



## Subset Sarima Modelling: An Alternative Definition and a Case Study

Ette Harrison Etuk<sup>1\*</sup> and Nathaniel Ojekudo<sup>2</sup>

<sup>1</sup>Department of Mathematics and Computer Science, Rivers State University of Science and Technology, Port Harcourt, Nigeria.

<sup>2</sup>Department of Computer Science, Ignatius Ajuru University of Education, Port Harcourt, Nigeria.

### Article Information

DOI: 10.9734/BJMCS/2015/14305

#### Editor(s):

(1) Dariusz Jacek Jakóbczak, Computer Science and Management Department, Koszalin University of Technology, Poland.

#### Reviewers:

(1) Anonymous, University of Kashan, Iran.

(2) Anonymous, Universiti Teknologi Mara, Kota Kinabalu, Sabah, Malaysia.

(3) Dan Dan Ekezie, Department of Statistics, Imo State University, Owerri, Nigeria.

Complete Peer review History: <http://www.sciencedomain.org/review-history.php?iid=728&id=6&aid=6921>

Received: 26 September 2014

Accepted: 24 October 2014

Published: 14 November 2014

### Method Article

## Abstract

**Aims:** To provide an alternative definition for subset Sarima modeling and demonstrate it by application to monthly internally generated revenue of Ikot Ekpene Local Government Area of Akwa Ibom State of Nigeria.

**Study Design:** The study design is theoretical as well as empirical.

**Place and Duration of Study:** Department of Mathematics and Computer Science, Rivers State University of Science and Technology, Port Harcourt, Nigeria.

**Methodology:** Based on the duality relationship between autoregressive (AR) and moving average (MA) models an alternative definition to subset sarima models is proposed. The seasonality of the above-mentioned time series is established. A non-seasonal differencing of the seasonal difference of the series yields a stationary series which is analysed by Sarima methods.

**Results:** Applying the duality relationship between AR and MA models, subset Sarima models may be defined in AR terms rather than exclusively in MA terms as earlier done. An analysis of a 120-point internally generated revenue series from 1998 to 2007 yields the additive model from the original SARIMA (1, 1, 0)x(1, 1, 0)<sub>12</sub> model. The additive model is found to be adequate.

\*Corresponding author: [ettetuk@yahoo.com](mailto:ettetuk@yahoo.com), [ettehetuk@gmail.com](mailto:ettehetuk@gmail.com);

**Conclusion:** Based on the new and equivalent definition the monthly internally generated revenue of Ikot Ekpene Local Government Area of Nigeria follows an additive Sarima Model.

Keywords: Sarima model, subset sarima model, additive sarima model, multiplicative sarima model, internally generated revenue, Ikot Ekpene LGA, Nigeria.

## 1 Introduction

Autoregressive (AR) models and modeling were introduced by Yule [1]. Box and Jenkins [2] followed this up with the proposal of the autoregressive moving average models and modeling, of which seasonal autoregressive integrated moving average (SARIMA) models and modeling are a special case. They defined SARIMA models as multiplicative and applied it on the airways data. They did not just define it but also justified the multiplicative approach by considering a movement of the series up the time axis from one season to the next. They did not mention or define any other category of SARIMA model type. They however briefly acknowledged the existence of the “non-multiplicative” type. The acknowledgement of the existence of the multiplicative sarima model naturally raises the question of the additive type. Suhartono [3] defined subset, multiplicative and additive Sarima Models in terms of the moving average (MA) model. His work is classical as it pioneers the discussion and application of the categories of Sarima models. His approach of using MA symbolism naturally raises the question of the exclusivity or otherwise of the use of MA terminology. Does the use of the MA in the definition not naturally imply the propriety of the use of the AR as an alternative given their duality?

In this work it is proposed that AR models could still have been used as basis of the definitions granted the duality relationships between the AR and the MA models. In section 2, arguments along duality lines were advanced in order to establish this fact. In section 3, the monthly internally generated revenue of Ikot Ekpene Local Government Area of Nigeria was used as a case study.

Sarima models and modeling have been applied to model seasonal time series. It is apparent that their popularity is increasing in recent times. Authors who have used these models of recent are Etuk [4], Eni et al. [5], Ekezie et al. [6], Otu et al. [7], Paul et al. [8], Suhartono and Lee [9], Shiri et al. [10], Prista et al. [11], Oduro-Gyimah et al. [12], Fannoh et al. [13], to mention a few.

## 2 Materials and Methods/Experimental Details/Methodology

The data used are 120 monthly values of internally generated revenue from 1998 to 2007 of the Ikot Ekpene Local Government Area (LGA) of Akwa Ibom State of Nigeria. The realization was used by Udoudo [14]. The data are given to the nearest ten thousands of naira. The data are retrievable from the records of the Research, Planning and Statistics Unit of the LGA. See the Appendix for the data.

