Forecasting Fish Production in Sri Lanka by Using ARIMA Model

Brintha Karunarathna[1], K. A. N. K. Karunarathna[2]

[1]Dept. of Crop Science, Faculty of Agriculture, Eastern University, Sri Lanka

[2]Dept. of Mathematics, Faculty of Science, Eastern University, Sri Lanka

Abstract: Fisheries sector plays an important role in Sri Lankan economy. Demand for fish is increasing with time. Therefore, forecasting the fish production is vital for better planning of fisheries development and management practices. This study aimed to find a suitable Autoregressive Integrated Moving Average (ARIMA) model to forecast annual fish production in Sri Lanka. National annual fish production data during the period from 1972 to 2016 were used for the study. During this period, fish production has increased from 101712 MT to 530920 MT with some fluctuations during the period. The minimum annual fish production of 100702 MT was recorded in 1973 while the maximum of 535050 MT has been recorded in 2014. Several ARIMA models were tested and the most appropriate model was selected based on the validity of assumptions and the accuracy of forecasts. ARIMA (1,1,1) model could be selected as the best model and it gave predictions for year 2016 with 1.35% forecasting error. Estimated production for year 2017 is 551046 MT. Developed model would help the decision makers to establish priorities in terms of fisheries management.

Keywords: Autoregressive Integrated Moving Average, forecast, error, model, production

INTRODUCTION

Fisheries sector socially and economically plays an important role in the Sri Lankans' life. Fish serves as the most important source of animal protein for the humans.

The demand for fish increases as the population increases. Better planning of fisheries development and application of management practices will be the key to the sustainable exploitation of the resources. Knowing the production in advance is important in these aspects. Statistical techniques are capable in providing forecasts with a reasonable accuracy.

Time series models have been used to forecast various phenomena in many fields like environment, economics, tourism, meteorology, agriculture, including fisheries. Among the stochastic time series models, Auto Regressive Integrated Moving Average (ARIMA) model, introduced by [1], is a powerful method of forecasting with low forecast errors. With compared to some statistical techniques, Univariate Box Jenkins method (UBJ) has more advantages in forecasting. UBJ models are derived with a solid foundation of classical probability theory. In univariate forecasting, UBJ models can handle various situations and provide more accurate short term forecasts. Stochastic time-series ARIMA models have the features of parsimonious, stationary, invertible, significant estimated coefficients, statistically independent and normally distributed residuals. It is an extrapolation method for forecasting

and it needs only the historical time series data on the variable under forecasting. Several scientists have used various ARIMA models to forecast agricultural productions [2-5].

In general, an ARIMA model is denoted by ARIMA (*p, d, q*), where *p, d* and *q* denote orders of auto-regression, integration and moving average, respectively and it is suitable for non-stationary series. Non-stationary data can often be made stationary by taking differences of the series.

In general, AR model of order (*p*) is of the form

$$Y_t = \mu + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots\ldots\ldots + \phi_p Y_{t-p} + \varepsilon_t$$

while MA model of order (*q*) is of the form

$$Y_t = \mu + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q} + \varepsilon_t.$$

Then, the ARIMA model of order (*p, d, q*) is of the form

$$Y_t = \mu + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots\ldots\ldots$$
$$+ \phi_p Y_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2-\ldots\ldots\ldots\ldots} + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

where;

Y_t – value at time t

Y_{t-1}, Y_{t-2}, Y_{t-p}– value at time at lag 1 to lag p respectively,

μ -constant,

ϕ s- coefficients of AR part

θ s - coefficients of MA part

ε_t s - random errors.

METHODOLOGY

National annual fish production of Sri Lanka during the period 1972-2016 was used for the analysis. Data from 1972 to 2012 were used for model fitting, while the data from 2013 to 2016 were used for the model validation. Models for data were fitted in four stages, named as identification, estimation, diagnostic, and forecasting stage.

