International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Impact Factor (2012): 3.358

Forecasting of Areca Nut (*Areca catechu*) Yield Using Arima Model for Uttara Kannada District of Karnataka

Srinath Reddy A. B.[1], Havaldar Y. N.[2], Pavan Kumar S.T[3], Adam Kamei[4]

[1]Research Associate A.R.S Ananthapur, Andhra Pradesh

[2]Associate Professor of Agriculture Statistics, University of Agricultural Sciences, Dharwad-580005

[3, 4]Ph.D Research Scholar, Department of Agricultural Statistics,
Bidan Chandra Krishi Vishwavidyala, Mohanpur, West Bengal-741252

Abstract: ***Arecanut (Areca catechu) also popular by name supari or betelnut. It is one of important commercial crop in Uttara Kannada district, contributing around 9 percent of area and 11 percent of production to the Karnataka state total. The present study is based on the secondary data of over 30 years collected from Directorate of Economics and Statistics. The prediction of arecanut production on a yearly time scales has been attempted by various research groups using different techniques models. Among the most effective approaches for analyzing time series data is the model introduced by Box and Jenkins, ARIMA (Autoregressive Integrated Moving Average). The selection of models was done based on their R^2 value and root mean square error (RMSE) value after analyzing the data, and further these models were used for prediction of arecanut production. From the different (p.d.q) models, ARIMA (1, 1, 5) was selected based on RMSE (0.325) and normalized BIC (-1.319) values for forecasting the production of arecanut in Uttara Kannada district. The model parameters were estimated using SPSS software and hence it was taken as best fitted model and forecasting has been done***

Keywords: Mean square error, Nonlinear, ARIMA, SPSS, AIC, BIC

1. Introduction

Many methods and approaches for formulating forecasting models are available in the literature. This research exclusively deals with time series forecasting model, in particular, the Auto Regressive Integrated Moving Average (ARIMA). These models were described by Box and Jenkins and further discussed in some other resources such as Walter. The prediction of production of arecanut on yearly time scales is not only scientifically challenging but is also important for planning and devising agricultural strategies. Time series analysis and forecasting has become a major tool in different applications in agriculture, horticulture, hydrology and environmental management fields.

In Karnataka, as per the State Horticulture Department source, around 4.55 lakh acres (1.84 lakh hectares) is under arecanut cultivation which forms around 46 percent of all India total. Its contribution to total production is around 2.24 lakh a ton that forms 47 percent of all India production in 2009-10. It is important to note that arecanut cultivation is undertaken with varying Extent in almost 28 out of 30 districts in Karnataka. Among which, Chikmagalur district stands first in both area (20 %) and production (17%), Shimoga stands second followed by Davanagere district. The top 7 districts viz. Chikmagalur, Shimoga, Davangere, Dakshina Kannada, Tumkur, Chitradurga and Uttar Kannada occupy 89 per cent of the area under arecanut and contribute around 91 per cent of areca produced in the state. Where Uttar Kannada contributes 7 percent out of 89 percent to tha Karnataka state. Small variations in the timing and the quantity of arecanut production have the potential to impact on agricultural output. Prior knowledge of production behavior will help Uttara Kannada farmers and also policy makers. So, the present study was taken to forecast the production well in advance.

2. Materials and Methods

The study was undertaken in Uttar Kannda district which is situated roughly in the mid North Western part of the State. The district lies between 13^0 55' and 15^0 31' north latitude and between 74^0 9' and 75^0 10' eastern longitude. The present study is based on the secondary data of over 30 years collected from Directorate of Economics and Statistics, Bangalore. Among the different time series analysis methods Box and Jenkins, ARIMA (Autoregressive Integrated Moving Average) was selected. ARIMA time-series models traditionally expressed as ARIMA (p,d,q) combine as many as 3 types of processes, *viz.* autoregression (AR) of order p; differencing d times to make a series stationary and moving average (MA) of order q. It may be pointed out that this methodology applies only to stationary data, a characteristic of which is that the mean and variance are constant over time. Three stages for carrying out the analysis are i) Identification ii) Estimation and iii) Diagnostic checking.
At the identification stage, two graphical devices, *viz.* estimated function (PACF) $\Box_{kk}$ are used with a view to tentatively select one or more candidate ARIMA models. Denoting the number of observations by n and the number of computable lags by $k_{(l,c)}$,

$$r_k = \sum_{t=1}^{n-k} (Z_i - \bar{Z})(Z_{ttk} - \bar{Z}) / \sum_{t=1}^{n-k} (Z_t - \bar{Z})^2$$

and

$$\phi_{kk} = [r_k - \sum_{j=1}^{k-1} \hat{\phi}_{k-1'j} r_{k-j}] / [1 - \sum_{j=1}^{k-1} \phi_{k-1'j} r_j]; k = 1,2,3.....$$

