Breaching Modeling Protocols- A Case of Sarima Class of Models

Aderibigbe O. Timothy, Bushirat T. Bolarinwa and Ismaila A. Bolarinwa*

Department of Statistics, The Federal Polytechnic, P.M.B. 55, Bida, Niger State, Nigeria
**Corresponding Author*

***Abstract:* - This article demonstrated how Box-Jenkins SARIMA modeling protocols can be bypassed to arrive at one of the best possible models for a series. Data on monthly average maximum temperature of Bida, Niger State (in ^oC), covering January 2002 to December 2016 was collected from the National Cereals Research Institute, Baddegi, Bida, Niger State, Nigeria. The number of non seasonal and seasonal AR and MA parameters was each set at 0, 1 and 2; the order of non seasonal and seasonal differencing was set at 0 and 1. All possible models arising from combinations of these parameters were examined. Since 4 of 364 combinations had no parameters, a total of 360 models were estimated by nonlinear least squares method. Models were compared on the basis of Akaike Information Criterion (AIC). The least AIC value of 2.3558 corresponding to SARIMA (1, 1, 2) *X* (2, 1, 2)$_{12}$ was obtained. Test of model adequacy of selected model produced a p-value of 0.000, implying a very high significance. It was recommended that the selected model be used for forecasting monthly average maximum temperature of Bida and tried on similar data for other locations in the country.**

***Keywords*: Temperature, SARIMA, AIC, Forecasts**

I. INTRODUCTION

Time series modeling is a statistical phenomenon that features wherever forecasting is of interest. Because forecasting is essential in various endeavours, the phenomenon has attracted a lot of attention from researchers. Time series analysis primarily aims at studying the behavior of past observations of a time series with a view to developing a model that adequately represents the process that generated the series. This developed model is then used to generate future values for the series, that is, make forecasts.

A few of areas in which forecasting is essential are meteorology, economics, marketing, finance, engineering, and sports. Owing to the role forecasting plays in several fields, efforts which led to the development of several forecasting models have been made by researchers. Such models include classical additive and multiplicative models, Holt-Winters model, Brown's model, SARIMA models by [1], self adaptive threshold autoregressive model and so on.

The SARIMA class of models is of primary importance to this article. Box-Jenkins methodology has been used by many researchers. Reference [2] predicted the variation in temperature in different places in the world by using different statistical approaches, including bivariate time-series models, and time-series smoothing both in the univariate and multivariate setting. Reference [3] did short term load forecasting using lifting scheme and ARIMA model; References [4] and [5] fitted SARIMA models to Nigerian Naira to US Dollar and Euro exchange rates. Reference [6] applied ARIMA models for predicting weekly rainfall data for period 1990 – 2011 from four rainfall stations in the Northwest of Iraq. Reference [7] applied SARIMA to model local temperature in Ashanti region of Northern Ghana.

The Box-Jenkins methodology is iterative in nature and entails the following steps: Postulation of general class of models, model identification, model estimation, model diagnostics, and forecasting. If an identified model fails diagnostic test, the procedure requires that another model be identified. This is repeated until adequate model is obtained.

It is worth mentioning that for a particular series, a large number of models may be found to be adequate. What this suggests is that a particular adequate model may not be the best of adequate models. Since the model parameters each has a small set of known possible values in practice, such information can be used to arrive at a model that is closer to the best than one ordinarily arrived at through the usual procedure by fitting a number of models using combinations of possible parameter values in practice and selecting the best of fitted models based on a valid criterion. This is the main objective of this article.

The remainder of the article is organized as follows: Section 2 presents the Methodology; Section 3 presents the Results and Discussion while Section 4 concludes the article and recommends.

II. METHODOLOGY

This section presents data collection, model form and the selection criterion.

A. Data

Data obtained are secondary data sourced from the National Cereals Research Institute, Baddegi, Bida, Niger State and represent the monthly average maximum temperature (in ^oC) covering January 2002 to December 20016.

