Andhra Pradesh J Agril. Sci : 2(2): 151-157, 2016

FORECASTING PRICES OF CHILLI (DRY) IN MAJOR MARKETS OF ANDHRA PRADESH AND TELANGANA

S.B. RAMYA LAKSHMI, I. BHAVANI DEVI AND G. MOHAN NAIDU*

Department of Agricultural Economics, S.V. Agricultural College, Tirupati – 517 502, A.P.

Date of Receipt: 02-02-2016 Date of Acceptance: 14-04-2016

ABSTRACT

In this study, chilli (dry) prices in Guntur and Khammam markets of Andhra Pradesh and Telangana were examined by using ARIMA, trend analysis, decomposition fit, moving average, SES, DES, WMM, ANN models, by using 156 months data of market prices of Red chillies the period from April 2000-01 to 2012-13 of the market prices of red chilli. The ARIMA (1,0,0) was the most suitable model to forecasts red chilli prices in Guntur, while ARIMA (1,0,1) in Khammam markets. According to the forecasts, the price of chilli per quintal would be ranging from Rs. 6,818 to 6,647 in Guntur, whereas Rs. 6,539 to Rs. 6,262 in Khammam markest for the months from April to July 2013.

KEYWORDS: Price forecasting, ARIMA, trend analysis, decomposition fit, Moving Average, SES, DES, WMM, ANN.

INTRODUCTION

Chilli known as "Universal Spice" is grown in almost all states in India. Among the spice crops grown in India, chilli is the major crop cultivated in 30 per cent of the area under spices (Ajjan *et al.* 2012). Andhra Pradesh, Telangana Maharashtra, Karnataka and Tamil Nadu account for about 75 per cent of total area as well as annual production of the country. During 2013-14 Andhra Pradesh and Telangana states occupied about 1/4th of the total area (26.46 per cent) contributing 63.70 per cent of production. India is also leading country in chilli exports.

Even though chilli occupies major position in Indian spices, there were very limited studies conducted in India and out of which most of those were concentrated on its cost of cultivation, price spread, marketing channels etc. Thus, keeping in view the importance of chillies there is a need to analyze the behaviour of prices**.**

Forecasting involves making estimates of the future values of variables of interest using past and current information. There are a number of methods to generate prediction using time series data. Mohan Naidu, Meena Kumari and Srikala (2014) studied behavior of market arrivals and prices of red chillies (*Capsicum annuum*). Kumuda Keerthi and Mohan Naidu (2013) applied ARIMA for forecasting monthly prices of tomato in Madanapalli market of Chittoor district. There exists a large number of papers studying time series forecasting and comparing the results obtained from different forecasting techniques. Box and Jenkins (1970) first introduced the ARIMA model building methodology. Attempts were made during present study to develop model for chilli prices in selected markets.

Forecasting involves making estimates of the future values of variables of interest using past and current information. There are a number of methods to generate prediction using time series data. ARIMA and other models (trend analysis, Decomposition fit, Moving Average, SES, DES, WMM, ANN) were used for forecasting chilli prices. Hence the present study was undertaken with an objective of forecasting the prices of chillies.

MATERIALS AND METHODS

The study was undertaken with an objective to forecast prices of red chilli in selected markets of Andhra Pradesh. Purposive sampling technique was used to select the markets. The secondary data pertaining to monthly prices (Rs. Qtls)$^{-1}$ of red chillies were collected from Guntur and Khammam Agricultural Market Committees. The data pertained to the period 2000-01 to 2012-13(13 years) and were analysed. To forecast future prices of chillies in the selected markets, ARIMA model, trend analysis, decomposition fit, moving average, single exponential smoothing, double exponential smoothing, Winter's multiplicative and ANN models were employed

Description of the model

In general, an ARIMA model is characterized by the notation ARIMA (p, d, q) where p, d and q denote orders of auto-regression, integration (differencing) and moving average respectively. In ARIMA, time series is a linear function of past actual values and random shocks. A stationary ARIMA (p, q) process is defined by the equation

$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_p y_{t-p} -$$
$$\theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - .... - \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (1)$$

where, y_t is the response (dependent) variable at time t.

