that can be used to test the nonlinearity is Ramsey's RESET test, White test and Terasvirta test. Ramsey's RESET test is a test used to detect nonlinearity using general tests for specification error (Gujarati, 1996). White Test is a test developed to detect nonlinearity of neural network models were invented by White (1989). Terasvirta test is a test used to detect nonlinearity were also developed from neural network models and are included in the test group developed type of Lagrange Multiplier with Taylor expansion (Terasvirta, 1993). The purpose of this study was to demonstrate that the data on the number of foreign tourists is the data nonlinear with nonlinearity is tested using three methods: Ramsey's RESET test, White test and Terasvirta test. In this study will use data on the number of foreign tourists at Juanda airport in 2000 until 2015.*

**Keywords:** *nonlinearity test, time series data, number of foreign tourists*

## I. INTRODUCTION

Time series data is a series of observations on a value taken at different times. Time series data is data in chronological order. Time Series is a series of variables that form the observation values observed from time to time and recorded in accordance with the sequence of events and the data is assumed to be interdependent with one another (dependent). Such data can be collected periodically at certain time intervals, such as daily, weekly, monthly, or yearly. Autoregressive which is one of the Time Series models, first introduced by [8] and later developed by [5]. Autoregressive models of order  $p$  or AR ( $p$ ) states that the value of observation all  $t$  depend on the values of  $p$  observations throughout the previous period.

But in some cases, the relationship between the data have shaped nonlinear tendencies. Based on these cases, necessary to test to show time series data used in the model included linear or nonlinear models. The test can be used to indicate that data be linear or nonlinear. There are some test that can be used to show the nonlinearity. Non linearity test used in this study were White test, Ramsey's RESET test and Terasvirta test. Ramsey's RESET test is a test used to detect nonlinearity using general tests for specification error [1]. White Test is a test developed to detect nonlinearity of neural network models were invented by [7]. Terasvirta test is a test used to detect nonlinearity were also developed from neural network models and are included in the test group developed type of Lagrange Multiplier with Taylor expansion [3].

several studies using non-linear models such, testing for nonlinearity in time series: the method of surrogate data [4] indicating that correctly identifies nonlinearity in several well-known examples of low-dimensional chaotic time series, event when contaminated with dynamical and observational noise.

linearity test data time series with reset test [6] showed the results of the data generated from nonlinear models produce nonlinearity significant at 5%.

based on previous studies, the test nonlinearity actually quite important in identifying the time series data that is used for helping to sort the data in the model is a linear or nonlinear models. if included in the linear model can be regressed using a parametric regression otherwise if included in the nonlinear model can be regressed using nonparametric regression or semiparametric regression. Therefore in this study, the test will be conducted non-linearity in the time series data. time series data used is data on the number of foreign tourists at the airport juanda in 2000 to 2015.

## II. AUTOREGRESSIVE MODEL

Autoregressive model is a model that describes the dependent variable influenced by the dependent variable itself in periods previously, or autocorrelation can be interpreted also as a linear correlation sequence periodically with time series itself with a time difference (lag) 0, 1, 2 or more periods. The general form autoregressive model with order p or written with AR (p) has the following equation:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + e_t \quad (1)$$

Where:

$Y_t$  = the value of a variable at t

$\phi_i$  = autocorrelation parameter i-th with  $i = 1, 2, \dots, p$

$e_t$  = error value in t

### A. Autoregressive 1

AR order are often used in time series analysis is  $p = 1$  is a model AR (1). AR (1) stated that the observed values to t depends on the values of the observations throughout the previous period. The general form autoregressive model with order 1 or written with AR (1) has the following equation:

$$Y_t = \phi_1 Y_{t-1} + e_t \quad (2)$$

Where:

$Y_t$  = the value of a variable at t

$\phi_1$  = autocorrelation parameter

$e_t$  = error value in t

## III. NONLINEARITY TEST

According to [2] some tests to detect non-linear relationships between variables in time series analysis. in this section the discussion focused on the detection nonlinearity on a time series model, particularly Ramsey's RESET test, test and test Terasvirta White. The following is an explanation for each of the nonlinearity test.

