Research Article *Journal of Plantation Crops, 2013, 41(3): 330-337*

Forecasting of arecanut market price in north eastern India: ARIMA modelling approach

Sandip Shil*, G.C. Acharya, C.T. Jose, K. Muralidharan, A.K. Sit and George V. Thomas

Central Plantation Crops Research Institute, Research Centre,
Kahikuchi, Guwahati, Assam - 781 017

(Manuscript Received: 21-06-13, Revised: 02-09-13, Accepted: 01-10-13)

Abstract

The paper deals with forecasting of minimum, maximum and average arecanut (*Areca catechu* L.) prices in the major arecanut markets of the Assam as well as Meghalaya based on the monthly price data. Monthly minimum, maximum, and average market price data of arecanut (in Rs./quintal) for the period May-2003 to March-2012 (for Assam) and February-2003 to March-2012 (for Meghalaya) were used. Box-Jenkins autoregressive integrated moving average (ARIMA) methodology was adopted for developing the models. An interrupted time-series model was also applied to resolve the problem of intervention point (October-2011) for Meghalaya price data. The proposed models were ARIMA (1, 0, 1), ARIMA (1, 1, 1), ARIMA (0, 1, 1) (for Assam market price data series) and, log ARIMA (0, 1, 1), log ARIMA (1, 0, 1) with linear trend and a man-made intervention (Oct-2011) and log ARIMA (0, 1, 1) with linear trend and a manmade intervention (Oct-2011) (for Meghalaya market price data series) for minimum, maximum, and average monthly price series, respectively.

Keywords: Arecanut, ARIMA, forecasting, interrupted time series model, intervention model

Introduction

Arecanut (*Areca catechu* L.) is one of the important plantation crops of north eastern states, especially in Meghalaya and Assam, and also plays significant role in the livelihood of the people. The arecanut palm is the source of the common masticator nut, popularly known as betel nut or supari and also popularly known as ‘Tamool’ (in Assam), ‘Kwai’ (in Meghalaya) and ‘Kuhva’ (Mizoram). In India, it is extensively used by all sections of people as masticator and for several religious and social ceremonies. India is the largest producer and consumer of arecanut in the world. It occupies a prominent place among the cultivated crops in the states of Kerala, Karnataka, Assam, Meghalaya, Tamil Nadu and West Bengal. According to Indian Horticulture Database 2010 (National Horticulture Board, 2010), the arecanut occupies about 400 thousand hectares area producing 478 thousands metric tons (MT) of processed nuts (chali) in India. Of which, Karnataka accounts for 47 per cent of the total arecanut production; Kerala 24 per cent, Assam 13 per cent and the rest of the production is distributed among other states (National Horticulture Board, 2010).

In north eastern part of India, it is mainly cultivated in Assam, Meghalaya, Mizoram, Tripura and to some extent in Nagaland with an area of 93.6 thousand hectares producing 97.7 thousands MT of chali. Assam stands first in area and production followed by Meghalaya, Mizoram, Tripura and Nagaland (National Horticulture Board, 2010). More than 85 per cent of the area under arecanut belongs to small and medium farmers with an average land holding less than 1 hectare. There are no large scale plantations in north eastern part of the country, there is a lack of organized marketing and data base which plays a negative impact on the market forecasting of arecanut. However, keeping the importance of the arecanut in the region, market and production forecasting will help in defining strategy for improvement of regularized market and post-harvest

Corresponding Author: sandip.iasri@gmail.com

operations. This will also create awareness among the growers to incline to the arecanut cultivation and subsequent marketing and income generation. This paper aims to develop forecasting models for arecanut price at major arecanut markets in Assam as well as Meghalaya by adopting Box-Jenkins autoregressive ARIMA time-series methodology.

Materials and methods

The monthly arecanut price data for the period of May-2003 to March-2012 (for Assam) and February-2003 to March-2012 (for Meghalaya) were downloaded (http://agmarknet.nic.in/). The raw data was available in the form of monthly raw nuts price (Rs. per quintal) for the major arecanut markets (Table 1) in Assam as well as Meghalaya. There were lot of inconsistency in the collected price data. The monthly prices were missing for some markets in some respective months. Hence, the minimum, maximum and average price for each month were computed based on the available market prices for that respective month. Thus, new price time series datasets for Assam as well as Meghalaya were generated for further analysis.