$y_{t-1}, y_{t-2},......., y_{t-p}$ are the responses with time lags t-1, t-2,…….,t-p respectively. $\phi_1, \phi_2,....., \phi_p$ are the coefficients to be estimated. ε_t is the error term at time t that represents the effects of variables not explained by the model; the assumptions about the error term are the same as those for the standard regression model. $\varepsilon_{t-1}, \varepsilon_{t-2},........., \varepsilon_{t-q}$ are errors that represent the effect of variables not explained by the model. The assumptions about the error term are the same as those for the standard regression model. $\theta_1, \theta_2,......., \theta_q$ are the coefficients to be estimated.

Identification

The foremost step in the process of modelling is to check for the stationarity of the series, as the estimation procedures are available only for the stationary series. There are two types of stationarity, viz., stationarity in 'mean' and stationarity in 'variance'. The visual examination of graph of the data, structure of autocorrelation and partial correlation coefficients help to check the presence of stationarity. Another way of checking for stationarity is to fit a first order autoregressive model for the raw data and test whether the coefficient 'ϕ_1' is less than one. If the model is found to be non-stationary , stationarity is achieved by differencing the series.

If 'X_t' denotes the original series, the non-seasonal difference of first order is

$$Y_t = X_t - X_{t-1} \quad (2)$$

The next step in the identification process is to find the initial values for the orders of non-seasonal parameters, p and q. They are obtained by looking for significance of autocorrelation and partial autocorrelation coefficients. There are no strict rules in choosing the initial values. Though sample auto correlation coefficients are poor estimates of population autocorrelation coefficients, still they are used as initial values while the final models are achieved after going through the stages repeatedly.

Estimation

At the identification stage, one or more models are tentatively chosen that seem to provide statistically adequate representations of the available data. Then precise estimates of parameters of the model are obtained by least squares. Standard computer packages like SAS,SPSS, MINITAB etc., are available for finding the estimates of relevant parameters using iterative procedures.

Diagnostics

Different models are obtained for various combinations of Auto Regressive and Moving Average individually and collectively .The best model is selected based on the following diagnostics:

a) Low Akaike Information Criteria (AIC)

b) Insignificance of auto correlations for residuals (Q-tests)

(a) Low AIC: Akaike Information Criteria is given by = (-2log L +2m)

Where m = p+q and L is the likely hood function. Since -2 log L is approximately equal to {n(1+log 2 π) + n log σ^2 }, where σ^2 is the model MSE, AIC is written as AIC = {n (1+log 2) + n log 2m}

And because first term in this equation is a constant, it is omitted while comparing between models. As an alternative to AIC, sometimes Schwarz Bayesian Information Criteria (SBC) is also used which is given by

SBC = log (m log n) /n.

(b) Non-significance of autocorrelations of residuals (Q tests)

After tentative model is fitted to the data, it is important to perform diagnostic checks to test the adequacy of the model and, if need be, to suggest potential improvements. One way to accomplish this is through the analysis of residuals. It has been found that it is effective to measure the overall adequacy of the chosen model by examining a quantity Q known as Box-Pierce statistic (a function of autocorrelations of residuals) whose approximate distribution is chi-square and is computed as follows:

$$Q = n\sum r^2(j) \quad (3)$$

Where summation extends from 1 to k with k as the maximum lag considered, n is the number of observations in the series, r(j) is the estimated autocorrelation at lag j; k can be any positive integer and is usually around 20. Q follows chi-square with (k-m_1) degrees of freedom, where m_1 is the number of parameters estimated in the model. A modified Q statistic is the Ljung-box statistic which is given by

$$Q = n(n+2)\sum r^2(j)/(n-j) \quad (4)$$

The Q statistic is compared to critical value from chi-square distribution. If model is correctly specified, residuals should be uncorrelated and Q should be small (the probability value should be large). A significant value indicates that the chosen model does not fit well.

