**State Space and Box-Jenkins Approaches: A Comparison of Models Forecasting Performance in Finance.**

\*1. O. D. Adubisi, 2. I. J. David 3. U. E. Awa

1, 2 Department of Mathematics and Statistics, Federal University Wukari, Taraba State, Nigeria

3. Department of Statistics, University of Port-Harcourt, River State, Nigeria

\* E-mail of corresponding author: adubisiobinna@fuwukari.edu.ng

**Abstract**

The goal of this study is to investigate the forecasting performance of two time series analysis methods. This paper describes a study that used data collected from the Central Bank statistical web database system in Nigeria to evaluate and compare the forecasting performance of the Box-Jenkins (ARIMA) model and State space model at different historic data periods. The comparison uses data series on inflation rates in Nigeria from January 2003 to December 2017. The residuals for the models showed that they were adequate techniques for use in inflationary analysis. The one year forecast evaluation results for different historic periods indicated that forecasts from both the seasonal ARIMA model and state space model are comparable. However, the state space model captures the dynamic structure of the inflationary series reasonably and requires no new cycle of identification and model estimation given the availability of new data.

**Keywords:** State Space Model, Filtering, Smoothing, Inflation rates, ARIMA.

1. **Introduction**

The most important area in financial planning research is inflationary analysis. This is due to the fact that sustainable economic growth in relation to low consumer price index is the major aim of macroeconomic policy makers. Inflation as measured by the consumer price index reflects the annual percentage change in the cost of acquiring goods and services over a specified intervals. The consumer price index (CPI) approach though it is the least efficient is used to measure inflation on monthly, quarterly and annually basis in Nigeria according to CBN (2015). Hence, in order to achieve and maintain price stability, the evolutionary dynamics and time dependent structure of the inflationary series must be studied using an appropriate stochastic modelling approach. The use of autoregressive integrated moving average (ARIMA) models introduced by Box and Jenkins (1976) has extensively been applied in inflationary modelling over the years: Etuk et al. (2012), Otu et al. (2014), Adebiyi et al. (2010) while authors like, Omekara et al. (2013) explored the application of the periodogram and Fourier series analysis in modelling and forecasting the monthly inflation rates in Nigeria. Salau (1998) noted that there is a growing concern among researchers that the ARIMA model may not always lead to an accurate forecast especially in health, finance, business and molecular biology where it is known that the level of randomness is high and parameters significance cannot be adequately guaranteed.