### A. Ramsey's RESET test

Ramsey has proposed a general test of specification error called RESET (regression specification error test). The general shape models describing the relationship among the independent variables (predictors) and the dependent variable (response) can be written:

$$Y = f(X) + \varepsilon \quad (3)$$

Hypothesis testing used in the test are the nonlinearity:

$H_0 : f(X)$  is a linear function of the X or the linear model

$H_1 : f(X)$  is a nonlinear function of the X or the nonlinear model

$H_0$  is rejected, which means non-linear model is appropriate, if the value of the F test meets namely the rejection region

$$F > F_{\alpha; (df_1, df_2)} \text{ atau } p\text{-value} < \alpha \quad (4)$$

The following steps in the RESET test by [1]:

1. Regression of  $Y_t$  on  $1, x_1, x_2, \dots, x_p$  and calculate the estimated values of the response variable  $\hat{Y}_t$ , so:

$$\hat{Y}_t = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p \quad (5)$$

Calculate the coefficient of determination of the regression, the  $R^2$  and further denote the  $R_{old}^2$ .

2. Regression of  $Y_t$  on  $1, x_1, x_2, \dots, x_p$  and 2 additional predictors that  $\hat{Y}_t^2$  and  $\hat{Y}_t^3$  calculate the estimated values of the response variable  $\hat{Y}_t$ , so:

$$\hat{Y}_t = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p \quad (6)$$

Calculate the coefficient of determination of the regression, the  $R^2$  and further denote the  $R_{old}^2$ .

3. Calculate F test score

$$F = \frac{(R_{new}^2 - R_{old}^2)/m}{(1 - R_{new}^2)/(n - p - 1 - m)} \quad (7)$$

Where:

m: additional predictors

p: early predictors

n: number of data in used

4. Based on the hypothesis of linearity, shows F test values approaching F distribution with degrees of freedom of m and (n-p-1-m). Conclusion  $H_0$  is rejected if  $F > F(\alpha, m, n-p-1-m)$  or p-value  $< \alpha$  (typically use the alpha value of 0.05).

#### B. White test

White test is non linearity detection test developed from neural models network raised by [7]. This test is included in the test group of type Lagrange Multiplier (LM). Hypothesis testing used in the test are the nonlinearitas:

$H_0 : f(X)$  is a linear function of the X or the linear model

$H_1 : f(X)$  is a nonlinear function of the X or the nonlinear model

$H_0$  is rejected, which means non-linear model is appropriate, if the value of the F test meets namely the rejection region

$$F > F_{\alpha; (df_1, df_2)} \text{ atau } p\text{-value} < \alpha$$

The following steps in the White test by [3]:

1. Regression of  $Y_t$  on  $1, x_1, x_2, \dots, x_p$ , calculate the residual value  $\hat{u}_t$  and calculate residual sum of squares:

$$SSR_0 = \sum \hat{u}_t^2 \quad (8)$$

2. Regression of  $Y_t$  on  $1, x_1, x_2, \dots, x_p$ , m additional predictors so calculate residual  $\hat{v}_t$  and calculate residual sum of squares:

$$SSR_1 = \sum \hat{v}_t^2 \quad (9)$$

3. Calculate F test score

$$F = \frac{(SSR_0 - SSR_1)/m}{SSR_1/(n - p - 1 - m)} \quad (10)$$

Where:

m: additional predictors

p: early predictors

n: number of data in used

- Based on the hypothesis of linearity, shows F test values approaching F distribution with degrees of freedom of m and (n-p-1-m). Conclusion  $H_0$  is rejected if  $F > F(\alpha, m, n-p-1-m)$  or p-value  $< \alpha$  (typically use the alpha value of 0.05).