Table 1. Major arecanut markets across north east India

Assam	Meghalaya	Arunachal Pradesh	Mizoram	Tripura
Bongiagaon	Baghmara	Pasighat	Aizawl	Bishalgarh
Cachar	Dadengiri			Panisagar
Dhekiajuli,	Garobadha			
Dhing,	Jowai			
Gauripur	Mawkyrwat			
Golaghat Goreswar	Nongpoh (R-Bhoi)			
Howly	Shillong			
Karimganj Kharupetia	Tura			
Lanka				
P.O. Uparhali Guwahati				
Silapathar Sivasagar				
Tinsukia				

The fundamental hypothesis of any time series modeling is that some aspects of the past pattern are continued to future. Time series process is often assumed to be based on the past values, tacitly assumed in form of numerical data, of the main variable but not on the explanatory variables (Venables and Ripley, 2002).

Forecasting methodology

The commodity price data are usually collected over time. Each observation of the data series, y_t was considered as a realization of a stochastic process $\{Y_t\}$, which is a family of random variables $\{Y_t, t \in T\}$, where $T = \{0, \pm 1, \pm 2 \pm, \ldots\}$. Standard time-series approach was applied to develop an ideal model, which adequately represent the set of realizations and also their statistical relationships in a satisfactory manner. In the present study, Box-Jenkins autoregressive integrated moving average (ARIMA) model (Kumar, 1990; Hossain *et al.*, 2006; Koutroumanidis *et al.*, 2009), was applied which is one of the most widely used time-series prediction methods. ARIMA models are, in theory, the most general class of models for forecasting a time series which can be stationarized by transformations such as differencing and logging. This method uses a systematic procedure to select an appropriate model from a rich family of ARIMA models. Such models amalgamate three types of processes, *viz*., autoregressive (AR) of order *p*, differencing of degree *d* to make the series stationary and moving average (MA) of order *q*, and is written as ARIMA (*p*, *d*, *q*). The general notation ARIMA (*p*, *d*, *q*) refers to the model with *p* order of autoregressive terms, *d* is the order of non seasonal differences and *q* order of moving average terms. This model contains the AR (p) and MA (q) models,

$$Y_t = \mu + \sum_{i=1}^{p} \phi_i Y_{t-i} + \sum_{i=1}^{q} \theta_i \in_{t-i} + \in_t \quad \text{.......(1)}$$

where, $\phi_1, \phi_2, \ldots, \phi_p$ are the parameters of the model, μ is a constant and $\in_t$ is white noise. The constant term is omitted by many authors for simplicity. The error terms $\in_t$ are generally assumed to be independent identically-distributed (i.i.d.) random variables sampled from a normal distribution with zero mean: $\in_t \sim N(0, \sigma^2)$ where σ^2 is the variance.

Choice of the most appropriate values for *p*, *d* and *q* is the major problem in ARIMA modeling technique. In our study, this problem was partially resolved by performing prediction through the following steps:

Model identification

(a) Identifying the order of differencing

The input data series for ARIMA needs to be stationary, *i.e.*, it should have a constant mean, variance, and autocorrelation through time (Box and Jenkins, 1970). Therefore, usually the series first needs to be differenced until it is stationary (this may also often requires log transforming the data to stabilize the variance). The number of times the series needs to be differenced to achieve stationarity is reflected in the *d* parameter. In order to determine the necessary level of differencing, we need to examine the plot of the data and autocorrelogram. Significant changes in level (strong upward or downward changes) usually require first order non seasonal (lag=1) differencing; strong changes of slope usually require second order non seasonal differencing. Seasonal patterns require respective seasonal differencing. If the estimated autocorrelation coefficients decline slowly at longer lags, first order differencing is usually needed. We examined each plot of the data and autocorrelogram to get the necessary level of differencing for each forecasting model for arecanut price data separately.

(b) Identifying the numbers of AR or MA terms

After a time series has been stationarized by differencing, the next step in fitting an ARIMA model is to determine whether AR or MA terms are needed to correct any autocorrelation that remains in the differenced series. By looking at the autocorrelation function (ACF) and partial autocorrelation (PACF) plots (Enders, 2004) of the differenced series, the numbers of AR and/or MA terms were determined for each forecasting model for arecanut price data.

