IDENTIFICATION MODEL OF LONG MEMORY IN USE ELECTRICITY CHARGES IN JAVA-BALI

by Walid Matematika

ICMSE 2015
INTERNATIONAL CONFERENCE ON MATHEMATICS, SCIENCE, AND EDUCATION

IDENTIFICATION MODEL OF LONG MEMORY IN USE ELECTRICITY CHARGES IN JAVA-BALI

Walid[1,a)],and Subanar[2,b)], Dedi Rosadi[3,c)] ,Suhartono[4,d)]

[1]*Department of Mathematics, State University of Semarang*
[2]*Department of Mathematics, University of Gadjah Mada*
[3]*Department of Mathematics, University of Gadjah Mada*
[4]*Department of Statistics, State Institute of Technology*

ABSTRACT

In this article will discuss the identification of long memory models using a modified method Rescale Range. This method implemented several measures, among others: make a plot of time series and choose the appropriate transformation to stabilize the variance in the event data is not stationary in variants. Analysis of the autocorelation function (acf) and partial autocorelation function (pacf) to study the behavior and properties of other statistics. Furthermore, estimates the value of the parameter differentiating d by determining the average value, the adjusted mean, standard deviation, deviation cumulative and cumulative deviation range and ends with the determination of the value of Hurst using Ordinary Least Square method (OLS).

Key Word: ACF and PACF, Long Memory, Ordinary Least Square (OLS), Stasionarity

INTRODUCTION

According to the Law number 20 of 2002 on electricity, electricity is a secondary form of energy is generated, transmitted and distributed to all kinds of purposes, excluding the electricity used for communications, electronics or gesture. The size of the electric power consumption is determined entirely by the customer, ie depending on how the customer will use the tools of electricity then the State Electricity Company which is abbreviated by PLN must follow the electric power needs of these customers in terms of adjusting the electrical power it generates from time to time, Marsudi (2006). PLN is a State-Owned Enterprises (SOEs), which served to provide electricity in Indonesia. The demand for electricity is dependent on the use of electrical equipment used by the community, so that PLN is required to adjust the needs of the community litrik from time to time. PLN tasked to estimate the electric power consumption per hour customer. This estimate is made based on the data usage of electric power at an earlier time.. The magnitude of the burden of the system can not be calculated with certainty because the amount depends on the customer's needs for electricity. Therefore, the magnitude of the burden of the system should be estimated and called the approximate load. There are three groups of load estimates, ie estimates of the burden of long-term, medium-term load forecasts and estimates of short-term load.

Estimate or forecast the amount of electric power used to optimize the use of power in society, so that there will be no wastage or power outage. According to F. Chan and C. Lim (2011) forecasting is an important tool in planning the fektif and efficient.So if it is based on a count of the Central Bureau of Statistics which is abbreviated by BPS, there are 50 billion SMEs, mean total losses reached 2.5 billion unavoidable due to blackouts by PLN. Not surprising because the electrical energy is indeed a prime mover in industrial activity and a wide range of human needs. Electricity has become a necessity indispensable to man as closely linked to the overall economic base. Ranging from the smallest to the greatest extent.Specifically in urban areas, dependence on electrical energy is very high and it concerns the core activities within the scope of the industry that allows can subdued overall economic security.

Other aspects that also suffered a setback as a result of planned blackouts by PLN is the effect on the decrease in the interest of investors to invest in Indonesia because he felt there was no guarantee of the power supply becomes the driving force of industrial production processes. According to Box and Jenkins (1976) there are several stages in forecasting, among others, model identification, parameter estimation, verification of models and forecasting. At the time series prediction models forecasting the future state be based on the past. Time series approach can use several methods, may include analysis using autocorrelation function and partial function autokoralasi to study changes with time series parametric models known as ARIMA time domain analysis.

Besides, there is an alternative approach that can be used as a method of spectral analysis. The spectral method is one form of the Fourier transform.To be able to forecast properly then one of the first steps is to identify the data model in order to know whether the data included in the data group which has a long memory categories or short memory. Identification of this initial data will provide important information relating to the next step of estimating the parameters terhadapa-existing parameters, then verify and new forecasting. Information good start would be an important contribution in providing a level of accuracy of forecasting.

