Factors that Determine Monthly Revenue Returns from Mobile Payments in Kenya

Kyalo Richard[1], Waititu Anthony[2], Wanjoya Anthony[3]

[1, 2, 3]Jomo Kenyatta University of Agriculture & Technology, Statistics and actuarial Science
P.O. Box 62,000 – 00200 Nairobi, Kenya

Abstract: ***This study examined factors that determine revenue collected from mobile payments in Kenya in terms of their effects on monthly revenue returns. Monthly data from March, 2007 to June, 2013 was analyzed. ARIMA diagnostic test which involved standardized residual and Ljung-Box test were performed to identify the correlation between each independent variable namely, Exchange rate of Kshs/USD, Number of customers, number of agents and the number of transaction against the value of transaction as the response variable. The model was used to quantify the correlation between each independent variable after validation process by use of Ljung-Box test, P-values, AIC and BIC and graphical performances which include ACF and PACF plots of residuals and significant variable identified using p value at 95% confidence interval were number of transaction was observed as the factor that significantly determines monthly revenue returns from mobile services operated in the country.***

Keywords: Mobile payments, ARIMA, Residual analysis, significance, generalized least square.

1. Introduction

Mobile payments in Kenya facilitate an average of $320 million per month in person-to-person transfers this is equivalent to roughly 10% of Kenya's GDP on an annualized basis. Extremely rapid uptake of mobile payment is a strong vote of assurance by local users in a new technology as well as an indication of significant latent demand for remittance services. In recent years a number of models for time dependent data have been developed and in this study, we shall use Auto Regressive Moving Average (ARIMA) models to ascertain factor(s) that may influence monthly revenue collected from mobile payment services. ARIMA model can be considered as a special type of regression model in which the dependent variable has been stationarised and the independent variables are all lags of the dependent variable and/or lags of the errors. The selection of a best model fit to historical data is directly related to whether residual analysis is done. Therefore, an indicative check such as the independence, homoscedasticity and normality of the residuals is the most important analysis of ARIMA model building.

Using non-stationary time series data in financial model produces unreliable and forged results and thus leads to poor forecasting as elucidated by Hyndman and Athanasopoulos (2013), "a stationary time series is one whose properties do not depend on the time at which the series is observed". This is done by transforming the time series data so that it becomes stationary.

2. Previous Research

Basically no much research has been done related to Mobile payment revenue bearing in mind that it's a new technology introduced few years back and by this our literature review is based on other studies related to financial and economic data at large. For instance McKerchar and Delleur (1974) used an ARIMA process to achieve stochastic modeling of monthly flows. McLeod et al. (1977) applied the ARIMA approach to average annual stream flows, annual sunspot number series and monthly airline passenger data and suggested a different ARIMA model for each data set. The impact of macroeconomic factors on the development of stock markets was the subject of numerous studies, such as Bodie (1976), Fama (1981), Pearce, Roley (1985). These studies provide clear evidence of strong causal correlation between macroeconomic factors and share yields.

Fernando and Jayawardena (1994) used various ARIMA models in forecasting monthly rainfall records. Venama et al. (1996) investigated climate change in the Senegal River basin via this approach. Correlation between macroeconomic factors and share prices in Asian market was studied by Mookerjee and Yu (1997) who realised positive correlation between the volume of foreign exchange reserves and money aggregates M1 and M2 and the development of the SSE index in Singapore.

Kasipillai et al. (2003) showed that other factor that causes tax evasion is tax culture. They evaluated the influence of education on tax compliance among undergraduate students in Malaysia, and found that there is a close relationship between education and tax compliance. Studies in India by Jain (1987) also found that complicated tax structure, dishonest staff, high tax rates and high tax rate on sales are factors that cause high black money in India.