Where

$$\hat{\phi}_{11} = r_1, \hat{\phi}_{kj} = \hat{\phi}_{k-i,j} - \hat{\phi}_{kk} \hat{\phi}_{k-1'k-j}; k = 3,4.... j = 2,3$$

If the r_k values tail off to zero rapidly, it indicates that the original series is stationary. Otherwise successive differences, are computed and the above procedure is continued till stationary it is achieved. Patterns based on spikes in □kk and $^r{}_k$ values are used to select the appropriate values of p and q which are, respectively, the orders of AR and MA. Finally, at the diagnostic checking stage, an appropriate model is selected based on the following goodness of fit statistics

i) Akaike's Information criterion (AIC):

AIC = ln V*(p,q) + (2/n)(p + q),

Where V* is an estimate of white noise variance obtained by fitting the corresponding ARIMA model. Mean Bayesian Information criterion (BIC): absolute error (MAE): Root mean squared error (RMSE):

BIC = lnV* (p,q) + (p + q) [ln(n)/n]

$$RMSE = [\sum_{t=1}^{n} (Z_t - \hat{Z}_t)^2 / n]^{1/2}$$

$$MAE = \sum_{t=1}^{n} | Z_t - \hat{Z}_t | / n$$

The lower the values of above statistics, the better are the model. A statistically adequate model is one whose random shocks are independent. Main stages in setting up a Box-Jenkins forecasting model are as follows Identification, Estimating the parameters, diagnostic checking and Forecasting

3. Identification of Models

A good starting point for time series analysis is a graphical plot of the data. It helps to identify the presence of trends. Before estimating the parameter (p, q) of model, the data are not examined to decide about the model which best explains the data. This is done by examining the sample ACF (Autocorrelation function) and PACF (Partial Autocorrelation function) of differenced series YB_{tB}. The sample auto correlations for k time lags can be found and denoted by $rB_{k\,B}$as follows.

$$\hat{\rho}(Y_t) = r_k(Y_t) \quad 3.11$$

$$= \frac{C_k(Y_t)}{C_0(Y_t)}$$

Where,

$$C_k(Y_t) = \frac{1}{n} \sum_{t-1}^{n-k} (Y_t - \bar{Y})(Y_{t+k} - \bar{Y})$$

K = 0, 1, 2,.... n
T = 1, 2,... n-k

$$Y_t = \frac{1}{n} \sum_{t=1}^{n} Y_t$$

n = Length of time period

Both ACF and PACF are used as the aid in the identification of appropriate models. There are several ways of determining the order type of process, but still there was no exact procedure for identifying the model.

4. Estimation of Parameters

After tentatively identifying the suitable model, next step is to obtain Least Square Estimates of the parameters such that the error sum of squares is minimum.

S (θ, ∅) = Σ eB_{tB}^2 (θ,∅) 3.12

Where,

t = 1, 2, 3... n

There are fundamentally two ways of getting estimates for such parameters. Trial and error: Examine many different values and choose set of values that minimizes the sum of squares residual Interactive method: Choose a preliminary estimate and let a computer programme refine the estimate interactively. The latter method is used in our analysis for estimating the parameters.

5. Diagnostic Checking

After having estimated the parameters of a tentatively identified ARIMA model, it is necessary to do diagnostic checking to verify that the model is adequate.

Examining ACF and PACF of residuals may show an adequacy or inadequacy of the model. If it shows random residuals, then it indicates that the tentatively identified model was adequate. When an inadequacy is detected, the checks should give an indication of how the model need be modified, after which further fitting and checking takes place.

