Journal of Statistical and Econometric Methods, vol.4, no.4, 2015, 41-53
ISSN: 1792-6602 (print), 1792-6939 (online)
Scienpress Ltd, 2015

Application of residual analysis in time series model selection

Ikughur, Atsua Jonathan[1] , Uba, Tersoo[2] and Ogunmola, Adeniyi Oyewole[3]

Abstract

In this study, five criteria of residual analysis in time series modelling and forecasting are evaluated using three study variables namely, Nigeria's Gross Domestic Product (GDP), Total Debts Accumulation (TDA) and Rate of Inflation (INFL). Considering five Auto Regressive Integrated Moving Average (ARIMA) specifications each for GDP and TDA and four ARIMA specifications for INFL, it was observed that four of the five criteria selected ARIMA(2,2,2) for the GDP I(2) while all the five criteria selected ARIMA(2,2,3) for TDA I(2) process. ARIMA(1,0,2) was also selected by all the criteria for INFL I(0) process. It is observed here that there is no particular criterion that clearly dominate others in the search for the "best" model specification and this suggests that modellers should consider the use of more than one criterion in model selection, especially when the family of ARIMA(p,d,q) models are of interest.

[1] Department of mathematics, Statistics and Computer Science Federal University of Agriculture Makurdi, Benue State-Nigeria. E-mail: atsua2004@yahoo.com
[2] Department of mathematics, Statistics and Computer Science Federal University of Agriculture Makurdi, Benue State-Nigeria.
[3] Department of Mathematics & Statistics, Federal University Wukari.

Article Info: *Received* : July 27, 2015. *Revised* : October 21, 2015.
Published online : December 1, 2015.

Keywords: Residual; ARIMA; dominate; model selection

1 Introduction

Several statistical methodologies can be applied to model a phenomenon. These methods include the regression analysis and analysis of variance. Specifically, a member of the family of regression models is useful in providing models when time series data are encountered. Whenever a model of such is fitted to the data and predictions are made, residuals are usually generated especially, when the data in question is a random sample drawn from a population. The process of modelling apart from obtaining the functional expression describing the data set requires modellers to assess the validity of the model, perform certain diagnostic testing and set up optimality and robustness criteria for which the 'best' model is determined.

In the theory of estimation and testing, residuals play a very important role especially, in drawing inference for linear models (Clarke, 2008). The analysis of residuals commences with the plot which may appear to exhibit non-normal pattern especially when a model is inappropriately specified or when there is non-homogeneity of error variance, or perhaps, the number of residuals is too small to provide a pattern of sufficient stability to permit valid statistical inference (Kleinbaum and Kupper, 1978).

In this study, we consider the family of linear stochastic time series model of the autoregressive integrated moving average (ARIMA) with the aim of identifying or defining residuals and their measures, review their usefulness in model diagnosis, validity check and of course determination of optimality criteria. Finally, empirical study is performed using three set of time-series data namely Nigeria's GDP series (1982-2011), Nigeria's Total Debts Outstanding (1982-2011) and Nigeria's rate of inflations series (1960-2011).

2 Time Series Model Specification

In this study, the Autoregressive Integrated Moving Average model is considered.

Definition 1. x_t is an ARIMA (p,d q) process if $\{x_t\}$ is stationary and if for every *t*,

$$\Phi(B)\nabla^d x_t = \theta(B)\varepsilon_t \tag{1}$$

which is further expressed as

$$\Phi(B)w_t = \Theta(B)\varepsilon_t \tag{2}$$

where w_t =$\nabla^d x_t$, ∇ denotes differencing whose order is denoted as d . The subscript *t* is used to denote the time period so that $\nabla^d x_t = x_t - x_{t-1}$, $x_t = \sum w_t$ reverts w_t to x_t while $\{\varepsilon_t\} \sim WN(0, \sigma^2)$ otherwise, called a white noise process.

$$\Phi(B) = \left(1 - \varphi_1 B^1 - \varphi_2 B^2 - \quad \ldots - \varphi_p B^p\right)$$

and

$$\Theta(B) = \left(1 - \theta_1 B^1 - \theta_2 B^2 - \quad \ldots - \theta_q B^q\right)$$

are transfer functions for Auto-Regressive (AR) and Moving-average (MA) portions respectively. When d = 0, $\{x_t\}$ is assumed stationary at its level so that

$$\Phi(B)x_t = \Theta(B)\varepsilon_t \tag{3}$$

The process defined in (2) above can be thought of as a pth order autoregressive process $\Phi(B)w_t = \varepsilon_t$ with ε_t following the qth order moving average process or, as $w_t = \Theta(B)\varepsilon_t$ with w_t following the pth autoregressive process. For d ≥1, $x_t = \Sigma^d w_t$ is called an invertible process. It is worth to note here $\Theta(B)$is invertible when the root of $\Theta(B) = 0$ lies outside the unit circle. Similarly, $\Phi(B)$ is assumed stationary with $\Phi(B) = 0$ lying outside the unit circle.

