The Efficacy of ARIMAX and SARIMA Models in Predicting Monthly Currency in Circulation in Ghana

Article · September 2013

3 authors:

Suleman Nasiru
C. K. Tedam University of Technology and Applied Sciences

Albert Luguterah
C. K. Tedam University of Technology and Applied Sciences

Lea Anzagra
University for Development Studies

Mathematical Theory and Modeling
ISSN 2224-5804 (Paper) ISSN 2225-0522 (Online)
Vol.3, No.5, 2013

The Efficacy of ARIMAX and SARIMA Models in Predicting Monthly Currency in Circulation in Ghana

Suleman Nasiru*, Albert Luguterah and Lea Anzagra
Department of Statistics, University for Development Studies,
P. O. Box, 24, Navrongo, Ghana, West Africa
*E-mail of Corresponding author: sulemanstat@gmail.com

Abstract
In this study, the efficacy of the ARIMAX model and SARIMA model in forecasting the Currency in Circulation in Ghana was compared. Both models appear to be adequate for forecasting the Currency in Circulation. Diagnostic tests of both models with the Ljung-Box test and ARCH-LM test revealed that both models were free from higher-order serial correlation and conditional heteroscedasticity respectively. The Diebold-Mariano test revealed that there is no significant difference in the forecasting performance of the two models. Hence, both models were proposed for predicting the Currency in Circulation. However, we recommend that continues monitoring of the forecasting performance of these models, review of market conditions and necessary adjustments are vital to make the use of these models more realistic.
Keywords: ARIMAX, SARIMA, Currency in Circulation, Ghana, forecasting

1. Introduction

The Currency in Circulation is one of the autonomous factors that propel money market liquidity. The variations in Currency in Circulation are vital indicators for monetisation and demonitisation of the economy. The share of the Currency in Circulation in money supply and its ratio to nominal Gross Domestic Product reveals its relative importance in any economy (Luguterah *et al*., 2013; Simwaka, 2006; Stavreski, 1998).
Several researches on Currency in Circulation have been done in both developed and developing countries. Balli and Elsamadisy (2011) modelled both daily and weekly Currency in Circulation for the state of Qatar using both regression and ARIMA models. Cabrero *et al*., (2002) modelled the daily series of bank notes in circulation in the context of managing the European monetary system. Also, Dheerasinghe (2006) modelled the currency in demand in Sri-Lanka with monthly, weekly and daily data set using time series models. Luguterah *et al*., (2013) modelled monthly Currency in Circulation in Ghana using SARIMA model. In another study, Luguterah *et al*., (2013) studied the effect of each month on the volume of monthly Currency in Circulation in Ghana.
Thus in this study, the forecasting accuracy of the ARIMAX model and the SARIMA model in predicting monthly Currency in Circulation in Ghana was compared.

2. Materials and Methods

This study was carried out in Ghana using data on reserve money growth, from January, 2000 to December 2011. The data was obtained from the website of the Bank of Ghana. The data was model using Seasonal Autoregressive Integrated Moving Average (SARIMA) model and ARIMAX model. Before modelling the data, preliminary tests were performed to determine evidence of seasonality and the order of non-stationarity of the data.

2.1 *Regression Analysis*

To investigate the evidence of seasonality in the Currency in Circulation, the data was logarithmically transformed and first differenced; before regressing on full set of periodic dummies. This was done to avoid spurious regression. The regression model is given by;

$$\Delta \ln CiC_t = \sum_{i=1}^{12} \alpha_i M_i + \varepsilon_t \qquad (1)$$

where M_i is a dummy variable taking a value of one for month i and zero otherwise (where i=1, 2,…,12), α_i are parameters to be estimated, and ε_t is the error term. The hypothesis tested is $H_0: \alpha_1 = \alpha_2 = \cdots = \alpha_{12} = 0$ against the alternative not all α_i are equal to zero. If the null hypothesis is rejected, then the data exhibit month-of-the-year seasonality.

