IOSR Journal of Agriculture and Veterinary Science (IOSR-JAVS)
e-ISSN: 2319-2380, p-ISSN: 2319-2372. Volume 11, Issue 5 Ver. II (May 2018), PP 22-28
www.iosrjournals.org

Forecasting Fish Production in Nigeria Using Univariate Arima Models

Ogunbadejo, HussainKehinde and Alliu, Kehinde Adewale
Corresponding Author: Ogunbadejo

Abstract: *Estimation and forecasting of fish production are crucial in supporting policy decisions regarding food security and development issues. The present study examines the current status of fish production in Nigeria. Univariate time series modeling using ARIMA model was developed for forecasting fish production. Box and Jenkins linear time series model, which involves autoregression, moving average,and integration, also known as ARIMA (p, d, q) model was applied. The annual production series of fish from 1971 to 2015 exhibited an increasing trend while prediction of fish production for 2016-2020 showed anincreasing trend. The study has shown that the best-fitted model for fish production series is ARMA (1, 1, 1). Themodel revealed a good performance in terms of explainingvariability and forecasting power.*
Keywords: *ARIMA, Univariate, fish production, forecasting*

Date of Submission: 09-05-2018 Date of acceptance: 26-05-2018

I. Introduction

The fisheries sector plays a vital role in the Nigerian national economy. It contributes to the national gross domestic product (GDP), as it also a source of employment, foreign exchange and a cheap source of protein supply in the country. Fish constitutes 60-70% of the national animal protein intake, with per capita consumption of 47.8 kg per year (FDF 2016).

Reliable forecasts are important for efficient planning of fish production and can possibly assist in making informed policy decisions. Therefore, it is vital to generate forecasts for fish production using rigorous statistical modeling techniques that will assist in making informed policy decisions. Univariate time series modeling is convenient in developing forecasting models for fish production. During the past few decades, several statistical forecasting models have been developed due to the advancements in computer technology. One of such models includes the AutoregressiveIntegrated Moving Average (ARIMA) model.

Using ARIMA model the following authors found robust results. Badmus and Ariyo (2011) determined the ARIMA (2,1,2) to forecast maize production in Nigeria. Awal and Siddique (2011) studied rice production in Bangladesh. The study revealed that production uncertainty of rice could be minimized if production were forecasted with accuracy. Biswas and Bhattacharyya (2013) forecasted production of rice in West Bengal.

The study found ARIMA (2,1,1) to be the best fitted model for rice production. Ehab and Frah (2016) modeled sorghum production in Sudan and the study revealed that growth in production is attributed to changes in harvested area land.

The objective of this study is to develop an Autoregressive Integrated Moving Average (ARIMA) model for forecasting Fishproduction in Nigeria and specifically to forecast fish production for the period up to year 2020. Estimation and forecasting of fish productionare crucial in planning and supporting policy decisionsregarding food security and development issues.

ARIMA MODEL

The development of ARIMA models is based on the methodology quantified in the classic work of Box and Jenkins. The autoregressive integrated moving average (ARIMA) model, represented by ARIMA (*p*, *d*, *q*), is given by:

ARMA models are obtained as combinations of the autoregressive (AR) and moving average (MA) models. Consider a stochastic process {Xt} specified as

$X_t = \alpha_1 X_{t-1} + \ldots + \alpha_p X_{t-p} + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_q \varepsilon_{t-q}$ ……………………………………..........(1)

where $\{\varepsilon_t\}$ is a purely random process with mean equal to zero and variance equal to σ^2.

This stochastic process is called an ARMA (p, q) process or model.

This class of models is of particular interest as the models result in a parsimonious representation of higher order AR(p) or MA(q) processes.

Using the lag operator, L this equation may be re-written as

$\varphi(L)X_t = \theta(L)\varepsilon_t$ ……………………………………………………………………………(2)

Where: φ(L) and θ(L) are polynomials of orders p and q, respectively, and are defined as
$\varphi(L) = 1 – \alpha 1L – \alpha 2L2 \dots - \alpha pLp$ ……………………………………………………………(3)
$\theta(L) = 1 + \beta 1L + \beta 2L2 + \dots + \beta qLq$. ………………………………………………………(4)
For stationarity, the roots of φ(L) = 0 must lie outside the unit circle and for invertibility of the MA component, the roots of θ(L) must again lie outside the unit circle. Thus, we have a combination of the "stability" conditions of both the autoregressive and the moving average processes.
The model building procedure involves the following three steps: namely identification, estimation of parameters, and diagnostic checking. Identification*:* Orders p, d, q of ARIMA models are specified to clarify the number of parameters to estimate.