Identification stage

In statistics, a stationary process is a stochastic process whose joint probability distribution does not change when shifted in time or space. Consequently, parameters such as the mean and variance also do not change over time or position. As a result, the mean and the variance of the process do not follow trends.

Stationarity of the series was examined through the time series plot, ACF and PACF plots. If ACF values either cuts off fairly quickly or dies down fairly quickly, time series is considered as a stationary one. If ACF dies down extremely slowly, then the time series should be considered as a non-stationary one. Series of fish production showed non stationarity. Therefore, non-stationarity was converted to stationarity by using the differencing technique.

The first order difference of data is, $Z_t = Y_t - Y_{t-1}$

where,

Z_t – series of first order differences Y_t, Y_{t-1} – response at time at lag 0 and lag 1.

Series of the first order differences was used for the models, because it showed a stationary pattern. The next step was to identify the initial values for the orders of the non seasonal parameters p and q through visual observation of ACF and PACF plots. MA components was identified by visual observation of spikes in ACF graph and the order of the AR component was identify by using PACF.

Estimation stage

Different ARIMA models were fitted for various combinations of p, d and q and the best model was selected based on validity of assumptions and accuracy of forecast.

Diagnostic stage

In this stage, it was confirmed whether model fits the data reasonably well through residual analysis. In order to confirm whether the residuals follow a white noise, the ACF of residuals and the Q statistic [6] were used.

The test statistic Q is given by

$$Q = \frac{n(n+2)\sum r_k^2}{n-k},$$

where,

r_k – residuals autocorrelation at lag k

n – number of residuals

If the P value associated with the Q statistics is small (i.e. $P < \alpha$), the model is considered as an inadequate model. Randomness of the residuals was also confirmed by ACF and PACF of the residuals. The selection of the best model was done based on the lowest accuracy measures from the models which fulfilled all other basic criteria.

Forecasting stage

Finally, selected ARIMA (p, d, q) model was used to forecast the fish production for the period from 2013 to 2022.

RESULTS AND DISCUSSION
Stationarity

Sationarity of the series was identified based on time series plot (Fig-1), ACF (Fig-2) and PACF (Fig-3). According to Fig-1, it is clear that there is an increasing trend in fish production. ACFs in Fig-2 do not drop to zero relatively quickly and they are significant for several lags. A significant spike can be observed in PACF also. These plots confirmed that series of fish production is not stationary.