Model estimation: Linear model coefficients were estimated using principles of least squares technique for each forecasting model for arecanut price data.

Model validation: Certain diagnostic methods like Unit Root Test (popularly known as Dickey-Fuller single mean test, 1979), face validity test, series plot of residuals, histogram of the residuals, goodness of fits *etc.* were used to test the suitability of each estimated model.

Forecasting: The best model chosen for each arecanut monthly minimum, maximum and average price data model for each state was used for forecast was using different model selection criteria like Akaike information criterion (AIC), root mean squared error (RMSE), mean absolute errors (MAE), mean percent forecast error (MPFE) *etc.*

Evaluation of chosen forecasting model

Diagnostic stage statistically determines adequacy of the fitted model. It is also necessary to ascertain whether or not the assumption of independence of the white noise residuals is met. If a model is an adequate representation of a time series, it should capture all the correlation in the series, and the white noise residuals should be independent of each other. Thus, any significant autocorrelation shown in the estimated white noise residuals at the ACF and/or partial autocorrelation function (PACF) indicates model inadequacy and suggests the model modification. With this concept, the residual analysis in our study was carried out through autocorrelation function, partial autocorrelation function and Box-Ljung test (Box *et al.*, 1994). To test the randomness of errors, residual analysis was also carried out using run test (Gujarati, 2003).

ARIMA model with intervention point

The aim to build an intervention model is to describe statistically changes in the mean level of a time series due to either natural or man-made causes. To forecast average and maximum market price of arecanut for Meghalaya, composite ARIMA models, called ARIMA model with intervention point, were applied. The composite ARIMA models are nothing, but a special kind of ARIMA model with input series called an intervention model or interrupted time series model. In time series analysis literature (Box *et al.,* 1994), an intervention event is an input series that indicates the presence or absence of an event. An intervention event causes a time series process to deviate from its expected evolutionary pattern. It is assumed that the intervention event occurs at a specific time, has a known duration, and is of a particular type. The time of the intervention is when the event begins to cause deviation. The duration of the intervention is how long the event causes deviation. This paper uses ζ_t to denote an intervention event. Point interventions begin to influence the recorded data at a specific time point and last for a specific duration. A point (pulse) intervention is a dummy regressor that takes a value of one (1) at the time of the intervention, and the rest of its values are zero (0). The duration of a point intervention is one time period. In other words,

$\zeta_t = 1, if\ t = time$

$\zeta_t = 0, otherwise$

The monthly minimum, maximum, and average arecanut market price data for the period of May-2003 to December-2009 (for Assam) and February-2003 to January-2010 (for Meghalaya) were used for forecasting model development and rest of the data (25% of total observations) for model validation. But, the entire period of February-2003 to October-2011 were considered to develop the forecasting model of the monthly maximum and average arecanut market price data for Meghalaya. The reason behind is that the month of October-2011 contained a manmade intervention data point. The presence of the particular data largely influenced the forecasting results. SAS v 9.2, software package was used to solve the specific time series problem (SAS Institute Inc., 1993).

Results and discussion

Forecast of arecanut major markets in Assam as well as Meghalaya

Box-Jenkins ARIMA methodology resolved the problem of deciding appropriate values for *p, d* and *q* for each of the developed forecasting model partially by following the steps described earlier. The preliminary step for fitting each of the ARIMA model started with the stationarity test. If market price data for a specific series was found non stationary, then stationarity for that series was achieved by taking the first difference (*d*=1). We calculated the statistics for Ljung-Box White noise test probability values, unit root test probability values and seasonal unit root test probability values for at least 10 lags for minimum, maximum, and average monthly price forecasting models respectively. We found that Ljung-Box White noise test probability values for each lag for all the models indicated that the null hypothesis (H_0) for Ljung-Box White noise test cannot be rejected; whereas unit root test probability values and seasonal unit root test probability values showed the results that the null hypothesis (H_0) for both the test can be rejected at 5 per cent significant level.