Forecasting

After the identification of the model and its adequacy check, it is used to forecast the prices of chilli for the next periods. Hence, we used the identified ARIMA model to forecast chilli prices for the months of April 2013 to July 2013.

Single Exponential Smoothing (SES)

Forecast requires only three pieces of data. The most recent actual, the most recent forecast and a smoothing constant. The smoothing constant (α) determines the weight given to the most recent past observations and therefore controls the rate of smoothing or averaging.

$$F_t = \alpha Y_{t-1} + (1-\alpha) F_{t-1}$$

The current forecast equals to a weighted average of the most recent actual value and the most recent forecast value. The first actual value is chosen as the forecast for the second period.

$$F_t = F_{t-1} + \alpha (Y_{t-1} - F_{t-1})$$

It means that the current forecast is equal to the old forecast plus a fraction (α) of the error in the previous forecast, where α always lies between 0 and 1 (0.1 and 0.9). The best α should be the one to minimize the SSE or RSE. If greater weight is given to the most recent actual value, a high smoothing constant is chosen. This refers to as low smoothing. When α = 1 provides no smoothing because the forecast equals to the most recent actual value. This refers to the zero smoothing and becomes a one period MA. That is

$$F_t = Y_{t-1} + (1-\alpha) F_{t-1}$$

$$F_t = Y_{t-1} \quad \text{(Naive model)}$$

Derivation of Exponential Weight for the past actual

$F_t = \alpha Y_{t-1} + (1-\alpha) F_{t-1}$; $F_{t-1} = \alpha Y_{t-2} + (1-\alpha) F_{t-2}$

$F_{t-2} = \alpha Y_{t-3} + (1-\alpha) F_{t-3}$; $F_{t-3} = \alpha Y_{t-4} + (1-\alpha) F_{t-2}$

After substitution

$$F_{t=} \alpha Y_{t-1} + (1-\alpha) [\alpha Y_{t-2} + (1-\alpha) F_{t-2}]$$

$$F_{t=} \alpha Y_{t-1} + (1-\alpha) \{\alpha Y_{t-2} + (1-\alpha) [\alpha Y_{t-3} + (1-\alpha) F_{t-3}]\}$$

Finally,

$$F_t = \alpha Y_{t-1} + \alpha (1-\alpha) Y_{t-2} + \alpha (1-\alpha)^2 Y_{t-3} + \alpha (1-\alpha)^3 Y_{t-4} + \ldots\ldots + (1-\alpha)^{``} F_{t-"}$$

where, $(1-\alpha)^{\infty} F_{t-\infty} = 0$

$$F_t = \alpha\sum_{S=0}^{\infty} (1-\alpha)^s Y_{t-s-1}$$

It simply states that the forecast in t^{th} period is equal to a weighted average of all past actual values and one initial forecast (closer to zero).

All weights will be summed to 1 since

$$\alpha\sum_{S=0}^{\infty} (1-\alpha)^s = \frac{\alpha}{1-\alpha} = 1$$

Brown's Double Exponential Smoothing

Let S' denotes a single smooth and S" denotes the double smoothed value

$$S'_t = \alpha Y_t + (1-\alpha) S'_{t-1}$$

$$S''_t = \alpha S'_t + (1-\alpha) S''_{t-1}$$

$$a_t = S'_t + (S'_t - S''_t) = 2 S'_t - S''_t$$

where, a_t is the smoothed value of the end of period t

$$b_t = \frac{\alpha}{1 - \alpha} (S'_t - S''_t)$$

where, b_t is the estimated trend of the end of the period t

Therefore, the m-period forecast is

$$F_{t+m} = a_t + m\, b_t$$

Same as the SES, the Double smoothing estimation also requires a starting value to initialize the formulae. The common methods for estimating the starting values for Browns exponential smoothing are as follows.