However, the Box-Jenkins ARIMA method can only be applied to time series that are stationary. Thus, the primary step in developing a Box-Jenkins model is to ascertain if the time series data is stationary. Gujarati and Porter (2009) argued that the main reason for requiring stationary data is that any model which is concluded from these data can itself be interpreted as stable or stationary,therefore providing a valid basis for forecasting.

Once stationarity has been addressed, the next phase is to determine the order (p and q) of theautoregressive and moving average terms. The basic tools for accomplishing this are the autocorrelation (ACF) and the partial autocorrelation (PACF) plots.

Estimation: E-views 9 .0 Software package was used to fit the ARIMA model. Akaike information criterion(AIC) and Bayesian information criterion (BIC) values were used for parameter estimation. A model with the smallest values of AIC, BIC and Q-statistics and with high R-square may be considered as an appropriate model for forecasting (Biswas and Bhattacharyya, 2013).

In an application, if computed p-value associated with the Q-Statistics is small (*p*-value < α), the modelis considered unsatisfactory (Gujarati and Porter 2009). Thus, one should repeat the analysis process until a suitable model has been obtained.

Stationarity

In order to use any time series data for analysis using Box-Jenkins methodology, the time series should meet the stationarity conditions. A time series *t X* is said to be weakly stationary or wide-sense stationary or covariance stationary if it fulfills the following three properties (Ljung, G.M.1978),
1. Mean is constant over time
2. Variance is constant over time;
3. Covariance between any two values of the seriesdepends only on their distance apart in time (*k*) not ontheir absolute location in time {t}.

II. Literature Review

Raymond Y.C. Tse, (1997) suggested that the following two questions must be answered to identify the data series in a time series analysis: (1) whether the data are random; and (2) have any trends? This is followed by another three steps of model identification, parameter estimation and testing for model validity. If a series is random, the correlation between successive values in a time series is close to zero. If the observations of time series are statistically dependent on each another, then theARIMA is appropriate for the time series analysis.

Meyler et al (1998) drew a framework for ARIMA time series models for forecasting Irish inflation. In their research, they emphasized heavily on optimizing forecast performance while focusing more on minimizing out-of-sample forecast errors rather than maximizing in-sample 'goodness of fit'.

Stergiou (1989) in his research used ARIMA model technique on a 45years' time series data (from 1964 to 1980 and 204 observations) of monthly catches of pilchard (Sardina pilchardus) from Greek waters for forecasting up to 12 months ahead and forecasts were compared with actual data for 1981 which was not used in the estimation of the parameters. The research found mean error as 14% suggesting that ARIMA procedure was capable of forecasting the complex dynamics of the Greek pilchard fishery, which, otherwise, was difficult to predict because of the year-to-year changes in oceanographic and biological conditions.

Contreras et al (2003) in their study, using ARIMA methodology, provided a method to predict next-day electricity prices both for spot markets and long-term contracts for mainland Spain and Californian markets. In fact a plethora of research studies is available to justify that a careful and precise selection of ARIMA model can be fitted to the time series data of single variable (with any kind of pattern in the series and with autocorrelations between the successive values in the time series) to forecast, with better accuracy, the future values in the series. This study is also an attempt to predict the future production values of fish production in Nigeria by fitting ARIMA technique on the time series data of past23 years' productions.

Model Identification

First stage of ARIMA model building is to identify whether the variable, which is being forecasted, is stationary in time series or not. By stationary we mean, the values of variable over time varies around a constant mean and variance. The ARIMA model cannot be built until we make this series stationary. We first have to difference the time series'd' times to obtain a stationary series in order to have an ARIMA (p,d,q) model with 'd' as the order of differencing used. Caution to be taken in differencing as over differencing will tend to increase in the standard deviation, rather than a reduction. The best idea is to start with differencing with lowest order (of first order, d=1) and test the data for unit root problems. So we obtained a time series of first order differencing.