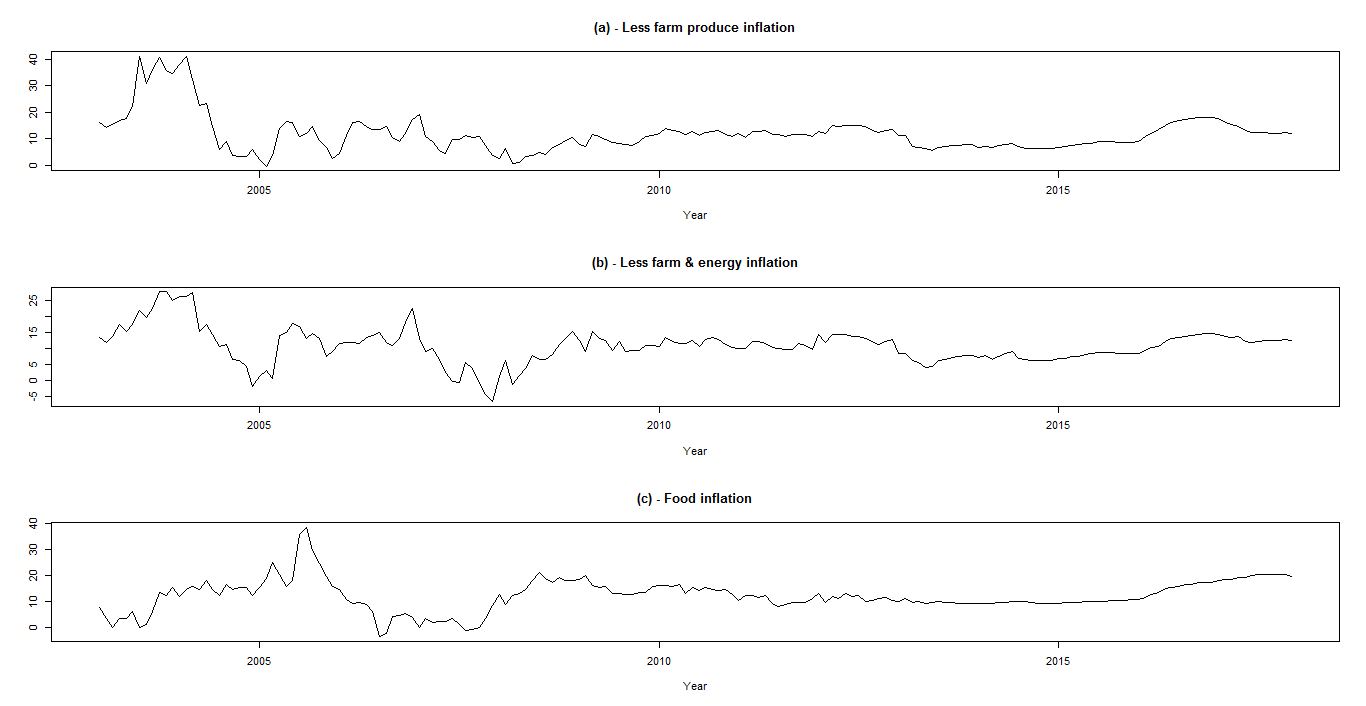
However, the state space method provides the key to exploring the different components of any time series as a structural model in time series analysis. This modelling approach allows a natural interpretation of time series as the combination of several components such as trend, cyclical, seasonal or regressive components. Hence, the method is used to model series in the presence of non-stationarity, structural changes, missing variables and irregular patterns unlike their counterpart, the Box-Jenkins method which requires at least a preliminary transformation of the series to get stationarity. The problems of estimation and forecasting in state space models are solved by recursively computing the conditional distribution of the quantities of interest, given the available information with the Kalman filter. Although, the literature on the dynamics of the state space models compared to the ARIMA models has increased over the years. There are still less empirical evidence on its performance relative to the Box-Jenkins approach in financial/economic applications.

Our aim is to develop a state space model for predicting inflation rates in Nigeria and also, to compare the predictive performance of the state space model with the ARIMA model for several forecasted periods.