2.2 *Augmented Dickey-Fuller Test*

The order of integration of data was investigated using the Augmented Dickey-Fuller (ADF) test. The regression model employed by Dickey and Fuller (1979) is given by;

$$\Delta Y_t = \alpha + \beta t + \delta Y_{t-1} + \sum_{i=1}^{p} \gamma_i \Delta Y_{t-i} + \varepsilon_t \qquad (2)$$

where α is a constant, β the coefficient on time trend series, $\sum_{i=1}^{p} \gamma_i \Delta Y_{t-i}$ is the sum of the lagged values of the dependent variable ΔY_t and p is the lag order of the autoregressive process. The parameter of interest in the ADF test is δ. For $\delta = 0$, the series contains unit root and hence non-stationary. The choice of the starting augmentation order depends on; data periodicity, significance of γ_i estimates and white noise residuals. The test statistic for the ADF test is given by

$$ADF = \frac{\hat{\delta}}{\mathrm{SE}(\hat{\delta})} \qquad (3)$$

where $\mathrm{SE}(\hat{\delta})$ is the standard error of the least square estimate of $\hat{\delta}$. The null hypothesis is rejected if the test statistic is greater than the critical value.

2.3 ***Diebold-Mariano Test***

The Diebold-Mariano test was used to assess whether the differences in the mean square errors of competing forecasts are statistically significant. The test statistic is given by;

$$\mathrm{S}_1 = \left[\hat{\mathrm{V}}(\bar{\mathrm{d}})\right]^{\frac{-1}{2}} \bar{\mathrm{d}} \qquad (4)$$

where $\bar{\mathrm{d}}$ is the mean of the coefficient of $\mathrm{d_t}$, which is the difference between the sets of squared forecast errors from two competing models, $\mathrm{d_t} = \mathrm{e}_{1t}^2 - \mathrm{e}_{2t}^2$ and $\hat{\mathrm{V}}(\bar{\mathrm{d}})$ is an estimate of the variance of $\bar{\mathrm{d}}$.

2.4 ***SARIMA Model***

The SARIMA model denoted by $\mathrm{ARIMA}(p,\ d, q) \times (P, D, Q)_s$ can be expressed using the lag operator as (Halim and Bisono, 2008);

$$\phi(L)\Phi(L^s)(1-L)^d(1-L^s)^D Y_t = \theta(L)\Theta(L^s)\varepsilon_t \qquad (5)$$
$$\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p$$
$$\Phi(L^s) = 1 - \Phi_1 L^s - \Phi_2 L^{2s} - \ldots - \Phi_P L^{Ps}$$
$$\theta(L) = 1 - \theta_1 L - \theta_2 L^2 - \ldots - \theta_q L^q$$
$$\Theta(L^s) = 1 - \Theta_1 L^s - \Theta_2 L^{2s} - \ldots - \Theta_Q L^{Qs}$$

where

$L^k Y_t = Y_{t-k}$

p, d, q are the orders of non-seasonal AR, differencing and MA respectively

P, D, Q are the orders of seasonal AR, differencing and MA respectively

Y_t represent the time series data at period *t*

The estimation of the model involves three steps, namely: identification, estimation of parameters and diagnostics. The identification step involves the use of the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) to identify the tentative orders of both the non-seasonal and seasonal components of the model. The second step involves estimation of the parameters of the tentative models that have been selected. In this study, the model with the minimum values of Akaike Information Criterion (AIC), modified Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) is adjudged the best model. The last stage which is the diagnostic stage involves checking whether the selected model adequately represents the Currency in Circulation. An overall check of the model adequacy was made at this stage using the Ljung-Box test and ARCH-LM test. These tests were performed to check for higher order autocorrelation and homoscedasticity respectively.

2.5 ***ARIMAX Model***

The ARIMAX model is simply an ARIMA model with additional or input variables. The model is an integration of a regression model with an ARIMA model. The result of this model covers the advantages of both models. The regression method describes the explanatory relationship while the ARIMA method takes care of the autocorrelation in the residuals of the regression model. The model is given by;

$$Y_t = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \frac{\theta(L)\Theta(L^s)}{\phi(L)\Phi(L^s)(1-L)^d(1-L^s)^D}\varepsilon_t \qquad (6)$$

where X_i represents the additional variable.

3. Results and Discussion

To confirm proper ordering of differencing filter, a unit root test was performed using ADF test. The ADF test confirms the existence of unit root under the situation where either a constant or constant with linear trend were included in the test. The results of the ADF test are shown in Table 1.