Test for stationarity: KPSS Test

Our null hypothesis (Ho) in the test is that the time series data is non-stationary while alternative hypothesis (Ha) is that the series is stationary. The hypothesis then is tested by performing appropriate differencing of the data in dth order and applying the KPSS test to the differenced time series data. First order differencing (d=1) means wegenerate a table of differenced data of current and immediate previous one (Xt = Xt –Xt-1). The KPSS test result, as obtained upon application, is shown below:

KPSS = 0.149876, Lag order = 4, p-value = 0.01

We, therefore, fail to accept the Ho and hence can conclude that the alternativehypothesis is true i.e. the series is stationary in its mean and variance. Thus, there isno need for further differencing the time series and we adopt d = 1 for ourARIMA(p,d,q) model.

This test enables us to go further in steps for ARIMA model development i.e. tofind suitable values of p in AR and q in MA in our model. For that, we need to examinethe correlogram and partial correlogram of the stationary (first order differenced) timeseries.

III. Results And Discussion

In modeling seasonal ARIMA process, the first step to determine whether the time series is stationary or non-stationary. There exists several tests of stationarity amongst others are ADF, PP and KPSS. Analysis of the descriptive statistics enables us to determine whether the data is normally distributed. The most common measures are mean, median skewness and kurtosis. In normally distributed data; the mean and median should be equal, for the variable in this study the mean and median is almost equal for IogFish.Skewedness is the tilt in the distribution and should be within the -2 and +2 range for normally distributed series. In a positively skewed distribution, the mean is typically higher than the median, whereas in negatively skewed distribution, the mean is lower than the median. Skewedness for a normal distribution is zero. In this study, Fish and I-Fish are within the above stated range thus normally distributed. Kurtosis is a measure of how outlier-prone a distribution is. Kurtosis for a normal distribution is 3. Distribution that is more outlier prone has kurtosis less than 3, meaning that the data is not normally distributed.

Table 1 Summary Statistics, using the observations 1971 - 2015

	FISH	LFISH
Mean	526.5358	13.10320
Median	486.3080	13.09460
Maximum	1123.000	13.93152
Minimum	242.5250	12.39886
Std. Dev.	215.2652	0.371887
Skewness	1.295973	0.415462
Kurtosis	4.139444	2.900118
Jarque-Bera	15.03098	1.313269
Probability	0.000545	0.518594
Sum	23694.11	589.6439
Sum Sq. Dev.	2038921.	6.085187
Observations	45	45

Table 2 Model 1: ARMA, using observations 1971-2015 (T = 45)

Dependent Variable: DLOG(FISH)
Method: ARIMA Maximum Likelihood (OPG - BHHH)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.020943	0.025019	0.837113	0.4075
AR(1)	0.248327	1.265715	0.196195	0.8455
MA(1)	-0.369476	1.245876	-0.296560	0.7683
SIGMASQ	0.021926	0.004305	5.093042	0.0000

Table 3 Model 2: ARIMA, using observations 1971-2015 (T = 45)

Dependent Variable: DLOG(LFISH)
Method: ARMA Maximum Likelihood (OPG - BHHH)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.001562	0.001959	0.797653	0.4298
AR(1)	0.285779	1.221745	0.233910	0.8162
MA(1)	-0.408885	1.190047	-0.343588	0.7330
SIGMASQ	0.000132	2.60E-05	5.060074	0.0000

Forecasts with Fish production with 95% confidence interval were generated by using ARIMA (1,1,1) model for the period of 2016 - 2020. Data presented in table show that the Fish production would increase and the predicted Fish production will range between 1,116,732metric tons in 2016 to 1,214,561 metric tons during the period 2020. This means that an increased quantity of Fish production will be available in future for domestic consumption and export.

The Actual and fitted values are presented in figure 1. A residual in excess of 2.5 standard errors were observed, however, the results reconfirmed that the estimated model is the best fit for forecasting Fish production in Nigeria.

Table 4 Forecast accuracy statistics

Root Mean Squared Error	0.468927
Mean Absolute Error	0.368229
Mean Absolute Percentage Error	2.858407
Theil's U	0.017643

Figure 1: Plot of Actual and Fitted logFish ARIMA (1, 1, 1)