A Delphi study carried out by (Shon and Swatman 1997) on effectiveness criteria for internet payment systems (IPS) revealed 15 factors distributed over six types of stakeholders: financial institutions, IPS providers, merchants, consumers, regulators and network providers. Security and reliability was important for almost all groups. Lower transaction costs were favored by merchants, consumers and financial institutions. Scalability and universality were important factors for network providers. Flexibility was also important for merchants.

Fernando and Jayawardena (1994) used ARIMA models in forecasting monthly rainfall records while Venama et al.

(1996) investigated climate change in the Senegal River basin via this approach. This is optimal indication that ARIMA model can be precisely applied in time dependent data to analyze factors that directly influence monthly returns on mobile payment service.

3. Materials ad Methods

For analyzing the impact of the highlighted factors that may precisely determine overall revenue returns from mobile payments, Time series data in conjugation with exchange rate of Ksh/USD March 2007 to June 2013 extracted from the Central Bank of Kenya website was used.

We consider a time series model to be useful in examining the dynamic determinants of financial series since its time dependent data thus suitable in estimating relationship between economic variables. This include Generalized least square (GLS), Ordinary least Square (OLS), Autoregressive integrated moving average (ARIMA), GARCH, and ARCH. Researcher have used these methods to estimate the relationship between independent and dependent variables in financial analysis. In this study, Box Jenkins method was used to build an autoregressive integrated moving average (ARIMA), this is a time series model, which is designed to examine sequentially lagged relationships for relationships that may not be apparent in data collected periodically.

Auto ARIMA function was used to identify the ARIMA model of the correct order. Ideal parameters were estimated by maximum likelihood method. Diagnostic tests, including p- values of Ljung-Box statistics and standardized residual analysis to ensure that the chosen model fits the data by residual turning out to be white noise. The general form of the ARIMA model is:

$$\widehat{Y}(t)-\phi Y(t-1)=\mu-\theta e(t-1) \quad (1)$$

Equation (1) above shows ARIMA model mathematical form where all the auto regressive (AR) terms and differences are collected on the left hand side of the equation while the moving average terms (MA) are collected on the right hand side. ARIMA model denoted as ARIMA (p, d, q)*(P, D, Q) that is a combination of past values and past residuals can be written as follows (Janacek and Swift, 1993; Ahmad et al., 2001; Sun and Koch, 2001):

$$\Theta(B)\phi F(B^s)(w_i-\mu)=C+\theta(B)\Theta(B^s)a_i \quad (2)$$

Where

$$w_i=(1-B)^d(1-B^s)^D \; xi \quad (3)$$

$$\Theta(B)=1-\Theta_1 B-\Theta_2 B^2-...-=\Theta_p B^p \quad (4)$$

$$\theta(B)=(1-\theta_1 B-\theta_2 B^2-...-=\theta_q B^q \quad (5)$$

$$\phi(B^s)=1-\phi_1 B^s-\phi_2 B^{2s}-...-=\phi_p B^{ps} \quad (6)$$

$$\theta(B)=1-\theta_1 B-\theta_2 B^{2s}-...-=\theta_Q B^{Qs} \quad (7)$$

Durbin Watson test was used to check whether the residuals are independent. The test statistic is shown in equation (8) below

$$k=\frac{\sum_{j=2}^{n}(e_j-e_{j-1}{}^2)}{\sum_{j=1}^{n}e_j{}^2} \quad (8)$$

Where $e_j=y_j-\widehat{y}_j$, y_j and $\widehat{y}_j$ are respectively the observed and predicted values of the response variable for individual *j*. *k* becomes smaller as the serial correlations increase. Durbin-Waston test is based on the assumption that the errors in the regression model are generated by a first-order autoregressive process observed at equally spaced periods as shown in equation (9)

$$\varepsilon_t=\rho\varepsilon_{t-1}+a_t \quad (9)$$

where ε_t is the error term in the model at time period t, a_t is an $NID(0,\sigma^2)$ random variable, and ρ ($|\rho|<1$) is the autocorrelation parameter. With any model we first check if there is good overall fit and then significance of each independent variable is determined using respective p-values, the test statistics hypothesis for testing the overall fit of the model is