Trend in series

Time series plot given in Fig-1, shows a positive trend in fish production. During the period from 1972 to 2016, fish production has increased from 101712 MT to 530920 MT with some fluctuations. There can be seen three segments in the series in periods 1972-1983, 1984-2004, and 2005-2016. During this studied the minimum annual fish production of 100702 MT was recorded in 1973 while the maximum of 535050 MT has been recorded in 2014. Reasonable drops in fish production can be observed in 1984 and 2005.
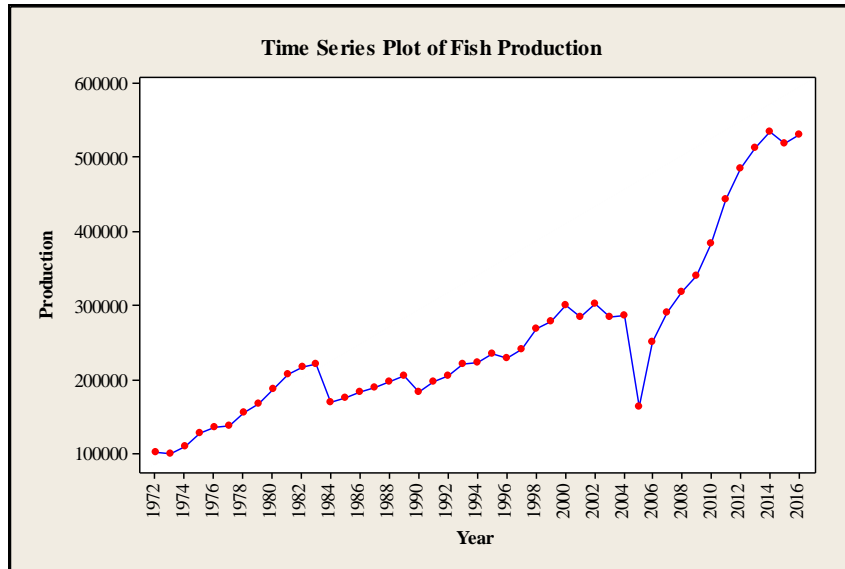
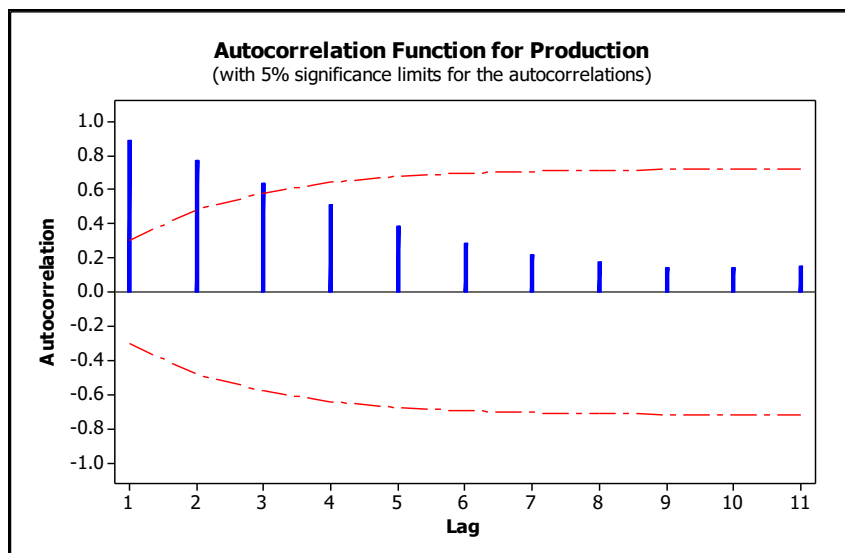**Fig-1: Time plot for the period 1972-2016**
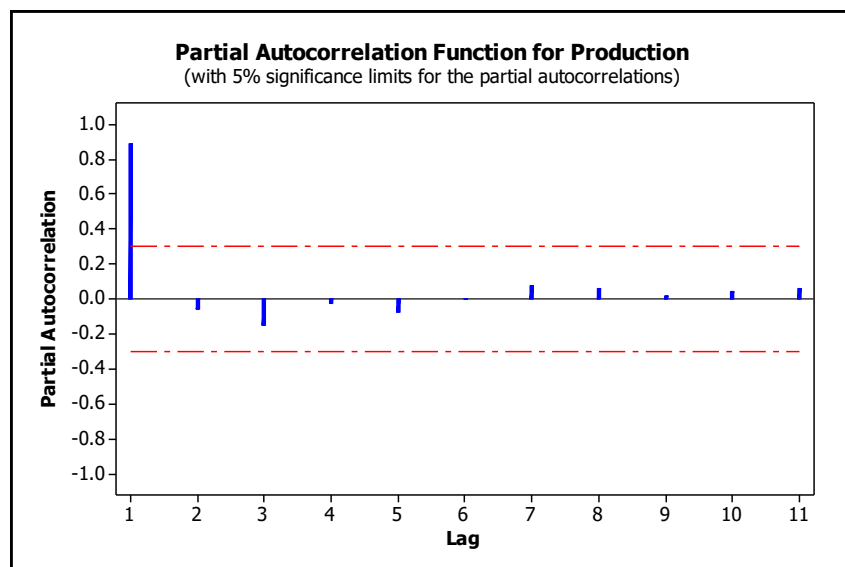