Based on the above background, this paper will focus on a study to identify the model of long memory using the rescale range. The data is the data used to simulate the electrical load is calculated every hour throughout the year. Issues to be studied is how does the identification of long memory models using the rescale range?

II. TIME SERIES ANALYSIS AND LONG MEMORY MODEL

Time series analysis by Chatfield (2001) has several purposes, namely forecasting, modeling, and control. Forecasting issues related to the establishment of models and methods that can be used to produce an accurate forecast. Modeling aims to obtain an appropriate statistical model to represent the behavior of a long-term time series data. Difference modeling with forecasting is forecasting more likely on a model of the "black-box" to get the forecast, while modeling tends to models that can be interpreted to explain what is going on with regard to the relationship between variables in a time series data. While the purpose of the control is widely used in engineering, especially signal processing.

Statistical modeling of time series analysis if traced back, preceded by Yule (1927) which introduced autoregresive linear models (AR) to predict yearly sunspot numbers. Since the publication concerned with the analysis of time series is growing rapidly. Until 1980, most of the research has focused on the model of linear time series, especially linear model class Autoregressive Integrated Moving Average (ARIMA). Box and Jenkins (1976) developed a complete procedure for ARIMA model methodology hitherto used as a standard procedure in the establishment linear time series models. Some literature which discusses many ARIMA model can be seen in Cryer (1986), Wei (1999) and Box et al. (1994). In addition, the properties related to statistical theory for the ARIMA model has also been extensively analyzed and developed by several researchers, among others, have been conducted by Brockwell and Davis (1991).

Time series is a series of observations Y_t on a variable Y, which each observation recorded at a certain time $t \in T$. In this case T is the set of time in which the observations made. If T is a discrete set, then $\{Y_t\ ,\ t \in T\}$ is a discrete time series. $\{Y_t\ ,\ t \in T\}$ is a notation whole time series, where Yt is the observation of $\{Y_t\ ,\ t \in T\}$ in time to t. In the case of discrete time series, observations are usually taken at the same time interval.

Statistical approach to the analysis of time series conducted using statistical models to describe the dynamic behavior of a time series. It is assumed that a time series generated from a mechanism or a stochastic model, which is usually defined by a stochastic difference equation. Stochastic difference equation that consists of an equation and some initial conditions. Results or a solution of this model is a stochastic process, which is a sequence of random variables $\{Y_t\}$ are defined on the probability space (Ω,ℜ,P). For certain ω Ω, $Y_t(\omega)$ is called a realization (sample path or trajectory) of $\{Y_t\}$ Each observation $Y_t(\omega)$ is a realizable value of the random variable Y_t whose values obtained in d-dimensional Euclidean space $\mathbb{R}^d$.

The most important assumptions in the analysis of time series is stationary data. Nonstationarity inspection data can be done with the help of time series plots using scatter plots. Nonstationarity in the mean can be resolved by a process of differentiating (differencing). According to Box and Jenkins (1976) stationary data in variance can be seen with a value of λ (lambda) estimate the Box-Cox transformation

Brockwell and Davis expressed a process $\{Y_t\}$ is said to be stationary stronger if $(y_{t_1}, y_{t_2}, \ldots, y_{t_k})'$ and $(y_{t_1+l}, y_{t_2+l}, \ldots, y_{t_{k+1}})'$ have the same joint distribution function for all integers k≥1dan for all $t_1, t_2, \ldots, t_k, k \in \mathbb{N}$. Processes with the first and second moments that are independent of time is also a concern in time series analysis. The following definitions relating to the concept of a weak stationary or stationary until the second order.