$$H_O:\beta_1=\beta_2=\beta_3=...=\beta_k$$

All independent variables are significant in predicting the response variable against

$$H_A: \text{ Atleast one } \quad \beta_k=0$$

The alternative is at least one independent variable is significant for predicting response variable. With independent variable if one of the variable does not contribute significantly to predict the value of dependent variable the coefficient will be zero. Thus the test statistics is to determine whether estimated coefficient is different from zero. The hypothesis for testing variables significance is:

$$H_O:\beta_j=0$$

The independent variable, xj, is not important for predicting the dependent variable

$$H_A:\beta_j>0 \text{ or } \beta_j<0 \quad \beta_j\neq 0$$

The independent variable, xj, is important for predicting y

With small p-value we rejects the null hypothesis that there is no independent variables which is significant. i.e., at least one of the independent variables is significant. In this study we used p-values at 95% confidence interval to study the variables that are significant in the model.

4. Results

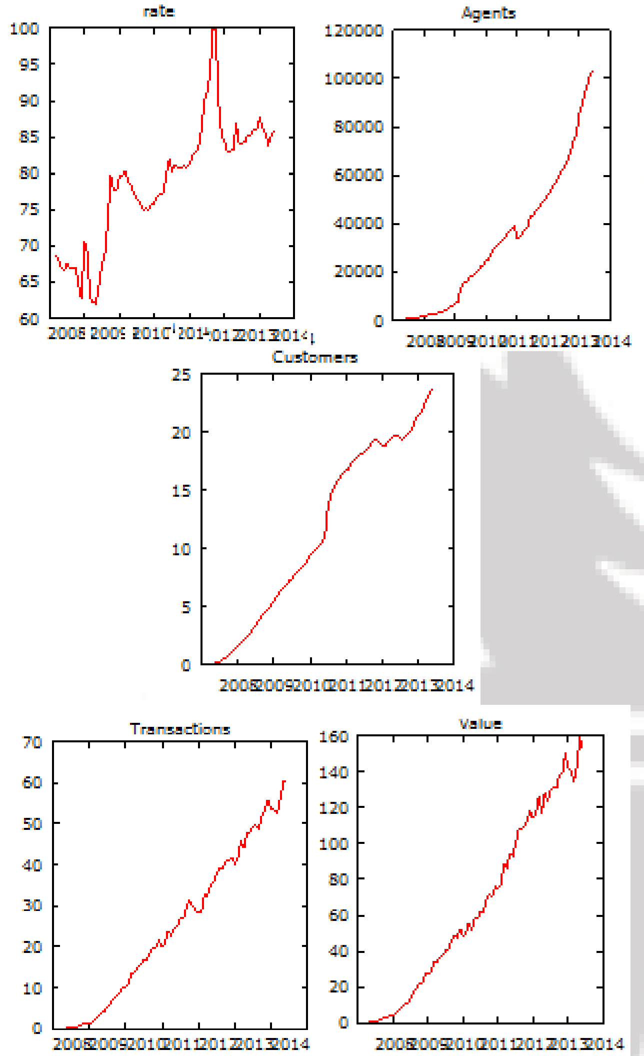

Figure 1: Dataset Graph

Apart from exchange rate we can recognize a linear trend with small drifts from the other variable and a fluctuating trend for the exchange rate. We started our analysis by fitting A Generalized Least Square using corARMA function in R a common application of to time-series regression, in which it is generally unlikely to assume that errors are independent. The plots of the ACF and the PACF for the monthly data sequence were drawn to gather information about the seasonal and nonseasonal AR and MA operators concerning the monthly series. On fitting Generalized Least Square on the data the ACF graph shown an attenuating sine wave pattern that reflected the random periodicity of the data and possible indication for the need for Non-seasonal and/or seasonal AR terms in the Model and hence opted for Auto ARIMA time series model instead.

