ISSN: 2348 9510
International Journal Of Core Engineering & Management (IJCEM)
Volume 3, Issue 4, July 2016

THE APPLICATION BOX JENKIN METHOD

Armi Bakar, SE.MM.,
Lecture Universitas Indraprasta PGRI

Teguh Sugiarto
Lecture Universitas Budi Luhur
DKI Jakarta, Indonesia

ABSTRACT
This research outlines the practical steps which need to be undertaken to use, autoregressive moving average (ARMA), autoregressive integrated moving average (ARIMA), autoregressive moving average exogenous variabel (ARMAX). A framework for ARMA, ARIMA, ARMAX and ARIMAX forecasting is drawn up. It considers two alternative approaches to the issue of identifying are models - the Box Jenkins approach and the objective penalty function methods. The emphasis is on forecast performance which suggests more focus on minimising out-of-sample forecast errors than on maximising in-sample 'goodness of fit'. Practical issues in are time series model that, forecasting are illustrated with reference to the harmonized index of consumer prices global FDI.

Keyword : Time series model, Inflation and Forecasting.

I. INTRODUCTION

In Bowerman et al (2005) dan Montgomery et al (1990) describes how its called forecast and forecasting. To do a thing to its future predictions of a thing and the conditions cannot be specified in the said by the name of divination (forecasts), whereas the science of making a prediction that will be done to a variable is then called with the forecasting (forecasting). Forecasting is a key to the elemen important in making a decision. This is useful later to reduce the risk in decision-making and reduce the possibility of the large number of costs that will be incurred or to minimize the possibility of cost.

Will happen. Most of the strength in reading divination is located on the strength of it's value in the period, the strength of a system of reading forecast and forecasting are usually categorized into function as long term and short term. Long term forecasting will to cover a value from 1 year up to 10 years in capacity to do an expansion over the long term policy towards the learning level of return on investment capital.

If the IEEE (1993) and Chow (2005) describes how and when we will do a forecast of its nature with regard to time, whether the time at mean short-term in nature which requires some knowledge of us in seeing and reading the predictions of a very short time as seconds until some day in the future. It can

ISSN: 2348 9510
International Journal Of Core Engineering & Management (IJCEM)
Volume 3, Issue 4, July 2016

make an information that later can be given to users of the decisions related to that time period. In terms of forecasting the short-term nature, indeed in the value of the results of the forecasting can be said in the substantial, as it pertains to the mood or circumstance of a scale or a system in the short term analysis, which requires the operation and deployment of better results. It's been a lot of research and writing that it incorporated the concept of good forecasting short-term, medium-term or long-term, and when to use the results of the forecast.

a. Time Series

Time series analysis can explain how to setup a data and explain/present it could be any time, as an example of an hour, daily or weekly. The basic idea of the forecasting was first built to equate and make accurate data available in order to become more available, and the value of forecasting can be used to model the time to the fore. General form and is often used to describe a characteristic equation has the form as follows in Janacek, G and L.Swift, (1993):

$$X(t) = T(t) + S(t) + R(t) \qquad t = \ldots -1,0,1,2,\ldots$$

Where T (t), shaped the customary trend or time series, S (t) is a seasonal time, and R (t) is a component rather than random (we can use Eviews to make data or Commandments).

b. Forecasting Method

Many models and forecasting methods which we know, where we will share or would classify both basic forecasting it into 2 types, the first qualitative and quantitative methods of the second. Forecasting methods with qualitative in General widely used as opinion and to forecast future predictions to read something objective, like the example we use historical data, that data is not avalaible, then we may be using subjective methods curva fittings, Delphi method and comparison of technology. Quantitative methods are typically a lot of make for forecasting macroeconomic data such as: regression analysis, decomposition methods, exponential smoothing, and the Box-Jenkins methodology.

c. Forecasting errors

In the General form and in the situation of all models forecasting can be used, but which have a level of certainty the exact error use. In forecasting error is used for forecasting over the variables $\hat{X}_t$ that have a real relationship with the value is:
In the General form and in the situation of all models forecasting can be used, but which has a level of assurance of the proper use $\hat{X}_t$ of the error.

$$e_t = X_t - \hat{X}_t$$

International Journal of Core Engineering & Management
ISSN: 2348 9510
International Journal Of Core Engineering & Management (IJCEM)
Volume 3, Issue 4, July 2016

To do positive and negative offset error, we need the top or use the absolute deviations. $|e_t| = |X_t - \hat{X}_t|$

d. The purpose of study

Research goals of this research are: 1) to see the use of a time series model in the application of macroeconomic data, 2) to see the efficiency of the time series model in forecasting macroeconomic data, 3) to compare between the time series model used in forecasting, and can draw the conclusion time series model which is better used for forecasting macroeconomic data.