The next step was to choose the most appropriate values for p and q for each developed ARIMA models. This problem was partially overcome by looking at ACF and PACF for each market price data series for minimum, maximum, and average monthly price forecasted models respectively with 5 per cent significant level. The residual ACF and PACF showed no significant values. Coupled with the results from residual ACF and PACF or IACF plots, it was concluded that the assumption of independence of error terms was not violated. Based on the overall results of all the models, we confirmed that each data series of developed models were stationary at 5 per cent significant level.

Assam market forecast

With the objective to forecast the Assam arecanut market price based on the respective monthly price data series, several forecasting model were applied. The candidate models for respective model are shown in Table 2. The price data of the period, May-2003 to December-2009, were used to develop each forecasting model and the data of the period, January-2010 to September-2011, were used to evaluate each model. The models ARIMA (1, 0, 1), ARIMA (1, 1, 1), ARIMA (0, 1, 1) were selected from the candidate models based on a selection of minimum RMSE criterion as well as prediction errors for minimum, maximum, and average monthly price data respectively with the precision of 5 per cent significant level. Table 4 depicts the various measures of goodness of fit for selected monthly price models of minimum, maximum and

Table 2. List of best four candidate models for markets of Assam. The selected model was highlighted with bold letters

Minimum-price series	Maximum-price series	Average-price series
I. Log simple exponential smoothing (RMSE- 3286.1)	I. Random walk with drift (RMSE- 2324.8)	I. Random walk with drift (RMSE- 2320.6)
II. ARIMA (1,0,0) (RMSE- 3093.4)	II. ARIMA (1,1,0) (RMSE-2507.5)	II. ARIMA (1,1,1) (RMSE-2339.0)
III. ARIMA (0,0,1) (RMSE-3228.3)	III. ARIMA (0,1,1) (RMSE- 2526.1)	III. ARIMA (1,1,0) (RMSE- 2384.0)
IV. ARIMA (1,0,1) (RMSE-3137.1)	**IV. ARIMA (1,1,1) (RMSE-2386.9)**	**IV. ARIMA (0,1,1) (RMSE- 2411.6)**

average market prices. The forecasts were generated using those respective models. Parameter estimates of those models with standard errors and significant values have been reported in Table 3 (minimum, maximum, and average monthly price model respectively). The graphical representations of fitted ARIMA models along with data points indicated that model fits well to the respective model's price data (Fig. 1, 2 and 3) for minimum, maximum, and average monthly price model respectively. Table 5 depicts the scenario of actual versus predicted price data for each respective model along with forecasts for October-2011 and November-2011. The forecasted results depicts that arecanut market price may vary Rs. 3149/- (minimum) to Rs. 11927/- (maximum) per quintal with average price Rs. 8724/- per quintal by next November-2011.

Meghalaya market forecast

Similarly, to forecast the Meghalaya arecanut market price based on the respective monthly price data series, several forecasting model were applied. Here, the price data of the period, May-2003 to January-2010, were used to develop each forecasting model and the data of the period, Febuary-2010 to October-2011, were used to validate each model. Among the candidate models (shown in Table 6), the best suited models, log ARIMA (0, 1, 1), log ARIMA (1, 0, 1) with linear trend and a manmade intervention (Oct-2011) and log ARIMA (0, 1, 1) with linear trend and a manmade intervention (Oct-2011) were selected for minimum, maximum, and average monthly price data respectively with the precision of 5 per cent significant level. Here, again the selections of minimum RMSE criterion as well

Table 3. Parameter estimates along with significant values for minimum, maximum and average-monthly price models for markets of Assam

Parameter	Minimum-price series		Maximum-price series		Average-price series	
	Estimate	Prob>|T|	Estimate	Prob>|T|	Estimate	Prob>|T|
Intercept	2073	0.0003	76.6201	0.0394	51.978	0.3430
Moving Average, Lag 1	0.5283	0.0196	0.8646	<.0001	0.6962	<.0001
Autoregressive, Lag 1	0.8006	<.0001	0.1279	0.3687	-	-
Model variance (σ2)	**3429639**		**3340493**		**2328058**	