If given $\{Y_t\}$ is a process with $E|Y_t|^2 < \infty$ for every $t \in \mathbb{N}$, then $\{Y_t\}$ is said to be stationary weak (weakly stationary) if $E(Y_t) = \mu$ for all $t \in \mathbb{N}$ and $Cov(Y_{r,}Y_s) = Cov(Y_{r+h,}Y_{s+h})$ for all $r, s, h \in \mathbb{N}$. If $\{Y_t\}$ is a process that is stationary weak (weakly stationary), then the function autokovarians $\gamma_y(.)$ of $\{Y_t\}$ defined $\gamma_y(h) = Cov(Y_{t,}Y_{t+h})$ to all $t, h \in \mathbb{N}$ (Brockwell and Davis, 1991).

Time series that have a long-term dependence, or better known as the long memory is a time series in which observations are far apart still have a high correlation. Autocorrelation of a process of long memory fell hyperbolic. A sequence $\{Y_t\}$ is said to follow the model Autoregresive Integrated Moving Average if the distinction to d namely $W_t = \nabla^d Y_t$ is a stationary ARMA process. If W_t is ARMA (p, q) then Y_t is ARIMA (p, d, q). In practice the value of d used is generally 1 or at most 3 (Wei, 1999).Model ARFIMA able to model the dependence of short-term and long-term. Observations generated by the structure of ARMA shows the dependence of short-term, while the distinction fractional parameter d, which caused values to fall hyperbolic ACF showed long-term dependence.

According to Wilfredo Palma (2007) definition of long memory can be understood as follows. Let $\gamma(h) = \langle y_t, y_{t+h} \rangle$ is the covariance function of the lag h of a stationary process $\{y_t : t\epsilon\mathbb{Z}\}$. The usual definition of long memory is

$$\sum_{h=-\infty}^{\infty} |\gamma\ (h)| = \infty \qquad (1)$$

However, there is an alternative definition. In particular, long memory can be defined by specifying the hyperbolic decline in autokovariansinya. Recalling that a function of measuring positively defined in some environments [a, ∞) infinite is said to change slowly in a sense because if and only if for every $c > 0, l(cx)/l(x)$ converges to 1 of x tends Infinity. Examples slowly changing function is $l(x) = \log(x)$ and $l(x) = b$, where b is a positive kosntanta. Furthermore notation $x_n \sim y_n$ which means that $\frac{x_n}{y_n} \rightarrow 1$ of $n \rightarrow \infty$, unless otherwise stated.

$$\gamma\ (h) \sim h^{2d-1} l_1(h) \qquad (2)$$

For $h \rightarrow \infty$, where d is called long memory parameters and $l_1(.)$ is a function that is changing slowly.

Another widely used definition of a strong dependence on the spectral domain is

$$f(\lambda) \sim |\lambda|^{-2d} l_2(1/|\lambda|) \qquad (3)$$

To λ in zero environments and $l_2(.)$ is a slowly changing function. Furthermore, an alternative definition of long memory on the rules that are based directly on the decomposition Wold

$$\psi_j \sim j^{d-1} l_3(j) \qquad (4)$$

for j> 0, where $l_3(.)$ is a slowly changing function. Model ARFIMA (p, d, q) developed Granger and Joyeuk (1980) is written as:

$$\phi(B)(1-B)^d(Y_t - \mu) = \theta(B)a_t \qquad (5)$$

With t: index of observation

d: distinguishing parameters (fractions)

μ: average observation

$a_t \sim \text{IIDN}\ (0, \sigma_a^2)$,

$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$ is a polynomial AR (p),

$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$ q are polynomials MA (q),

$(1 - B)^d = \nabla^d = \sum_{k=0}^{\infty} \binom{d}{k} (-1)^k B^k$ differentiator fractional operator,

$$\sum_{k=0}^{\infty} \binom{d}{k} (-1)^k B^k = F(-d, 1; 1; B)$$

$$= \sum_{k=0}^{\infty} \frac{\Gamma(k-d)}{\Gamma(k+1)\Gamma(-d)} B^k$$

F is the hypergeometric function is defined as follows.

$$F(a, b; c; d) = 1 + \frac{a.b}{1.c} B + \frac{a(a+1)b(b+1)}{1.2.c(c+1)} B^2 + \cdots \quad (6)$$

a, b and c are real numbers and B is the operator Backshift.

