Package iClick OSSR for Tracking The Best ARIMA Model

Gumgum Darmawan[1], Nuning Kurniasih[2], Darmawan Napitupulu[3], Dahlan Abdullah[4], Nanik Kurniawati[5], Ansari Saleh Ahmar[6]
{ gumgum@unpad.ac.id[1], nuning.kurniasih@unpad.ac.id[2], darwan.na70@gmail.com[3], dahlan@unimal.ac.id[4], nesnes2017@yahoo.com[5], ansarisaleh@unm.ac.id[6] }

Department of Statistics, Faculty of Mathematics and Natural Science, Universitas Padjadjaran, Bandung, Indonesia[1]
Faculty of Communication Sciences, Library and Information Science Program, Universitas Padjadjaran, Bandung, Indonesia[2]
Research Center for Quality System and Testing Technology, Indonesian Institute of Sciences, Jakarta Indonesia[3]
Department of Informatics, Universitas Malikussaleh, Aceh, Indonesia[4]
The National Archive of The Republic of Indonesia (ANRI)[5]
Department of Statistics, Faculty of Mathematics and Natural Science, Universitas Negeri Makassar, Makassar, Indonesia[6]

Abstract. Package iClick is one of many packages in OSSR (opensource Software R) that have many syntaxes. There is five principle iclick syntax available in this package. GUI features of this package make researcher can easily use this package. In this paper, we study about getting the best ARIMA (Autoregressive Moving Average) model. We use IBM data which are available in this package. To track the best ARIMA model, order of ARIMA is tried from 1,2 and 3. According to the result, ARIMA with order 2 is suitable for IBM data because of the significance of parameter and white noise.

Keywords: ARIMA,Package iClick, OSSR.

1 Introduction

Package iClick is an external package to support the analysis financial-economic time series data based on Graphical User Interface (GUI). Published in 2015, package iClick is very useful for modeling time series data. Button-Based GUI package, make this package easy to use for researchers.

This package can be used to times series modeling, Autoregressive Moving Average (ARIMA)[1]–[3] and General Autoregressive Conditional Heteroscedastic (GARCH). ARIMA and GARCH are two time series models that commonly used to forecast economic-financial data. For instance,[4] used this model together with GSTAR to model consumer price index phenomenon at several cities in North Sumatera,[5] forecasted consumer price index by Seasonal Autoregressive Integrated Moving Average (SARIMA). Moreover, ARIMA can be mixed by other model like Singular Spectrum Analysis (SSA), for example[6]–[8].

Because of widely usage, SARIMA and GARCH had been making in many packages by the researcher. Packages ARIMA are [9], [10], [11], [12] and [13], for generating SARIMA time series data, we may use package [14] and [15]. Then, Packages GARCH in are [16], [17], [18], [19], [20], [21] and [22].

However, for getting ARIMA model, we must identify the order of the ARIMA models in the plot of autocorrelation (ACF) and partial autocorrelation (pack). This step is difficult enough for researchers to identify the model because of the subjectivity of ACF and pack visualization interpretation.

Sometimes, it is hard to identify the order of ARIMA because the visualization of this plot is not suitable for theory[23]. In that cases, we should trial and error way to get the proper ARIMA order from these figures.

In this paper, we describe step by step how to modeling time series data with ARIMA models by package iClick. By package iClick, researchers can easily get the best ARIMA models without choosing order as in classical way/procedure. Moreover, by package iClick, we can easily click the part of results by GUI features. Data used in this paper are available in this package, IBM.

2 Methodology

2.1 Headings, tables and figures

There are five steps in ARIMA modeling, from determining general models to the last step Fixed model. In the first step, time series data are visualized in three graphs, time series plot, Autocorrelation plot (ACF) and Partial Autocorrelation plot (ACF). These models are beneficial to identify the pattern of time series data such us trend, stationarity (mean and variance), seasonal pattern. Moreover, the plot of ACF and pacf can be used to determine the order of general model. General models are a collection of temporary ARIMA models that have specification by visualizing [23].

