

Modelling Monthly Headline Consumer Price Index (HCPI) through Seasonal Box-Jenkins Methodology

Emerson A. Jackson¹✍, Abdulai Sillah¹, Edmond Tamuke¹

¹Model Building and Analysis Section, Research Department, Bank of Sierra Leone

DISCLAIMER STATEMENT: The views expressed in this work are those of the authors and do not reflect that of the Bank of Sierra Leone.

Abstract: In this empirical work, cognisance has been given to providing a review of literature on the seasonal Box-Jenkins modelling, particularly with reference to a univariate model. Seasonal pattern of Headline Consumer Price Index (HCPI) has been produced for Sierra Leone and with EVIEWS making use of best model selection of (6,0)(0,0). Data were seasonally adjusted with iteration and sufficient diagnostic test outcomes showing that forecast using Static method yielded best outcome, with Year-on-Year inflation over the three monthly period forecasted outcomes. The correlogram of the resultant series revealed very stable outcome of the results, while MAPE for the forecast evaluation revealing marginal error for the outcome, indicating that the model is quite adequate with the chosen methodology.

Keywords: ARIMA Model, HCPI, Time Series, Sierra Leone

1. Introduction

In Sierra Leone, monitoring HCPI is very key in terms of addressing the central bank's objective in maintaining price stability. There are various methodologies used by professionals and institutions across the world to monitor this and one such approach is through the use of Box-Jenkins ARIMA method. The word methodology as used here is rather scientific given the tendency of economic sciences to follow the doctrines of Popper's falsification approach of testing concept, as a way of proving outcomes in support of theoretical postulation (Jackson, 2016). The Box-Jenkins method is based on the use of time series data which are sequentially collected over a period of time, as in the case with Headline Consumer Price Index (HCPI), which is to be used in this study. Values collected on time series events are said to exhibit tendencies of autocorrelation which is a function of the lag values separated by the correlated values (also referred to as 'Autocorrelated Function, acronymed ACF) and they are said to be stationary where the mean and variance are constant (Etuk et al, 2013).

2. Aim and Objective

The main aim of this empirical work is to forecast monthly HCPI using Sierra Leone data which are mostly required by institution like the central bank in tracking its core objective of price stability. In this regard, the main objective is to produce an out-of-sample monthly forecast for HCPI data using the Box-Jenkins ARIMA methodology.

3. Seasonal Time Series Literature

In order to address the going-forward situation in keeping up with the bank's objectives (price stability), time series is normally used and in this case the univariate approach can be applied. In this regard, Time Series can be defined as a sequence of observed values ordered in time and with the assumption that future predicted values are dependent on historical sequences of the observed variable under study (Etuk et al, 2013 and Taneja et al, 2016). The stationarity of a series (X_t) would normally follow an autoregressive moving average model of the orders p and q , normally with the designate ARMA(p,q) and which is said to satisfy the following difference equation:

This article is published under the terms of the Creative Commons Attribution License 4.0

Author(s) retain the copyright of this article. Publication rights with Alkhaer Publications.

Published at: <http://www.ijsciences.com/pub/issue/2018-01/>

DOI: 10.18483/ijSci.1479; Online ISSN: 2305-3925; Print ISSN: 2410-4477



Selemun Hagos (Correspondence)



selemun21@gmail.com



+

$$X_t - \alpha_1 X_{t-1} - \alpha_2 X_{t-2} - \dots - \alpha_p X_{t-p} = \epsilon_t + \beta_1 \epsilon_{t-1} + \beta_2 \epsilon_{t-2} + \dots + \beta_q \epsilon_{t-q} \quad (\text{eq.1})$$

OR

$$A(L)X_t = B(L)\epsilon_t \quad (\text{eq2})$$

In this case, ϵ_t is said to be a sequence of random variables, with zero mean and constant variance, normally referred to as a *white noise* process (α and β being constant). This means that with $p = 0$, eq. 1 becomes a *moving average model of the order q* and with a designated MA(q). With $q = 0$, then the process will now become an *autoregressive* process of the order p, designated as AR(q).

