IJTA © Serials Publications

National Academy of Agricultural Science (NAAS) Rating : 3.03

Study on Modeling and Forecasting of Coconut Production in India

Prema Borkar*

ABSTRACT: *The paper describes an empirical study of modeling and forecasting time series data of coconut production in India. Yearly coconut production data for the period of 1950-1951 to 2012-2013 of India were analyzed by time-series methods. Autocorrelation and partial autocorrelation functions were calculated for the data. The Box Jenkins ARIMA methodology has been used for forecasting. The diagnostic checking has shown that ARIMA (1, 0, 0) is appropriate. The forecasts from 2013-2014 to 2019-2020 are calculated based on the selected model. The forecasting power of autoregressive integrated moving average model was used to forecast coconut production for seven leading years. These forecasts would be helpful for the policy makers to foresee ahead of time the future requirements of coconut production, import and/or export and adopt appropriate measures in this regard.*

Key words: *ACF - autocorrelation function, ARIMA - autoregressive integrated, Moving average, Forecast, PACF - partial autocorrelation function, Coconut.*

The coconut is a benevolent tree, a nature's gift to mankind, as it is a source of food, beverage, oilseed, fibres, timber, health products and also associated with mystery and omen in the life of people. The coconut tree provides clothing utensils and dwellings, therefore, is an important source of earning livelihood to the people of coconut growing states, especially in the coastal areas. The coconut tree therefore, is eulogized, reverently as "Kalpavruksha" or tree of life by the people.

The coconut crop is grown in eighteen States and three Union Territories covering an area of 2.136 million hectares of land, with a production of 22,680 million nuts in the country. The major coconut crop acreage is concentrated on the West Coast region of the country comprising the states of Kerala, Karnataka and Maharashtra, followed by East Coast of Tamil Nadu, Andhra Pradesh, Orissa and Pondicherry. The coconut cultivation areas also traditionally located in the coastal region of Gujarat, Goa, West Bengal, Islands of Andaman & Nicobar and Lakshadweep. About 90 per cent of the area of coconut cultivation and equally the same per cent of production of coconut are from the four Southern states, viz. Kerala, Karnataka, Tamil Nadu and Andhra Pradesh. Kerala is considered as the land of coconut and holds the key for the development of coconut production and marketing in the country.

Since, the production and marketing scenario of coconut in the country has witnessed a phenomenal development, particularly in the field of production such as development of improved high yielding dwarf varieties of crossbred coconut palm, traditional, non-traditional, commercial and industrial coconut product. Hence the present study was undertaken to forecast the coconut production in India.

Forecasts have traditionally been made using structural econometric models. Alteration has been given to the univariate time series models known as auto regressing integrated moving average (ARIMA) models, which are primarily due to the work of Box and Jenkins (1970). These models have been extensively used in practice for forecasting economic time series, inventory and sales modeling (Brown, 1959; Holt *et al.*, 1960) and are generalization of the exponentially weighted moving average process. Several methods for identifying special cases of ARIMA models have been suggested by Box and Jenkins and others. Makridakis *et al.*, (1982), and Meese and Geweke (1982) have discussed the methods of identifying univariate models. Among others Jenkins and Watts (1968), Yule (1926, 1927), Bartlett (1964), Quenouille (1949), Ljune and Bos (1978) and Pindyck and Tubinfeld (1981) have also emphasized the use of ARIMA models.

* Assistant Professor, Gokhale Institute of Politics and Economics (Deemed to be a University), Pune – 411004. Maharashtra, India.

In this study, these models were applied to forecast the production of coconut crop in India. This would enable to predict expected coconut production for the years from 2014 onward. Such an exercise would enable the policy makers to foresee ahead of time the future requirements for coconut production, import and/or export of coconut thereby enabling them to take appropriate measures in this regard. The forecasts would thus help save much of the precious resources of our country which otherwise would have been wasted.