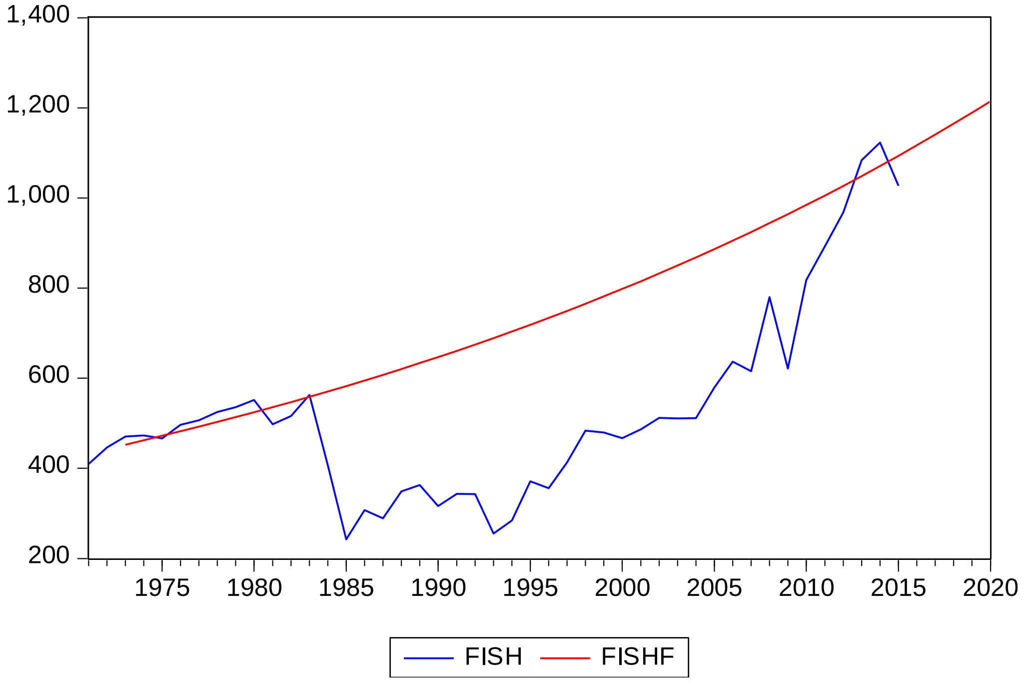
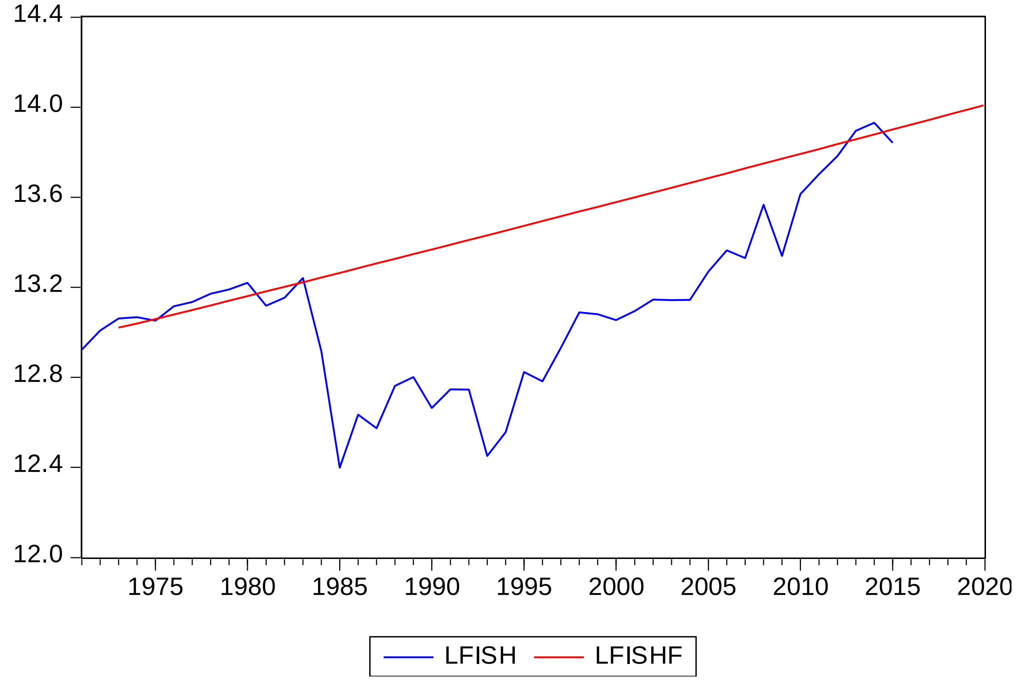


Table 5 Fish Production Forecasts (in thousand tons) for 5 years

Year	Forecast
2016	1116.732
2017	1140.424
2018	1164.62
2019	1189.328
2020	1214.561

Figure 2: The plots of the forecasts for the ARIMA models (1, 1 ,1)