One of the procedures for diagnostic checking mentioned by Box-Jenkins is called over fitting *i.e.* using more parameters than necessary. But the main difficulty in the correct identification is not getting enough clues from the ACF because of inappropriate level of differencing. The residuals of ACF and PACF considered random when all their ACF were within the limits.

$$\mu \pm 1.96 \sqrt{1/(n-12)}$$

The minimum Akike Information Coefficient (AIC) criterion is used to determine both the differencing order (d, D) required attaining stationarity and the appropriate number of AR and MA parameters, it can be computed as follows.

$$AIC_{(p+q)} = \{(1+\log 2\pi)+n\log\sigma^2+2m\}$$

Where,

σ^2 = Estimated MSE

n = Number of observations

m = p + q + P + Q

This diagnostic checking helps us to identify the differences in the model, so that the model could be subjected to modification, if need be.

6. Forecasting

After satisfying about the adequacy of the fitted model, it can be used for forecasting. Forecasts based on the model.

(1-□B) (1-□B)P^{sP} YB_{tB} = (1-θB) (1-(H)P^{sP}B) eB_{tB}

were computed for upto 4 months (m) ahead. The above model gives the forecasting equation as

YB_{tB} = □YB_{t-1B} + □YB_{t-12B} - □□YB_{t-13B} + eB_{tB}-θeB_{t-1B} – (H) eB_{t-12B} + θ (H) eB_{t-13B}

Given data upto time‘t’ the optional forecast of Y (also called Ex-Ante forecast) model at the t is the conditional expectation of YB_{t+1B}. It follows, in particular, that

$$e_t = Y_t - Y_{t-1}$$

The errors eB_{tB} in model are in fact that forecast errors for unit lead time. That for an optimal forecast these ‘one step ahead’ forecast errors ought to form an uncorrelated series is otherwise obvious. Suppose, if these forecast errors were autocorrelated and then it could be possible to forecast the next forecast error in which case it could not be optimal.

The required expectations are easily found because

$$E(Y_{t+m}) = Y_t(m), E(e_{t+m}) = 0$$

Where,

m = 1, 2, 3..... n

$$E(Y_{t-m}) = Y_{t-m} E(e_{t-m}) = a_{t-m} = Y_{t-m} - Y_{t-m-1}$$

Where, m = 0, 1, 2... n

For instance, to determine the three month ahead (1-3) forecast for series YB_{tB} (use equation

YB_{t+1B} = YB_{t+3B}

= ∅YB_{t+2B} + ϕYB_{t-9B} - ∅ϕYB_{t-10B} + eB_{t+13B} - θeB_{t-2B} – (H) eB_{t-9B} + θ(H) eB_{t-10B}

Taking conditional expectations at time t,

YB_{tB} (1) = YB_{tB} (3)

= ∅ $YB_{t(2)B}$ + ϕyB_{t-9B}-∅ ϕYB_{t-10B}+0-θ (0)-(H) (YB_{t-9B}-YB_{t-10B})+θ(H) (YB_{t-10B}–YB_{t-11B})

Because, $E(e_{t+1)} = 0, E(e_{t-1}) = Y_{t-1} - \hat{Y}_{t-1} = e_{t-1}$

i.e. YB_{tB} (3) = 0 YB_{tB} (2)

The forecast $YB_{t B}$(2) can be obtained in a similar way in terms of $YB_{t B}$(1) from E (YB_{t+2B}). Similarly YB_{tB} (1) can be obtained from E (YB_{t+1B}). In practice it is very easy to compute the forecast YB_{tB} (1), YB_{tB} (2), YB_{tB} (3) *etc.* recursively using the forecast function.

E (YB_{t+1B}) = E(YB_{t+1B}-1 + QB_{t+1B}- eB_{t+1B}-1) - θ eB_{t+1B}-1 – (H) eB_{t+1B}-12 + θ (H) eB_{t+1B}-13 and using 3.18 and 3.19.

However, using these methods, Ex-post forecasts can also be calculated for comparing with the value actually realized. The accuracy of forecasts for both Ex-ante and Ex-post were tested using the following tests MAPE and RMSE.

7. Results and Discussion

As Box-Jenkin model was preferred to the multiplicative time series model for forecasting purpose. It was used for forecasting of arecanut production. The results were presented below. The detailed analysis of forecasting of production of arecanut in Uttara Kannada district has been presented as under.