### 2.1 Sarima Models

A stationary time series  $\{X_t\}$  is said to follow an *autoregressive moving average model of order p and q* denoted by ARMA (p, q) if

$$X_t - \alpha_1 X_{t-1} - \alpha_2 X_{t-2} - \dots - \alpha_p X_{t-p} = \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2} + \dots + \beta_q \varepsilon_{t-q} \quad (1)$$

where the left hand side (LHS) of (1) is the AR component and the right hand side (RHS) is the MA component. The  $\alpha$ 's are the AR coefficients and the  $\beta$ 's their MA counterparts. They are constants such that the model is stationary as well as invertible. Suppose that model (1) is put as

$$A(L)X_t = B(L)\epsilon_t \tag{2}$$

where the AR operator  $A(L) = 1 - \alpha_1L - \alpha_2L^2 - \dots - \alpha_pL^p$  and the MA operator  $B(L) = 1 + \beta_1L + \beta_2L^2 + \dots + \beta_qL^q$  where  $L^k X_t = X_{t-k}$ . For stationarity  $A(L) = 0$  must have roots with moduli greater than one. For invertibility  $B(L) = 0$  must have such roots too.

Suppose a series is non-stationary. Box and Jenkins [2] proposed that such a series may be made stationary after a number of differencing. Suppose for  $\{X_t\}$  the minimum number of differencing necessary for stationarity is equal to  $d$ , if the  $d^{\text{th}}$  difference  $\{\nabla^d X_t\}$  follows an ARMA( $p, q$ ) the original series  $\{X_t\}$  is said to follow an *autoregressive integrated moving average model of order  $p, d$  and  $q$* , denoted by ARIMA( $p, d, q$ ). Here  $\nabla = 1 - L$ .

For a seasonal series of period  $s$ , Box and Jenkins [2] moreover proposed that it may be modelled by the multiplicative model.

$$A(L)\Phi(L^s)\nabla^d\nabla_s^D X_t = B(L)\Theta(L^s)\epsilon_t \tag{3}$$

where the LHS and the RHS of (3) are respectively the AR and the MA components and  $\Phi(L)$  and  $\Theta(L)$  are the respective seasonal operators whose coefficients are such that stationarity and invertibility of the entire model are guaranteed. Here  $\nabla_s = 1 - L^s$  and  $D$  is the degree of seasonal differencing. Model (3) is called a *multiplicative seasonal autoregressive integrated moving average model of order  $p, d, q, P, D, Q$  and  $s$*  and denoted by SARIMA( $p, d, q$ )x( $P, D, Q$ ) $_s$  model.

## 2.2 Duality Relationship between AR and MA models

The dual of the model (2) may be defined as the model.

$$B(L)X_t = A(L)\epsilon_t \tag{4}$$

(McLeod [15])

AR and MA models are well known to have duality relationships. These include the following facts:

- 1) A finite-order model of one type is equivalent to an infinite order model of the other type; (Box and Jenkins [2])
- 2) An AR model is always invertible whereas a MA model is always stationary. Stationarity and invertibility are twin and dual requirements ensuring a one-to-one correspondence between a particular autocorrelation structure and a model (Priestley [16]).
- 3) The autocorrelation function (ACF) of a model of one type is the inverse of the inverse autocorrelation (IACF) of the other type. The same applies between partial autocorrelation function (PACF) and the inverse partial autocorrelation function (Cleveland [17], Hipel et al. [18]).

- 4) The ACF of one model behaves as the PACF of the other. In particular, the PACF of an AR(p) cuts off at lag p; the ACF of an MA(q) model cuts off at lag q.
- 5) The spectrum of an AR is the inverse of its MA dual and vice versa, and so on. (Box and Jenkins [2])

The duality is such that the model types may be used interchangeably. For instance, the IACF of a stationary time series may be estimated by fitting an AR model of sufficiently long order and obtaining the IACF accordingly i.e. the inverse of the ACF of the dual MA model. A well-known approach to AR model fitting is by the Yule-Walker equations via the ACF. AR-MA duality has been employed to apply the same linear optimization approach to model MA via the IACF (Oyetunji [19]).

### **2.3 Suhartono[3]’s Proposal for Subset, Multiplicative and Additive Sarima Modelling**

Suhartono [3] proposed an algorithm for fitting a subset, multiplicative and an additive Sarima model thus:

Fit the model

$$X_t = \varepsilon_t + \beta_1\varepsilon_{t-1} + \beta_s\varepsilon_{t-s} + \beta_{s+1}\varepsilon_{t-s-1} \tag{5}$$

*If  $\beta_{s+1}$  is not statistically significant, the model is additive. Otherwise, Check whether  $\beta_{s+1} = \beta_1\beta_s$ . If so, the model is multiplicative. Otherwise, the model is subset.*

Etuk et al. [20,21] and Etuk [22] have applied this proposal to fit additive SARIMA models.

### **2.4 Sarima Model Fitting**

The general Sarima (p, d, q)x(P, D, Q)<sub>s</sub> model may be fitted by first of all determining the orders. The seasonal order s might be suggestive from a knowledge of the seasonal nature of the series or from the seasonal pattern of the time-plot or correlogram. Otherwise, a preliminary inspection of the series could indicate a seasonal pattern and the period s accordingly determined. The AR orders: non-seasonal p and seasonal P may be estimated by the non-seasonal and the seasonal cut-off points of the PACF respectively. Similarly, the MA orders q and Q may be estimated by the non-seasonal and the seasonal cut-off lags of the ACF respectively. The difference orders d and D often need not add up to more than 2 for stationarity to be achieved. At each point of the differencing process there will be need to test for stationarity. This shall be done by the Augmented Dickey Fuller (ADF) Test.