#### C. Terasvirta test

Terasvirta test included in the group Lagrange Multiplier (LM) test with a Taylor expansion approach that uses a test statistic  $\chi^2$  with degrees  $v$ . Terasvirta test procedure is described as follows [3]:

- Regression of  $Y_t$  on  $1, x_1, x_2, \dots, x_p$ , and calculate the residual value  $\hat{u}_t$ .
- Regression of  $Y_t$  on  $1, x_1, x_2, \dots, x_p$ , and m additional predictors which is the result of Taylor expansion approach.
- Calculate the coefficient of determination ( $R^2$ ) and regression in the previous step.
- Calculate statistics test  $\chi^2 = nR^2$  with n is number of data. Hypothesis testing used in the test are the nonlinearitas:

$H_0 : f(X)$  is a linear function of the X or the linear model

$H_1 : f(X)$  is a nonlinear function of the X or the nonlinear model

- Based on the hypothesis of linearity, shows  $\chi^2$  test values approaching  $\chi_v^2$  distribution. Conclusion  $H_0$  is rejected if p-value from  $\chi^2$  test values  $< \alpha$  (typically use the alpha value of 0.05).

### IV. METHODOLOGY

Data used in this study are secondary data the number of foreign tourists at the airport Juanda obtained from BPS. Data used in this study is the monthly data, the period to be examined is January 2000 to December 2015. This study begins with a description of the data that will be used to determine the amount of data to be used as well as other descriptions of the data. Then the data will be plotted using a time series plot to show data on the number of foreign tourists pattern at Juanda airport and final testing will be performed on the data using nonlinear three nonlinearitas test is Ramsey's RESET test, White test and Terasvirta test. This research using minitab program to descriptive data and time series plot, whereas for the non linearity test using the assistance program R.

### V. RESULTS AND DISCUSSION

The first step in this research was to determine the amount of data on the number of foreign tourists at the airport Juanda started January 2000 to December 2015. The following description is shown in the table data to be used:

TABEL 1. DESCRIPTIVE STATISTICS

Variable	Total count	Mean	SE Mean	St Dev
Juanda	192	11923	330	4569

Based on Table 1, indicated that the data used in this study as many as 192 data with mean 11923 and standard deviation 4569. Since determining much of the data used, the next step is to see patterns in the data on the number of foreign tourists. then be shown a pattern of data using time series plot in the figure below:

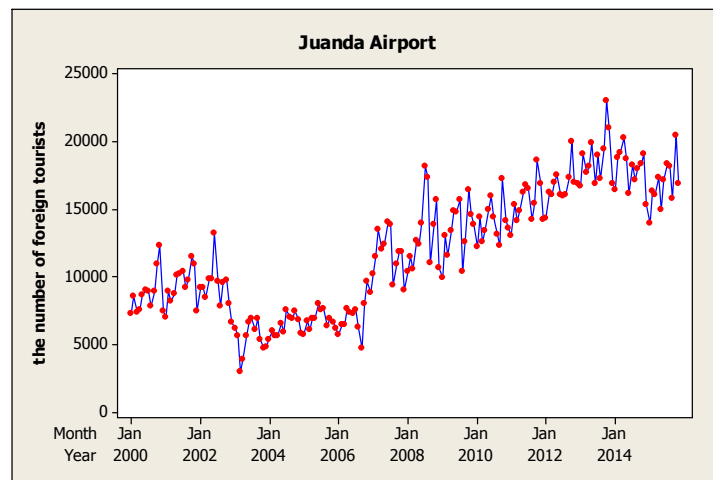


FIGURE 1. TIME SERIES PLOT OF JUANDA

Shown in Figure 1, data on the number of foreign tourists have a pattern that is up and down. The next will be tested for nonlinearity because such data is not always a linear pattern. before conducting the test, will be determined in advance time series model to be used in testing. the model used in this study is a model AR (1), this model is a time series model with a univariate predictor variables. The following models of the AR (1) to be used in this study:

$$Y_t = Y_{t-1} \quad (11)$$

after determining the models to be tested, last step the model will be tested using three test program nonlinearity using R. after the test is done using R obtained the following results:

```
#packages in used
library(lmtest)
library(tseries)

#Load Data
data=read.csv("d://wisman.csv",sep=";",header=T)
attach(data)

#Ramsey's RESET Test
model = Yt~Yt.1
reset.test = reset(model, type='regressor',data=data)
reset.test$p.value

#Terasvirta Test
teras.test=teravirta.test(x=Yt.1, y=Yt, type = "F",scale = T)
teras.test$p.value

#White Test
Whi.test=white.test(x=Yt.1, y=Yt, type = "Chisq",scale = T)
Whi.test$p.value
```