B. Model

The model is seasonal ARIMA model. That is, SARIMA. SARIMA $(p, d, q) \times (P, D, Q)_s$ is symbolically represented as:

$$\phi_p(B)\Phi_P(B^S)\nabla^d\nabla_S^D X_t = \theta_q(B)\Theta_Q(B^s)\varepsilon_t$$

where

p and *P* are the orders of non-seasonal and seasonal AR respectively;

q and *Q* are the orders of non-seasonal and seasonal MA respectively;

d and *D* are the orders of non-seasonal and seasonal differencing respectively;

s is seasonality

and

$$\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 .. \phi_p B^p$$

$$\Phi_P(B^s) = 1 - \phi_1 B^s - \phi_2 B^{2s} .. \phi_P B^{Ps}$$

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

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

The nomenclature below was adopted for both non-seasonal (p, d, q) and seasonal (P, D, Q) model parameter settings:

$$\begin{array}{ll} 1: & (0,0,0) \\ 2: & (0,1,0) \\ . & . \\ . & . \\ . & . \\ 18: & (2,1,2) \end{array}$$

Model (1, 2) hence, corresponds to SARIMA (0, 0, 0) X $(0, 1, 0)_{12}$ while model (2, 18) corresponds to SARIMA (0, 1, 0) X $(2, 1, 2)_{12}$ Parameters *p*, *q*, *P*, and *Q* were each set at 0, 1, and 2 while orders *d* and *D* were set at 0 and 1. A total of 320 models were fitted.

C. Model Estimation

Model estimation was carried out by a method known in the literature as *conditional least squares*.

D. Akaike Information Criterion

Model comparisons were carried out based on Akaike Information Criterion (AIC). Reference [8] proposed AIC is of the form

$$AIC = -2\,Log_e(L) + K$$

where *L* is the likelihood function and *K* is the number of parameters in the model. The practice is to pick model with lowest AIC.

III. RESULTS AND DISCUSSION

Tables 1a and 1b (attached as APPENDIX I) present the AICs of the 320 fitted models. Values are missing for 4 combinations that do not contain any parameter. The least of AICs is 2.3558, corresponding to Model (16, 18). That is, SARIMA (1, 1, 2) X $(2, 1, 2)_{12}$. The model was hence, used for forecasting (Forecasts attached as APPENDIX III). The software output for selected model estimation is attached as APPENDIX II. The selected model is symbolically, represented as follows:

$$(1+0.473)(1+0.726B^{12}+0.098B^{24})(1-B)(1-B^{12})X_t =$$
$$0.004+(1+0.374B+.626B^2)(1-0.033B^{12}+0.864B^{24})$$

The constant part of model, AR(1), SAR(12), MA(2) and SMA(24) parameters are all significant while other parameters, that is, those corresponding to SAR(24), MA(1) and SMA(12) are all not significant at 5% level. The F-test of overall regression is highly significant with p-value of 0.000 (See APPENDIX II). The selected model is hence, unsurprisingly, highly significant. This model is definitely one of the best models that could be fitted to this particular data. This approach ensures that, instead of simply following the methodology and arriving at a model that may be quite far from being the best, though adequate, a model that is quite close to the best is obtainable.

IV. CONCLUSION AND RECOMMENDATION

This article has demonstrated how one of the best SARIMA models for a particular series can be obtained through breaching of Box-Jenkins modeling methodology protocols. The approach has produced a highly significant model.

The following recommendations are hereby, made:

i) The selected model could be used for forecasting monthly maximum temperature of Bida.
ii) The model could be tried on similar data for other locations in the country.
iii) The approach employed could be used on other meteorological data like rainfall, relative humidity and so on.

ACKNOWLEDGEMENT

We express huge appreciation to the Management of National Cereals Research Institute, Baddegi for making the data for the research available.

REFERENCES

[1]. Box, G.E.P. & Jenkins, G. (1976). *Time series analysis, Forecasting and Control*. San Frasisco: Holden-Day.
[2]. Mann, M.E. (2008). Smoothing of climate time series revisited. *Geographical Research Letters*, *35*(L16708), 1-5.
[3]. Lee, C. & Ko, C. (2011). Short-term load forecasting using lifting scheme and ARIMA models. *Expert Systems with Applications*, *38*, 5902-5911.
[4]. Etuk, E.H. (2012). Forecasting Nigerian Naira-US Dollar exchange rates by a seasonal ARIMA model. *American Journal of Scientific Research*, *59*, 71-78.
[5]. Etuk, E.H. (2013). The fitting of SARIMA model to monthly Naira-Euro exchange rates. *Mathematical Modeling*, *3*(1), 17-26.
[6]. Zakaria, S.Y., Mohammed, D., Ali, S., Al-Ansari, N., & Knutson, S. (2013). Estimation of annual harvest runoff at Sulaymniyah Governorate, Kurdistan Region of Iraq. *Natural Science*, *5*(12), 1272-1283.