MATERIALS AND METHODS

Respective time series data for this study were collected from various government publications of India. Box and Jenkins (1976) linear time series model was applied. Auto Regressive Integrated Moving Average (ARIMA) is the most general class of model for forecasting a time series. Different series appearing in the forecasting equations are called "Auto-Regressive" process. Appearance of lags of the forecast errors in the model is called "moving average" process. The ARIMA model is denoted by ARIMA (*p*, *d*, *q*),

Where,

"*p*" stands for the order of the auto regressive process,

"*d*" is the order of the data stationary and

"*q*" is the order of the moving average process.

The general form of the ARIMA (*p*, *d*, *q*) can be written as described by Judge, *et al*., (1988).

$$* \quad \Delta^d y_t = \delta + \theta_1 \Delta^d y_{t-1} + \theta_2 \Delta^d y_{t-2} + \ldots + \theta_p y_{t-p} + e_{t-1} \alpha e_{t-1} - \alpha_2 e_{t-2} \alpha_q e_{t-2} \quad \ldots (1)$$

Where,

Δ^d denotes differencing of order d, i.e.,

$\Delta y_t = y_t - y_{t-1}$,

$\Delta_2 y_t = \Delta y_t - \Delta_{t-1}$ and so forth,

$Y_{t-1} \ldots y_{t-p}$ are past observations (lags),

$\delta, \theta_1 \ldots \theta_p$ are parameters (constant and coefficient) to be estimated similar to regression coefficients of the Auto Regressive process (AR) of order "*p*" denoted by AR (*p*) and is written as

$$Y = \delta + \theta_1 y_{t-1} + \theta_2 y_{t-2} + \ldots + \theta_p y_{t-p} + e_t \quad \ldots (2)$$

Where,

e_t is forecast error, assumed to be independently distributed across time with mean θ and variance $\theta_2 e$, $e_{t-1}, e_{t-2} \ldots e_{t-q}$ are past forecast errors,

$\alpha_1, \ldots \alpha_q$ are moving average (MA) coefficient that needs to be estimated.

While MA model of order *q* (i.e.) MA (*q*) can be written as

$$Y_t = e_t - \alpha_1 \alpha_{t-1} - \alpha_2 e_{t-2} \ldots \alpha_q e_{t-q} \quad \ldots (3)$$

The major problem in ARIMA modeling technique is to choose the most appropriate values for the *p*, *d*, and *q*. This problem can be partially resolved by looking at the Auto correlation function (ACF) and partial Auto Correlation Functions (PACF) for the series (Pindyk & Rubinfeld, 1991). The degree of the homogeneity, (*d*) i.e. the number of time series to be differenced to yield a stationary series was determined on the basis where the ACF approached zero.

After determining "*d*" a stationary series $\Delta d y_t$ its auto correlation function and partial autocorrelation were examined to determined values of p and q, next step was to "estimate" the model. The model was estimated using computer package "SPSS".

Diagnostic checks were applied to the so obtained results. The first diagnostic check was to draw a time series plot of residuals. When the plot made a rectangular scatter around a zero horizontal level with no trend, the applied model was declared as proper. Identification of normality served as the second diagnostic check. For this purpose, normal scores were plotted against residuals and it was declared in case of a straight line. Secondly, a histogram of the residuals was plotted. Finding out the fitness of good served as the third check. Residuals were plotted against corresponding fitted values: Model was declared a good fit when the plot showed no pattern.

Using the results of ARIMA (*p*, *q*, *d*), forecasts from 2014 up to 2020 were made. These projections were based on the following assumptions.

- Absence of random shocks in the economy, internal or external.
- Agricultural price structure and polices will remain unchanged.
- Consumer preferences will remain the same.

RESULTS AND DISCUSSION

Building ARIMA model for coconut production data in India

To fit an ARIMA model requires a sufficiently large data set. In this study, we used the data for coconut production for the period 1950-1951 to 2012-2013. As we have earlier stated that development of ARIMA model for any variable involves four steps: identification, estimation, diagnostic checking and forecasting. Each of these four steps is now explained for coconut production. The time plot of the coconut production data is presented in Figure 1.