The model (5) is a SARIMA(0, 0, 1)x(0, 0, 1)<sub>s</sub> in {X<sub>t</sub>}. The only order is the period s of seasonality. The ACF of such a model is such that the autocorrelation at lag s is significant and the ones at lags s-1 and s+1 are equal. Therefore the ACF structure suggestive of the model (5) should be such that the value at lag s is significant and those at lags s-1 and s+1 are comparable; at least they should be of the same signs.

After order determination the model coefficient parameters may be estimated by a non-linear optimization approach like the maximum likelihood or the least squares approach. This is because

the presence in the model of items of a white noise process necessitates the adoption of such non-linear techniques. After model fitting the fitted model is subjected to diagnostic checking by residual analysis. Uncorrelatedness of the residuals and/or better still their normality confirms the adequacy of the model. In this write-up the econometric and statistical package Eviews shall be used for all the analytical work. It is based on the least squares criterion for the estimation process.

### **3 Results and Discussion**

#### **3.1 Proposed Alternative or Complementary Algorithm for Fitting Subset, Multiplicative and Additive Sarima Model**

Based on the above discussion of the AR-MA duality we hereby propose that, depending on the empirical autocorrelation structure an alternative or a complement to Suhartono [3]'s proposal is

Fit

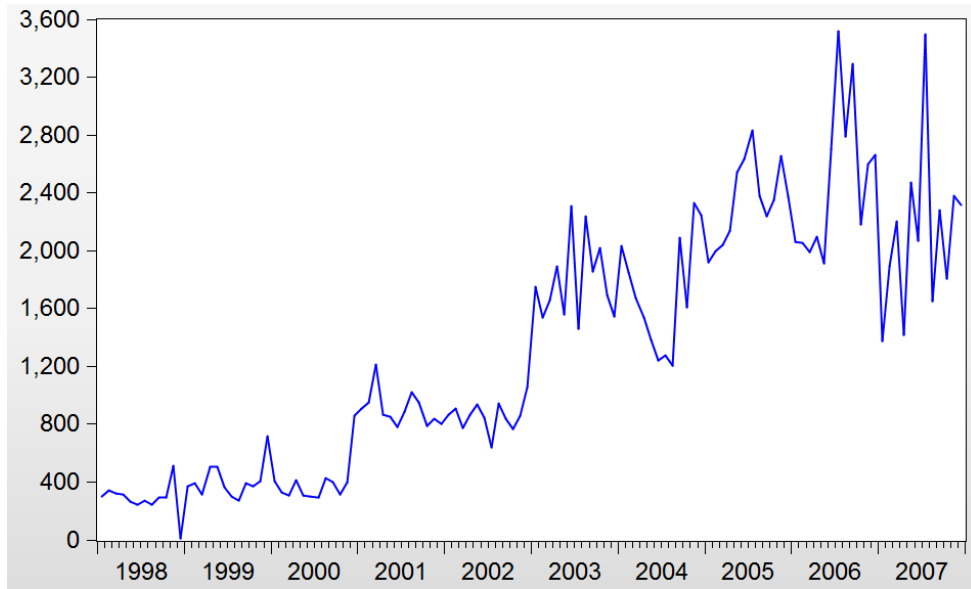
$$X_t = \varepsilon_t + \beta_1 X_{t-1} + \beta_s X_{t-s} + \beta_{s+1} X_{t-s-1} \quad (6)$$

If  $\beta_{s+1} = 0$  the model is an additive SARIMA model. If not, if  $\beta_{s+1} = \beta_s \beta_1$  then the model is multiplicative. Otherwise it is subset.

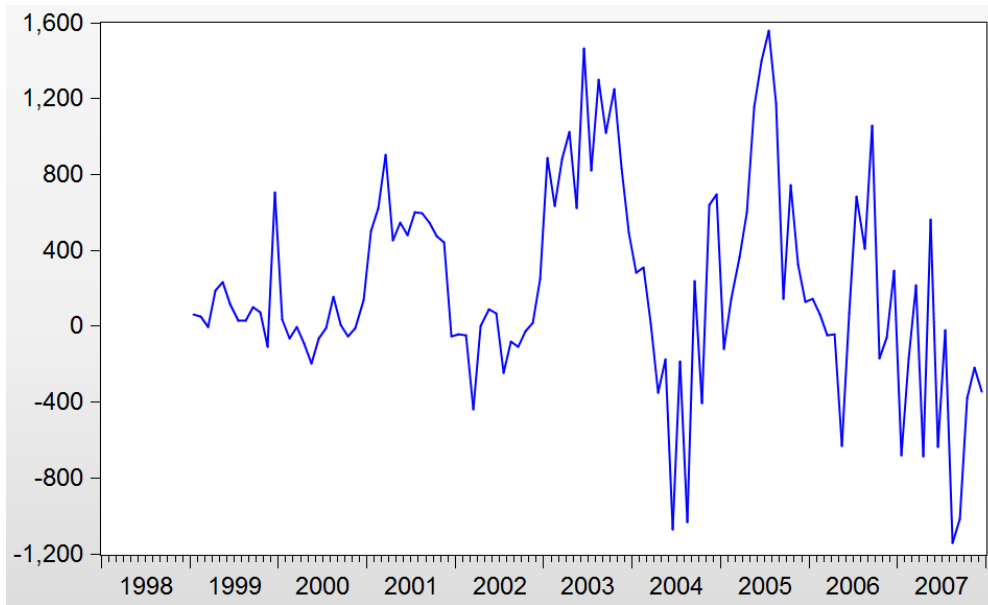
Model (6) shall be suggestive if the PACF is such that at lag  $s$  the partial autocorrelation is significant and the values at lags  $s-1$  and  $s+1$  are comparable.