For d is value fraction, fractional differencing operator $(1 - B)^d$

Defined as

$$(1 - B)^d = 1 + \sum_{k=1}^{\infty} \frac{\Gamma(-d+k)}{\Gamma(-d)k!} B^k \quad (7)$$

If the equation $\lambda_k(d) = \frac{\Gamma(-d+k)}{\Gamma(-d)k!}$ in equation (7) described for various values of k then:

For k = 1, is obtained $\frac{\Gamma(-d+1)}{\Gamma(-d)1!} = \frac{(-d)!}{(-d-1)!1!} = -d$,

For k=2, diperoleh $\frac{\Gamma(-d+2)}{\Gamma(-d)2!} = \frac{(-d+1)!}{(-d-1)!2!} = \frac{-d(1-d)}{2}$,

k=3, diperoleh $\frac{\Gamma(-d+3)}{\Gamma(-d)3!} = \frac{(-d+2)!}{(-d-1)!3!} = \frac{-d(1-d)(2-d)}{6}$ and so on so that the equation (3.3) can be written back into

$$(1 - B)^d = 1 + \sum_{k=1}^{\infty} \lambda_k B^k \quad (8)$$

with $\lambda_0(d) = 1$, $\lambda_1(d) = -d$ $\lambda_2(d) = -\frac{1}{2}d(1-d), \lambda_3(d) = -\frac{1}{6}d(1-d)(2-d)$ and so on. So that equation (3.4) above can be written back into

$$(1 - B)^d = 1 - dB - \tfrac{1}{2}d(1-d)B^2 - \tfrac{1}{6}d(1-d)(2-d)B^3 - \cdots \quad (9)$$

III. RESEARCH METHODS

The data used in this study is the electricity consumption data which is secondary data from the PLN in Java and Bali. Data is data consumption every half-hour electricity load Se Bali-Java, taken from January 1 to the date of the end of December 2010. Furthermore, the data were divided into two groups, as the data in the sample is observed during the period 1 January until the end of November 2010 and Data out Sample observation for December

2010. Through the program Minitab and Matlab searchable form plot (scatter) data. Furthermore, the data in the analysis to determine the magnitude of the value difference d (difference) method rescale range (R / S). This process begins with determining the average value of the data, the adjusted mean and standard deviation.

Furthermore, the determination of the value of the cumulative deviation and range of the cumulative deviation, and the final stage is to determine the statistical value of Hurst (H) with melogaritmakan statistics (R / S) and assess the value of H through Ordinary Least Square method (OLS).

IV. RESULT

Results of original data plot in figure 1 shows the data is relatively flat, indicating that the stationary time series in the mean and variance. Plots containing long time series of memory tend to be stationary in the average, because the value of d is small.

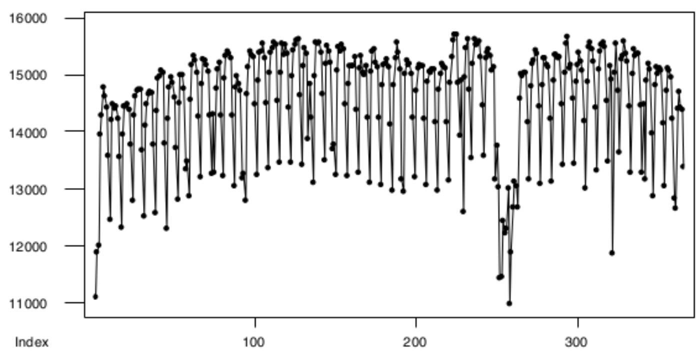


Figure 1. Plot the original data

Box-Jenkins method modeling ARFIMA through several stages, namely the identification of the model, parameter estimation, verification and forecasting. Basically the method used had the same stage with ARIMA models but each stage has its own distinction. To better show the difference will be given a sample application using electrical data are created every hour. Build an ARIMA model is based on the procedure Box-Jenkins (Box, Jenkins 1994), begins with the identification of the sequence of the data stationary models. Figure 1 and Figure 2 shows that the data is stationary. Results of the autocorrelation function graph (acf) and partial autocorrelation function (pacf) can be seen in the following figure.