After determining general models by ACF and pacf plot, we choose temporary model one by one to estimate its parameter. Estimation parameters are run in Maximum Likelihood method. If one or more parameters are insignificant, then the temporary model is dropped and choose another temporary model in the collection of general models.

The next step is diagnostic check, in this step all models are evaluated the fit of the model to data. Its indication showed by its residual.

In the Diagnostic check, the residual must have two conditions. The first, the residuals must follow the Normal distribution with stationary mean and homoscedastic variance. The second condition, the residuals should zero correlation among others. If its residuals are white noise, then the model go to fixed model.

Fixed model is an ARIMA model that parameters are significant, and the residuals are white noise. Some time, fixed models are more than one, so, we choose one is the best. For

choosing the best model, we can evaluate fit model by MAPE (Mean absolute percentage error), MSE (Mean Square Error) or AIC (Akaike Information Criterion).

After getting the best ARIMA model, we use this model for forecasting observation in the future. To distinguish the data and the forecast result, we should plot both in a different color by software R.
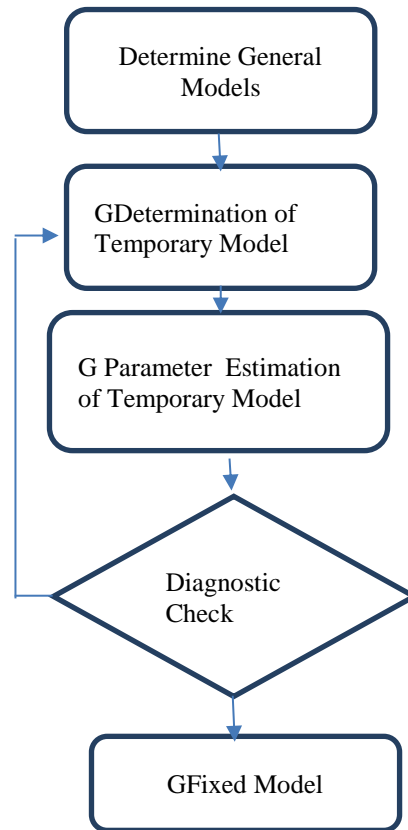

Fig 1 Flowchart of ARIMA Procedure

Figure 1 shows the procedure of ARIMA modeling from time series data. This procedure consists of five-step from determining general model to fixed model. Fixed model is the best ARIMA model chosen for forecast time series observation in the future.

2.2 Operating Package

To operate the package iClick, we must download and install this package in the R Console or R Studio. After installing the package, we must 'call' this package by syntax

> library(iClick)

Before using package iClick, we identify the time series data by plotting in time series plotting, ACF, and pacf. Although, the iClick will automatically choose ARIMA model, we can predict the order of ARIMA after.
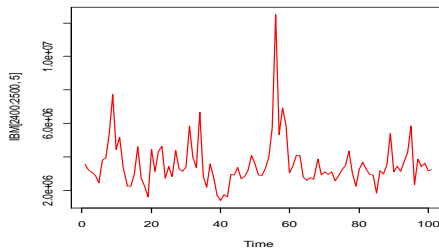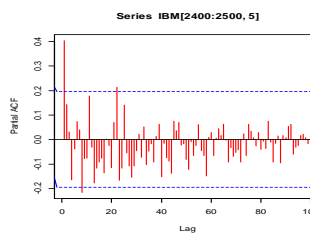

Fig 2. Time series plot

From the figure 2, we can see that time series data is relatively stationer in mean so the element of integrated will not include in the model. If the model is not stationer in variance, then we must transform by Box-cox transformation.
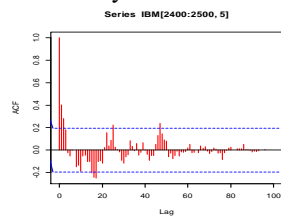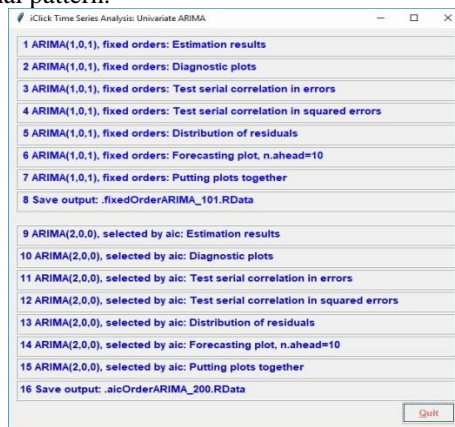