$$\text{Let } S'_t = S''_t = Y_1$$

$$a_1 = Y_1$$

$$b_1 = \frac{(Y_2 - Y_1) + (Y_4 - Y_3)}{2}$$

Artificial Neural Networks (ANN) Model

An ANN is a mathematical structure designed to mimic the information processing functions of a network of neurons in the brain. Each neuron, individually, functions in a quite simple fashion. It receives signals from other cells through connection points (synapses), averages them and if the average over a short of time is greater than a certain value, the neuron produces another signal that is passed on to other cells. It is the high degree of connectivity rather than the functional complexity of the neuron itself that gives the neuron its computational processing ability. Neural networks are very sophisticated modeling techniques, capable of modeling extremely complex functions. The neural network user gathers representative data and then invokes training algorithms to automatically learn the structure of the data.

The methodology consisted of an Artificial Neural Network, trained with the well-known error back propagation learning algorithm, which has been successfully used in a number of forecasting applications. This is especially true in light of the fact that when we analyze a relatively small sample. Multilayer Perceptrons (MLPs) use a linear PSP function (i.e., they perform a weighted sum of their inputs) and a non-linear activation function. The standard activation function (sigmoid) was used in the study. To determine the appropriate configuration of the feed-forward network, several parameters have been varied. The number of neurons in the hidden layer was determined automatically by adopting network complexity

Table 1. Residual analysis of monthly prices of chilli in selected markets

Market	Model	AIC	SBC
Guntur	100	2305.78	2311.88
Khammam	101	2285.87	2295.02

Table 2. ACF and PACF values of monthly prices of red chilli in Guntur and Khammam markets

Lags	Guntur		Khammam	
	ACF	PACF	ACF	PACF
1	0.036	0.036	0.035	0.035
2	0.131	0.130	0.085	0.084
3	0.065	0.057	0.095	0.090
4	-0.046	-0.068	-0.019	-0.032
5	-0.013	-0.027	-0.070	-0.086
6	-0.107	-0.098	-0.035	-0.036
7	-0.076	-0.061	-0.048	-0.029
8	-0.030	0.000	-0.080	-0.059
9	-0.015	0.015	-0.047	-0.035
10	-0.068	-0.068	0.007	0.021
11	-0.032	-0.039	0.022	0.037
12	-0.078	-0.076	0.098	0.096
13	0.041	0.049	-0.074	-0.103
14	-0.047	-0.039	0.097	0.071
15	-0.041	-0.050	-0.039	-0.055
16	-0.076	-0.096	-0.115	-0.114

RESULTS AND DISCUSSION

Residual analysis, ACF, PACF and also forecasted values in Guntur and Khammam markets are shown in Tables 1 and 2. The actual and forecasted values are shown in Figures 1 and 2.