[7]. Asamoah-Boaheng, M. (2014). Using SARIMA to forecast monthly mean surface air temperature in the Ashanti Region of Ghana. *International Journal of Statistics and Applications*, *4*(6), 292-298.

[8]. Akaike, H. (1974). A new look at the statistical model identification. *IEE Transactions on Automatic Control*, *19*(6), 716-723.

APPENDIX I

RESULTS OF ANALYSIS

TABLE Ia AICs FOR FITTED SARIMA MODELS

	1	2	3	4	5	6	7	8	9
1	-	-	3.037032	2.926197	2.923133	2.837741	4.073351	2.598093	2.605540
2	-	-	3.482624	3.402012	3.352671	3.259151	3.841531	3.035483	3.029606
3	4.110783	3.054587	3.026628	2.907191	2.898662	2.801354	3.664356	2.564711	2.576724
4	4.145840	3.395794	3.337215	3.179204	3.164691	3.031822	3.849653	2.829903	2.834250
5	3.850642	3.067448	3.036587	2.915037	2.906075	2.793333	3.618027	2.571084	2.578210
6	4.156678	3.188913	3.169474	3.028189	3.027892	2.881632	3.863752	2.654620	2.664119
7	4.205758	3.053365	3.026245	2.906987	2.899607	2.808214	3.724904	2.564304	2.576244
8	4.148570	3.081832	3.055133	2.906163	2.906146	2.780241	3.851973	2.541310	2.550929
9	3.959133	3.040032	3.014198	2.874814	2.873705	2.755554	3.642239	2.530976	2.539943
10	4.153673	3.072475	3.044531	2.922791	2.921095	2.791469	3.749541	2.552823	2.562884
11	3.575109	3.056126	3.030768	2.889149	2.893677	2.766650	3.386221	2.545638	2.555478
12	4.129895	3.084528	3.052816	2.939126	2.939285	2.809098	3.842507	2.565678	2.576449
13	3.994977	3.065229	3.080000	2.915099	2.907391	2.816979	3.678008	2.573021	2.578952
14	4.158038	3.071970	3.046830	2.920735	2.915560	2.793192	3.859632	2.550984	2.562979
15	3.931831	3.0048373	3.025440	2.887250	2.895124	2.769138	3.648735	2.542952	2.551917
16	4.164896	3.030877	3.011240	2.897252	2.895894	2.786373	3.830293	2.558685	2.568757
17	3.291267	3.012061	2.989295	2.868769	2.866595	2.763282	3.012907	2.552094	2.561595
18	4.135987	3.034417	3.006157	2.908262	2.907708	2.813661	3.774731	2.568668	2.579569

TABLE Ib CONTINUATION OF TABLE Ia

	10	11	12	13	14	15	16	17	18
1	2.616153	2.627534	2.651971	3.58639	2.592699	2.599477	2.628419	3.050495	2.453393
2	3.045648	3.099057	3.069101	3.534964	3.046528	3.039656	3.057177	3.050495	2.814308
3	2.582086	2.576724	2.582086	3.283885	2.568160	2.573574	2.594938	2.604245	2.392298
4	2.835819	2.843311	2.825346	3.496109	2.841478	2.845523	2.846766	2.854259	2.643008
5	2.586309	2.596793	2.587238	3.274853	2.576561	2.583351	2.598961	2.609460	2.394832
6	2.670539	2.682465	2.661256	3.490098	2.660724	2.670056	2.682724	2.694657	2.507059
7	2.583570	2.593300	2.622604	3.331779	2.565633	2.571059	2.596386	2.606112	2.395574
8	2.552252	2.564721	2.569401	3.499458	2.550118	2.559531	2.564421	2.577331	2.378254
9	2.538784	2.550845	2.563035	3.293883	2.538215	2.546915	2.551111	2.563197	2.361501
10	2.568755	2.581259	2.561294	3.292267	2.561978	2.571910	2.581022	2.593517	2.386188
11	2.556165	2.568700	2.563320	3.218744	2.553197	2.562847	2.568928	2.581457	2.400276
12	2.581253	2.593509	2.580938	3.470086	2.573570	2.584397	2.593549	2.605816	2.407429
13	2.592416	2.601824	2.631570	3.297726	2.575809	2.581118	2.605231	2.614640	2.406155
14	2.562802	2.574993	2.581471	3.364387	2.560133	2.569255	2.574894	2.587104	2.376118
15	2.551654	2.563739	2.576791	3.299348	2.550108	2.558837	2.563988	2.576094	2.372237
16	2.574868	2.587445	2.563692	3.305050	2.567011	2.576896	2.587701	2.600285	2.355800
17	2.563994	2.573453	2.543242	2.946738	2.558884	2.568114	2.576711	2.589115	2.382014
18	2.588694	2.601093	2.585274	3.491806	2.576261	2.587202	2.601311	2.613725	2.377346