II. REVIEW OF THE LITERATURE

a. Linear Regression Method

This regression shaped model equation as follows:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon$$

Where is the dependent variable, y, xi is the variable and independent or impacting against Y, βi is parameter regression regression parameters having links with xi, and ε is the error term. For this model always gives the assumption that the error of the time (error term) ε has a value of average or larger than zero and variance value are constan.
After the parameters of βi is known, then we'll do an estimate of the value of the observation y and xi. with the bi (i = 0,1, 2, ..., k) we estimated into the form of βi (i = 0,1, ..., 2 k). This equation has the form of error rates up to 50% of change could give positive and negative values, we will calculation into the equation parameters are explained:

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_k x_k$$

Then we use the bias value estimation by the method of least square estimates method minimizes our value by looking at the sum of squared residuals (SSE) in equation parameters b_i:

$$\underline{B} = [b_0 \quad b_1 \quad b_2 \cdots b_k]^T = (\underline{X}^T \underline{X})^{-1} \underline{X}^T \underline{Y}$$

Where Y and X is spelled out in the form of a column vector and matrix :

$$\underline{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad and \quad \underline{X} = \begin{bmatrix} 1 & x_{11} & x_{12} \cdots x_{1k} \\ 1 & x_{21} & x_{22} \cdots x_{2k} \\ \vdots & \vdots & \vdots \quad\quad \vdots \\ 1 & x_{n1} & x_{n2} \cdots x_{nk} \end{bmatrix}$$

After our calculation parameters, then this model will we can use to do this, the model predictions more accurate predictive value if the value of y is smaller than the standard error s.

$$s = \sqrt{\frac{SSE}{n-(k+1)}}, \quad SSE = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2, \quad y_i : observed, \quad \hat{y}_i : estimated$$

b. AR: Auto Regressive

In addition to MA processes, we can also make assumptions if the data we have to follow the process of AR is stated as follows:

$$y_t = \mu + \sum_{i=0}^{p} \phi_i y_{t-i} + u_t$$

Where p indicates the order of the AR (p).

c. Moving Average (MA)

In the time series modeling (univariate) we can assume that data will follow the process rentet MA. This process formed from white noise process :

$$y_t = \mu + \sum_{i=0}^{q} \theta_i u_{t-i} + u_t$$

To reveal the order of the MA can be expressed with the MA (q). Although the series that follows the process of MA formed from white noise series but if two or more of β her is not equal to zero then the series is not stationary.

For this moving average has order q (i.e., MA (q)), this model can be written with the other equation as follows :

$$X_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q}$$

ISSN: 2348 9510
International Journal Of Core Engineering & Management (IJCEM)
Volume 3, Issue 4, July 2016

Although the series that follows the process of MA formed from white noise series but if two or more of her β is not equal to zero then the series is not stationary, then it can be rendered with the form :

$$X_t = \theta(B)\varepsilon_t$$

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

d. ARMA(Box-Jenkins): Auto Regressive Moving Average

Time series modeling could also follow a combined process of AR and MA, their definitions are as follows :

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q}$$

where, ϕ_i and θ_j ordinary people might call autoregressive and moving average parameters, and in this case, to model the ARMA (p, q) model Methodology of ARMA popularized by Box & Jenkins (1970), and this model is often called the with Box-Jenkins models.

If we have data that are not yet integrated model can be applied to data of ARMA and ARIMA model (p, I, d) where I showed in order a the observed series of integrated. The methodology suggested by Box and Jenkins is a systematic way to determine the order of an ARMA series.

in addition to using the adjusted R-squared for model selection criteria, information criterion is also used for the selection of the model. The best model is a model that could be minimised criteria information. General information criteria used are:

AIC (Akaike Information Criterion)

$$AIC = \ln(\hat{\sigma}^2) + \frac{2k}{T}$$

SIC (Schawarz's Information Criterion)

$$SIC = \ln(\hat{\sigma}^2) + \frac{k}{T}\ln(T)$$

The criteria of the information each has drawbacks, so we can't use one criteria information only. The best way is to do a comparison between different criteria.

ISSN: 2348 9510
International Journal Of Core Engineering & Management (IJCEM)
Volume 3, Issue 4, July 2016

e. **ARIMA Autoregressive Integrated Moving:**

Average is often written as ARIMA ARIMA (p, d, q) meaning that p is the order of the autocorrelation coefficient, d is differentiation and q is the order in the coefficient moving average (moving average) can be seen in Hamilton (1994), Hayashi, f. (2000), Box, Jenkin et al (1976) and Box, Jenkin et al (2008).