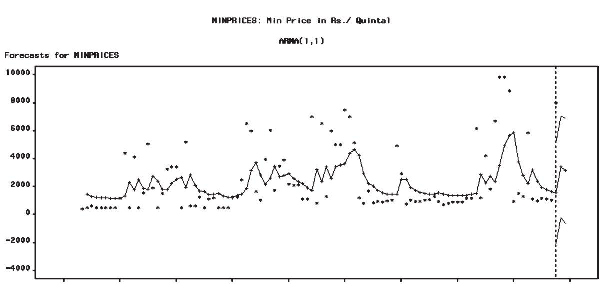

Fig. 1. Fitted ARIMA (1, 0, 1) model with actual data for minimum price model for markets of Assam

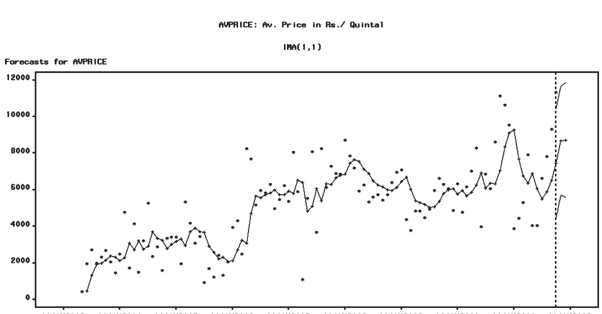

Fig. 2. Fitted ARIMA (0, 1, 1) model with actual data for average price model for markets of Assam

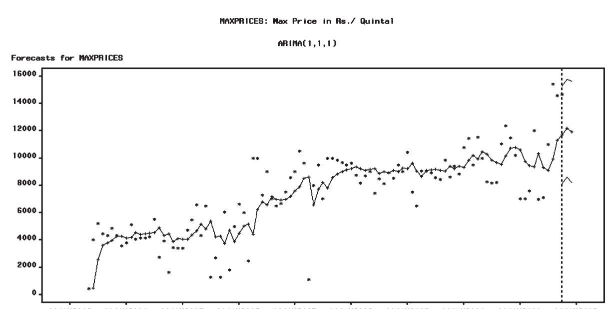

Fig. 3. Fitted ARIMA (1, 1, 1) model with actual data for maximum price model for markets of Assam

Table 4. Various measures of goodness of fit for monthly price models for markets of Assam

Measure	Calculated value for		
	minimum price forecast model	maximum price forecast model	average price forecast model
Akaike Information Criterion (AIC)	360.25	348.22	346.67
Root Mean Square Error (RMSE)	3137.1	2386.9	2411.6
R-Square	0.059	0.105	-0.072

Table 5. Assam arecanut market forecast results

Month	Actual value for average price	Predicted value for average price	Actual value for minimum price	Predicted value for minimum price	Actual value for maximum price	Predicted value for maximum price
November-09	6058	6007	828	1405	9398	9214
April-10	7009	5872	1148	1465	11539	9927
August-10	6051	6372	1850	2765	8199	9683
September-11	9322	6541	1018	1650	14594	11289
October-11	-	8672	-	3417	-	12178
November-11	-	8724	-	3149	-	11927

as prediction errors were used to select best models from candidate respective models (Table 6) along with various measures of goodness of fit (Table 8) for those selected models. Parameter estimates of those selected models with standard errors and significant values (Table 7), prediction results (Table 9) along with graphical representations (Fig. 4, 5 and 6) of fitted models were given for minimum, maximum and average monthly price model respectively.

But during data collection from Meghalaya market, it has been observed that among the major arecanut markets in Meghalaya (Table 1), the recorded arecanut price in the following markets, Garobadha, Jowai, Shillong and Tura, were Rs. 6000/-, 32400/-, 17254/- and 8633/- per quintal, respectively, in the month of October 2011, whereas in the month of September 2011, the market prices were Rs. 4350/-, 16219/-, 16739/- and 7657/- per quintal, respectively. Further, the market prices were Rs. 2020/-, 11520/-, 11000/- and 1233/- per quintal in the month of March 2012, respectively. The price was quite high particularly in the Jowai market than other arecanut markets of Meghalaya. A man made

Table 6. List of best five Candidate models for markets of Meghalaya. The selected model was highlighted with bold letters