The evidence of seasonality was investigated by regressing the first differenced, logarithmically transformed data on full set of periodic dummies. The regression model was significant with an *F*- statistic of 15.7664 and a *p*-value of 0.0000. The results as depicted in Table 2 revealed that there is pronounced month-of-the-year seasonality.

Estimating the SARIMA Model

The logarithmically transformed data was both seasonally and non-seasonally differenced to make the data stationary. The ADF test in Table 3 affirms that the transformed seasonal and non-seasonal differenced Currency in Circulation is stationary.

After obtaining the order of integration of the Currency in Circulation, the order of the Autoregressive and Moving Average for both seasonal and non-seasonal components was determined. This was obtained from the ACF and PACF plots based on the Box-Jenkins (1976) approach. From Figure 1, the ACF plot have significant spike at the non-seasonal lag 1 and seasonal lag 12, with significant spikes at other non-seasonal lags. The PACF plot also has significant spikes at the non-seasonal lags 1 and 2 and seasonal lags 12 and 24. The PACF plot also has significant spike at other non-seasonal lags.

Using the lower significant lags of both the ACF and PACF and their respective seasonal lags, tentative models were identified for the Currency in Circulation (Table 4). Among these possible models presented in Table 4, ARIMA $(0, 1, 1)(0, 1, 1)_{12}$ was chosen as the appropriate model that fit the data well because it has the minimum values of AIC, AICc and BIC compared to other models.

Using the method of maximum likelihood, the estimated parameters of the derived model are shown in Table 5. Observing the *p*-values of the parameters of the model, it can be seen that both the non-seasonal and seasonal Moving Average components are highly significant at the 5% level.

To ensure the adequacy of the estimated model, the ARIMA $(0, 1, 1)(0, 1, 1)_{12}$ was diagnosed. As shown in Figure 2, the standardised residuals revealed that almost all the residuals have zero mean and constant variance. Also, the ACF of the residuals depict that the autocorrelation of the residuals are all zero that is they are uncorrelated. Finally, in the third panel, the Ljung-Box statistic indicates that there is no significant departure from white noise for the residuals as the *p*-values of the test statistic clearly exceeds the 5% significant level for all lag orders.

To buttress the information depicted in Figure 2, the ARCH-LM test and t-test were employed to test for constant variance and zero mean assumption respectively. The ARCH-LM test result shown in Table 6, failed to reject the null hypothesis of no ARCH effect in the residuals of the selected model. Also, the t-test gave a test statistic of -1.3281 and a *p*-value of 0.1865 which is greater than the 5% significance level. Thus, we fail to reject the null hypothesis that the mean of the residuals is approximately equal to zero. Hence, the selected model satisfies all the assumptions and it can be concluded that ARIMA $(0, 1, 1)(0, 1, 1)_{12}$ model provides adequate representation of the Currency in Circulation.

Estimating the ARIMAX Model

Dropping the insignificant variables at the 5% level of significance in Table 2, a new regression model was fitted with the transformed, non-seasonal, first differenced series. As shown in Table 7, all the variables were significant. The R-squared for this model is about 59.8% and the Durbin-Watson statistic of 2.5103 indicated that the model was free from first order serial correlation.

An examination of the ACF and PACF plot of the residuals shown in Figure 3 revealed that the model residuals were not free from higher order serial correlation but were stationary.

An appropriate ARIMA (p, 0, q) model was therefore developed for the residuals using lower significant lags of the ACF and PACF. As shown in Table 8, ARIMA (0, 0, 1) model appears to be the best model for the residuals as it has the least AIC, AICc and BIC values.

Since the best model has been identified for the residuals, the next step was to concatenate the regression model with the ARIMA (0, 0, 1) model for the residuals. As shown in Table 9, all the parameters of the integrated model were significant.

To ensure the adequacy of the model, the ARIMAX model was diagnosed. From the diagnostic plot shown in Figure 4, the residuals of the model can be said to have zero mean and constant variance. In addition, the ACF plot of the residuals showed that the residuals are uncorrelated. Finally, the Ljung-Box statistic shown in the third panel depicts that there is no significant departure from white noise for the residuals as the *p*-values of the test statistic clearly exceeds the 5% significance level for all lag orders.