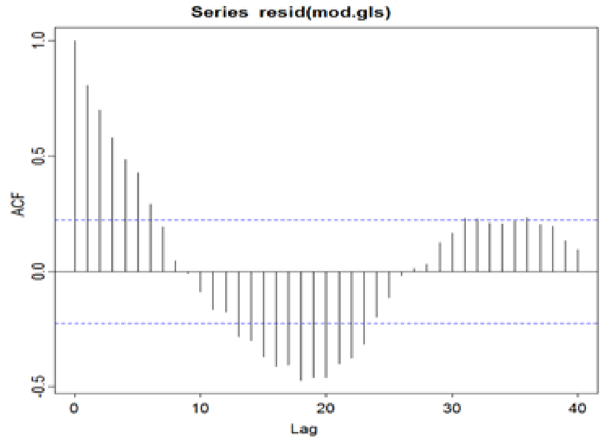

Figure 2: GLS ACF plot

Based on goodness of fit the Auto ARIMA function fitted ARIMA (0,1,0)(0,1,1) to the data with Akaike criterion (266.6817), test for normality of residual was normally distributed with the test statistics chi square(2)=19.3535 and p-value of 6.27264e-005.The selected best models were consistent with the independence assumption for all tests. Table 1 show a summary ARIMA model fit to the data

Table 1: ARIMA model summary

	Coefficient	*Std. error*	*Z*	*p-value*
Const	0.053309	0.166081	0.3210	0.7482
Theta	-0.42630	0.130484	-3.267	0.0011
Rate	-0.0.0264	0.077430	-0.342	0.7323
Agents	-0.00018	0.000126	-1.463	0.1435
Customers	-0.34002	0.504228	-0.674	0.5001
Transaction	2.80757	0.174278	16.11	2.18e-05*

P value <0.05 was considered statistically significant

For the selected best models, the results related to the normality of residuals using Kolmogorov Smirnoff test, skewness tests were smaller than the critical values at the 5% level of significance. These results suggest that the residuals of the best models are normally distributed. In addition to these tests, the two plots shown in figure 3 and figure 4 are the histogram and QQ-plot of the residuals. Both of these plots are used to test for Normality. We can see that the QQ-plot approximately follows the QQ-line visible on the plot and the histogram shows a bell-shaped distribution. These are good indicators of Normality within the residuals. A KS test may also be used here to test for a normal distribution. For this set of residuals, in summary the graph shows no obvious violations of the normality assumption as.