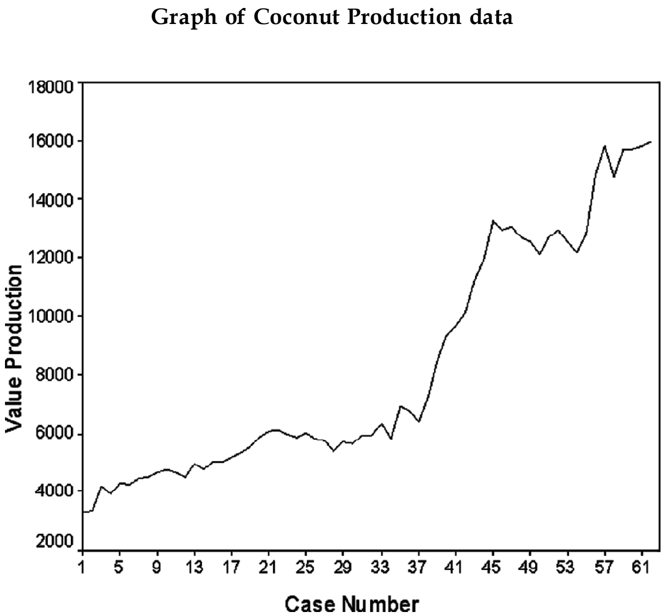

Figure 1: Time plot of coconut production data

The above time plot indicated that the given series is non-stationary. Non-stationarity in mean is corrected through appropriate differencing of the data. In this case difference of order 1 was sufficient to achieve stationarity in mean.

The newly constructed variable X_t can now be examined for stationarity. The graph of X_t was stationary in mean. The next step is to identify the values of p and q. For this, the autocorrelation and partial autocorrelation coefficients of various orders of X_t are computed (Table 1). The ACF and PACF (Fig. 2 and 3) shows that the order of p and q can at most be 1. We entertained three tentative ARIMA models and chose that model which has minimum

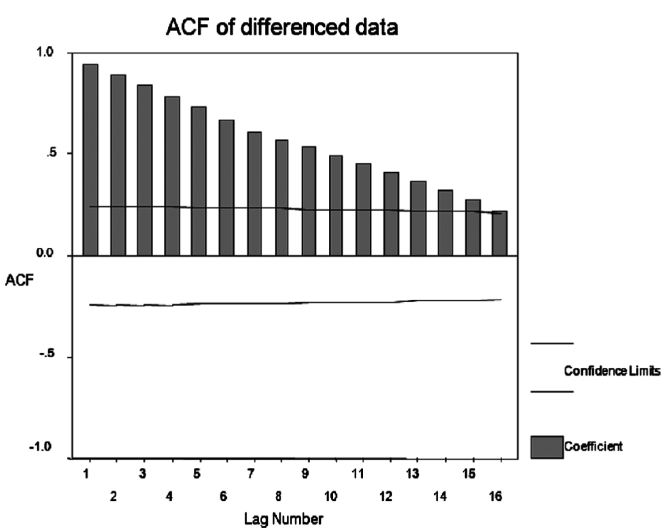

Figure 2: ACF of differenced data

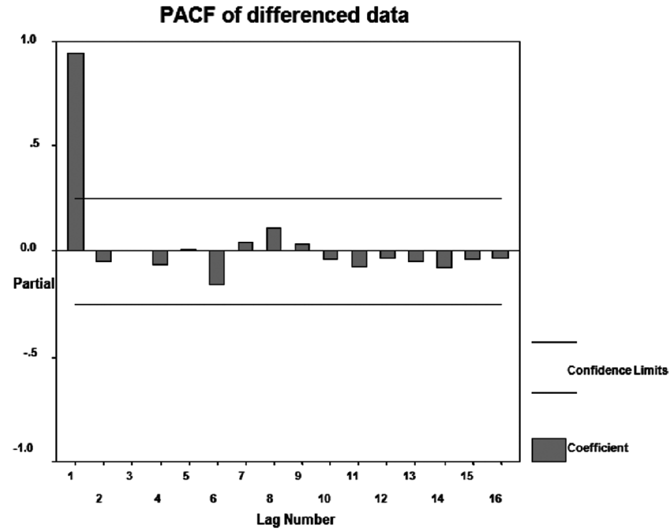

Figure 3: PACF of differenced coconut data

AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion). The models and corresponding AIC and BIC values are