Fig 3 ACF Plot   **Fig 4** PACF

These figures show that the data are no clearly seen the order of ARIMA. So, we may trial and error way to track the order of model. In that figure, we can see also that the time series data do not have a seasonal pattern.

Fig 5. Output GUI of ARIMA By iClick Package

The figure 5 is the result of ARIMA model by package iClick, the syntax are ,
>iClick.ARIMA(IBM[2400:2500,5],AR=1,MA=1,n.ahead=10,ic="aic")
>iClick.ARIMA(IBM[2400:2500,5],AR=2,MA=2,n.ahead=10,ic="aic")
 After clicking this syntax above, the figure will show in GUI form. We can choose the best ARIMA model in this iClick output.
The detailed output will appear in R Console after clicking one of the result lists. Moreover, the results of all output are available in Save output button.

3   Result and Discussion

According to the output of iClick in figure 5, There are two ARIMA models i.e. ARIMA (1,0,1) and ARIMA(2,0,0). The value of AIC for ARIMA(2,0,0) is smaller then ARIMA(1,0,1), A good ARIMA model is the one that has minimum AIC among all the other ARIMA models. The last row is the test serial correlation in residuals.

Table1. Diagnostics Check

ARIMA(1,0,1)	ARIMA(2,0,0)
Coefficients:	Coefficients:
ar1 ma1 intercept	ar1 ar2 mean
0.6213 -0.2639 3549900.5	0.3424 0.1394 3551065.3
s.e. 0.1507 0.178 249390.7	s.e. 0.0978 0.097 247466.8
sigma^2 estimated as 1.705e+12: log likelihood = -1565.73, aic = 3139.45	sigma^2 estimated as 1.752e+12: log likelihood=-1565.58 AIC=3139.17
Tests for residual serial correlation	Tests for residual serial correlation
1. Box-Pierce test	1. Box-Pierce test
data: out.fixed$residuals	data: out.auto$residuals
X-squared = 9.4623, df = 12, p-value = 0.663	X-squared = 9.6252, df = 12, p-value = 0.6488
2. Box-Ljung test	2. Box-Ljung test
data: out.fixed$residuals	data: out.auto$residuals
X-squared = 10.399, df = 12, p-value = 0.581	X-squared = 10.575, df = 12, p-value = 0.5656
Two-sample Kolmogorov-Smirnov test	Two-sample Kolmogorov-Smirnov test
data: z and z1	data: z and z1
D = 0.207, p-value = 0.025	D = 0.158, p-value = 0.158

alternative hypothesis: two-sided		alternative hypothesis: two-sided	

According to the test of residuals dependency, residuals of ARIMA(1,0,1) is white noise because p-value of the Box-Pierce test and Box-Ljung test have bigger than 5%. Similarly, the value of the Box-Pierce test and Box-Ljung test of ARIMA(2,0,0) have bigger than 5 %.

However, in the first model, its residuals are no white noise, because the value Kolmogorov-Smirnov test is smaller than 5%, it is mean that its residuals don't N changeormal Distribution. So we choose the second model. This result is appropriate from the first step above that the ARIMA model will not be included difference parameter.

The forecast and mathematical model of ARIMA(2,0,0) can be seen below,

A mathematical model of ARIMA(2,0,0)

$$\overset{\otimes}{x}_t = \phi_1 \overset{\otimes}{x}_{t-1} + \phi_2 \overset{\otimes}{x}_{t-2} + a_t \qquad (1)$$

Where;

ϕ_1, ϕ_2 = parameters of Autoregression,

a_t = residual at-t,

$\overset{\otimes}{x}_t$ = Time series data with mean μ.