IV. Conclusion

As the population increases overtime gradually, it is necessary to plan to meet requirements of nation. For this purpose, forecasting is the key tool to alarm about the need of the nation in advance. Fish is the basic need for any country all over the world. In this study, we developed time-series model to forecast Fish production in Nigeria on the basis of historical data. i.e. 1971-2015. Best model is selected on the basis of model selection criteria. i.e. AIC and SBIC. In this study, the ARIMA (1,1,1) was the best candidate model selected for making predictions for one year for Fish production in Nigeria using a 45 year's times series data. ARIMA was used for the reasons of its capabilities to make predictions using a time series data within any kind of pattern and with autocorrelations between the successive values in the time series. The prediction for 2016 - 2020 fish production was about 1,116,732 metric ton to 1,214,561 metric ton respectively at 95% confidence level.

References

[1]. Awal, M. A., Siddique, M. A. B. (2011). Rice production in Bangladesh employing by ARIMA model. Bangl. J. Res., 36(1), 51–62.
[2]. Badmus, M. A., Ariyo, O. S. (2011). Forecasting cultivated area and production of Maize in Nigerian using ARIMA Model. Asian J. Agric. Sci., 3(3), 171–176.
[3]. Biswas, R., Bhattacharyya, B. (2013). ARIMA modeling to forecast area and production of rice in West Bengal. J. Crop Weed, 9(2), 26–31.
[4]. Box, G. E. P., Jenkin, G. M. (1976). Time Series of analysis: Forecasting and control. Jon Wiley & Sons.
[5]. Contreras, J., Espinola, R., Nogales, F.J., Conejo, A.J., (2003), ARIMA Models to Predict Nextday Electricity Prices, IEEE Transactions on Power Systems, Vol. 18, No. 3, pp. 1014-1020.
[6]. Gujarati, D. N., Porter, D. C. (2009). Basic Econometrics, 5thEdition. New York: McGraw-hill, USA.
[7]. Ljung, G.M. and Box, G.E.P. (1978), On a measure of lack of fit in time series models, Biometrika, Vol. 67, pp. 297-303.
[8]. Meyler, Aidan; Kenny, Geoff and Quinn, Terry (1998), Forecasting Irish inflation using ARIMAmodels, Central Bank and Financial Services Authority of Ireland Technical PaperSeries, Vol. 1998, No. 3/RT/98 (December 1998), pp. 1-48.
[9]. Milton Iyoha and Patricia Adamu (2015) forecasting Currency in Circulation inNigeria. West African Financial and Economic Review (WAFER). Vol. 12, No.1
[10]. Raymond Y.C. Tse (1997), An application of the ARIMA model to real-estate prices in HongKong, Journal of Property Finance, Vol. 8, No. 2, pp.152 – 163.
[11]. Stergiou, K. I. (1989), Modeling and forecasting the fishery for pilchard (Sardina pilchardus) inGreek waters using ARIMA time-series models, ICES Journal of Marine Science, Volume 46, No. 1, pp. 16-23.

Appendix 1: Model estimation range: 1985 - 2007

	l_fish	fitted	residual	
1985	5.49113	6.03893	-0.547807	*
1986	5.72704	5.69240	0.0346464	
1987	5.66681	5.80905	-0.142244	
1988	5.85504	5.77824	0.0768023	
1989	5.89360	5.88100	0.0125989	
1990	5.75688	5.91657	-0.159685	
1991	5.83875	5.86467	-0.0259208	
1992	5.83796	5.91205	-0.0740857	
1993	5.54330	5.92374	-0.380435	*
1994	5.64897	5.79764	-0.148666	
1995	5.91634	5.84466	0.0716752	
1996	5.87473	5.97041	-0.0956768	
1997	6.02391	5.96719	0.0567211	
1998	6.18101	6.04655	0.134457	
1999	6.17308	6.13348	0.0395967	
2000	6.14654	6.15202	-0.00547405	
2001	6.18685	6.16092	0.0259261	
2002	6.23778	6.19674	0.0410338	

2003	6.23590	6.23727	-0.00136685
2004	6.23711	6.25573	-0.0186152
2005	6.36223	6.27458	0.0876499
2006	6.45653	6.34490	0.111629
2007	6.42245	6.40441	0.0180462

Ogunbadejo "Forecasting Fish Production in Nigeria Using Univariate Arima Models." IOSR Journal of Agriculture and Veterinary Science (IOSR-JAVS) 11.5 (2018): 22-28.