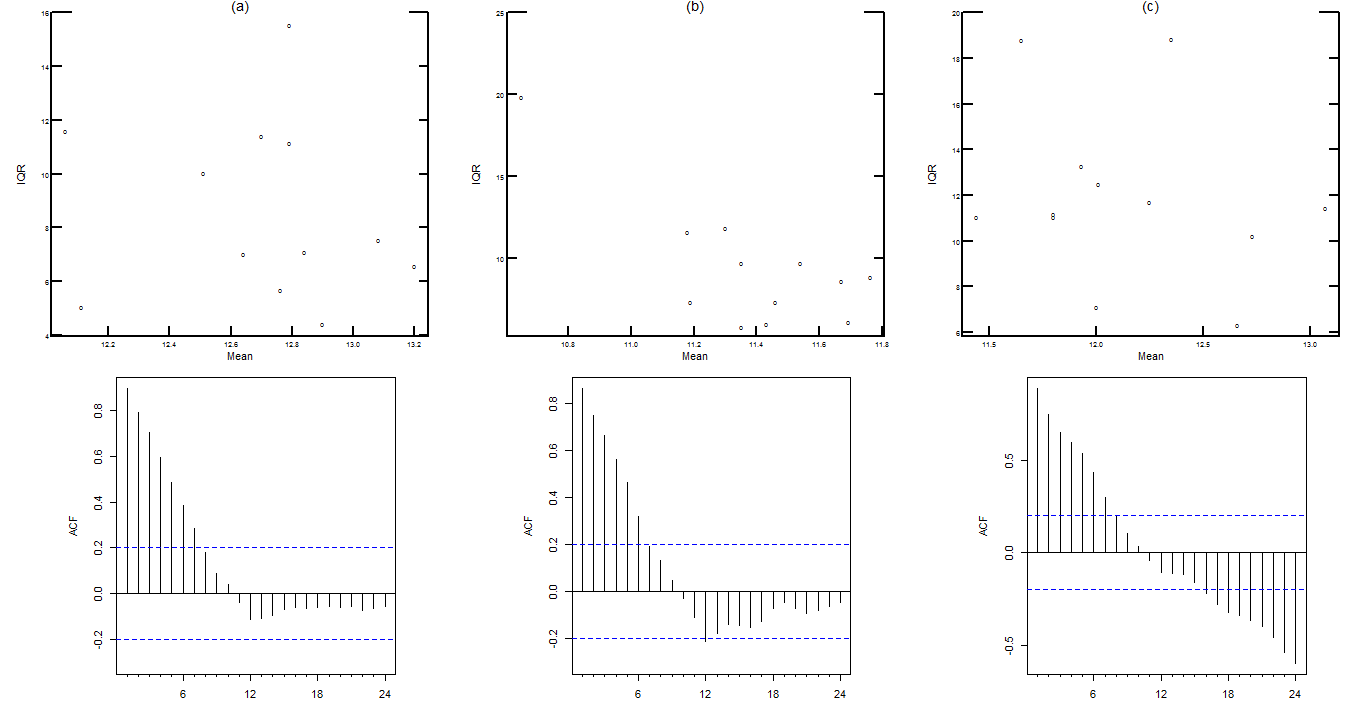
1. **Methods**

**2.1 Data source and descriptive analysis**

The monthly (core & non-core) inflation rates for Nigeria were used for the comparison of the state space model to the ARIMA model. Data were obtained from the Central Bank of Nigeria web database. The SAS Econometric Time Series (SAS/ETS) Software is used for modelling and forecasting the data series. All graphs and plots were generated using R-project statistical software. The time series plots in Fig. 1 depict the inflationary series for the period January 2003 to December 2017. These time plots show that the inflationary series are non-stationary. Additional, the plots display a level series with no marked significant seasonality in the data series. The main aim of this paper is to assess the predictive performance of the models for several forecasted periods. The period of January 2003 to December 2017 was chosen and divided into seven periods of twelve months corresponding to the fifteen years of complete inflationary series.



**Fig. 1:** Monthly Inflation rates for (a) less farm produce, (b) less farm & energy and (c) Food for Nigeria.



**Fig. 2:** Mean IQR plots and autocorrelation function (ACF) for the period 01/2003 - 12/2010.

(a) Less farm produce inflation mean IQR & ACF, (b) Less farm and energy inflation mean IQR & ACF and (c) Food inflation mean IQR & ACF.

**2.2 Modelling Techniques**

**2.2.1 Box-Jenkins Approach**

The ARMA model generalizes at once the autoregressive and moving average models as a mixed model introducing the advantage of flexibility in its applications to better approximate real stationary series. However, it is apparent that for most real time series, the stationarity hypothesis is not appropriate. The analysis of nonstationary time series with ARMA models which requires at least a preliminary transformation of the data to get stationarity leading to the development of the autoregressive integrated moving average (ARIMA) model was addressed by Box and Jenkins (1976) and Box et al. (1994). Stationarity in the series, according to Box et al (1994) is normally achieved by the removal of the trend and seasonal influence through differencing of the series. The first-order difference of the series is given by  or in terms of the Lag operator  as and so the *dth*order differencing is expressed as . The seasonal difference of the series with *Dth* order differencing is expressed in Lag operator form as , where is the data seasonal period.

Suppose the process  is a stationary observed inflationary series, then the ARMA models for forecasting inflation rates is expressed as

  (1)

Where,  is the autoregressive operator, is the moving-average operator of the model and  is a white noise. The Box-Jenkins approach is an iterative building process to identify the best model though model identification, model parameters estimation and residuals diagnostic checking. During the model identification stage, the estimated autocorrelation function (ACF) and partial autocorrelation function (PACF) plots are used to check for series stationarity and tentative autoregressive moving average (ARMA) orders in other to determine the model structure while the model parameter estimation stage involves estimation of the parameters after identifying the candidate model. The diagnostic stage examines the residuals to determine the adequacy of the fitted model. The model is used for forecasting if deemed adequate otherwise the whole model building process is restarted from the beginning until an adequate model is found.