ARIMA (p, d, q)	*AIC*	*BIC*
1 0 0	971.21	975.46
1 1 1	973.17	979.51
1 1 0	972.13	976.36

So the most suitable model is ARIMA (1, 0, 0) this model has the lowest AIC and BIC values.

Table 1
Autocorrelations and partial autocorrelations

Lag	*Autocorrelation*	*Std. error*	*Lag*	*Partial Autocorrelation*	*Std. error*
1	0.947	0.124	1	0.947	0.127
2	0.892	0.123	2	-0.050	0.127
3	0.839	0.122	3	-0.002	0.127
4	0.783	0.121	4	-0.062	0.127
5	0.731	0.120	5	0.008	0.127
6	0.667	0.119	6	-0.154	0.127
7	0.610	0.118	7	0.043	0.127
8	0.569	0.117	8	0.111	0.127
9	0.533	0.116	9	0.029	0.127
10	0.496	0.114	10	-0.043	0.127
11	0.453	0.113	11	-0.074	0.127
12	0.412	0.112	12	-0.029	0.127
13	0.371	0.111	13	-0.049	0.127
14	0.324	0.110	14	-0.079	0.127
15	0.275	0.109	15	-0.036	0.127
16	0.223	0.108	16	-0.030	0.127

Model parameters were estimated using SPSS package. Results of estimation are reported in Table 2. The model verification is concerned with checking the residuals of the model to see if they contain any

Table 2
Estimates of the fitted ARIMA model

		Estimates	Std Error	t	Approx sig.
Non- Seasonal lag	AR1	0.99498	0.00853	116.5847	0.0000
	AR2	0.34106	0.12038	2.83312	0.0062
	MA1				
Constant		9437.18	5379.02	1.75444	0.08446
Number of Residuals		63			
Number of Parameters		2			
Residual df		60			
Adjusted Residual Sum of Squares		21608240.2			
Residual Sum of Squares		26545707.4			
Residual Variance		334366.45			
Model Std. Error		578.24428			
Log-Likelihood		-483.60618			
Akaike's Information Criteria (AIC)		971.21236			
Schwarz's Bayesian Criterion (BIC)		975.46662			

systematic pattern which still can be removed to improve on the chosen ARIMA. This is done through examining the autocorrelations and partial autocorrelations of the residuals of various orders. For this purpose, the various correlations up to 16 lags were computed and the same along with their significance which is tested by Box-Ljung test are provided in Table 3. As the results indicate, none of these correlations is significantly different from zero at a reasonable level. This proves that the selected ARIMA model is an appropriate model. The ACF and PACF of the residuals (Fig. 4 and 5) also indicate 'good fit' of the model.

The last stage in the modeling process is forecasting. ARIMA models are developed basically to forecast the corresponding variable. There are two