FIGURE 2. SYNTAX NONLINEARITY TEST

The syntax is obtained based on [2], with some simple changes. After the test is done using R obtained the following results:

```

> #Ramsey's RESET Test
> model = Yt~Yt.1
> reset.test = reset(model, type='regressor',data=data)
> reset.test$p.value
[1] 0.01006504
>
>
> #Terasvirta Test
> teras.test=teravirta.test(x=Yt.1, y=Yt, type = "F",scale = T)
> teras.test$p.value
[1] 0.009820545
>
>
> #White Test
> Whi.test=white.test(x=Yt.1, y=Yt, type = "Chisq",scale = T)
> Whi.test$p.value
[1] 0.007855382

```

FIGURE 3. RESULT NONLINEARITY TEST

Syntax in R is based on the steps [1], [7] and [3]. Based on the results of running using the R found that the three test produces a value less than the p-value of 0.05. Ramsey's RESET test the values obtained data on the number of foreign tourists at the airport juanda of 0.01006504, with a p-value of 0.05 so that it can be seen that the number of foreign tourists at the airport juanda smaller than the p-value. White test the values obtained data on the number of foreign tourists at the airport juanda of 0.009820545, with a p-value of 0.05 so that it can be seen that the number of foreign tourists at the airport juanda smaller than the p-value. Terasvirta test the values obtained data on the number of foreign tourists at the airport juanda of 0.007855382, with a p-value of 0.05 so that it can be seen that the number of foreign tourists at the airport juanda smaller than the p-value.

## VI. CONCLUSIONS

Based on the results and discussions can be concluded that the time series data, especially data on the number of foreign tourists at Juanda airport is a nonlinear model. Because it is based on three trials showed nonlinearity p-value less than 0.05. as shown in the table below:

	Ramsey's RESET test	White test	Terasvirta test
p-value	0.01006504	0.009820545	0.007855382

After finding out that the time series data can be non-linear form, this data can be used for regression semiparametric or nonparametric regression.

## REFERENCES

- [1] Gujarati, D.N., "Basic Econometric 5th Edition", New York: Mc Graw Hill International, page 521-523, 1996.
- [2] Suhartono, "Statistical Data Analysis with R", "Analisis Data Statistik dengan R", Surabaya:ITS, 2008.
- [3] Terasvirta, T., Linc, F and Granger, C.W.J., "Power of The Neural Networks Linearity Test", Journal of Time Series Analysis, vol.14 pp.159-171, 1993.
- [4] Theiler, J., Eubank, S., Longtin, A., Galdrikian, B., and Farmer, D., "Testing for Nonlinearity in Time Series: The Method of Surrogate Data" North Holland: Physica D, vol.58 pp.77-94, September 1992.
- [5] Walker, G., "On Periodicity in Series of Related Terms", proceeding of the royal society of london, ser.A, vol. 131, pp 518-532, 1931.
- [6] Warsito and Ispriyati, "Linearity Test Data Time Series With Reset Test", "Uji Linieritas Data Time Series Dengan Reset Tes", journal of mathematic and computer, vol.3 no.3, pp. 36-44, December 2004.
- [7] White, H., "An additional hidden unit test for neglected nonlinearity in multilayer feedforward networks", Proceedings of The International Joint Conference on Neural Networks, Washington, DC, pp.451-455, 1989.
- [8] Yule, G. Udny, "Why Do We Sometimes Get Nonsense-Correlations Between Time Series? A Study In Sampling And The Nature Of Time Series", journal of the royal statistical society vol.89, no.1, pp.1-43, January 1926.