The ARCH-LM test and t-test were employed to test for constant variance and zero mean assumptions respectively. As shown in Table 10, the ARCH-LM test failed to reject the null hypothesis of no ARCH effect in the residuals of the model. Also, the t-test gave a test statistic of 1.9588 with *p*-value of 0.0521 which is greater than the 5% level of significance. Hence, the null hypothesis of zero mean of the residuals was not rejected. The model satisfies all the important assumptions of modelling and therefore can be said to be an adequate representation of the Currency in Circulation.

Comparative Analysis of Models

Since both models developed were adequate for representing the data, the Diebold-Mariano test, was used to compare the predictive accuracy from both models. From the results (Table 11), the test revealed that there was no significant difference in the forecast accuracy of the two models as all the *p*-values for the test statistics were larger than the 0.05 significance level. Hence, we fail to reject the null hypothesis that the two models have the same forecast accuracy.

4. Conclusion

In this study, the forecasting accuracy of the ARIMAX model and SARIMA model for forecasting Currency in Circulation in Ghana was compared. The Diebold-Mariano test indicated that there was no significant difference in the forecasting accuracy of the two models. Thus, we proposed both the ARIMAX model and SARIMA model for forecasting Currency in Circulation in Ghana. However, since Currency in Circulation is volatile and subject to several unobservable development in the economy, sole reliance on these forecasting models to predict the Currency in Circulation for the purpose of liquidity management by the Bank of Ghana is not advisable. Therefore continues monitoring of the forecasting performance of these models, review of market conditions and necessary adjustments are required to make the use of these models more realistic.

References

Balli, F., and Elsamadisy E. M., (2011). Modelling the Currency in Circulation for the State of Qatar. Central Bank of Qatar. http://mpra.ub.uni muenchen. de/20159/1/ Qatarcirculation. pdf.

Bank of Ghana (2012). Monetary Time Series Data. www.bog.gov.gh. Date Accessed, 10th November, 2012.

Bell, W. R., and Hillmer, S. C., (1983). Modelling Time Series with Calendar Variation. *Journal of the American Statistical Association*, 78: 526-534.

Box, G. E. P., and Jenkins, G. M., (1976). Time Series Analysis: Forecasting and Control. Holden-Day, San-Francisco.

Cabrero, A., Camba-Mendez, G., Hirsch, A., and Nieto, F., (2002). Modelling the Daily Banknotes in Circulation in the Context of the Liquidity Management of the European Central Bank. European Central Bank Working Paper No. 142.

Dheerasinghe, R., (2006). Modelling and Forecasting Currency in Circulation in Sri Lanka. Central Bank of Sri Lanka Staff Papers No. 144.

Dickey, D. A., and Fuller, W. A., (1979). Distribution of the Estimators for Autoregressive Time Series with a Unit-root. *Journal of the American Statistical Association*, 74: 427-431.

Diebold, F. X., and Mariano, R., S., (1995). Comparing Predictive Accuracy. *Journal of Business and Economic Statistics*, 13: 253-263.

Halim, S., and Bisono, I. N., (2008). Automatic Seasonal Autoregressive Moving Average Models and Unit Root Test Detection. *International Journal of Management Science and Engineering Management*, 3(4): 266-274.

Halvacek, M., Michael K., and Josef C., (2005). The Application of Structural Feed-Forward Neural Networks to the Modelling of Daily Series Currency in Circulation. Czech National Bank Working Paper.

Ljung, G. M., and Box, G. E. P., (1978). On A Measure of Lack of Fit in Time Series Models. *Biometrika*, 65: 297-303.

Luguterah A., Suleman N., Anzagra L., (2013). A Predictive Model for Monthly Currency in Circulation in Ghana. *Mathematical Theory and Modeling*, 3(4): 43-52.

Luguterah A., Anzagra L., Suleman N., (2013). Monthly Effect on the Volume of Currency in Circulation Ghana. *Research Journal of Finance and Accounting*, 4(5): 132-137.

Simwaka, K., (2006). The Determinants of Currency in Circulation in Malawi. Research and Statistics Department, Reserve Bank of Malawi.