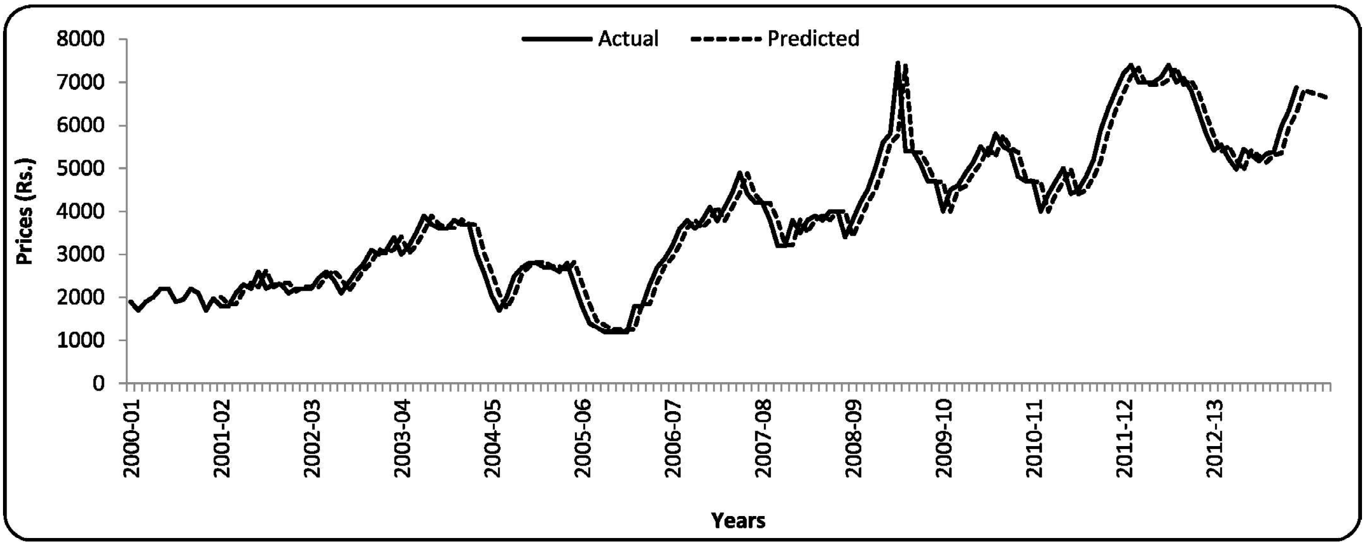

Fig. 1. Actual and predicted (forecasted) of chilli prices in Guntur market

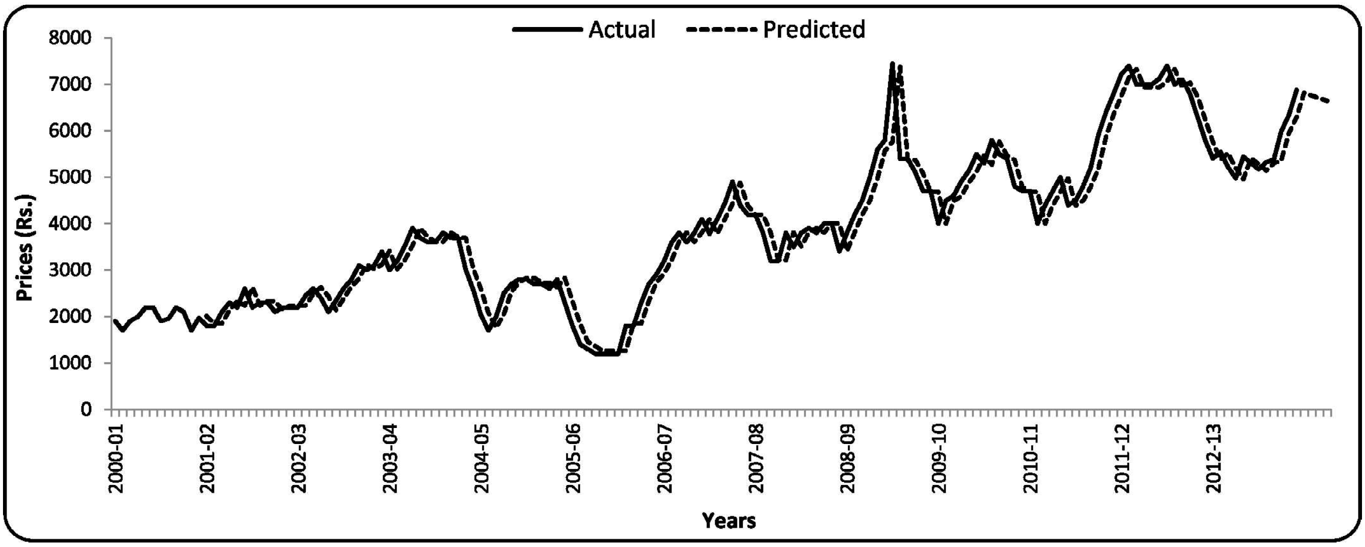

Fig. 2. Actual and predicted of chilli prices in Khammam market

Prices of chilli in Guntur Market

ARIMA model was estimated after transforming the price data of chillies into stationary series. The ACF and PACF values are presented in Table I. Based on ACF and PACF many models were tested. Finally model (1,0,0) was identified as the best model to forecast prices in Guntur market. The forecasted prices obtained from the model with the least difference between AIC and SBC were considered.

Both actual and predicted were done and it was compared with actual observations. The prices were forecasted up to July 2013. The results of actual and predicted prices are illustrated in Fig 1. As can be seen from the graph the actual and forecasted prices of chillies were more or less closer. According to the forecasts, the price of chillies per quintal would be ranging from Rs. 6,818 to Rs.6, 647 per quintal for the months from April to July 2013 in Guntur market.

Prices of Chilli in Khammam Market

The particulars of forecasting of prices of chilli in Khammam market are presented here under. The ACF

Table 3. Comparison of forecasted values with real time prices for chilli in Guntur market

Month	Real time prices	Forecasted values								Deviation (%)							
		ARIMA	T.A	D.F	M.A	S.E.S	D.E.S	W.M.M	ANN	ARIMA	T.A	D.F	M.A	S.E.S	D.E.S	W.M.M	ANN
April 2013	6192	6818	6261	5770	5730.25	6882	6940	5076	7119	-10.11	-1.11	6.81	7.46	-11.14	-12.08	18.02	-14.97
May 2013	6050	6759	6291	5877	5730.25	6882	6984	4965	7332	-11.72	-3.99	2.86	5.29	-13.75	-15.44	17.93	-21.19