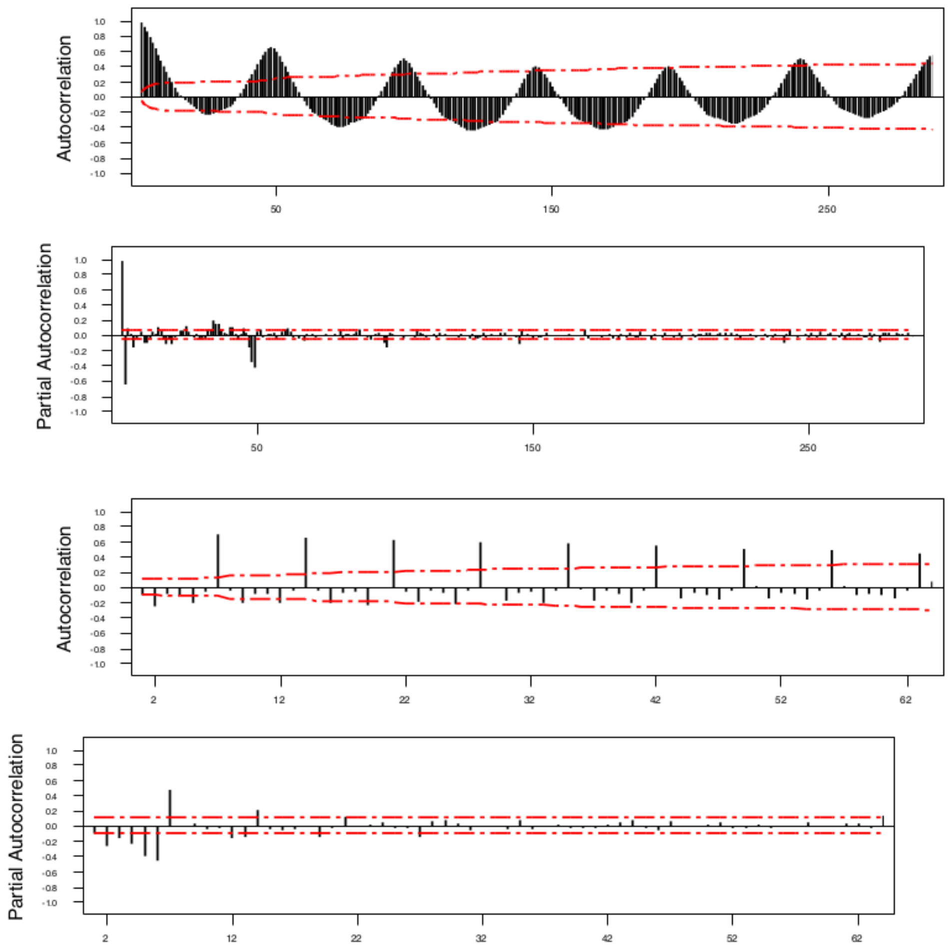


Figure 4. acf and pacf original data and the data first difference

ACF and PACF plot based on the original data, and stationary data showed that the dissolution was based on the second lag, with a value of 0.98 down to 0.65 fakp. Visible differences in age seasonal (lag 12). After seasonal differences in age, ACF and PACF plots. ACF plot indicates that the ACF on a regular lag down very slowly and shows that the sequence differences could be necessary. Furthermore, differences in the data sequence has a regular and daily seasonal ACF and PACF plots. ACF plot results showed that the 168 and 336 lag significantly and tends stopped down very slowly. Therefore, it requires to use a sequence of weekly seasonal difference (lag 24), (lag 28) and (lag56).