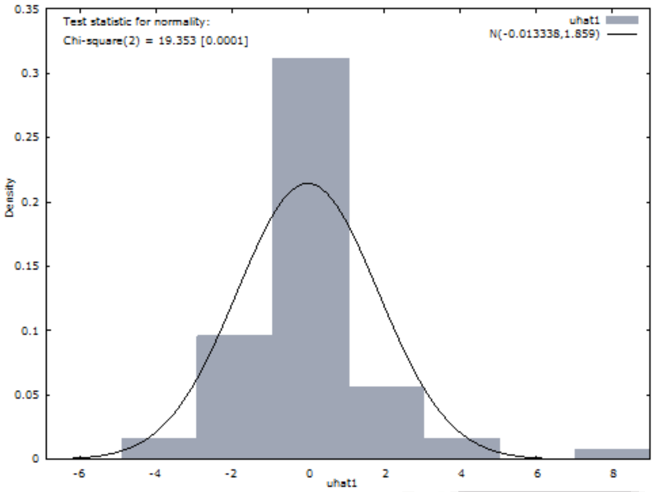

Figure 3: Normality test plot

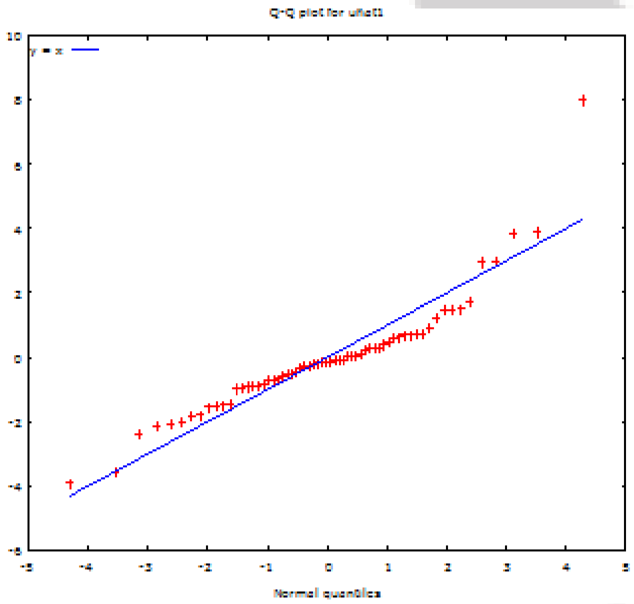

Figure 4: QQ Plot

An overall diagnostic test for the model gave the results shown in figure 5

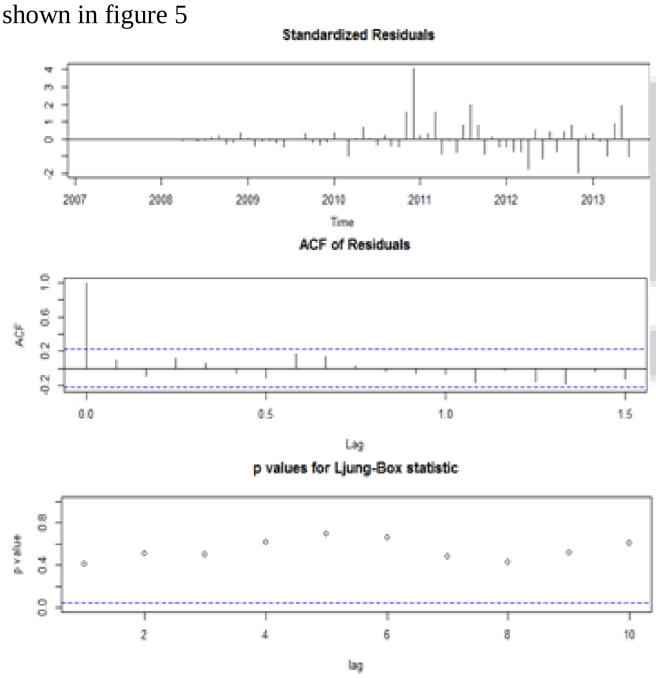

Figure 5: Diagnostic plots

Observation from overall diagnostic test signifies the following;

- The standardized residuals don’t show cluster of volatility
- The autocorrelation function (ACF) show no significant autocorrelation between residuals
- The p-values for the Ljung-Box statistics are all large, indicating that the residuals are pattern less meaning that the residual are white noise.
- This forms a good conclusion that the residuals of the