Box and Jenkins(1976) presented the algorithm for estimating the parameters of an ARIMA process with w_t =$\nabla^d x_t$ otherwise, an integrating process. This occurs in three stages thus:

(i) The AR parameters $\varphi_1, \varphi_2, \ \ldots, \varphi_p$ are estimated from the autocovariances denoted as $C_{q-p+1,} \ \ldots, C_{q+p}$;

(ii) Using the estimate of $\hat{\varphi}$ obtained in (i) above, the first q+1 autocovariances denoted as $C_j', (j = 1,2, \dots, q)$of the derived series $w_t' = w_t - \hat{\varphi}_1 w_{t-1} - \cdots - \hat{\varphi}_p w_{t-p}$ are calculated;

(iii) Thirdly, the autocovariances $C_0', C_1', \dots, C_q'$ are used in an iterative calculation to compute initial estimate of the MA parameters $\theta_1, \theta_2 \dots, \theta_q$ and the residual variance, σ_ε^2.

According to Pindykt and Rubbinfed (1981), estimates of the model's parameters can be obtained for the p-autoregressive and q-moving average parameters by choosing parameter values that will minimize the sum of squared differences between the actual time series $w_t = \nabla^d x_t$ and the fitted time series w_t in terms of the residual error from ARIMA process. Thus,

$$\varepsilon_t = \Theta^{-1}(B)\Phi(B)w_t \qquad (4)$$

So that the estimate of $\Phi = (\phi_1, \phi_2, \dots, \phi_p)$ and $\Theta = (\theta_1, \theta_2, \dots, \theta_p)$ are obtained by

$$S(\theta, \phi) = \sum_i \varepsilon_t^2 \qquad (5)$$

The expression in (5) is non-linear in parameters if MA terms are present. For this reason, an iterative method of non-linear estimation is used to estimate the model's parameters.

3 Residuals analysis

Given the model defined in (1) above, the residuals generated by the model for the corresponding values of $w_t = \nabla^d x_t$ are denoted by ε_t , t = 1,2, ...,n as in (4) above. Here, it is assumed that the unobserved residuals are normally distributed with zero mean and common variance, that is, $\varepsilon_t \backsim N(0, \sigma^2)$. The first of its significance is that it provides a diagnostic procedure for checking whether the initial specification of the model is correct. The expectation is that the residuals should resemble a white noise process which by assumption, are un-autocorrelated. If they are autocorrelated, new specifications are given for p, d and q and another diagnostic check is performed. In this study, the following methods, otherwise rules of residual analysis namely, Durbin Watson (DW) Test, Ljung-Box-Pierce (Q) Test, Akaike Information Criteria (AIC), Standard Error (SE) of the Regression (otherwise, the time series model) and Mean Absolute Percentage Error (MAPE) are considered and applied to determine the specification that best model the series under study.

Rule 1: Durbin Watson Test:- This test proposed by Durbin and Watson (1951) considers the test of the Null hypothesis Ho: ρ = 0 and the test statistic is based on residuals from the Ordinary Least Squares (OLS) regression procedure and is defined as :

$$d = \frac{\sum_{t=2}^{N}(\hat{\varepsilon}_t - \hat{\varepsilon}_{t-1})^2}{\sum_{t=1}^{N}\hat{\varepsilon}_{t-1}^2} \tag{6}$$

Chatfield(1982) observes that the coefficient d is related to the first order autocorrelation coefficient of the residual so that the numerator in expression (6) above can be represented as

$$\sum_{t=2}^{N}(\hat{\varepsilon}_t - \hat{\varepsilon}_{t-1})^2 = 2\sum_{t=1}^{N}\hat{\varepsilon}_{t-2}^2 - 2\sum_{t=2}^{N}\hat{\varepsilon}_t\hat{\varepsilon}_{t-1}$$

For which

$$DW = 2(1 - r_1) \tag{7}$$

where

$$r_1 = \frac{\sum_{t=2}^{N}\hat{\varepsilon}_t\hat{\varepsilon}_{t-1}}{\sum_{t=1}^{N}\hat{\varepsilon}_{t-1}^2} \tag{8}$$

Heinnushek and Jackson(1977) identifies two functions of Durbin Watson as firstly testing for serial correlation and secondly, a way to estimate this correlation which can be used to obtain generalised least squares (GLS) estimates.