Table 2. Forecast Result and ARIMA (2,0,0)

Forecast Result [10]		Mathematical Model
3,385,041	3,539,433	
3,447,745	3,544,301	
3,492,544	3,547,128	$\overset{\otimes}{x}_t = 0.342 \overset{\otimes}{x}_{t-1} + 0.139 \overset{\otimes}{x}_{t-2} + a_t$
3,516,624	3,548,774	
3,531,114	3,549,732	

In Table 2, a mathematical model of ARIMA(2,0,0) generates 10-time series forecast in the left of the table. All of the parameter and the result of the forecast are positive values.

Next tracking, the second syntax is

>iClick.ARIMA(IBM[2400:2500,5],AR=2,MA=2,n.ahead=10,ic="aic") .
The result of this syntax is ARIMA(2,0,2), its diagnostic check can be seen in table 3.
According to test of residuals dependency, residuals of ARIMA(2,0,2) is white noise because p-value of the Box-Pierce test and Box-Ljung test have bigger then 5%. Normality test of its residuals is insignificant, it shows that residuals are Normal Distribution. Moreover, if we check the value of t-values, all of t-values of all parameter are significant (The value of t-table = -1.98), i.e

t1=(1.7287/0.1246) = 13.874,

$$t2 = (-0.7788/0.1179) = -6.605,$$

$$t3 = (-1.4815/0.192) = -7.716,$$

$$t4 = (0.4815/0.1905) = 2.527.$$

These equations refer to [20], all of the equation result shows that all parameters are significant, so we can continue to diagnostics check. The significance of the hypothesis is done by comparing the value of t-table =-1.98 and the values of all t-table above.

Table 3 Diagnostic check for ARIMA(2,0,2)

Coefficient	Tests for residual serial correlation	Normality Test
ar1 ar2 ma1 ma2 intercept 1.7287 -0.7788 -1.4815 0.4815 3556812.63 s.e. 0.1246 0.1179 0.1920 0.1905 40254.27 aic = 3135.98	1.Box-Pierce test data: out.fixed$residuals X-squared = 9.2402, df = 12, p-value = 0.6823 2. Box-Ljung test data: out.fixed$residuals X-squared = 10.181, df = 12, p-value = 0.6	Two-sample Kolmogorov-Smirnov test data: z and z1 D = 0.14851, p-value = 0.2153 alternative hypothesis: two-sided

Table 4. Forecast Result and Mathematical Model

Forecast Result [10]		Mathematical Model
3,503,330	3,718,781	
3,579,874	3,720,701	$\overset{\&}{x}_t = \phi_1 \overset{\&}{x}_{t-1} + \phi_2 \overset{\&}{x}_{t-2} - \theta_2 a_{t-2} - \theta_1 a_{t-1} + a_t$
3,638,333	3,713,989	
3,679,780	3,700,890	
3,705,902	3,683,474	

In Table 4, a mathematical model of ARIMA(2,0,2) generates ten time series forecast in the left of the table. All of forecast results are positive values.

After comparing many ARIMA models in this results, we choose ARIMA(2,0,2) because many reasons. The first, the value of AIC is the smallest of all ARIMA models, the parameters are significant and white noise.

4 Conclusion

Package iClick is relatively easy to operate, and Graphical User Interface in this package makes user can interpretation model easily. All output of ARIMA models are showed in this

GUI, so we can choose the best model in the output released. However, this package can not be used to analyze seasonal time series.

References

[1]   A. S. Ahmar *et al.*, "Implementation of the ARIMA(p,d,q) method to forecasting CPI Data using *forecast package* in R Software," *J. Phys. Conf. Ser.*, vol. 1028, no. 1, p. 012189, Jun. 2018.

[2]   A. S. Ahmar *et al.*, "Modeling Data Containing Outliers using ARIMA Additive Outlier (ARIMA-AO)," *J. Phys. Conf. Ser.*, vol. 954, no. 1, 2018.