**2.2.2 State Space Approach**

The structural time series model as noted by Harvey (1989) is based on certain probabilistic assumptions which attempt to capture the essential characteristics of the data generating process. The basic structural model for representing an observed time series  is

 (2)

Where, is the stochastic trend,  is the periodic seasonal component and  is the observational disturbance. Using the concept of random walk, the linear trend model may be obtained recursively from

 (3)

Where  and  are mutually uncorrelated random disturbances and is the slope of the trend. The seasonal effect  can be modelled by adding a seasonal component to the structural model. When the seasonal pattern is constant over time, the seasonal effect denoted by  is

, so that  (4)

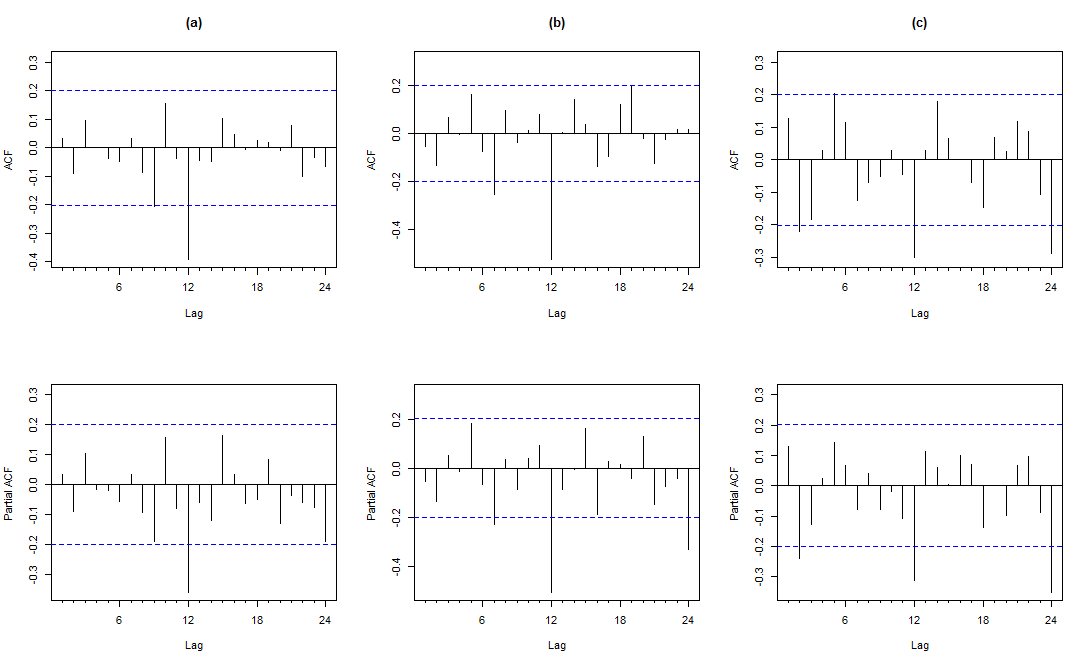
The introduction of a disturbance term  into the right hand side of (4), allows the seasonal pattern to change over time, giving the dummy seasonal specification.

 (5)

Once the components basic structures are defined, they can be assembled into an aggregate form, expressed as a state space model (SSM). A linear Gaussian state space model for forecasting the inflationary series is generally expressed in this form

 (6)

Where  is an unobserved  vector of system parameters at time, defined as , the matricesare initially assumed known, is a Gaussian white noise, The initial state vector is a random vector distributed as , independent of . The first equation in (6) is the state or evolution equation describing the process of variation of the system parameters along with time, including their time-varying dependence while the second is the measurement equation with a linear regression structure where the coefficient vector varies with time. The relationship between the state vectors  and the observed series  is specified by the state space model. For more details on the mathematical theory associated with state space models, see Harvey (1989) and Durbin and Koopman (2001), Commandeur and Koopman (2007).



**Fig. 3:** Correlogram plots after first differencing for the period 01/2003 - 12/2010: (a) less farm produce inflation (b) less farm and energy inflation & (c) food inflation.