As emphasised by Etuk et al (2013), apart from satisfying the stationarity condition, *invertibility* is also very key for time series; this ensures the uniqueness of model covariance structure and which makes way for expression of current events in terms of past history of the series. In this vein, AR(p) model can be written as:

$$X_t + \alpha_{p1} X_{t-1} + \alpha_{p2} X_{t-2} + \dots + \alpha_{pp} X_{t-p} = \epsilon_t$$

In this case, the coefficient of the last sequence (α_{ii}) is the partial autocorrelation function (PACF) of X_t . The ACF of an MA(q) model cuts off after lag q, whereas that of an AR(q) model is a combination *sinusoidals* dying off slowly (Etuk et al, 2013). The PACF of an MA(q) model will gradually die off slowly, whereas that of an AR(q) models cuts off after lag p. These two terms, AR and MA seem to portray some form of duality relationship. The condition for stationarity and invertibility for either eq.1 or 2 is such that $A(L) = 0$ and $B(L) = 0$ should have root outside the unit root circle respectively. Based on Box and Jenkins (1976), differencing of appropriate order automatically render a non-stationary series (X_t) stationary.

It is assumed that the degree of differencing required for stationarity be 'd' and in this case, the series (X_t) can be represented as: $A(L)\tilde{N}^d X_t = B(L)\epsilon_t$ (eq.3)

Where $\tilde{N} = 1$ and therefore, $A(L)\tilde{N}^d = 0$, with unit root d time and which means that differencing to d renders the series stationary. On this note, eq.3 is said to be an ARIMA of orders p,d and q, with designated ARIMA(p,d,q).

The time series (X_t) would follow a multiplicative (p,d,q) X (P,D,Q) seasonal ARIMA model if:

$$A(L)\phi(L^s)\tilde{N}^d\tilde{N}^{D_s}X_t = B(L)\Theta(B^s)\epsilon_t \quad (\text{eq. 4})$$

where ϕ and Θ are polynomials of order P and Q respectively. That is,

$$\phi(L^s) = 1 + \phi_1 L^s + \dots + \phi_p L^{sp} \quad (\text{eq. 5})$$

$$\Theta(L^s) = 1 + \Theta_1 L^s + \dots + \Theta_q L^{sq} \quad (\text{eq. 6})$$

Where the ϕ_i and the Θ_j are constants such that the zeros of the equations (5) and (6) are all outside the

unit circle for stationarity or invertibility respectively. Equation (5) represents the autoregressive operator whereas (6) represents the moving average operator. Here $\tilde{N}_s = 1 - L^s$.

4. Materials and Methods

The data for this work involve monthly HCPI from 2000M01 to 2017M09 obtained from the Central Statistical Office in Sierra Leone. The Box and Jenkins methodology is applied, a type of ARIMA model.

4.1. Determination of the orders d, D, P, q and Q:

It is very important that seasonal differencing is used to get rid of seasonal trend; with the presence of trend, non-seasonal differencing will be necessary to remove all seasonal trend. For simplicity of the model, orders of differencing d and D must add up to at most to 2, that is, " $d + D < 3$ ". An AR seasonal component is relevant where the ACF of the differenced series has a positive spike at the seasonal lag. In the case of a negative spike at seasonal lag, seasonal MA term is recommended. An AR(p) model has a PACF which truncates at lag p and an MA(q) has an ACF which truncates at lag q. In practice, $\pm 2/\sqrt{n}$, where n is the sample size (Etuk et al, 2013).

4.2. Model Estimation

The influence of a 'white nose' in ARIMA model means the presence of a non-linear iteration in the process of estimation of the parameters. Model optimization criterion like least error sum of squares, maximum likelihood or Automatic model specification is used. An initial estimate is usually chosen and followed by subsequent iterative steps with the aim of improving the estimated output for optimality. In the case of pure AR and pure MA models, linear optimization techniques exist as researched / outputted by Box and Jenkins (1976), Oyetunji (1985) and Etuk (1987). Eviews software which normally make use of least squares approach involving nonlinear iterative techniques is to be used.

4.3. Diagnostic Checking

The estimated model should be tested for goodness-of-fit. To start with, the variable is to be checked for unit root, either through use of Augmented Dicky-Fuller or Phillip-Perrons. Following on, analysis of the residuals of the model can be checked through the Autocorrelation results and also stability table output. In the event that the model is correct, the residuals would exhibit uncorrelated feature, thereby following a normal distribution with mean zero and constant variance.