Minimum-price series	Maximum-price series	Average-price series
I. Simple exponential smoothing (RMSE- 1449.8)	I. Log ARIMA (1,1,1) without intercept (RMSE- 2489.7)	I. ARIMA (0,1,1) without intercept and with a manmade intervention (Oct-2011) (RMSE- 1193.1)
II. ARIMA (1,1,0) (RMSE- 1453.4)	II. Log ARIMA (0,1,1) without intercept (RMSE- 2488.3)	II. Log ARIMA (0,1,1) without intercept and with a manmade intervention (Oct-2011) (RMSE- 1234.3)
III. ARIMA (0,1,1) (RMSE- 1453.7)	III. Log ARIMA (1,1,0) without intercept (RMSE- 2502.6)	III. ARIMA (0,1,1) with linear trend and a manmade intervention (Oct-2011) (RMSE- 1180.1)
IV. ARIMA (1,1,0) with log transformation (RMSE- 1636.0)	IV. Linear (Holt) exponential smoothing (RMSE- 2460.0)	**IV. Log ARIMA (0,1,1) with linear trend and a manmade intervention (Oct-2011) (RMSE- 1305.8)**
V. Log ARIMA (0,1,1) (RMSE- 1623.2)	**V. Log ARIMA (1,0,1) with linear trend and a manmade intervention (Oct-2011) (RMSE- 1888.5)**	V. Linear (Holt) exponential smoothing (RMSE- 1289.9)

Table 7. Parameter estimates along with significant values for minimum, maximum, and average-monthly price models for markets of Meghalaya

Parameter	Minimum-price series		Maximum-price series		Average-price series	
	Estimate	Prob>|T|	Estimate	Prob>|T|	Estimate	Prob>|T|
Intercept	0.03898	0.3640	7.6816	<.0001	0.0965	0.1011
Moving Average, Lag 1	0.1698	0.1411	0.0261	0.8750	0.1570	0.1197
Autoregressive, Lag 1	-	-	0.6216	<.0001	-	-
Linear Trend	-	-	0.0162	<.0001	-0.0011	0.2500
Point:OCT2011	-	-	0.7811	0.0441	0.4016	0.2530
Model variance ($\sigma 2$)	**0.2105**		**0.1405**		**0.1170**	

Table 8. Various measures of goodness of fit for monthly price models for markets of Meghalaya

Measure	Calculated value for		
	minimum price forecast model	maximum price forecast model	average price forecast model
Akaike Information Criterion (AIC)	314.47	1594.1	1500.3
Root Mean Square Error (RMSE)	1632.2	1888.5	1305.8
R-Square	0.199	0.791	0.617

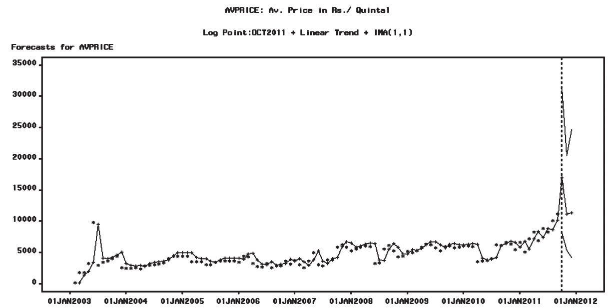

Fig. 6. Fitted Log ARIMA (0, 1, 1) with linear trend and a manmade intervention (Oct-2011) model with actual data for average price model for markets of Meghalaya

Table 9. Meghalaya arecanut market forecast results

Month	Actual value for average price	Predicted value for average price	Actual value for minimum price	Predicted value for minimum price	Actual value for maximum price	Predicted value for maximum price
January-10	6067	6283	6067	6254	6067	7394
September-10	6226	6178	1450	1672	9390	10201
March-11	7279	5563	1300	1409	11462	10438
September-11	11242	10235	4350	5522	16740	16377
October-11	16072	17040	6000	5181	32400	34758
November-11	-	11164	-	6333	-	14891
December-11	-	11380	-	6807	-	14738

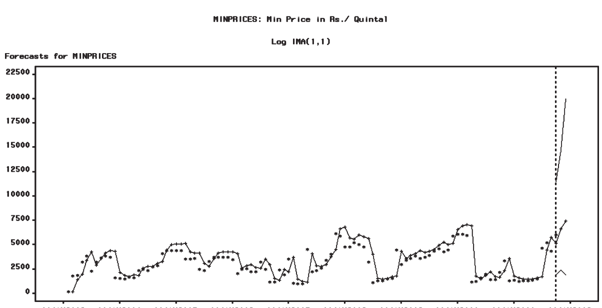

Fig. 4. Fitted Log ARIMA (0, 1, 1) model with actual data for minimum price model for markets of Meghalaya