Stavreski, Z., (1998). Currency in Circulation. National Bank of the Republic of Macedonia Working Paper No. 1.

Suleman, N., and Sarpong, S., (2012). Modeling the Pattern of Reserve Money Growth in Ghana. *Current Research Journal of Economic Theory*, 4(2): 39-42.

Table 1: ADF test of Currency in Circulation in level form

Test	Constant		Constant+ Trend	
	Test Statistic	*P*-value	Test Statistic	*P*-value
ADF	5.4972	1.0000	5.2189	1.0000

Table 2: Regression parameters of the transformed first differenced series

Variable	Coefficient	Standard error	T-statistic	*P*-value
January	-0.0777	0.0187	-4.1648	0.0001*
February	-0.0516	0.0178	-2.9040	0.0043*
March	-0.0021	0.1780	-0.1186	0.9058
April	0.0196	0.0179	1.0989	0.2738
May	0.0129	0.0179	0.7216	0.4718
June	0.0015	0.0180	0.0857	0.9318
July	0.0340	0.0180	1.8890	0.0611
August	0.0050	0.0181	0.2785	0.7811
September	0.0300	0.0181	1.6567	0.1000
October	0.1243	0.0182	6.8391	0.0000*
November	0.0380	0.0182	2.0860	0.0389*
December	0.1706	0.0183	9.3292	0.0000*

*: Means significant at the 5% level of significance

Table 3: ADF test of seasonal and non-seasonal differenced series

Test	Constant		Constant+ Trend	
	Test Statistic	*P*-value	Test Statistic	*P*-value
ADF	-5.0165	0.0000	-4.9081	0.0001

Table 4: Tentative SARIMA models

Model	AIC	AICc	BIC
ARIMA $(1, 1, 1)(1, 1, 1)_{12}$	-368.92	-368.44	-354.54
ARIMA $(1, 1, 1)(2, 1, 1)_{12}$	-367.05	-366.37	-349.80
ARIMA $(2, 1, 1)(1, 1, 1)_{12}$	-366.92	-366.25	-349.67
ARIMA $(1, 1, 0)(1, 1, 0)_{12}$	-353.80	-353.62	-345.18
ARIMA $(0, 1, 1)(0, 1, 1)_{12}$	-372.16*	-371.97*	-363.53*

*: Means best based on the selection criteria

Table 5: Estimates of parameters for ARIMA $(0, 1, 1)(0, 1, 1)_{12}$

Variable	Coefficient	Standard error	*z*-statistic	*P*-value
θ_1	0.3809	0.0786	4.8400	0.0000*
Θ_1	0.7109	0.0849	8.3570	0.0000*

*: Means significant at the 5% level of significance

Table 6: ARCH-LM test of residuals of ARIMA (0, 1, 1)(0, 1, 1)$_{12}$

Lag	Test statistic	df	*P*-value
12	2.8814	12	0.9963
24	4.7132	24	1.0000
36	6.3775	36	1.0000

Table 7: Parameters of regression model on selected periodic dummies

Variable	Coefficient	Standard error	T-statistic	*P*-value
January	-0.0810	0.0168	-4.8342	0.0000*
February	-0.0547	0.0160	-3.4079	0.0009*
October	0.1209	0.0160	7.5383	0.0000*
November	0.0346	0.0160	2.1554	0.0329*
December	0.1671	0.0160	10.4173	0.0000*

*: Means significant at the 5% level of significance

Table 8: Tentative models for the residuals

Model	AIC	AICc	BIC
ARIMA (1, 0, 1)	-416.0500	-414.9800	-392.2900
ARIMA (1, 0, 0)	-415.4200	-414.5900	-394.6300
ARIMA (0, 0, 1)	-417.3900*	-416.5700*	-396.6000*

*: Means best based on the selection criteria

Table 9: Parameters of the ARIMAX model

Variable	Coefficient	standard error	z-statistic	*P*-value
θ_1	0.3174	0.0813	3.9024	0.0001*
January	-0.0808	0.0164	-4.9178	0.0000*
February	-0.0543	0.0151	-3.6041	0.0003*
October	0.1306	0.0154	8.4684	0.0000*
November	0.0346	0.0159	2.1807	0.0292*
December	0.1671	0.0159	10.5396	0.0000*