Fig-2:ACF for data from 1972-2016

Fig-3: PACF for data from 1972-2016

Series of first order differences is shown in Fig-4. Series of first order differences fluctuates around zero without any pattern confirming its stationarity.
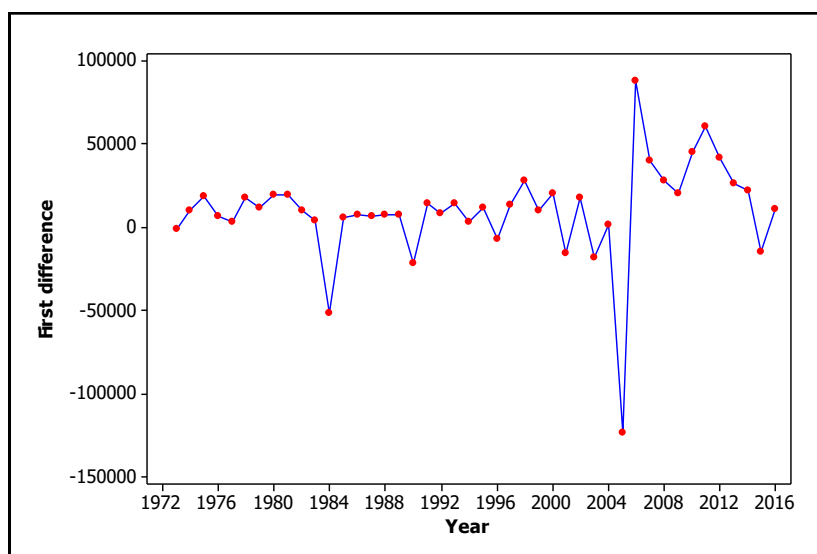

Fig-4: Time plot for series of 1st order differences

Estimation of parameters

Estimates of parameters of the models were obtained by using least squares as advocated by Box and Jenkins using Minitab package. ARIMA (1,1,1) model was selected through several ARIMA models with different combinations of *p*, *d* and *q* parameters, after evaluating residuals and accuracy measures. Parameter estimates for the selected model are given in Table 1.

Table-1: Estimates of parameters

Model	Type	Coefficients	P value
ARIMA (1,1,1)	AR(1)	0.9989	0.000
	MA(1)	0.9561	0.000

Diagnostics of the selected models

Modified Box-Pierce (Ljung-Box) Chi-Square statistic showed that errors are uncorrelated for the selected models and results are shown in Table 2.

Table-2: Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Model	P-values at Lags		
	12	24	36
ARIMA (1,1,1)	0.892	0.954	0.998

ACF for residuals strengthens that errors are uncorelated (Fig-5). This proves that selected ARIMA model is appropriate model to forecast the annual national fish production.
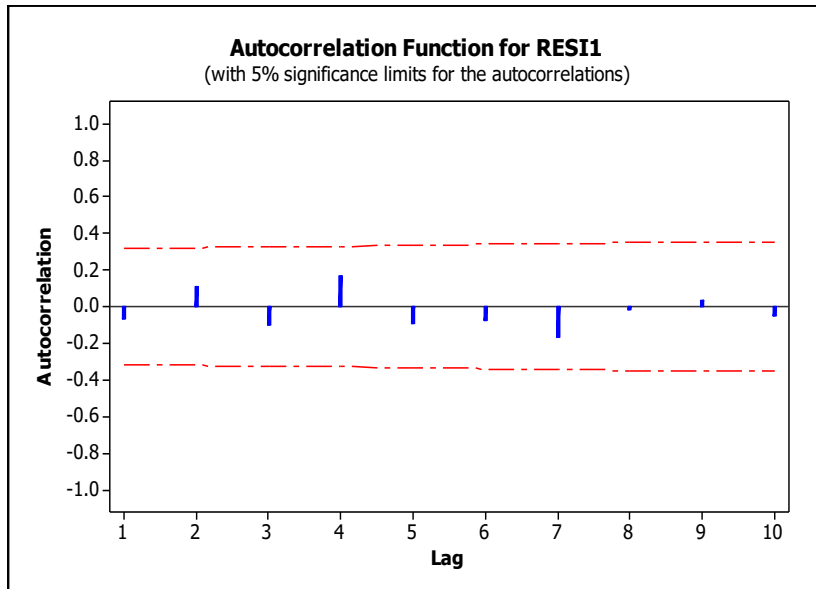
Fig-5: ACF of residual for ARIMA (1,1,1)

Forecasts

Observed and forecasted values of the fish production for the period 2013-2022 are given in Table 3. Forecasted errors for 2013, 2014, 2015 and 2016 are 2.66%, 4.28%, 0.95%, and 1.35% respectively.