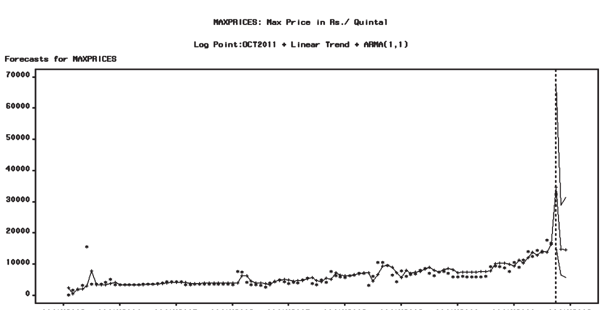

Fig. 5. Fitted Log ARIMA (1, 0, 1) with linear trend and a manmade intervention (Oct-2011) model with actual data for maximum price model for markets of Meghalaya

intervention in the supply chain of arecanuts or excessive storage by traders was observed in Jowai market in the month of October-2011. Due to presence of this particular price value, our data was largely affected, specially the computation of maximum as well as average price for that particular month. The forecasted results for maximum and average price were influenced. There were several ways to solve this problem. First was to delete the records of Jowai market and proceed for further analysis. Another was to use a point intervention forecasting model. The whole price data of May-2003 to October-2011 was considered for developing models for maximum and average price series. The proposed models were log ARIMA (1, 0, 1) with linear trend and a manmade intervention (Oct-2011) and log ARIMA (0, 1, 1) with linear trend and a manmade intervention (Oct-2011) for maximum and average price series, respectively.

The forecasted results depicts that arecanut market price may vary Rs. 6807/- (minimum) to Rs. 14738/- (maximum) per quintal with average price Rs. 11380/- per quintal by next December-2011.

Conclusion

Adopting the Box-Jenkins ARIMA time-series methodology, minimum, maximum, and average monthly price forecasting models for the major markets of Assam as well as Meghalaya were developed. This type of market price forecasting may be important to both government and industrial sector to predict future policy making decisions. Since uncertainty increases as prediction is made further from the data we have, the standard errors associated with predictions increases. It is advisable to use ARIMA methodology for short-term forecast only.

References

Box, G.E.P. and Jenkins, G.M. 1970. *Time Series Analysis: Forecasting and Control*. Holden-Day, Oakland.

Box, G.E.P., Jenkins, G.M. and Reinsel, G.C. 1994. *Time Series Analysis*. Prentice Hall, New Jersey.

Dickey, D.A. and Fuller, W.A. 1979. Distribution of the estimators for autoregressive time series with a unit root. *Journal of American Statistical Association* **74**: 427-31.

Enders, W. 2004. *Applied Econometric Time Series*, Second Edition, Wiley. pp. 48-238.

Gujarati, D.N. 2003. *Basic Econometrics*, Second Edition, McGraw-Hill Companies Inc., New York. pp. 465-467.

Hossain, M.Z., Samad, Q.A. and Ali, M.Z. 2006. ARIMA model and forecasting with three types of pulse prices in Bangladesh: a case study. *International Journal of Social Economics* **33**(4): 344-53.

Koutroumanidis, T., Ioannou, K. and Arabatzis, G. 2009. Predicting fuel wood prices in Greece with the use of ARIMA models, artificial neural networks and hybrid ARIMA-ANN model. *Energy Policy* **37**: 3627-34.

Kumar, K. 1990. Some recent developments in time series analysis. *Singapore Journal of Statistics* **1**: 45-73.

National Horticulture Board 2010. *Indian Horticulture Database 2010*, NHB Publication, India. pp. 6.

Venables, W.N. and Ripley, B.D. 2002. *Modern Applied Statistics with S*. Fourth Edition, Springer.

SAS Institute Inc. 1993. *SAS/ETS User's Guide, Version 6*, Second Edition, Cary, NC: SAS Institute Inc. pp. 1-1169.