#### **3.2 A Practical Example**

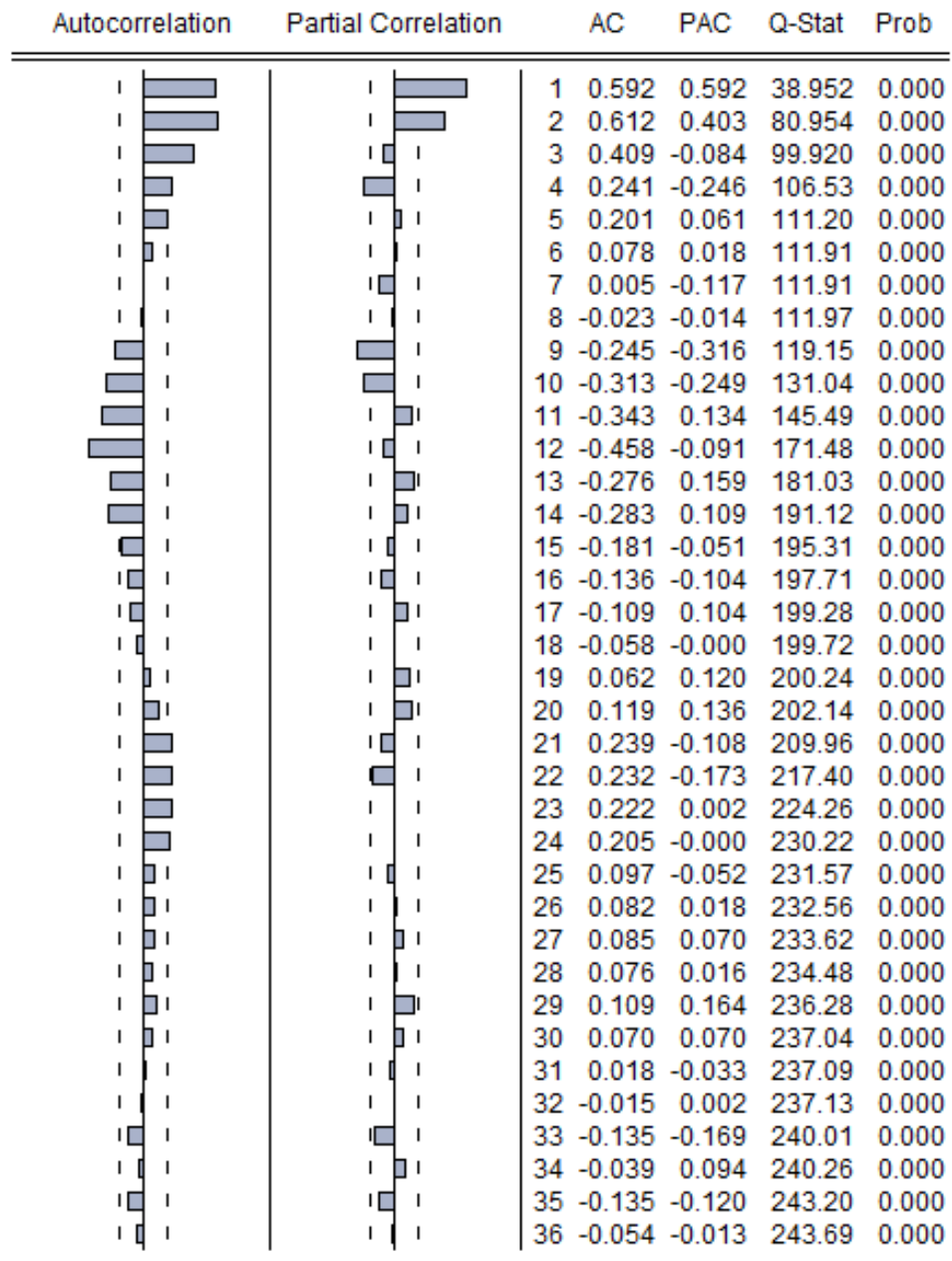
The realization of monthly internally generated revenue of Ikot Ekpene LGA covering 1998 to 2007 shall be herein called IIGR. The 120-point data has a time-plot that shows an upward secular trend and some seasonality rising in amplitude over time (Fig. 1). A preliminary examination of the data suggests the presence of a seasonal tendency of period 12 months; eight of the ten yearly minimums lie between May to October and seven of the ten maximums fall into the complementary interval between September to the next April. A 12-monthly differencing yields the series SDIIGR which has a generally horizontal trend (Fig. 2) and a correlogram which shows the presence of some seasonality of period 12 months (Fig. 3). A non-seasonal differencing of SDIIGR yields the series DSDIIGR which has a horizontal trend (Fig. 4) and a correlogram in (Fig. 5) which shows seasonality of period 12 months and the involvement of seasonal AR and MA components of order one each. (Table 1) shows that the ADF test adjudges both IIGR and SDIIGR as non-stationary but DSDIIGR as stationary.