Forecasting model using ARIMA can be done with the formula:

$$Y_t = \gamma_0 + \partial_1 Y_{t-1} + \partial_2 Y_{t-2} + \ldots + \partial_n Y_{t-p} - \lambda_1 e_{t-1} - \lambda_2 e_{t-2} - \lambda_n e_{t-q}$$

Description :

B	: Koefisien Regression
Y_T	: Dependent Variabel in time t
$Y_{t-1} \ldots Y_{t-p}$	: Variabel lag
e_t	: Residual term
$W_1 \ldots W_q$	: Weight
$e_{t-1} \ldots e_{t-p}$	: Residual value

ARIMA model assumes that the input data must be stationary. If the input data is not a stationary adjustment needs to be done to generate data is stationary. One way that is commonly used is the method of differentiation (*differencing*). This method is done by reducing the value of the data at any given period to the value of the previous period data

f. Others(ARMAX,ARIMAX)

ARMAX and ARIMA is used only in time and to the reading of input parameters. Since the data is read in general are used every time and day, exogenous variable which is often included in the ARIMAX and ARMAX models in J.Y. Fan and J.D. McDonald (1994). In the use and implementation of the evolution of evolutionary programming (EP) and fuzzy logic (FL) in a conventional time series model is not always used.

III. RESEARCH METHOD

Data and Time Research

Research as for the data in this study using data on FDI as much as 204 countries time lapse by 2014. The study lasted for 2 months from August to October 2015.

ISSN: 2348 9510
International Journal Of Core Engineering & Management (IJCEM)
Volume 3, Issue 4, July 2016

Table 1 : FDI data all country hidden

No.	COUNTRY	FDI
1	Afghanistan	49.000
2	Albania	1.149
3	Algeria	1.505
201	West Bank and Gaza	127.000
202	Yemen, Rep.	-738.000
203	Zambia	1.508
204	Zimbabwe	545.000

Sources : proceed by author

Technique of Data Analisys

The data obtained will be analyzed using time series data analysis as has been outlined in previous explanations. But before the data is analyzed using the methods of time series, but the author assume the data is used for research are qualified test of time series data. The model of time series analysis conducted in this study using the ARMA model analysis of the (1,1), ARIMA (1,1), (1.1) ARMAX and ARIMAX (1,1).

IV.RESULT AND DISCUSSION

ARMA (1,1) : Auto Regressive Moving Average

The form of the equation from the model ARMA is as follows, assuming the variable y :

$$Yt = Bo + BtYt\text{-}1 + ao + a1ut + a2ut\text{-}1$$

If the results of the model output for ARMA we can see using AR model (1) and MA (1).

Table 2 : result ARMA model

Dependent Variable: FDI

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	146.2434	15.43913	9.472253	0.0000
AR(1)	0.606178	0.336622	1.800767	0.0732
MA(1)	-0.690041	0.307708	-2.242520	0.0260

R-squared	0.014618	Mean dependent var	143.8414
Adjusted R-squared	0.004764	S.D. dependent var	275.2537
S.E. of regression	274.5973	Akaike info criterion	14.08316
Sum squared resid	15080733	Schwarz criterion	14.13212
Log likelihood	-1426.440	Hannan-Quinn criter.	14.10296
F-statistic	1.483465	Durbin-Watson stat	1.886216
Prob(F-statistic)	0.229336		

Inverted AR Roots	.61
Inverted MA Roots	.69

Sources : proceed by author

ISSN: 2348 9510
International Journal Of Core Engineering & Management (IJCEM)
Volume 3, Issue 4, July 2016

For [Table: 2] AR and MA of the model, then the model for ARMA also so, as with the model equations above can be inferred on the model in the proposal as follows: significant for the model MA and AR, we can see from the probabilistic value and the value of the t-test statistics are higher than t table, and a significant level of less than 5% alpha for AR and MA. The value of akaike information criterian and also great information swarch criterian, this indicates that the model in the proposal is good enough.

ARIMA: Autoregressive Integrated Moving-Average

When we use time series data that already in different as much as a times, then process the data will become a stationer with applying model ARMA (p, q) then the equation we will create, will be the ARIMA (p, d, q). ARIMA stands for Autoregressive Integrated Moving Average. The following results display model ARIMA by using variable that is still the same.