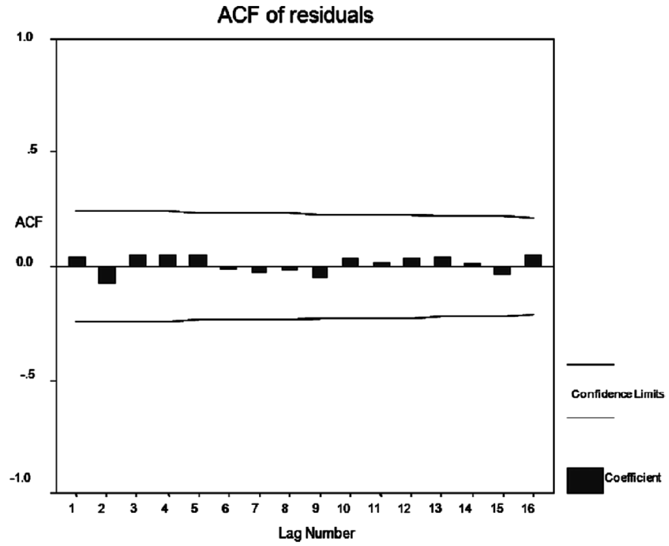

Figure 4: ACF of residuals of fitted ARIMA model

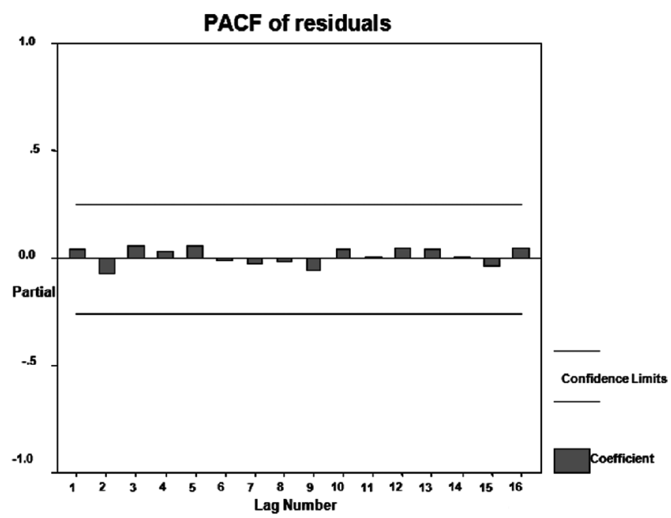

Figure 5: PACF of residuals of fitted ARIMA model

Table 3
Autocorrelations and partial autocorrelations of residuals

Lag	*Autocorrelation*	*Std.error*	*Box- Ljung*	*Df*	*Sig.*	*Lag*	*PartialAutocorrelation*	*Std.error*
1	0.042	0.124	0.113		0.737	1	0.042	0.127
2	-0.070	0.123	0.441		0.802	2	-0.072	0.127
3	0.047	0.122	0.590		0.899	3	0.054	0.127
4	0.044	0.121	0.724		0.948	4	0.035	0.127
5	0.049	0.120	0.889		0.971	5	0.053	0.127
6	-0.006	0.119	0.892		0.989	6	-0.008	0.127
7	-0.024	0.118	0.932		0.996	7	-0.020	0.127
8	-0.013	0.117	0.944		0.999	8	-0.019	0.127
9	-0.049	0.116	1.126		0.999	9	-0.055	0.127
10	0.035	0.114	1.217		1.000	10	0.038	0.127
11	0.013	0.113	1.230		1.000	11	0.006	0.127
12	0.034	0.112	1.321		1.000	12	0.048	0.127
13	0.037	0.111	1.431		1.000	13	0.037	0.127
14	0.006	0.110	1.434		1.000	14	0.009	0.127
15	-0.029	0.109	1.505		1.000	15	-0.035	0.127
16	0.051	0.108	1.727		1.000	16	0.045	0.127

kinds of forecasts: sample period forecasts and post-sample period forecasts. The former are used to develop confidence in the model and the latter to generate genuine forecasts for use in planning and other purposes. The ARIMA model can be used to yield both these kinds of forecasts. The residuals calculated during the estimation process, are considered as the one step ahead forecast errors. The forecasts are obtained for the subsequent agriculture years from 2013-14 to 2019-2020.

In our study, the suitable model for coconut production was found to be ARIMA (1, 0, 0). The forecasts of coconut production, lower control limits (LCL) and upper control limits (UCL) are presented in table 4. The validity of the forecasted values can be checked when the data for the lead periods become available. The model can be used by researchers for forecasting of coconut production in India. However, it should be updated from time to time with incorporation of current data.