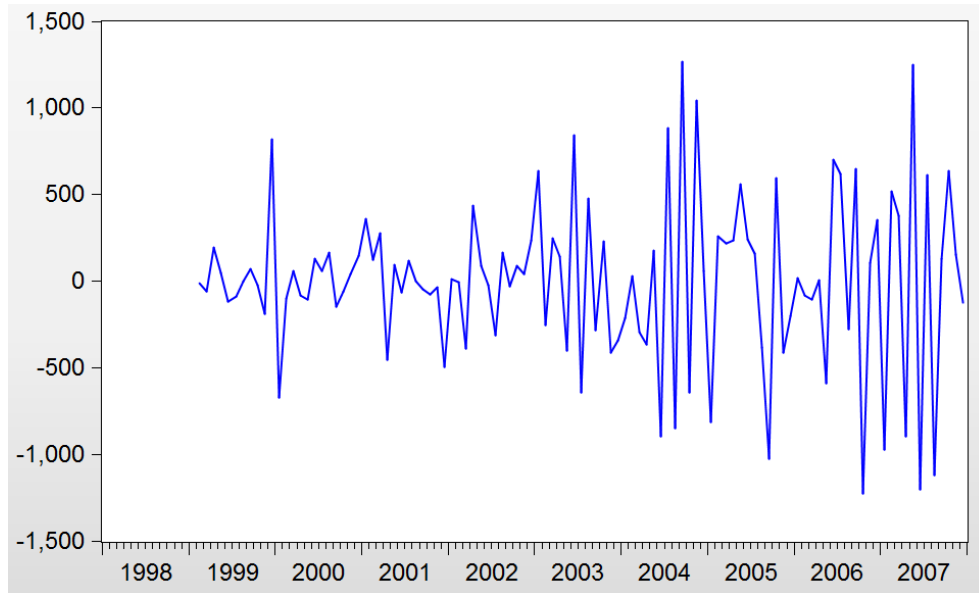
**Fig. 1. The time plot of IIGR**



**Fig. 2. The time plot of SDIIGR**



**Fig. 3. The correlogram of SDIIGR**



**Fig. 4. The time plot of DSIIGR**

The autocorrelation structure, inter alia, suggests the SARIMA(1, 1, 0)x(1, 1, 0)<sub>12</sub> model for IIGR. Applying this new algorithm, the model (6) yields the model estimated in (Table 2) as

$$X_t + .5493X_{t-1} + .5135X_{t-12} + .1899X_{t-13} = \varepsilon_t \quad (7)$$

$(\pm.0877) \quad (\pm.1037) \quad (\pm.1145)$

Clearly the lag 13 coefficient of model (7) is not statistically significant. Hence the adoption of the additive model estimated in (Table 3) as

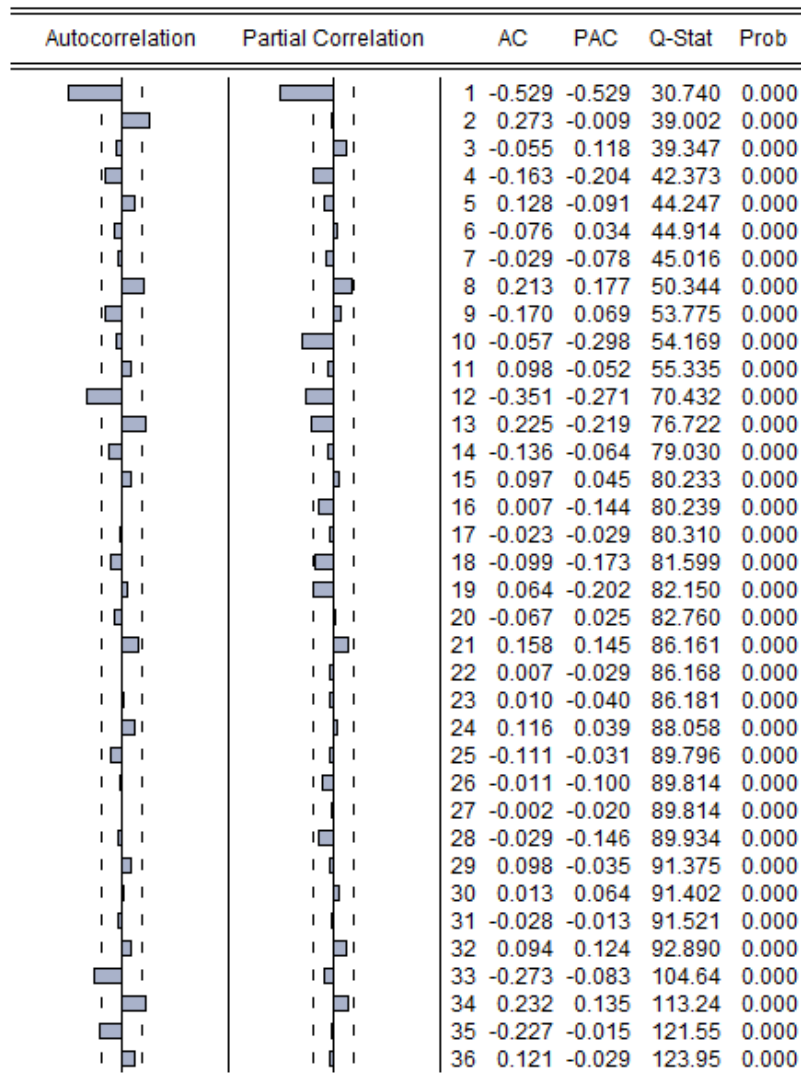
$$X_t + .4748X_{t-1} + .4356X_{t-12} = \varepsilon_t \quad (8)$$