**2.3 Model validation and forecasts comparison**

The goodness-of fit of the models were evaluated using the Akaike information criterion (AIC) and Bayesian information criterion (BIC), which gives a fair comparison between models with different number of parameters. The model with the lowest AIC or BIC value is normally preferred. The forecast periods were evaluated by computing the various measures of accuracy for each of the different type of inflationary series at different forecasted periods. The measures of accuracy used were the mean absolute error (MAE), mean absolute percentage error (MAPE) and root mean square percentage error (RMSPE).

**Table 1**

Forecasting performance for the three inflationary series. \*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Forecasting period | Accuracy parameters for each interval model | | | | | |
| MAE | | MAPE | | RMSPE | |
| SSM | SARIMA | SSM | SARIMA | SSM | SARIMA |
| Less farm produce inflation | | | | | | |
| 2011 | **0.776** | 0.936 | **6.353** | 7.855 | **8.131** | 10.226 |
| 2012 | **2.655** | 4.058 | **18.593** | 28.680 | **19.879** | 29.835 |
| 2013 | 6.494 | **2.628** | 91.946 | **37.288** | 99.336 | **43.447** |
| 2014 | **1.273** | 2.948 | **19.431** | 44.675 | **21.937** | 53.607 |
| 2015 | 1.836 | **1.479** | 21.719 | **17.343** | 22.871 | **19.785** |
| 2016 | 6.289 | **4.868** | 38.317 | **30.121** | 41.214 | **31.630** |
| 2017 | 4.806 | **0.522** | 37.880 | **3.877** | 41.718 | **4.815** |
|  |  |  |  |  |  |  |
| Less farm & energy inflation |  |  |  |  |  |  |
| 2011 | **0.857** | 0.925 | **8.068** | 8.383 | **8.951** | 9.326 |
| 2012 | **2.734** | 3.971 | **20.120** | 29.521 | **21.256** | 30.580 |
| 2013 | 6.719 | **3.495** | 111.152 | **57.823** | 125.227 | **66.362** |
| 2014 | **1.262** | 4.010 | **18.868** | 57.608 | **21.957** | 60.175 |
| 2015 | 1.592 | **1.346** | 19.113 | **16.792** | 20.047 | **18.032** |
| 2016 | **3.827** | 5.124 | **28.298** | 38.417 | **31.207** | 41.289 |
| 2017 | **2.038** | 2.072 | **16.101** | 16.322 | **17.679** | 19.029 |
|  |  |  |  |  |  |  |
| Food inflation |  |  |  |  |  |  |
| 2011 | 2.193 | **1.361** | 0.232 | **0.136** | 0.274 | **0.167** |
| 2012 | 1.188 | 1.124 | *0.101* | *0.101* | 0.131 | 0.129 |
| 2013 | **0.829** | 2.315 | **0.086** | 0.242 | **0.101** | 0.294 |
| 2014 | **0.639** | 2.951 | **0.065** | 0.309 | **0.107** | 0.363 |
| 2015 | **0.979** | 3.105 | **0.097** | 0.308 | **0.124** | 0.350 |
| 2016 | 4.566 | **1.376** | 0.286 | **0.086** | 0.318 | **0.099** |
| 2017 | **2.333** | 5.988 | **0.117** | 0.302 | **0.129** | 0.323 |

\*Boldface means the approach ranks first; italic show similar behaviour.

**MAE:** Mean absolute error.

**MAPE:** Mean absolute percentage error.

**RMSPE:** Root mean square percentage error.

**SARIMA:** Seasonal autoregressive integrated moving average.

**SSM:** State space method.

Using the historic data, we generated one-year-ahead forecast data for each inflationary historic period. The resultant residuals obtained by subtracting the forecast data from the historic holding data were used to compute the values for MAE, MAPE and RMSPE. The SARIMA model was used as the benchmark for the comparison due to its reliable forecasting power in inflationary analysis over the years.