The tentative models were first identified based on the Auto Correlation Function (ACF) and Partial Auto Correlation Function (PACF) then the suitable model was selected. The computed values of ACF and PACF of arecanut production were 23 lags. Since the coefficient dropped to zero after the first or second lag. Each individual coefficient of ACF and PACF were tested for their significance using‘t’ test. The absence of peak at first values clearly indicate suitability of the choice of d = 1, to accomplish stationary series. Hence, based on ACF and PACF many models were tried, finally model (1, 1, 5) was tentatively identified based on R^2 value (0.979) for production of arecanut. From the different (p.d.q) models, ARIMA (1, 1, 5) was selected based on MAPE (1.626) and normalised BIC (-1.319) values for forecasting the production of arecanut in Uttara Kannada district. The model parameters were estimated using SPSS software and hence it was taken as best fitted model and forecasting has been done. The results were presented in the Table.

Results of ARIMA model to the production of Arecanut

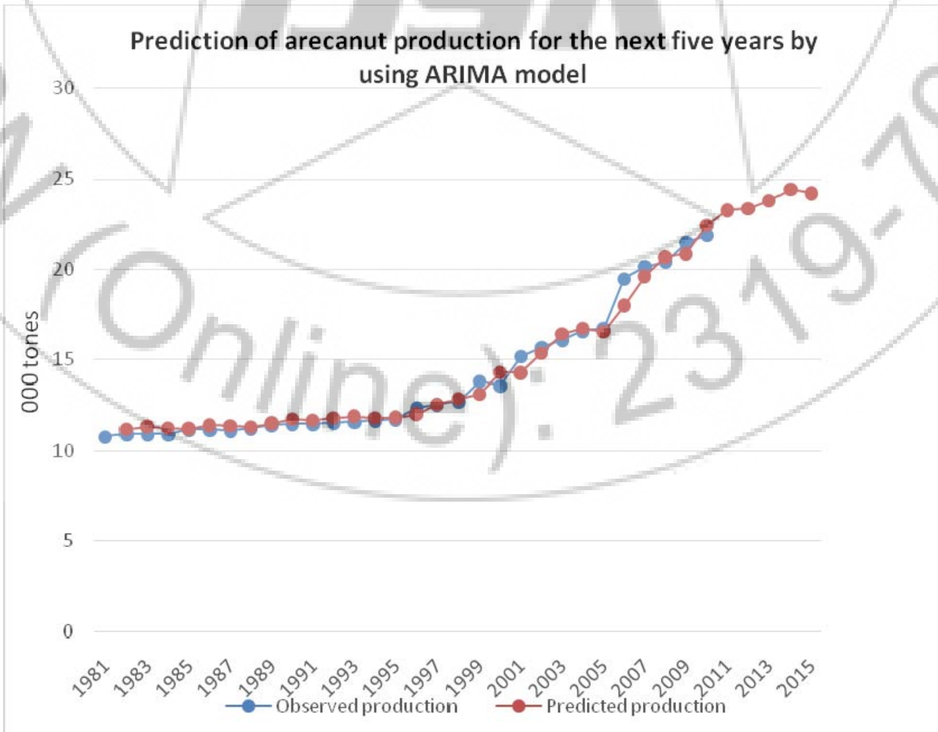
Fitted Models	R^2	RMSE	MAPE	MAE	MaxAPE	MaxAE	Normalized BIC
ARIMA (1,1,5)	0.979	0.325	1.626	4.716	0.211	0.688	-1.319
ARIMA (0,1,2)	0.974	0.329	1.791	5.419	0.234	0.822	-1.827