$(\pm.0794) \quad (\pm.0931)$

where in model (7) and model (8),  $X = \text{DSDIIGR}$ . The model (8) might be preferred to model (7) on the following grounds

- 1) Parametric parsimony: The non-significance of the lag 13 coefficient of model (7) makes model (8) preferable as it has fewer parameters.
- 2) Schwarz criterion: The Schwarz criterion has a smaller value for model (8) than for model (7).
- 3) Residual uncorrelatedness shown in (Fig. 6): All the autocorrelations of the residuals of model (8) are non-significant.
- 4) Jarque-Bera Residual normality test of (Fig. 7). The null hypothesis of normality of the residuals is not rejected ( $p = 0.9474$ ).





**Fig. 5. The correlogram of DSDIIGR**

**Table 1. Augmented dickey fuller tests for non-stationarity**

Series	Test statistic	1% critical value	5% critical value	10% critical value	Conclusion
IIGR	-1.4314	-3.4866	-2.8861	-2.5799	Nonstationary
SDIIGR	-2.8707	-3.4931	-2.8889	-2.5815	Nonstationary*
DSDIIGR	-18.3637	-3.4931	-2.8889	-2.5815	Stationary

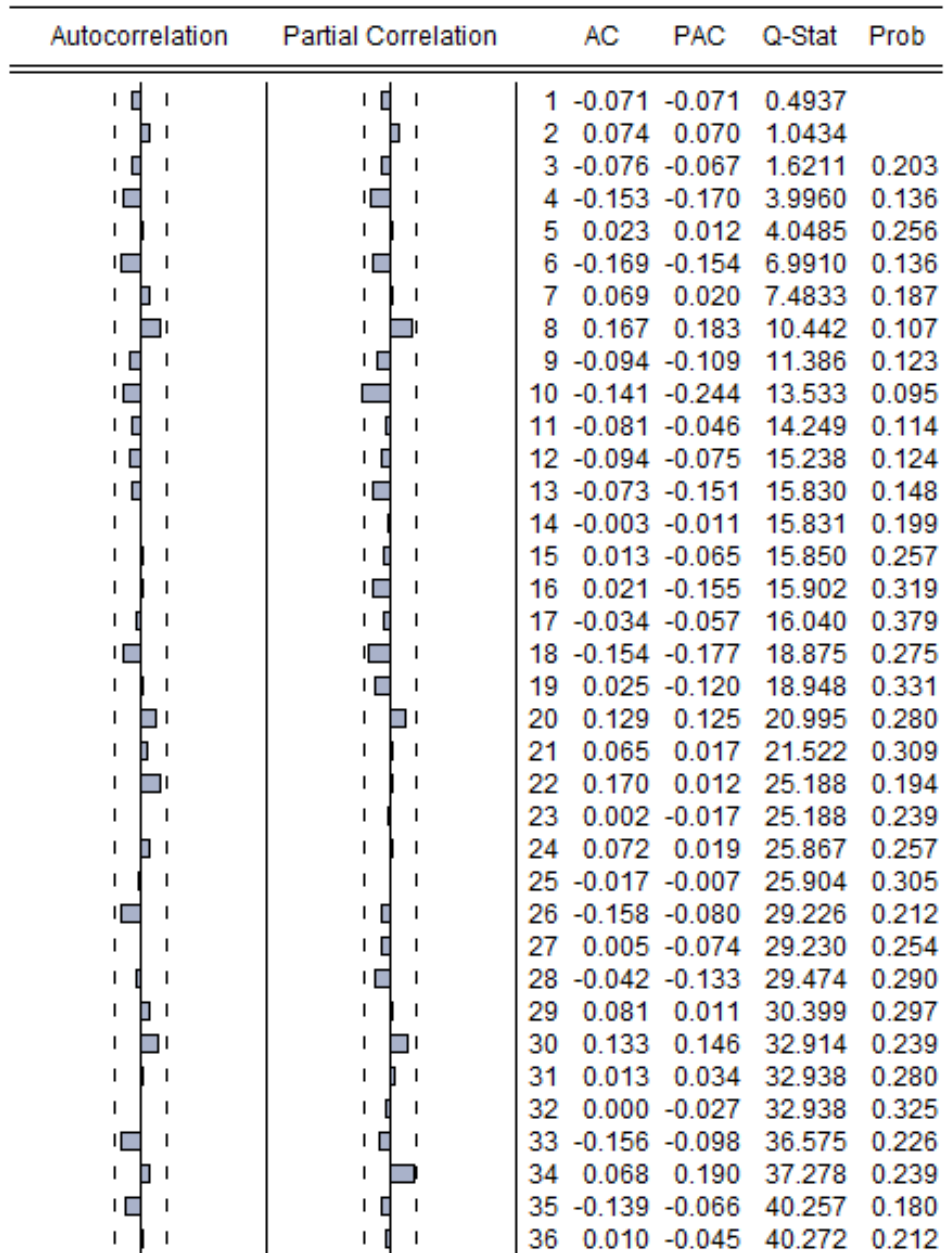
\*Test significant at 5% level but non-significant at 1% level.