1. **Results and discussion**

**3.1 Descriptive results**

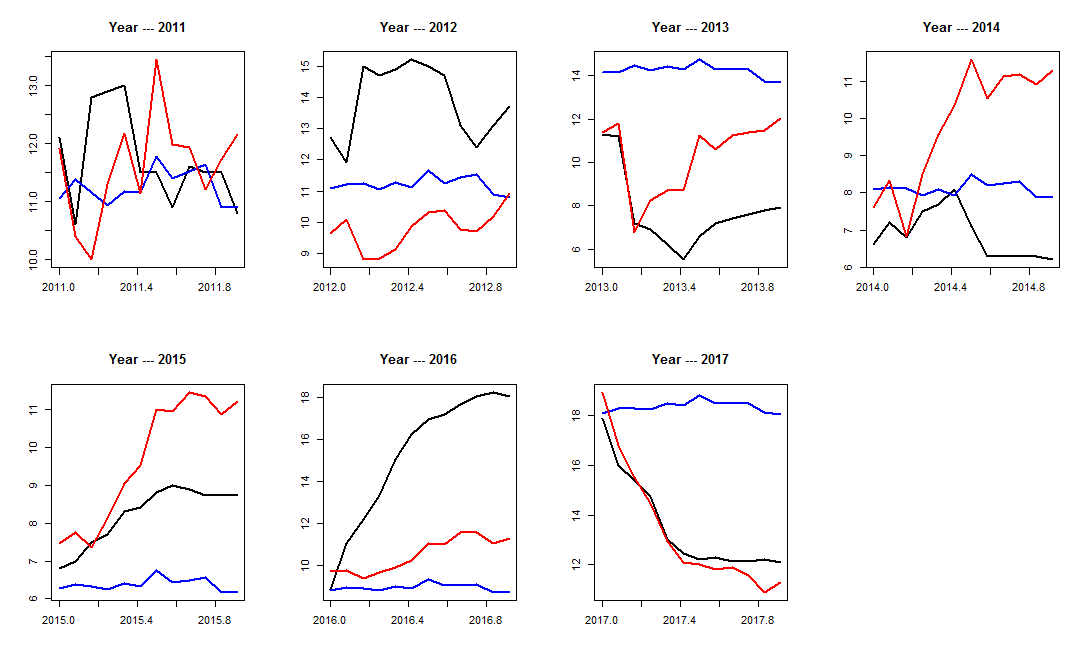
The data series for the various inflation rates were analysed for possible identification of non-homogeneous variance using the mean interquartile range (IQR) scatter plot. The interquartile range is plotted against the mean for each seasonal period. A random scatter about a straight line in the scatter plot would suggest a logarithmic transformation. The scatter plots for the inflationary series in Figs. 2(a)-2(c) depicts an independent relationship between the seasonal period IQR and mean, suggesting variance homogeneity in the series, so no transformation is needed. For each modelling time interval (01/2003 to 12/20nn, nn = 10, 11… 17), the series were evaluated and the ACF used to identify the degree of trend and seasonality differencing required.

The estimated ACF for the different types of inflationary series in the first interval (01/2003 – 12/2010) are depicted in Figure 2(a)-2(c), the presence of non-stationary mean is clearly visible in the ACF plots as observed from the time series plots. Once stationarity was achieved by first-order differencing, the behaviours of the estimated ACF and the PACF of the stationary series in Figure 3(a)-(c) were used to identify possible tentative models and the parameters were estimated using maximum likelihood method. The same iterative process was carried-out for other intervals for the three types of inflationary series. The examination of the estimated residuals, ACF and PACF and the normality plots from each interval model did not show the presence of stochastic structure and were approximately normally distributed for the different inflationary series.