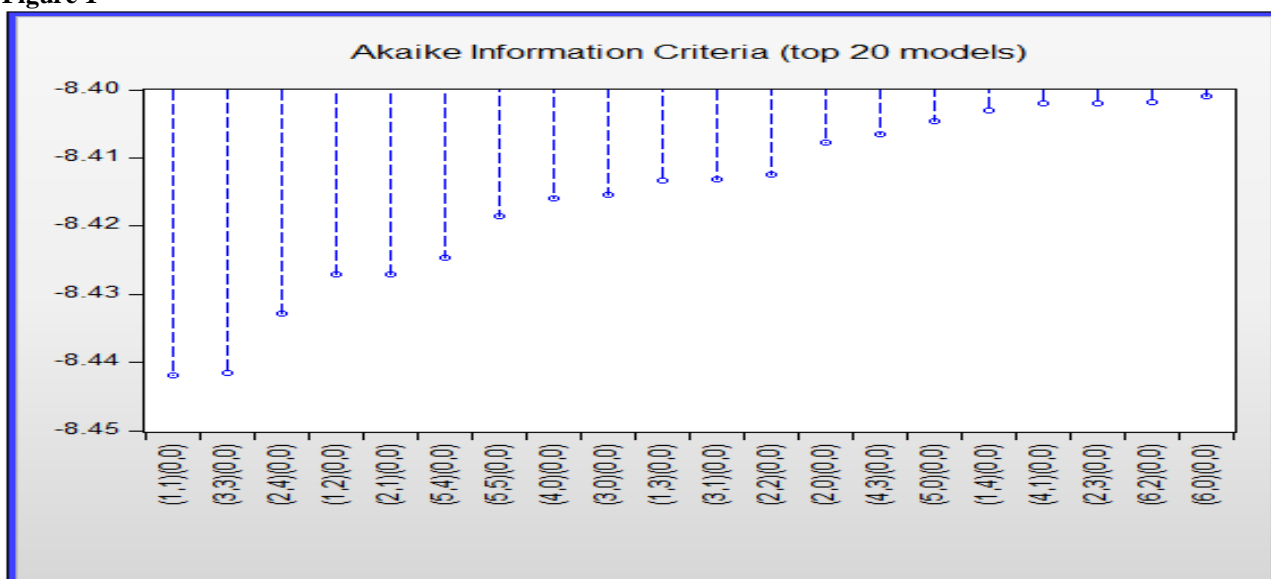
5. Result and Discussion

This was based on a univariate forecast of historical HCPI data. The process commenced by making use of high frequency data ranging from 2007M1 to 2017M10. Based on initial Unit Root diagnostics (Table 1), it became established that the data is not an ARMA process due to it not being stationary at

levels. Going forward, Automatic ARIMA function in EViews was applied, which gives an ARIMA(6,0) (0,0) output as shown in Figure 1. With iteration, an ARIMA(4,0)(0,0) model was outputted because it carries the lowest AIC value [Reference to Figure 1) .

Table 1: Phillips-Perron test statistic			Adj. t-Stat	Prob.*
			-5.629466	0.0000
Test critical values:	1% level		-4.031309	
	5% level		-3.445308	
	10% level		-3.147545	

Figure 1



This provide evidence of EViews best model choice, which was the first iteration on the estimation output table. As seen in Table 2 below, the iteration process suggested a more reduced form of the original model suggested by EViews (4,0)(0,0) which is said to have produced the least AIC result and also with an Inverted AR Roots result which is within the accepted limit value, meaning that the model is stable.

Table 2: Model Estimation

Dependent Variable: D(HCPI_SA)
 Method: ARMA Maximum Likelihood (OPG - BHHH)
 Date: 12/06/17 Time: 12:53
 Sample: 2007M02 2017M10
 Included observations: 129
 Convergence achieved after 36 iterations
 Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.931163	0.382374	2.435216	0.0163
AR(4)	0.125753	0.057098	2.202419	0.0295
AR(2)	0.254703	0.078015	3.264784	0.0014
AR(1)	0.472282	0.053113	8.892047	0.0000
SIGMASQ	0.245527	0.018674	13.14840	0.0000
R-squared	0.601837	Mean dependent var		0.911797