**Table 2. Estimation of sarima (1, 1, 0)X(1, 1, 0)<sub>12</sub> Model**

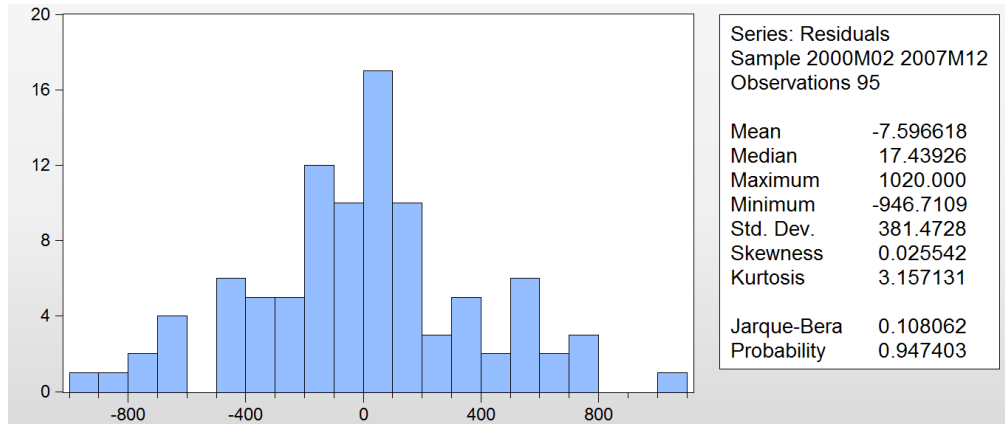
Dependent variable: DSDIIGR Method: Least Squares Date: 09/01/14 Time: 16:34 Sample (adjusted): 2000M03 2007M12 Included observations: 94 after adjustments Convergence achieved after 3 iterations				
Variable	Coefficient	Std. error	t-Statistic	Prob.
AR(1)	-0.549309	0.087658	-6.266482	0.0000
AR(12)	-0.513477	0.103721	-4.950534	0.0000
AR(13)	-0.189913	0.114462	-1.659181	0.1005
R <sup>2</sup>	0.435551	Mean dependent var		-2.957447
Adjusted R <sup>2</sup>	0.423146	S.D. dependent var		499.6014
SE of regression	379.4518	Akaike info criterion		14.74673
Sum squared resid	13102515	Schwarz criterion		14.82790
Log likelihood	-690.0962	Hannan-Quinn criter.		14.77951
Durbin-Watson stat	2.001126			
Inverted AR Roots	.90-.24i .23+.91i -.37, -.94+.24i	.90+.24i .23-.91i -.69-.66i	.66+.67i -.26+.91i -.69+.66i	.66-.67i -.26-.91i -.94-.24i

**Table 3. Estimation of the additive model (8)**

Dependent variable: DSDIIGR Method: Least Squares Date: 09/01/14 Time: 11:34 Sample (adjusted): 2000M02 2007M12 Included observations: 95 after adjustments Convergence achieved after 2 iterations				
Variable	Coefficient	Std. error	t-Statistic	Prob.
AR(1)	-0.410726	0.079372	-5.981830	0.0000
AR(12)	-0.435619	0.093097	-4.679201	0.0000
R <sup>2</sup>	0.410726	Mean dependent var		-4.000000
Adjusted R <sup>2</sup>	0.404390	S. D. dependent var		497.0407
SE of regression	383.5951	Akaike info criterion		14.75788
Sum squared resid	13684503	Schwarz criterion		14.81165
Log likelihood	-698.9993	Hannan-Quinn criter.		14.77961
Durbin-Watson stat	2.127339			
Inverted AR roots	.87+.24i .21+.89i -.71+.65i	.87-.24i .21-.89i -.71-.65i	.63+.66i -.28-.89i -.95+.24i	.63-.66i -.28+.89i -.95-.24i



**Fig. 6. The correlogram of the additive sarima residuals**



**Fig. 7. The histogram of the additive sarima residuals**

## 4 Conclusion

A new alternative or complementary algorithm for SARIMA fitting has been proposed. The monthly internally generated revenue of Ikot Ekpene LGA of Akwa Ibom State of Nigeria has been modeled as an additive SARIMA model (8). This means that a current value depends on its last value and its value of a year ago. Forecasts may be based on the fitted model.

## Authors' Contributions

Author EHE designed the study, performed the statistical analysis, wrote the protocol, and the first draft of the manuscript, managed the analyses of the study and managed the literature searches. Both authors read and approved the final manuscript.

## Competing Interests

Authors have declared that no competing interests exist.

## References

- [1] Yule G. On a method of investigating the periodicities in disturbed series, with special reference to Wolfer's sunspot numbers. *Phil. Trans. Roy. Soc. London.* 1927;(A):226-267.
- [2] Box GEP, Jenkins GM. *Time series analysis, forecasting and control.* San Francisco: Holden-Day; 1976.
- [3] Suhartono. Time series forecasting by using seasonal autoregressive integrated moving average: Subset, multiplicative or additive model. *Journal of Mathematics and Statistics.* 2011;7(1):20-27.