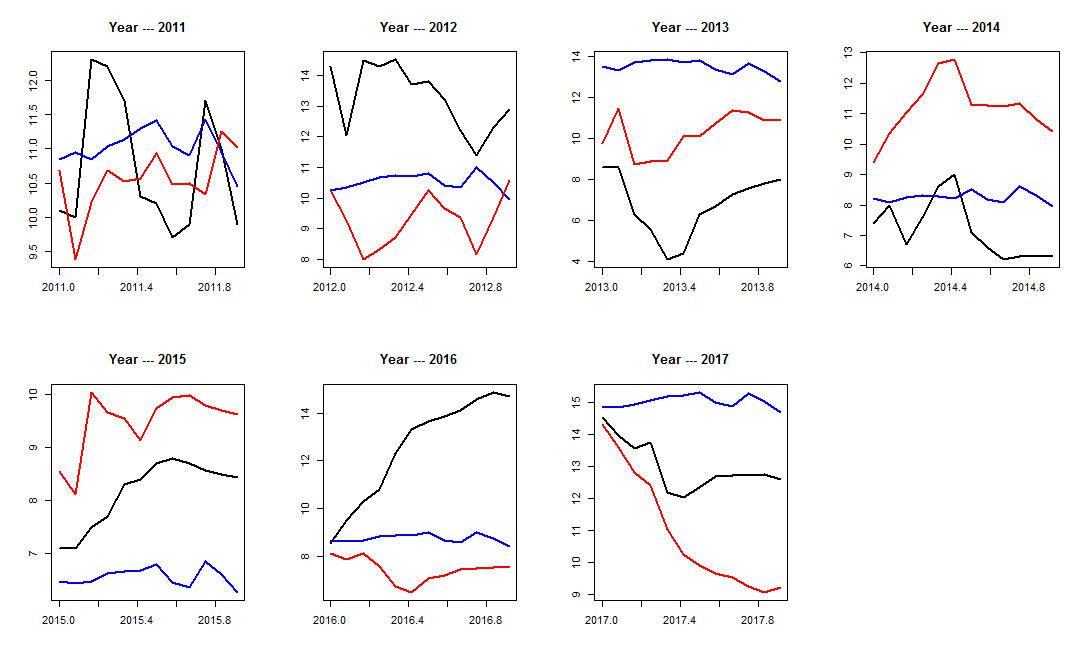
**3.2 Modelling and forecasting results**

The results from the post-sample forecast and the hold-up data for model validation for the two forecasting approaches for three inflationary series at different modelling time periods are presented in Table 1. Some tentative conclusion is drawn from the forecast performance table which shows that no approach dominates the other. Also, given that the MAE and RMSPE depend on the mean value of the series, which hinders comparisons of models for predicting series that have different mean values, MAPE would be understandable for examining the model forecast accuracy.

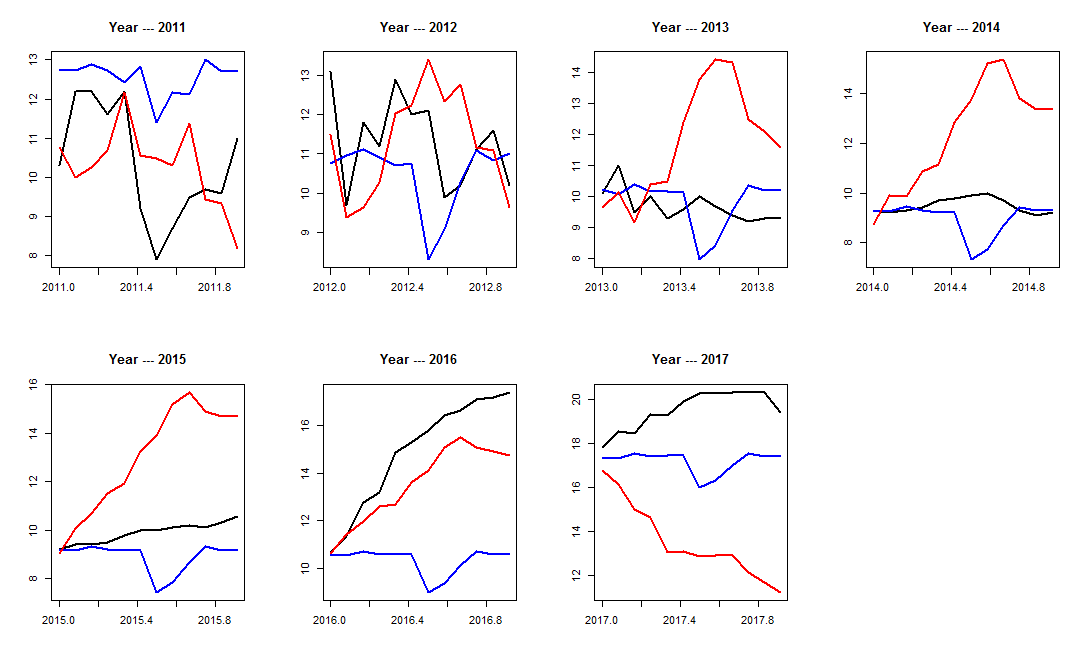
The SSM has smaller forecasting errors for three forecasted years of less farm produce inflation (2011, 2012 and 2014), five forecasted years for less farm and energy inflation (2011, 2012, 2014, 2016 and 2017) and four forecasted years for food inflation (2013, 2014, 2015 and 2017) suggesting a better performance in these time periods. The SARIMA model has smaller errors for four forecasted years of less farm inflation (2013, 2015, 2016 and 2017), two smaller errors for less farm and energy inflation (2013 and 2015) and two forecasted years for food inflation (2011 and 2016) showing better performance in the time periods. Additional, for one forecasted period of food inflation (2012), the results are inconclusive given the same error provided by the MAPE, though the MAE and RMSPE states that the SARIMA performed better. The graphical representations of the historic and forecast data for each inflationary series at different time period are presented in Fig. 3, Fig. 4 and Fig. 5. The plots provided comparisons of the predicted inflationary series with the historic data for 12 months at different time periods of the series.