APPENDIX II

Model Estimation Output SARIMA (1, 1, 2) X (2, 1, 2)$_{12}$

MA Backcast: 2003M01 2005M02				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.003824	0.000872	4.384655	0.0000
AR(1)	-0.472583	0.221357	-2.134937	0.0346
SAR(12)	-0.726213	0.083142	-8.734593	0.0000
SAR(24)	-0.097727	0.079741	-1.225556	0.2225
MA(1)	-0.373762	0.203083	-1.840440	0.0679
MA(2)	-0.625840	0.194033	-3.225425	0.0016
SMA(12)	0.032878	0.030437	1.080213	0.2820
SMA(24)	-0.863735	0.029416	-29.36256	0.0000
R-squared	0.738862	Mean dependent var		0.000000
Adjusted R-squared	0.725221	S.D. dependent var		1.458650
S.E. of regression	0.764615	Akaike info criterion		2.355800
Sum squared resid	78.34125	Schwarz criterion		2.522326
Log likelihood	-159.2618	Hannan-Quinn criter.		2.423470
F-statistic	54.16280	Durbin-Watson stat		1.900613
Prob(F-statistic)	0.000000			

APPENDIX III

Forecasts for January 2017 to December 2019

		95 Percent Limits	
Period	**Forecast**	**Lower**	**Upper Actual**
Jan.2017	36.3588	34.7236	37.9939
Feb.2017	38.1832	36.5245	39.8418
Mar.2017	39.5574	37.8987	41.2160
Apr.2017	38.2181	36.5501	39.8861
May 2017	35.1551	33.4856	36.8247
Jun.2017	32.8956	31.2205	34.5708
Jul.2017	32.0839	30.4060	33.7618
Aug.2017	30.8560	29.1737	32.5383
Sep.2017	31.6438	29.9582	33.3294
Oct.2017	33.4834	31.7938	35.1730
Nov.2017	35.9788	34.2856	37.6719
Dec.2017	35.9547	34.2578	37.6516
Jan.2018	35.9372	34.2077	37.6667
Feb.2018	38.3796	36.6436	40.1157
Mar.2018	39.8117	38.0720	41.5514
Apr.2018	38.3305	36.5855	40.0755
May 2018	35.6041	33.8549	37.3533
Jun.2018	33.6708	31.9167	35.4249
Jul.2018	32.1647	30.4062	33.9232
Aug.2018	30.8709	29.1077	32.6341
Sep.2018	32.3056	30.5379	34.0733
Oct.2018	33.8939	32.1216	35.6662
Nov.2018	35.9404	34.1636	37.7173
Dec.2018	36.2732	34.4918	38.0546
Jan.2019	36.5930	34.8115	38.3745
Feb.2019	38.7051	36.9202	40.4900
Mar.2019	39.9977	38.2084	41.7871
Apr.2019	38.8656	37.0725	40.6587
May 2019	35.7874	33.9901	37.5846
Jun.2019	33.4846	31.6834	35.2858
Jul.2019	32.4710	30.6657	34.2762
Aug.2019	31.3397	29.5305	33.1489
Sep.2019	32.3120	30.4988	34.1251
Oct.2019	34.1062	32.2890	35.9233
Nov.2019	36.3610	34.5399	38.1821
Dec.2019	36.5202	34.6952	38.3452