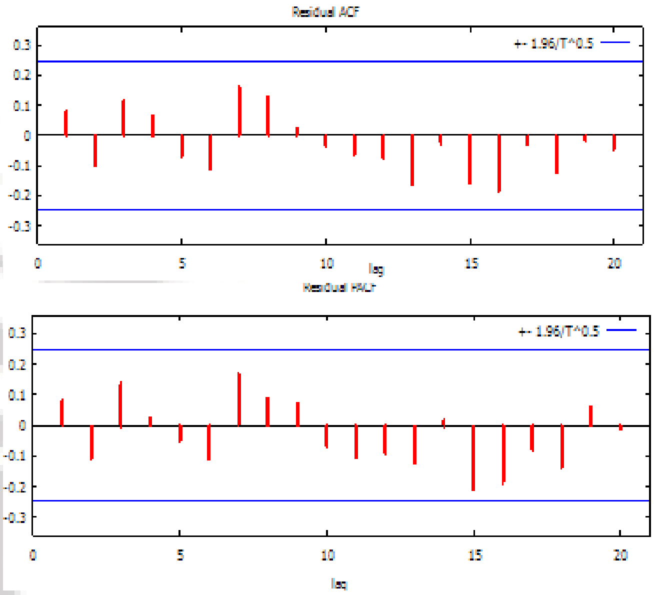

Figure 6: ACF and PACF plot

As shown in figure 6 the ACF and PACF values are all within the 95% zero bound indicating that there is no correlation amongst the residuals. The function (ACF) of the series was examined and found to be different from zero, this implies that there is dependence between observations. With ACF plots, diagnostic test and test for normality we had a sufficient information to proof that ARIMA(0,1,0)(0,1,1) best fitted the mobile payment data and that we can use the results to analyze the factors that determines the revenue collected from this service in a given time.

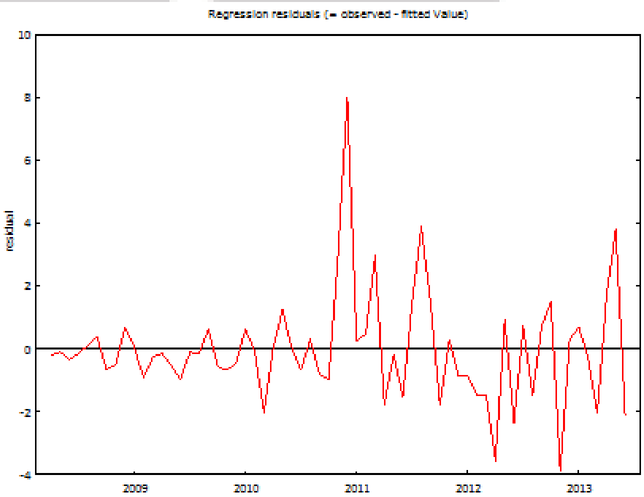

Figure 7: Residual plot against time

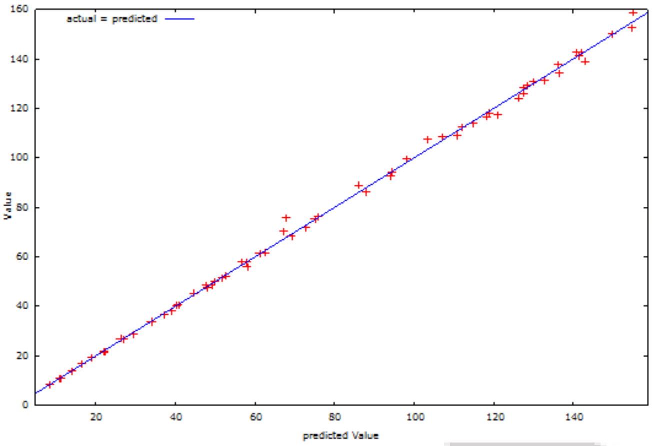

Figure 8: Actual vs. Fitted plot

5. Summary and Conclusion

This study was concerned with identifying factor(s) that affect revenue turn out from mobile payment services by fitting an ARIMA model and testing residuals from the model, Independence analysis of the residuals was examined by using the Ljung-Box Q statistic. ARIMA(0,1,1)(0,1,1) best fitted the dataset and using P values from Table 1 To determine which independent variable was significant the exchange rate and number of customers with (p-values 0.7323 and 0.5001 respectively) >0.05 are not significant at 95% confidence interval in the model, However Number of agents with p-value 0.1435>0.05 at 95% CI was not significant this means that the number of transaction with p-value 2.18e-05* was the only significant variable in the model.

From the fitted model we can conclude that number of transactions determines the total revenue to be collected in any given month this is a good model in the sense that as the number of transaction increases we expect the revenue to increase unlike other predictors such as number of agents which may not influence revenue turn out in any way.