Adjusted R-squared	0.588993	S.D. dependent var	0.788332
S.E. of regression	0.505398	Akaike info criterion	1.519361
Sum squared resid	31.67295	Schwarz criterion	1.630206
Log likelihood	-92.99877	Hannan-Quinn criter.	1.564400
F-statistic	46.85761	Durbin-Watson stat	2.046465
Prob(F-statistic)	0.000000		

Inverted AR Roots .91 .07-.48i .07+.48i -.59

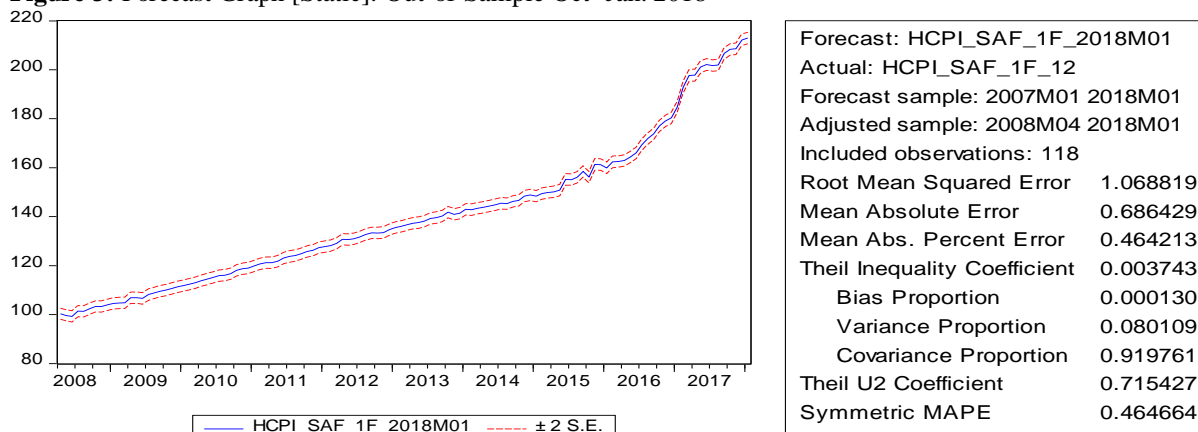
Figure 2

Date: 12/06/17 Time: 12:55
 Sample: 2007M01 2017M10
 Included observations: 129
 Q-statistic probabilities adjusted for 3 ARMA terms

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	-0.027	-0.027	0.0973	
		2	-0.036	-0.037	0.2714	
		3	0.048	0.046	0.5854	
		4	-0.027	-0.025	0.6811	0.409
		5	0.004	0.006	0.6830	0.711
		6	-0.064	-0.068	1.2448	0.742
		7	0.126	0.127	3.4614	0.484
		8	0.057	0.058	3.9185	0.561
		9	0.008	0.027	3.9272	0.687
		10	-0.083	-0.097	4.9021	0.672
		11	-0.053	-0.056	5.3011	0.725
		12	0.009	-0.004	5.3121	0.806
		13	-0.065	-0.042	5.9239	0.822
		14	0.066	0.058	6.5731	0.833
		15	0.072	0.059	7.3521	0.833
		16	0.041	0.040	7.6078	0.868
		17	-0.041	-0.033	7.8624	0.896
		18	0.094	0.120	9.2014	0.867
		19	-0.073	-0.075	10.011	0.866
		20	0.128	0.164	12.539	0.766
		21	-0.006	-0.041	12.545	0.818
		22	0.079	0.100	13.521	0.811
		23	-0.033	-0.108	13.690	0.846
		24	-0.131	-0.092	16.469	0.743
		25	0.033	-0.012	16.642	0.783
		26	-0.105	-0.071	18.445	0.733
		27	0.127	0.129	21.111	0.632
		28	0.003	-0.001	21.112	0.686
		29	-0.032	-0.029	21.284	0.727

The above Autocorrelation table shows a perfect diagnostic check for the model which means the process is within the accepted range.

This then made it possible to run the estimation with HCPI being the dependent variable of its ARMA processes as indicated below using 128 observations, with sample period ranging between 2007M1 – 2017M10. We commenced the forecasting process with both Dynamic and Static processes, but noticed that the Static forecasts (Figure 3) were more closed to the actual HCPI data as given.