Forecasts of the ARIMA (1,1,1) model is shown in Fig-6 with the observed values. Forecasted values lie closer to observed values and their deviations are small. This graph shows the accuracy of forecasts of the selected model.

Table-3: Observed and forecasted values for period from 2013 to 2022

| Year	Obseved value	Forecast values	95% limits		% Error
			Lower	Upper	
2013	512840	499173	436412	561933	2.66
2014	535050	512161	421485	602838	4.28
2015	520190	525137	411716	638558	0.95
2016	530920	538098	404386	671811	1.35
2017		551046	398467	703625	
2018		563980	393446	734515	
2019		576901	389023	764779	
2020		589808	385009	794607	
2021		602702	381277	824127	
2022		615582	377735	853429	
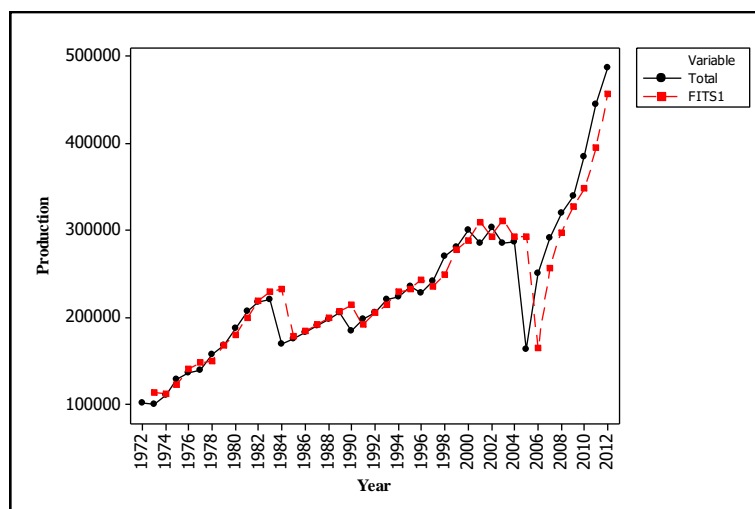
Fig-6: Actual and forecast of national fish production by ARIMA (1,1,1)

CONCLUSION

The most appropriate ARIMA model for predicting the annual fish production of Sri Lanka, was found to be ARIMA (1,1,1). Obtained model predicted the annual fish production of 538098 MT in 2016 with 1.35% of forecast error. Forecast value for year 2017 is 551046 MT with 95 % CI is 398467 MT-703625 MT.

REFERENCES

1. Box GEP, Jenkins GM. Time series analysis: forecasting and control. Holden Day, San Francisco, CA; 1970.
2. Jai Sankar T, Prabakaran R, Senthamarai K, Suresh S. Statistical modeling for cattle production forecasting. Journal of modern Mathematics and statistics. 2010; 4(2): 53-57.
3. Farhan A, Hassnain S, Irum R, Abdul S. Forecasting milk production in Pakistan. Pakistan J. Agric. Res. 2011; 24(1-4):82.
4. Badmus MA, Ariyo OS. Forecasting Cultivated Areas and Production of Maize in Nigerian using ARIMA Model. Asian Journal of Agricultural Sciences. 2011; 3(3):171-176.
5. Jai Sankar T, Prabakaran R. Forecasting milk production in Tamilnadu. International Multidisciplinary Research Journal. 2012; 2(1):10-15.
6. Ljung GM, Box GEP. On a measure of lack of fit in time series models. Biometrika. 1978; 65:297–303.