The statistic DW is asymptotically equivalent to the test on r_1. If $r_1 = 0$, then $DW = 2$. Positive serial correlation is associated with DW< 2. Generally, the range of DW is $0 < DW < 4$ so that the values of DW near 2 indicates no first order serial correlation.

It is worth to note here that the distribution of DW depends on the sample size, number of coefficients being estimated and also the sample values of the explanatory variables.

Rule 2: Akaike Information Criteria. This is given by

$$AIC = log\hat{\sigma}_k^2 + \frac{n+2k}{n} \quad (9)$$

where $\hat{\sigma}_k^2 = \frac{SSE_k}{n}$ which is also based on residuals of the ARIMA model.

Rule 3: Standard Error of the Regression. This is the summary measure based on the estimated variance of the residuals and it is given by

$$SE\ Reg = \left(\frac{\sum_{t=1}^{T} \varepsilon_t^2}{T-k}\right)^{1/2} \quad (10)$$

where T is the length of time and k is the number of estimated ARIMA parameters.

Rule 4: Box – Pierce Test. Box and Pierce (1970) considered the large sample properties of all the residual autocorrelated coefficients for any ARIMA process. Their results showed that $1/\sqrt{N}$ supplies the upper bound for the standard error of the autocorrelation coefficients up to lag k, r_k computed from the residuals. Thus, when r_k's are computed, the values that lie outside the range $\pm 2/\sqrt{N}$ are certainly significant different from zero.

Ljung, Box and Pierce (1978) also described what they called a portmanteau lack of fit test. Thus, instead of looking for r_k separately, a group up to the first lag k are considered using the test statistic given as;

$$Q = \frac{n(n+2)}{(n-k)} \sum_{k=1} r_k^2 \quad (11)$$

Where n is the number of terms in the difference series and the test statistics has a chi-squared distribution with k-p-q degree of freedom denoted as $Q \sim \chi_{k-p-q}^2$

Rule 5: Mean Absolute Percentage Error (MAPE). This measure the percentage departure of an observation from its forecasted values and is given by

$$MAPE = 100 * \sum_{t=T+1}^{T+h} \left|\frac{\hat{y}_t - y_t}{y_t}\right| /h \quad (12)$$

where *h (>0)* denotes the forecast length.

3.1 Dominating Criterion

Definition 2. Let R_1, R_2, ...,R_5 be the model selection rule, then a specification say, S_i dominates another specification S_j, when $S_i < S_j$ for all *i* and $S_i \leq S_j$ for some *i* (*i*=1,2,...,5).

Considering the models specified for the data set, one common and interesting feature of these criteria is the fact that they select the model with the least value of the statistic except in the case of DW statistic which seeks for values that are very close to 2, and the Q statistic which checks for the value of autocorrelations of the residuals close to zero. Although, it is always difficult to have all the criteria agree on same decision line however, it is possible to have more of these statistics agreeing on a certain decision. Thus, it will be convenient to conclude by selecting that model with most agreeable decisions rule. In other word, decision on the "bestness" of a model specification should be determined by considering that specification with dominating criteria.

4 Data for Analysis

In this study, three data sets namely gross domestic product (GDP), total debt accumulation (TDA) of Nigeria for the period 1981 to 2009 and Nigeria's rate of inflation (INFLA) for the period 1961 to 2008 are considered. Statistical analyses of interest in this study are the various residual analyses performed on the Auto Regressive Integrated Moving Average models specified for the series under study and the associated tests including the unit roots test as shown on Table 1 below using the Eviews-5 statistical package commonly used for the analysis of econometrics and time series problem. Results are shown on Tables 2 to 4 below for GDP, TDO and INFLA series.

Table 1: Characterization of GDP, TDA and INFLA series under unit root hypothesis.

S/No	Variable	ADF-Level 0	ADF-Level 1	ADF-Level 2	Order of Stationarity I(k)
1	GDP	0.380	-2.451	-4.013*	I(2)
2	TDA	-2.438	-3.477	-6.637	I(2)
3	INFLA	-3.7579			I(0)

* Shows level at which the series is stationary (no unit root). I(k) shows the order of integration, k=0,1,2.

The result above suggests that GDP and TDA are non-stationary series and stationarity can only be induced when the series are differenced twice. However, INFLA series shows an I(0) stationary process indicating that the series is stationary at its level. The essence of this investigation is to help modeller to determine the order of integration for the ARIMA model. Certainly, a modeller is interested in that model that has significant parameters in the first place and also, satisfying certain optimality conditions. In the works of Box and Jenkins(1976) and furthered by various scholar including Chatfield(1982) among others, it has been shown that it is possible to have a set of ARIMA specifications with significant parameters and so, it becomes necessary to perform further diagnostic and optimality checks based on residual analysis to select the ‘best’ model.