Table 3 : result ARIMA model

Dependent Variable: D(FDI)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.318691	0.439924	-0.724423	0.4697
AR(1)	-0.023029	0.071402	-0.322526	0.7474
MA(1)	-0.985900	0.006682	-147.5507	0.0000

R-squared	0.501158	Mean dependent var	2.692332
Adjusted R-squared	0.496145	S.D. dependent var	391.7245
S.E. of regression	278.0569	Akaike info criterion	14.10827
Sum squared resid	15385813	Schwarz criterion	14.15740
Log likelihood	-1421.935	Hannan-Quinn criter.	14.12815
F-statistic	99.96202	Durbin-Watson stat	1.994309
Prob(F-statistic)	0.000000		

Inverted AR Roots	-.02
Inverted MA Roots	.99

Sources : proceed by author

If we look at [Table: 2], if the previous model we discussed i.e. AR, MA, ARMA, and then for the model of ARIMA is also not the case, by looking at the equation from the previous model, then the model a in the form of ARIMA time is insignificant, for C and AR, while for a significant level of MA probailistik 5%, for model AR value probabilistic and t-test values higher than his statistics t table , and the level of significance of 5% on alpha. The value of akaike information criterian and swarch information criterian is also great. The overall model of ARIMA model better than the previous model. Because the value of AIC and SIC him pretty high.

ARMAX and ARIMAX (X/Exogenus)

ISSN: 2348 9510
International Journal Of Core Engineering & Management (IJCEM)
Volume 3, Issue 4, July 2016

ARMAX this model is the time series of expansion ARMA/ARIMA and SARIMA. ARMAX model in the ARIMAX and SARIMAX factor that affects the dependent variable Y at time t is influenced not only by the function of the variable Y in time but also by the other independent variables in the variable time to-t in general. If the model is the following are data time series and assumed the stationer, then the model can be used by adding the ARIMAX/SARIMAX component model integrated into the Yt, whereas if Yt stationer but variable Xn 1.2 etc not stationer then the model can be directly used for ARMAX model/ARIMAX.

First results out ARMAX models and put the second result output model ARIMAX.

Table 4 : result ARMAX model

Dependent Variable: INFLASI

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.743668	0.095748	7.766930	0.0000
MA(1)	-0.205843	0.140076	-1.469510	0.1443
R-squared	0.128423	Mean dependent var		0.578943
Adjusted R-squared	0.121220	S.D. dependent var		0.881730
S.E. of regression	0.826562	Akaike info criterion		2.473044
Sum squared resid	82.66783	Schwarz criterion		2.518770
Log likelihood	-150.0922	Durbin-Watson stat		2.018842
Inverted AR Roots	.74			
Inverted MA Roots	.21			

Sources : proceed by author

The ARMAX models in produce at [Table: 2] above can be ensured that the probability of significant, however the numbers R Squared is quite low, while the value of AIC and SIC. So the AR coefficients can be summed up significantly with t = 7.77 far from the critical value of 1.96, while for significant t value MA = + 1.47 below critical values so it is not significant at the 5% level of alpha.

Table 5 : result ARIMAX model

Dependent Variable: D(FDI)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	-0.020229	0.071398	-0.283330	0.7772
MA(1)	-0.985472	0.008747	-112.6647	0.0000
R-squared	0.499849	Mean dependent var		2.692332
Adjusted R-squared	0.497348	S.D. dependent var		391.7245
S.E. of regression	277.7246	Akaike info criterion		14.10099

ISSN: 2348 9510
International Journal Of Core Engineering & Management (IJCEM)
Volume 3, Issue 4, July 2016

Sum squared resid	15426186	Schwarz criterion	14.13374
Log likelihood	-1422.200	Hannan-Quinn criter.	14.11424
Durbin-Watson stat	1.994510		
Inverted AR Roots	-.02		
Inverted MA Roots	.99		

Sources : proceed by author

If we look at [Table 5] above, the variable to be included in the model is not significant at the 5% level of alpha. To see the test result diagnosis checking against the ARIMAX model, can be performed using the test Plot and Ljung Box ACF/PACF, assuming that lag length according to specify that we want. Here the author is not continuing the test, because the time model is less good or nice. This happens because the value of R Squared low.

V.CONCLUSION

These assumptions have important consequences in interpreting the data and economic models. This is because the data is essentially stationary does not have too great variations during the period of observation and has a tendency to approach its average value (Insukindro, 1994; Gujarati, 1995; Engle and Granger, 1987). Second, in the regression of a variable with variable time series other time, a researcher wants that the coefficient of determination R2 has a high value but often there is no relation between the two variables meaning. This situation indicates a spurious regression problems (spurious regression), as a result, among others, the coefficient of regression number inefficient, test the raw regression coefficients to be invalid. Third, the regression models with data time series is often used for the purposes of forecasting or predictions. The results of the prediction will not be valid if the data used is not stationary. A few possibilities why not used forecasting techniques that do not use structural model, where the equation shows the relationship between the variables based on economic theory and logic.