The policy implication of this study is that factor influencing returns on monthly revenue collection can be used to forecast monthly revenue returns from mobile payments services, which is certainly useful for various financial players such as government and policy makers of the country and thus time series analysis with ARIMA model is an accurate method of modeling a time dependent data.

References

[1] T. Abe, Y. Tokuda, S. Ohde, S. Ishimatsu, and R. B. Birrer, "The influence of meteorological factors on the occurrence of trauma and motor vehicle collisions in Tokyo," Emergency Medicine Journal, vol. 25, no. 11, pp. 769–772, 2008.
[2] Hipel, K.W., McLeod, A.I. and Lennox, W.C., "Advances in Box-Jenkins Modeling 1. Model Construction", Water Resources Research, 13, 567-575, 1977.
[3] Zhang, G.P., "Time Series Forecasting Using a Hybrid ARIMA and Neural Network Model", Neuro computing, 50, 159-175, 2003.
[4] McLeod, A.I., "Diagnostic Checking Periodic Autoregression Models with Application", The Journal of Time Series Analysis, 15, 221-233, 1995.
[5] Dickey D.A. and Fuller, W.A., "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root", Econometrica, 49, 1057-1072, 1981.
[6] Cromwell, J.B., Labys, W.C. and Terraza, M., Univariate Tests for Time Series Models, A Sage Pub-lications, Series/number: 07-99, London, 1994.
[7] Hipel, K.W. and McLeod, A.I., Time Series Modelling of Water Resources and Environmental Systems. Development in Water Science, vol. 45, Elsevier Scientific Publishing Company, Amsterdam, The Netherlands, 1994.
[8] FAMA, E. F. Stock returns, real Activity, inflation and money. The American Economic review: 1981, 71(4): 45-565. R. Caves, Multinational Enterprise and Economic Analysis, Cambridge University Press, Cambridge, 1982. (book style)
[9] Litterman, R., 1986. "Forecasting with Bayesian Vector Autoregressions - Five Years of Experience", Journal of Business and Economic Statistics, January, No. 1, Vol. 4, pp. 25-38.
[10] G. E. P. Box and D. R. Cox, "An analysis of transformation (with discussion)," The Royal Statistical Society, vol. 26, pp. 211–252, 1964.
[11] T. Abe, S. Ohde, S. Ishimatsu et al., "Effects of meteorological factors on the onset of subarachnoid hemorrhage: a time- series analysis," Journal of Clinical Neuroscience, vol. 15, no. 9, pp. 1005–1010, 2008.
[12] Montgomery, D. C., Peck, E. A. and Vining, G. G. (2001). Introduction to Linear Regression Analysis. 3rd Edition, New York, New York: John Wiley & Sons.
[13] Cromwell, J.B., Labys, W.C. and Terraza, M., Univariate Tests for Time Series Models, A Sage Publications, Series/number: 07-99, London, 1994.
[14] Box, G.E.P and G.M. Jenkins., 1976. Time Series Analysis; Forecasting and Control, Holden-Day Inc., USA, pp: 574.
[15] Hampe, J. F. and P. M. C. Swatman (2001). Mobile payment: opportunities, challenges and solutions. Panel at the Bled Electronic Commerce Conference, Bled, Slovenia.

Author Profile

Richard Kyalo Holds a BSc degree in Mathematics and Computer science from Jomo Kenyatta University of Agriculture and Technology. Currently he is finalising his MSc. Applied Statistics in JKUAT.

Anthony G. Waititu Holds a PhD in Applied Statistics and currently he is a senior lecturer in the department of Statistics and Actuarial Sciences, Jomo Kenyatta University of Agriculture and Technology.

Wanjoya Anthony Holds a PhD in Applied Statistics from Università degli Studi di Padova and currently he is a senior lecturer in the department of Statistics and Actuarial Sciences, Jomo Kenyatta University of Agriculture and Technology.