For GDP series, six models are specified as shown in Table 2 below. These are ARIMA(1,0,0), ARIMA(1,1,0), ARIMA(1,1,2), ARIMA(1,0,2), ARIMA(2,2,2) and ARIMA(2,0,1). For these specifications, the parameters of the models are all significant and hence the model. This provides a class of models satisfying the criteria of significant parameters. However, the choice of the “best” model criteria has to be determined, and of course the consistency of these criteria as shown on Table 2 below.

Table 2: Various ARIMA specifications for GDP series

Criterion	ARIMA (1,0,0)	ARIMA (1,1,0)	ARIMA (1,1,2)	ARIMA (1,0,2)	ARIMA (2,2,2)	ARIMA (2,0,2)
SE Reg.	15736.22	15607.55	14405.59	12937.57	13959.21*	13959.4
AIC	22.1991	22.1840	22.0896	21.8714	22.066*	22.066
DW	1.01694	2.4038	2.0546*	2.1029	2.322	2.322
Q	6.97 (p<.05)	(p>.05)	(p>.05)*	(p>.05)*	(p>.05)*	6.83 (p<.05)
MAPE	52.58	70.88	48.91	53.78	26.47*	26.47

* The model gives the best specification in terms of model residuals. P <.05 suggests significant autocorrelation in the residual for at least one lag.

Considering the SE of the ARIMA models in Table 2 above, ARIMA(1,0,2) specification has the least SE value, followed by ARIMA(2,2,2) specification. However, from the AIC, it is clear that ARIMA(2,2,2) has the least AIC value of 20.066 than ARIMA(1,0,2). Although ARIMA(2,2,2) and ARIMA(2,0,2) have the same value of AIC, ARIMA(2,2,2) has the smallest SE of Regression when compared with ARIMA(2,0,2). Apart from ARIMA(1,0,0) that exhibit weak positive autocorrelation, all other specifications exhibit weak negative autocorrelation. Thus, in the class of weak negative autocorrelation, it is evidenced that ARIMA(2,2,2) and ARIMA(2,0,2) have the smallest MAPE of 26.47% each. The value of the Q-statistics is suggests non-significant autocorrelation of the residuals for ARIMA(1,1,0), ARIMA(1,1,2) ARIMA(1,0,2) and ARIMA(2,2,2). The search for the best model is therefore narrowed down to ARIMA(2,2,2) with several criteria suggesting that it dominate all other specifications, even as evidenced in Table 1 which shows that GDP series is an I(2) process.

Table 3: Various ARIMA specifications for TDA series

Criterion	ARIMA (1,0,0)	ARIMA (0,0,1)	ARIMA (0,0,2)	ARIMA (0,0,3)	ARIMA (2,2,0)	ARIMA (2,2,3)
SE Reg.	873768	1879153	1433952	1181176	1005077	836713.9*
AIC	30.233	31.763	31.254	30.89	30.553	30.283*
DW	1.865	0.4514	1.269	1.990	1.977	1.994*
Q	(p>.05)*	(p<.05)	(p<.05)	(p<.05)	(p>.05)*	(p>.05)*
MAPE	87.9	100	100	100	86.53	85.83*

* The model gives the best specification in terms of model residuals. P <.05 suggests significant autocorrelation in the residual for at least one lag.

TDA series has six possible ARIMA(p,d,q) specifications with significant parameters as shown in Table 3 above. These are ARIMA(1,0,0), ARIMA(0,0,1), ARIMA(0,0,2), ARIMA(0,0,3), ARIMA(2,2,0) and ARIMA(2,2,3). Again, when SE of ARIMA models are considered, the ARIMA(2,2,3) specification has the least SE value of 836713.9 followed by ARIMA(1,0,0) with SE of 873768. It is clear that ARIMA(1,0,0) has the least AIC value of 30.23 than ARIMA(2,2,3) with AIC value of 30.283. Although ARIMA(1,0,0) appeared to have the least value of AIC, ARIMA(2,2,3) has the smallest SE. Both specifications have positive autocorrelation with ARIMA(2,2,3) having almost zero autocorrelation. In terms of suitability, ARIMA(2,2,3) possesses the desirable qualities in terms of DW. In terms of MAPE, ARIMA(2,2,3) has MAPE of 85.83% and is followed by ARIMA(2,2,0) with MAPE of 86.53%. Again, the value of Q-statistics suggests non-significant autocorrelation of the residuals for ARIMA(1,0,0), ARIMA(2,2,0) and ARIMA(0,0,3) so that the search for the best model is pointing at ARIMA(2,2,3) which dominates other specification. Similarly, a critical examination of TDA series suggests a non-stationary process of I(2) like GDP series. Thus, it will be sufficient to recommend ARIMA(2,2,3) specification as the best for the TDA series.