REFERENCES

- B.L.Bowerman, R.T.O'Connell, and A.B.Koehler, "*Forecasting, Time Series, and Regression: An Applied Approach*," 4th ed. California: Thomson Brooks/Cole, 2005.
- D.C.Montgomery, L.A.Johnson, J.S.Gardiner, "*Forecasting & Time Series Analysis*," 2th ed. New York: McGraw-Hill, 1990.
- *IEEE Transactions on Power Systems*, Vol. 8, No. 1, February 1993
- J.H.Chow, F.F.Wu, J.A.Momoh, *Applied Mathematics for restructured electric power systems*, New York: Springer, 2005
- G.Janacek, L.Swift, *Time series: forecasting, simulation, applications*, West Sussex: Ellis Horwood Limited, 1993.
- Moghram and S.Rahman, "*Analysis and evaluation of five short-term load forecasting techniques*," *IEEE Trans. on Power Systems*, vol.4, no.4, pp. 1484-1491, Oct.1989.

- J.Y. Fan and J.D. McDonald. *A Real-Time Implementation of Short-Term Load Forecasting for Distribution Power Systems*. *IEEE Transactions on Power Systems*, vol.9, no.2, pp.988-994, 1994.
- Pindyck, Robert S. and Daniel L. Rubinfeld. *Econometric Models and Economic Forecasts*, 4thedition, New York: McGraw-Hill. 1998.
- Tong, H. and K. S. Lim. "Threshold Autoregression, Limit Cycles and Cyclical Data," *Journal of the Royal Statistical Society. Series B (Methodological)*, 42, 245–292. 1980.
- Baille, Richard (1996). "Long Memory Processes and Fractional Integration in Econometrics," *Journal of Econometrics*, 73, 5–59..
- Box, George E. P. and Gwilym M. Jenkins (1976). *Time Series Analysis: Forecasting and Control*, Revised Edition, Oakland, CA: Holden-Day.
- Box, George E.P., Jenkins, Gwilym M., and Gregory C. Reinsel (2008). *Time Series Analysis: Forecasting and Control*, Fourth Edition, Hoboken, New Jersey: John Wiley & Sons.
- Doornik, Jurgen A. and Marius Ooms (2003). "Computational Aspects of Maximum Likelihood Estimation of Autoregressive Fractionally Integrated Moving Average Models," *Computational Statistics & Data Analysis*, 42, 333–348.
- Fair, Ray C. (1984). *Specification, Estimation, and Analysis of Macroeconometric Models*, Cambridge, MA: Harvard University Press.
- Geweke, J. F. and S. Porter-Hudak (1983). "The Estimation and Application of Long Memory Time Series Models" *Journal of Time Series Analysis*, 4, 221–238.
- Granger, C. W. J. and Roselyne Joyeux (1980). "An Introduction to Long-Memory Time Series Models and Fractional Differencing," *Journal of Time Series Analysis*, 1, 15–29.
- Greene, William H. (2008). *Econometric Analysis*, 6th Edition, Upper Saddle River, NJ: Prentice-Hall.
- Hamilton, James D. (1994). *Time Series Analysis*, Princeton University Press.
- Hayashi, Fumio. (2000). *Econometrics*, Princeton, NJ: Princeton University Press.
- Hosking, J. R. M. (1981). "Fractional Differencing," *Biometrika*, 68, 165–176.
- Johnston, Jack and John Enrico DiNardo (1997). *Econometric Methods*, 4th Edition, New York: McGraw-Hill.
- Kohn, Robert and Craig F. Ansley (1985). "Efficient Estimation and Prediction in Time Series Regression Models," *Biometrika*, 72, 694–697.
- Rao, P. and Z. Griliches (1969). "Small Sample Properties of Several Two-Stage Regression Methods in the Context of Auto-Correlated Errors," *Journal of the American Statistical Association*, 64, 253–272.
- Sowell, Fallaw (1992). "Maximum Likelihood Estimation of Stationary Univariate Fractionally Integrated Time Series Models," *Journal of Econometrics*, 53, 165–188.
- Sowell, Fallaw (1992a). "Modeling Long-run Behavior with the Fractional ARIMA Model," *Journal of Monetary Economics*, 29, 277–302.