This paper forecast future coconut production based on the data from 1950-51 to 2012-13, using ARIMA model. The forecast will help policy makers to design future coconut production strategies.

Table 4
Forecasts for Coconut Production (2013-14 to 2019-2020)

(Nuts per hec)

Years	*ForecastedProduction*	*Lower limit*	*Upper limit*
2013-2014	15898.09	14262.88	17533.32
2014-2015	15865.65	13865.77	17865.54
2015-2016	15833.37	13527.38	18139.36
2016-2017	15801.25	13226.73	18375.78
2017-2018	15769.29	12953.04	18585.55
2018-2019	15737.49	12699.89	18775.11
2019-2020	15705.86	12463.11	18948.62

REFERENCES

Bartlett M. S., (1964), "On The Theoretical Specification of Sampling Properties of Autocorrelated Time Series", *J. Roy. Stat. Soc.,* **B8**: 27-41.

Box G. E. P., and G. M. Jenkins, (1970), Time series analysis: forecasting and Control, *Holden Day, San Francisco, CA.*

Box G. E. P., and G. M. Jenkins, (1976), Time Series Analysis: Forecasting and Control, *Rev. Ed. San Francisco. Holden-Day.*

Brockwell P. J., and Davis R. A., (1996), Introduction to time series and Forecasting, *Springer.*

Brown R. G., (1959), Statistical Forecasting For Inventory Control. *New York,* McGraw-Hill.

Iqbal N., K. Bakhsh, A. Maqbool, and A. S. Ahmad, (2005), Use of the ARIMA Model for forecasting Wheat area and production in Pakistan", *Journal of Agriculture & Social Sciences*., **2**: 120-122.

Jenkins G. M., and D. G. Watts, (1968), Spectral Analysis and its Application, Day, San Francisco, *California, USA.*

Ljunge G. M., and G. E. P. Box, (1978), "On a Measure of Lack of Fit in Time Series Models", *Biometrika*, **65**: 67-72.

Kendall M. G., and A. Stuart, (1966), The advanced theory of statistics. Vol. 3, Design and Analysis and Time-Series, Charles Griffin & Co. Ltd., London, United Kingdom.

Holt C. C., F. Modigliani, J. F. Muth, and H. A. Simon, (1960), Planning, Production, Inventores, and Work Force. *Prentice Hall, Englewood Cliffs, NJ,* USA.

Muhammad F., M. S. Javed, and M. Bashir, (1992), Forecasting Sugarcane Production in Pakistan using ARIMA Models, *Pak. J. Agri. Sci.,* **9**(1): 31-36.

Makridakis S., A. Anderson, R. Filds, M. Hibon, R. lewandowski, J. Newton, E. Parzen, and R. Winkler, (1982), The Accuracy of Extrapolation (time series) methods: Results of a Forecasting Competition, *Journal of Forecasting Competition*, *J. Forecasting,* **1**: 111-153.

Meese R., and J. Geweke, (1982), A Comparison of Autoregressive Univariate Forecasting Procedures for Macroeconomic Time Series, Unpublished Manuscript, University of California, Berkeley, CA, USA.

Prindycke R. S., and D. L. Rubinfeld, (1981), Econometric Models and Economic Forecasts, *2nd Ed. New York, McGraw-Hill.*

Quenouille M. H., (1949), "Approximate Tests of Correlation in Time- Series, *J. Roy. Stat. Soc.,* **B11**: 68-84.

Saeed N., A. Saeed, M. Zakria, and T. M. Bajwa, (2000), Forecasting of Wheat Production in Pakistan using ARIMA models, *International Journal of Agriculture and Biology*., 4: 352-353.

Yule G. U., (1926), "Why Do We Sometimes Get Nonsence-corrleations Between Times Series. A study in Sampling and the Nature of Series", *J. Roy. Stat. Soc*., **89**: 1-69.

Yule G. U., (1927), "On a method of Investigation Periodicities in Disturbed Series, With Specia; Reference To Wolfer's Sunspot Number", *Phill. Trans*., **A226**: 267-298.