Table 4: Various ARIMA specifications for INFLA series

Criterion	ARIMA (1,0,0)	ARIMA (0,0,1)	ARIMA (1,0,2)*	ARIMA (0,0,2)	ARIMA (1,0,1)*
SE Reg.	15.59	17.7	14.283**	16.686	15.597
AIC	8.352	8.6	8.237**	8.508	8.37
DW	1.957	1.510	1.868**	1.838	2.28
Q	P<.05	P<.05	p>.05**	p>.05**	P<.05
MAPE	108.46	100	277.47	100	117.47

* The model has a constant term.

** is the model that gives the best specification in terms of model residuals. P <.05 suggests significant autocorrelation in the residual for at least one lag.

For INFLA series, there are five possible ARIMA(p,d,q) specifications whose parameters are significant as shown in Table 4 above. These are ARIMA(1,0,0), ARIMA(0,0,1), ARIMA(1,0,2), ARIMA(0,0,2) and ARIMA(1,0,1). Using the SE of ARIMA models criterion, ARIMA(1,0,2) has the smallest SE and AIC of 14.28 and 8.237 respectively. In terms of DW, ARIMA(102) specification among others have weak positive auto-correlation except for ARIMA(1,0,1) which also has negative auto-correlation. ARIMA(1,0,2) has MAPE of 277.47% which is higher than all other specifications. The Q-statistics suggests non-significant autocorrelation of the residuals for ARIMA(1,0,2), and ARIMA(0,0,2) so that the search for the best model is pointing at ARIMA(1,0,2) which dominates other specification for more than 50% of the criteria under consideration.

5 Concluding Remark

The process of time series modelling has been described by Box and Jenkins among others and several methods of model selection have been suggested. The plot of Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) do not give sufficient information on the most suitable model specified hence the need to utilize every meaningful statistical procedure to identify the most suitable model for any series among the entertained models.

In order to fit a suitable stochastic model for each of the time series namely, GDP, TDA and INFLA, this study utilized the Augmented Dickey-Fuller Test to examined the series for stationarity and hence order of integration to be specified and thereafter, entertained several specifications for each series.

Using the specified methods of residual analysis, it was found that it is not always sufficient to utilize a single method of residual analysis to select the 'best' model hence, the need to consider several methods and identify the specification that dominates others in terms of the selection criteria. In this study, it has been found that the SE of Regression, AIC and Q statistics are frequently in agreement and in some cases, the DW and MAPE tests leading to the selection of ARIMA(2,2,2), ARIMA(2,2,3) and ARIMA(1,0,2) respectively for GDP, TDA and INFLA series.
The study concludes by suggesting the joint use of SE of regression, AIC and Q statistics as important criteria in determining the most suitable model for any specified series especially when the class of ARIMA models are to be entertained.

References

[1] Box G.E.P and Jenkins, *Time Series analysis, Forecasting and Control. Holden day*, Mc-Graw Hill books, New York, Revised Edition, 1976.

[2] Box G.E.P and pierce D.A., Distribution of residual auto correlations in auto-regressive integrated moving Average Time Series Model, *J. American Statistical Association*, **70**, (1970), 1509-1526.

[3] Chatfield C., *The Analysis of Time Series. An introduction*, Champman and Hall London. 2nd Edition, 1982.

[4] Durbin J. and Watson G.S., Testing for Serial Correlation in Least square Regression, *Biometrika*, **38**, (1951), 159-178.

[5] Evans J.R. and Lynsday W.M., *The Management and Control of Quality*. West publishing company, N.Y., third Edition, 1996.

[6] Henushek E.A. and Jackson J.E., *Statistical Methods For Social Scientists*, Academic press inc. New York, 1977.

[7] Kleinbaum D.G. and Kupper L.L., *Applied Regression Analysis and other Multivariable Method*, Duxbury press, New York, 1978.

[8] Ljung G.M. and Box G.E.P., On a Measure of Lack of fit in Time Series Models, *Biometrika*, **65**, (1978), 297-304.

[9] Pindyck R.S. and Rubbinfield D.L., *Econometric Models and Economic Forecasting*, Mc-Graw Hill, New York, second Edition, 1981.