Model ARFIMA identification of possible patterns of long memory time series Y_t can be seen from the value of Hurst (H), performed by the method rescale range (R / S) (Yu WC and Zivot, 2009). Methods rescale range (R / S), commonly often used to test the effects of long memory. The method was introduced by Harold Edwin Hurst in hydrological studies in 1951. Test of R / S is based on the range of R (k, T) of the sample standard deviation S (k, T), with some of the following steps.

Determination of average value, ajusted mean and standard deviation of the time series. Those values can be given as in the formula below

$$\bar{y}_t = \frac{1}{t}\sum_{i=1}^{t} y_i \quad \text{, for } t = 1, 2, \dots, T \tag{10}$$

$$y_{t,i}^{adj} = y_i - \bar{y}_t \quad \text{, for } t = 1, 2, \dots, T \tag{11}$$

The next step is to determine the cumulative deviation values and value ranges and cumulative deviation of the time series. Those values can be given as in the formula below

$$y_t^* = \sum_{i=1}^{t} y_i^{adj} \text{, with } t = 1,2, \dots, T \tag{12}$$

$$R_t = \text{Maks}\,(y_1^*, y_2^*, \dots, y_t^*) - \text{Min}\,(y_1^*, y_2^*, \dots, y_t^*) \tag{13}$$

$$S_t = \sqrt{\frac{1}{t}\sum_{i=1}^{t}\left(y_{t,i}^{adj}\right)^2} \quad \text{, for } t = 1, 2, \dots, T \tag{14}$$

The final step is to determine the value of Hurst (H) through statistical [ratio $(\frac{R}{S})_t = \frac{R_t}{S_t}$, referred to rescale the adjusted range or statistics (R/S) (Beran, 1994).The value of Hurst (H) is determined through a statistical R/S of the time series with the formula.

$$(\mathrm{R} / \mathrm{S})_\mathrm{t} = \mathrm{c.t^H} \text{, for } \mathrm{t} = 1,2, \dots \mathrm{T} \tag{15}$$

with:c: constantH: Statistics HurstBy taking transformation $J_t = (\frac{R}{S})_t = \frac{R_t}{S_t}$, created Rescale adjusted range $\frac{R_t}{S_t}$. According to Hurst, $\lim_{n\to\infty} E\left(\frac{R_t}{S_t}\right) = c\, t^H$, where c is a constant and H is the Hurst statistics. So that the value of H can be determined by statistically $\frac{R_t}{S_t}$ of time series with the formula.

$$\frac{R_t}{S_t} \approx \mathrm{c\, t^H} \quad \mathrm{t} = 1,2, \dots \mathrm{T} \tag{16}$$

In this case, the value of Hurst (H) with logarithm statistical (R/S) to the values of t, and assess the value of H through Ordinary Least Square method (OLS), (Hurst: 1951). Logarithm equation (16) is

$$\log J_t = \log(\mathrm{c.t^H})$$

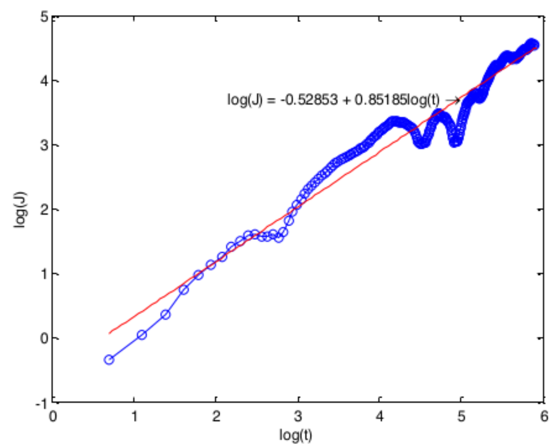
$$\text{atau } \log J_t = \mathrm{c} + \mathrm{H}.\log \mathrm{t} \;, \quad \text{untuk} \quad \mathrm{t} = 1,2, \dots \mathrm{T} \tag{17}$$

Observations show that for large values of t, log (R/S) spread around a straight line with a slope coefficient (slope) is greater 0.5. In this probabilistic terminology fulfilled that for large t, valid

$$\log \mathrm{E}\,(J_t) \approx a + H.\log\ t\ \text{, dengan } H > \frac{1}{2} \quad (18)$$