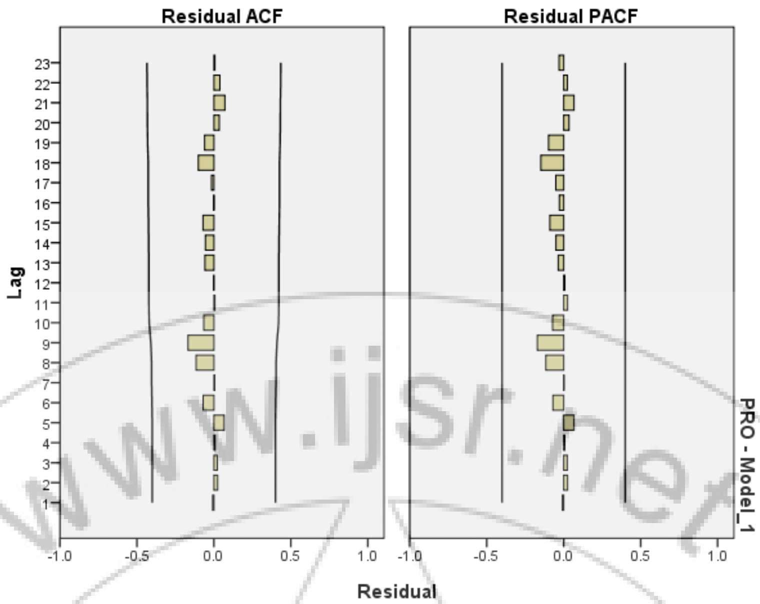
The predicted values of production of arecanut for the next five years (2011, 2012, 2013, 2014 and 2015) were represented graphically and shown in the figure. In this study we gave large scale comparison of different models in order to know the best model for the forecasting of the arecanut production. For the model comparison, yearly production of arecanut was considered. The comparison of all the 12 models was carried out in the process based on the MAPE and RMSE values which w.ere considered to be least. According to the Table 4.21 which represents the MAPE and RMSE values, the ARIMA model with least MAPE value of 1.626 per cent and RMSE value of 0.325 was considered as best fit among all the models considered. The suitable model was ARIMA(1,1,5) for forecasting the production of arecanut from (1981-2015).Based on ACF and PACF many models were tried, finally model (1, 1, 5) was tentatively identified based on R^2 value (0.979) for forecasting of arecanut production in Uttara Kannada district. The results were presented in the Table.

Forecasting of production was done based on the model (1, 1, 5). The predicted values of production of arecanut for the next five years (2011, 2012, 2013, 2014 and 2015) were represented graphically and shown in the figure Similar study was done by Hamilton (1).

Prediction of arecanut production for the next five years by using ARIMA model

Year	Observed production (in 000 tones)	Predicted production (in 000 tones)
1981	10.804	
1982	10.952	11.18644
1983	10.907	11.32721
1984	10.915	11.2382
1985	11.147	11.20174
1986	11.158	11.40418
1987	11.139	11.37861
1988	11.256	11.28056
1989	11.444	11.51006
1990	11.492	11.75461
1991	11.514	11.68785
1992	11.572	11.79112
1993	11.592	11.90131
1994	11.669	11.80186
1995	11.762	11.82161
1996	12.373	12.00959
1997	12.562	12.56094
1998	12.727	12.84208
1999	13.808	13.08949
2000	13.59	14.29218
2001	15.169	14.26708
2002	15.668	15.35433
2003	16.072	16.41453
2004	16.552	16.73028
2005	16.701	16.53599
2006	19.47	17.9902
2007	20.135	19.6236
2008	20.373	20.67204
2009	21.47	20.87289
2010	21.854	22.45392
2011		23.2954
2012		23.39068
2013		23.81723
2014		24.43335
2015		24.23789





ACF and PACF of residuals of fitted ARIMA model for arecanut production

References

[1] Hamilton, J.D., 1994. Time Series Analysis. Princeton University Press, Princeton, pp: 7-71.

[2] Ahmad, B., A. Ghafoor and H. Badar, 2005. Forecasting and growth trends of production and export of kinnow from Pakistan. J. Agric. Soc. Sci., 1: 20-24

[3] Box, G.E.P. and G.M. Jenkin, 1976. Time Series of Analysis, Forecasting and Control, Sam Franscico, Holden-Day, California. USA.

[4] Granger, C. W. T. and P. Newbold.1986. Forecasting Economics Time Series, Second Edition, Academic Press, Inc., Orlando, Florida.

[5] Gujarati, D. N., 2003. Basic Econometric, Fourth Edition, MeGraw-Hill Inc., New York. Pp.23- 26.

[6] Haque, M. E., M. I. Hossain and K. M. M. Rahman. 2004. Searching for the best fitting deterministic model for innovative growth analysis and forecasting of rice production in Bangladesh. The Bangladesh Journal of Agricultural Economics. 27(1):1 5-35..

[7] Mastny, V., 2001, The use of Box-Jenkins methodology to predict the price development of agricultural commodities. *Acta Unierisatatis Agriculturae et Silviculturae- Mendelianae-Brunensis*, **49**(2): 165-172.