**Fig. 3:** Forecast comparison for less farm produce inflation (Historic data ----- SSM forecast ---- SARIMA forecast -----).



**Fig. 4:** Forecast comparison for less farm & energy inflation (Historic data ----- SSM forecast ---- SARIMA forecast -----).



**Fig. 5:** Forecast comparison for food inflation (Historic data ----- SSM forecast ---- SARIMA forecast -----).

1. **Conclusion**

In this paper, we developed both the state space model and seasonal ARIMA model to forecast different types of inflationary series at different time periods. The ARIMA approach is based on stationarity of the series whilst the state space is not. We examined the forecast performance by comparing a year ahead forecast with one year holding period at different time periods of the data. The models were able to forecast the direction and magnitude of the inflationary series with reasonable monthly forecast errors. Based on the three forecast error measures and the plots of the forecasted values, no one approach dominates the other in terms of inflationary series analysis. However, the state space approach performed much better in most of the time periods for less farm and energy inflation and food inflationary series unlike in the less farm produce inflation.

**References**

Adebiyi, M. A., Adenuga, A. O., Abeng, M. O., Omanukwe, P. N., and Ononugbo, M. C. (2010). Inflation forecasting models for Nigeria. Central Bank of Nigeria Occasional Paper No. 36, Abuja, *Research and Statistics Department*.

Box G. E. P., Jenkins G. M., and Reinsel G. C. (1994). Time series Analysis Forecasting and Control, 3rd Edition, Prentice Hall, Englewood Cliffs, New Jersey-USA.

Box G. E. P. and Jenkins G. M. (1976). Time series Analysis Forecasting and Control, Revised Edition, Holden-Day. ISBN: 0816211043, San Francisco-USA.

CBN (2015). Financial Policies by the Apex bank of Nigeria. Central Bank of Nigeria publications. Assessed on 30/01/2018.

Commandeur J. J. F and Koopman S. J. (2007). An introduction to state space time series analysis, 1st edition, Oxford University Press Inc, Oxford, UK. ISBN 978-0-19-922887-4.

Durbin J. and Koopman S. J. (2001). Time series analysis by state space methods, 1st edition, Oxford University Press Inc, Oxford, UK. ISBN 0-19-852354-8.

Etuk E. H., Uchendu B., and Victoredema U. A. (2012). Forecasting Nigeria Inflation Rates by a Seasonal ARIMA Model. *Canadian Journal of Pure and Applied Sciences*, 6 (3), 2179-2185.

Harvey A. C. (1989). Forecasting structural time series models and the Kalman filter, 1st edition, CambridgeUniversity Press, Edinburge Building, Cambridge CB2 8RU, UK. ISBN 978-0-521-40573-7.

Omekara, C. O., Ekpenyong, E. J. and Ekerete, M. P. (2013). Modeling the Nigeria Inflation Rates using Periodogram and Fourier Series Analysis. *CBN Journal of Applied Statistics*, 4 (2): 51-68.

Otu, A. O., Osuji, G. A., Opara, J., Mbachu, H. I. and Iheagwara, A. I. (2014). Application of SARIMA models in modeling and forecasting Nigeria’s Inflation Rates. *American Journal of Applied Mathematics and Statistics*, 2 (1): 16-28.

Salau M. O. (1998). Arima modelling of Nigeria crude oil export, AMSE, modelling, measurement & control. 8(1): 1 – 20.