If $\{y_t\}$ H = {0,5} then {y_t shows a pattern of short memory, whereas if 0 <H <0.5, then $\{y_t\}$ shows an intermediate pattern memory, and if 0.5 <H <1 then $\{y_t\}$ indicate long pattern memory. H values in the regression equation to equation (9) can be estimated through odinary Least Square method (OLS). If $\| H| = \frac{1}{2}$ then $\{y_t\}$ indicate a pattern of short memory, whereas if $0 < |H| < \frac{1}{2}$ then $\{y_t\}$ shows an intermediate pattern memory, and if $\frac{1}{2} < |H| < 1$ then $\{y_t\}$ show a pattern of long memory. Differensi fractional d parameter estimates given by the estimated value of $\frac{1}{2} < |H| < 1$.

Results of analysis with matlab program for Hurst value can be seen in the following figure.



Based on calculations obtained $\bar{y} = 14487{,}6$ and Ordinary Least Square method with X = log(t) and Z = log (R/S) obtained estimated values H = 0.85185 so that d = 0.35185. These results indicate that the data load electricity consumption in Java and Bali following the pattern of long memory.

V. CONCLUSIONS AND RECOMMENDATIONS

V.1. knot Based on the analysis and discussion of the identification of long memory models using a modified method of data load rescale range on electricity per hour are taken over the past year indicate that the data pattern is a long memory with the estimated value of H = 0.85185 so that d = 0.35185.

V.2. Suggestion Development of a model for the identification of the data memory long progress, one with a whittle test.

REFERENCE

Beran J. (1994), Statistcs for Long –Memory Processes, New York, Chapman&Hall

Box, E.P. George, dkk. (1994). Time Series Analysis, Forecasting and Control. Prentice. Hall, Inc., New Jersey.

Box, G.E.P., dan Jenkins, G.M. (1976). Time Series Analysis, Forecasting and Control, 3rd edition. Englewood Cliffs, Prentice Hall.

Brockwell, P.J. and Davis, R.A. (1991). Time Series Theory and Methods, 2nd edition. New York, Springer Verlag.

Chatfield, C. (2001). Time Series Forecasting. Chapman & Hall, London.

Cryer, J.D. (1986). Time Series Analysis. Boston: PWS-KENT Publishing Company.

F.Chan, dan C Lim. (2011). Spektral Analysis of Seasonality In Tourism Demand, International Journal of Mathematics and Computers in Simulation, Vol 81, 1409-1418

Granger, C.W.J. and Joyeuk R. (1980). An Introduction to Long Memory Time Series Models and Fractional Differencing, Journal of Time Series Analysis, vol. 1, pp. 15-29

Marsudi. (2006). Operasi Sistem Tenaga Listrik. Yogyakarta, Graha Ilmu

Samarasinghe S. (2007), Neural Networks for Applied Science and Engineering, Auerbach Publication Taylor& Francis Group

Wei. (1999), Statistical for Time Series, Wiley Series Probability and Statistics

Wilfredo Palma. (2007), Long-Memory Time Series Theory and Methods, Canada, John Wiley&Sons

Yu WC., dan Zivot E. (2009), Forecasting the Term Structures of Treasury and Corporate Yields: Dynamic Nelson-Siegel Models Evaluation, Department of Economics, Winona State University, Somsen 319E, Winona, MN 55987. Email: wyu@winona.edu.

IDENTIFICATION MODEL OF LONG MEMORY IN USE ELECTRICITY CHARGES IN JAVA-BALI

ORIGINALITY REPORT

6%	4%	2%	2%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

MATCH ALL SOURCES (ONLY SELECTED SOURCE PRINTED)

2%

★ Palma. "Long-Memory Processes", Wiley Series in Probability and Statistics, 03/03/2007

Publication

Exclude quotes	On	Exclude matches	< 15 words
Exclude bibliography	On		