[3]   U. Khair, H. Fahmi, S. Al Hakim, and R. Rahim, "Forecasting Error Calculation with Mean Absolute Deviation and Mean Absolute Percentage Error," *J. Phys. Conf. Ser.*, vol. 930, no. 1, p. 012002, Dec. 2017.

[4]   H. Bonar, B. N. Ruchjana, and G. Darmawan, "Development of generalized space time autoregressive integrated with ARCH error (GSTARI–ARCH) model based on consumer price index phenomenon at several cities in North Sumatera province," in *AIP Conference Proceedings*, 2017, vol. 1827, no. 1, p. 20009.

[5]   D. A. Lubis, M. B. Johra, and G. Darmawan, "Peramalan Indeks Harga Konsumen dengan Metode Singular Spectral Analysis (SSA) dan Seasonal Autoregressive Integrated Moving Average (SARIMA)," *J. Mat. MANTIK*, vol. 3, no. 2, pp. 74–82, 2017.

[6]   Q. Zhang, B.-D. Wang, B. He, Y. Peng, and M.-L. Ren, "Singular spectrum analysis and ARIMA hybrid model for annual runoff forecasting," *Water Resour. Manag.*, vol. 25, no. 11, pp. 2683–2703, 2011.

[7]   M. Zokaei, R. Mahmoudvand, and N. Najari, "Comparison of Singular Spectrum Analysis and ARIMA Models."

[8]   E. S. Silva and H. Hassani, "On the use of singular spectrum analysis for forecasting US trade before, during and after the 2008 recession," *Int. Econ.*, vol. 141, pp. 34–49, 2015.

[9]   I. Svetunkov, "Statistical models underlying functions of'smooth'package for R," 2017.

[10]   D. Stoffer and M. D. Stoffer, "Package 'astsa,'" *Blood*, vol. 8, p. 1, 2017.

[11]   A. I. McLeod and M. A. I. McLeod, "Package 'FitARMA,'" 2013.

[12]   A. I. McLeod, Y. Zhang, C. Xu, and M. A. I. McLeod, "Package 'FitAR,'" *Dept. Stat. Actuar. Sci., Univ. West. Ontario, London, ON, Canada, Tech. Rep*, 2013.

[13]   J. Gabry and B. Goodrich, "rstanarm: Bayesian applied regression modeling via Stan. R package version 2.11. 1." 2016.

[14]   O. J. T. Briët, P. H. Amerasinghe, and P. Vounatsou, "Generalized seasonal autoregressive integrated moving average models for count data with application to malaria time series with low case numbers," *PLoS One*, vol. 8, no. 6, p. e65761, 2013.

[15]   T. Kharrat, G. N. Boshnakov, and M. G. N. Boshnakov, "Package 'StableEstim,'" 2016.

[16]   B. Pfaff and M. B. Pfaff, "Package 'gogarch,'" 2009.

[17]   T. Nakatani, "ccgarch: An R package for modelling multivariate garch models with conditional correlations, 2010," *R Packag. version 0.2. 0, URL< http//CRAN. r-project/package= ccgarch*.

[18]   A. Ghalanos, M. A. Ghalanos, and L. Rcpp, "Package 'rugarch,'" 2018.

[19]   A. Ghalanos, "rmgarch: Multivariate GARCH models," *R Packag. version 0.98*, 2012.

[20]   R. Cesur and J. J. Sabia, "When war comes home: The effect of combat service on domestic violence," *Rev. Econ. Stat.*, vol. 98, no. 2, pp. 209–225, 2016.

[21]   D. Wuertz *et al.*, "Package 'fGarch,'" Technical report, working paper/manual, 09.11. 2009. URL http://cran. r-project. org/web/packages/fGarch/fGarch. pdf, 2013.

[22]   D. Ardia and M. D. Ardia, "Package 'bayesGARCH,'" 2017.

[23]   W. Wei, *Time- Series Analysis*. 2006.