- [4] Etuk EH. Modelling of daily Nigerian Naira-British pound exchange rates using Sarima methods. *British Journal of Applied Science and Technology*. 2014;4(1):222-234.
- [5] Eni D, Adeyeye FJ, Duke SOO. Modeling and forecasting maximum temperature of Warri City – Nigeria. *Research Journal of Engineering and Applied Sciences*. 2013;2(5):370-375.
- [6] Ekezie, DD, Jude O, Idochi O. Modelling and forecasting malaria mortality rate using Sarima models (a case study of Aboh Mbaise general hospital, Imo State Nigeria). *Science Journal of Applied Mathematics and Statistics*. 2014;2(1):31-41.
- [7] Otu AO, Osuji GA, Opara J, Mbachu HI, Iheagwara AI. Application of Sarima models in modelling and forecasting Nigeria’s inflation rates. *American Journal of Applied Mathematics and Statistics*. 2014;2(1):16-28.
- [8] Paul RK, Panwar S, Sarkar SK, Kumar A, Singh KN, Farooqi S, Choudhary VK. Modelling and forecasting of meat exports from India. *Agricultural Economics Research Review*. 2013;26(2):249-255.
- [9] Suhartono Lee MH. Forecasting of tourist arrivals using subset, multiplicative or additive seasonal Arima Model. *Matematika*. 2011;27(2):169-182.
- [10] Shiri G, Salahi B, Samdzadeh R, Shiri M. The investigation and forecasting of relative humidity variation of pars Abad-e-Moghan, North West of Iran by Arima model. *Research Journal of Applied Sciences*. 2011;6(2):81-87.
- [11] Prista N, Diawara N, Costa MJ, Jones C. Use of Sarima models to assess data-poor fisheries: A case study with a sciaenid fishery off Portugal. *Fishery Bulletin*. 2011;109(2):170-185.
- [12] Oduro-Gyimah FK, Harris E, Darkwah KF. Sarima time series model application to microwave transmission of Yeji-Salaga (Ghana) line-of-sight link. *International Journal of Applied Science and Technology*. 2012;2(9):40-51.
- [13] Fannoh R, Orwa GO, Mung’atu JK. Modeling the inflation rates in Liberia Sarima approach. *International Journal of Science and Research*. 2014;3(6):1360-1367.
- [14] Udoudo UP. Time series analysis of revenue data of Ikot Ekpene L. G. A. from 1998 – 2008. An unpublished thesis submitted to department of mathematics/computer science, Rivers State University of Science and Technology, Port Harcourt, Nigeria in partial fulfillment of the requirements for the award of Masters of Science (M. Sc.) Degree in Applied Statistics.
- [15] McLeod AI. Duality and other properties of multiplicative seasonal autoregressive-moving average models. *Biometrika*. 1984;71(1):207-211.
- [16] Priestley MB. *Spectral analysis and time series*. London: Academic Press; 1981.
- [17] Cleveland WS. The inverse autocorrelations of a time series and their applications. *Technometrics*. 1972;14:277-297.

- [18] Hipel KW, McLeod AI, Lennox WC. Advances in Box-Jenkins modelling 1. *Water Resources Research*. 1977;13:567-575.
- [19] Oyetunji OB. Inverse autocorrelations and moving average time series modelling. *Journal of Official Statistics*. 1985;1:315-322.
- [20] Etuk EH, Wokoma DSA, Moffat IU. Additive Sarima modelling of monthly Nigerian Naira-CFA franc exchange rates. *European Journal of Statistics and Probability*. 2013;1(1):1-12.
- [21] Etuk EH, Aboko IS, Victor-Edema U, Dimkpa MY. An additive seasonal Box-Jenkins model for Nigerian monthly savings deposit rates. *Issues in Business Management and Economics*. 2014;2(3):54-59.
- [22] Etuk EH. An additive Sarima model for daily exchange rates of the Malaysian Ringgit (MYR) and Nigerian Naira (NGN). *International Journal of Empirical Finance*. 2014;2(4):193-201.

## APPENDIX

<b>Year</b>	<b>Monthly revenue data (N'0000)</b>											
1998	304	343	319	315	268	246	269	243	294	293	516	009
1999	369	394	313	504	503	366	300	273	394	369	406	717
2000	406	329	309	415	306	300	291	428	401	318	399	859
2001	906	953	1212	867	851	779	890	1025	949	792	839	805
2002	864	908	776	867	939	846	642	942	839	769	856	1056
2003	1752	1540	1658	1891	1560	2309	1461	2241	1856	2018	1691	1548
2004	2034	1850	1673	1539	1384	1239	1278	1208	2092	1611	2329	2243
2005	1917	1995	2037	2140	2543	2637	2836	2381	2237	2354	2657	2370
2006	2061	2055	1988	2099	1910	2706	3522	2791	3294	2185	2597	2661
2007	1379	1889	2202	1415	2474	2070	3500	1651	2281	1810	2378	2318

© 2015 Etuk and Ojekudo; This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Peer-review history:**

The peer review history for this paper can be accessed here (Please copy paste the total link in your browser address bar)

[www.sciencedomain.org/review-history.php?iid=728&id=6&aid=6921](http://www.sciencedomain.org/review-history.php?iid=728&id=6&aid=6921)