AIC=-417.39 AICc=-416.57 BIC=-396.6

*: Means significant at the 5% level of significance

Table 10: ARCH-LM test of the ARIMAX model

Lag	Test statistic	df	*P*-value
12	5.2238	12	0.9501
24	7.2102	24	0.9996
36	11.1008	36	1.0000

Table 11: Diebold-Mariano test

Forecast horizon	Test statistics	*P*-value
1	-1.2050	0.2282
2	-0.9675	0.3333
3	-0.8653	0.3869
4	-0.8033	0.4218
5	-0.7501	0.4532

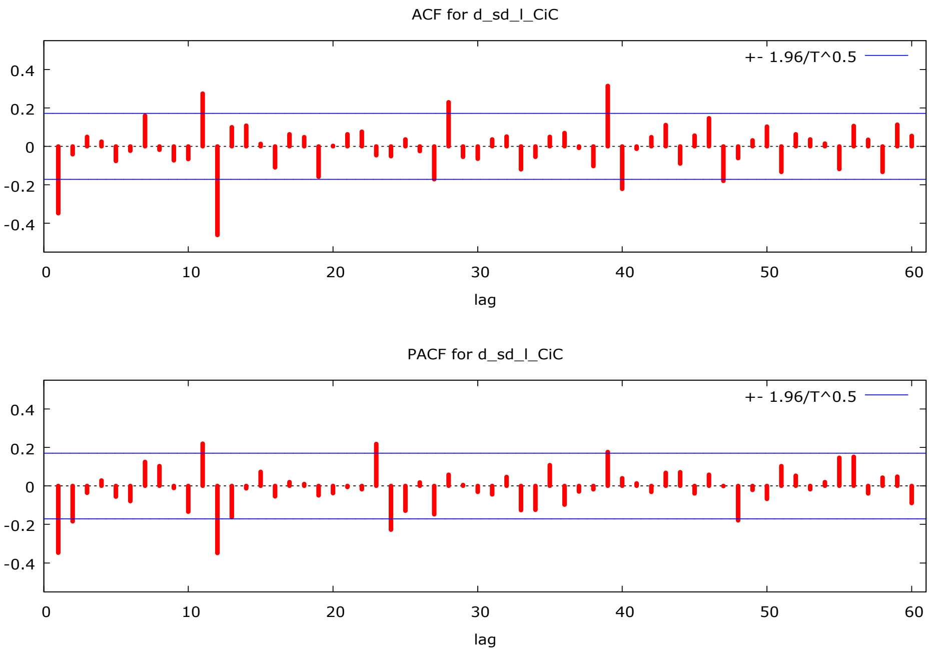

Figure 1: ACF and PACF plot of differenced series

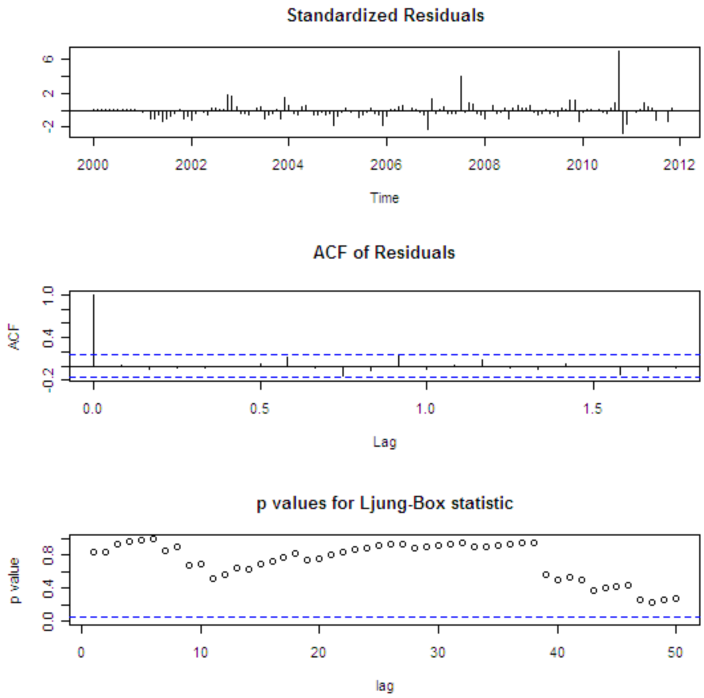

Figure 2: Diagnostic plot of ARIMA (0, 1, 1)(0, 1, 1)$_{12}$

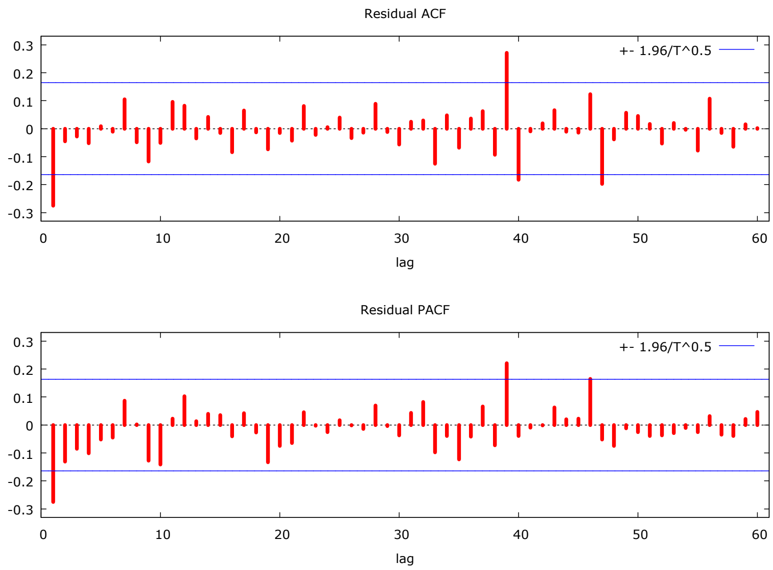

Figure 3: ACF and PACF plot of model residuals

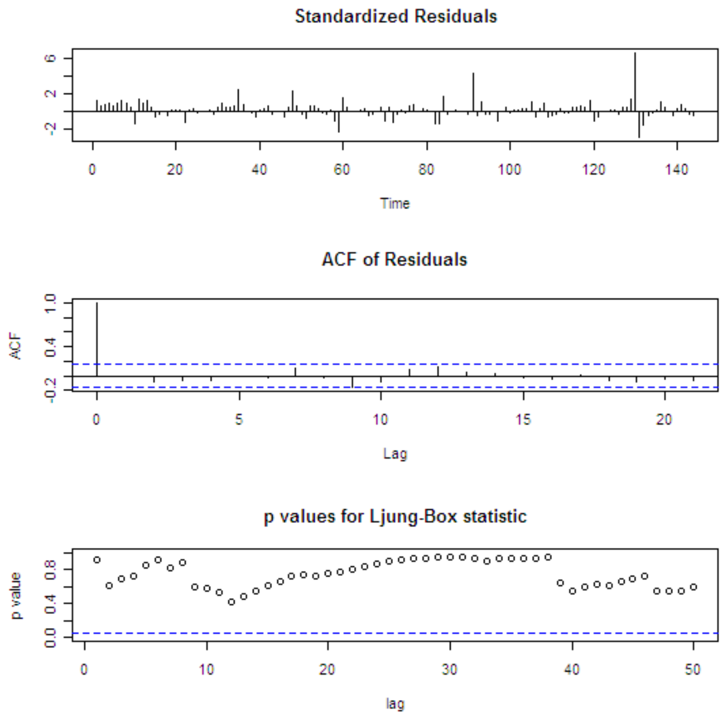

Figure 4: Diagnostic plot of the ARIMAX model

1. Suleman Nasiru is a Senior Research Assistant at the Department of Statistics, University for Development Studies. He is currently a final year master's student in Applied Statistics at the University for Development Studies.

2. Dr. Albert Luguterah is a lecturer at the Department of Statistics, University for Development Studies. He obtained his PhD in Applied Statistics at the University for Development Studies. He is currently the head of the Statistics Department.

3. Mrs. Lea Anzagra is an Assistant lecturer at the Department of Statistics, University for Development Studies. She obtained her master's degree in Applied Statistics at the University for Development Studies.