Figure 3: Forecast Graph [Static]: Out-of-Sample Oct- Jan. 2018

Both in-sample and out-of-sample forecast were carried out by way of 'Statistic' method as it seem to have been the best in terms of produce a more stable forecast result than the Dynamic method as in the Mean Absolute Percentage Error (MAPE) which is quite small comparatively. This means that degrees of error in the forecast process is quite small comparatively.

Table 3: PROJECTIONS FOR INFLATION RATE USING EIEWS

Year	Linear Model		
	Forecast HCPI	Forecast Y-O-Y [%]	Monthly Change
Nov. 2017	208.5715	14.13565	0.33264
Dec. 2017	212.187	13.23285	1.733482
Jan. 2018	212.9938	11.17167	0.380209
R ²	0.601		
MAPE	0.464		

The above forecast evaluation (table 3) indicate reduction in Year-on-Year forecast for the monthly HCPI for November 2017, December 2017 and January 2018. The R² value is also high and as already mentioned, a low MAPE value.

5.1. Intuitive interpretation for the Sierra Leone Economy

This section of the discussion is based on experience of the Sierra Leone economy, despite the result of the scientific process has revealed a positive outcome in terms of a reduction in *Year-on-Year* inflation forecast.

5.1.1. Identified Downside Risks to the Forecast

The country at the moment is in the harvest period (Quarter 4), mostly around the months of November – March; positive impact for bumper supply of food commodities in the domestic market. This also have the effect of reducing food commodity prices as indicated in the forecast results for Q4 and also leading into the first month of 2018Q1. At the time of producing this empirical work, there has been indicative signs of (relative) stability in the country's exchange rate, which is reflected positively on the inflation forecast outcome as shown in Table 3. Stability in the exchange rate can also be attributed to the country's increase appetite for

(mineral) commodity export in a current stable global market.

5.1.2. Identified Upside Risks to the Forecast

The thirst of the government to finance domestic debt through its central bank wing is in itself a threat to inflationary pressure, through glut of money supply in the domestic economy, which may also have the effect of filtering into economic agents' appetite for increased spending. Where other external shocks are likely to surface in the economy, for example, increase in energy price, it is likely that such outcome may take a higher price toll on the domestic economy, with some level of pronounced pass-through effects on different retail outlets in the economy. Equally, rushed intervention by the government authorities to curb the situation through increase spending may likely have a knock-on effect on the budget deficit.

6. CONCLUSION [BEST CHOICE]

From the result, there seem not to be any residual autocorrelation, with a relatively good out-of-sample **STATIC** forecast. The ARIMA model as established in this forecast process indicate a relatively satisfactory outcome, with small forecast error as indicated by the very low MAPE. The short-term forecast of ARIMA model using STATIC process

seemed quite ideal in this situation, but in going forward, a more appropriate model and forecast method will be explored, particularly in using other variables in determining their impact on CPI. Such futuristic approach will seek to explore the use of exogenous variable(s) in the ARIMA process, thereby resulting in an ARIMAX model outcome.

References

1. Box, G.E.P. and Jenkins, G.M. (1976). "Time Series Analysis, Forecasting and Control", Holden-Day, San Francisco.
2. Etuk, E. H. (1987). *On the Selection of Autoregressive Moving Average Models*. An unpublished Ph. D. Thesis, Department of Statistics, University of Ibadan, Nigeria.
3. Etuk, E.H., Moffat, I.U and Chims, B.E. (2013). Modelling Monthly Rainfall Data of Port Harcourt, Nigeria by Seasonal Box-Jenkins Methods. *International Journal of Science*, Vol. 2 (July): pp. 60-67.
4. Jackson, E. A. (2017). Economic Methodology: Paradox of Ceteris Paribus (CP) Law in the Context of Sierra Leone. Method(e)s: *African Review of Social Sciences Methodology*, Vol. 2(1-2): pp. 31-4. DOI: 10.1080/23754745.2017.1354553.
5. Oyetunji, O. B. (1985). "Inverse Autocorrelations and Moving Average Time Series Modelling". *Journal of Official Statistics*, 1, pp. 315 – 322.
6. Taneja, K., Ahmad, S., Ahmad, K. and Attri, S. D. (2016). Time series analysis of aerosol optical depth over New Delhi using BoxeJenkins ARIMA modeling approach. *Atmospheric Pollution Research*, Vol. 7 (2016): pp. 585-596.