June 2013	6233	6702	6322	5931	5730.25	6882	7028	4914	7362	-7.52	-1.43	4.85	8.07	-10.41	-12.75	21.16	-18.11
July 2013	6281	6647	6353	6301	5730.25	6882	7071	4997	7221	-5.83	-1.15	-0.32	8.77	-9.57	-12.58	20.44	-14.97

Table 4. Comparison of forecasted values with real time prices for chilli in Khammam market

Month	Real time prices	Forecasted values								Deviation (%)							
		ARIMA	T.A	D.F	M.A	S.E.S	D.E.S	W.M.M	ANN	ARIMA	T.A	D.F	M.A	S.E.S	D.E.S	W.M.M	ANN
April 2013	7100	6539	5843	6364	5148	6613	6658	5429	6980	7.90	17.71	10.36	27.49	6.86	6.23	23.54	1.69
May 2013	4500	6443	5871	5748	5148	6613	6700	5040	7512	-43.18	-30.46	-27.74	-14.41	-46.96	-48.89	-12.00	-66.93
June 2013	4454	6351	5899	5642	5148	6613	6742	4638	7415	-42.59	-32.43	-26.66	-15.59	-48.47	-51.37	-4.13	-66.48
July 2013	4100	6262	5927	5608	5148	6613	6784	4545	7182	-52.73	-44.55	-36.78	-25.57	-61.29	-65.46	-10.85	-75.17

and PACF vales are presented in Table 1. Based on ACF and PACF values many models were tested. Finally model (1,0,1) was identified as the best model to forecast prices in Khammam market.

The results of Actual and predicted prices are illustrated in Fig.2. As can be observed from the graph, there was not much variation between actual and forecasted prices in Khammam market. The price of chilli per quintal would be Rs. 6,539 to Rs. 6,262 for the period commencing from April to July 2013 in Khammam market.

Comparison of Forecasted Values with Real Time Prices for Chillies in Selected Markets

A comparision was made between the forecasted values obtained through various models *viz.* ARIMA, Trend Analysis, Decomposition Fit, Moving Average, Single Exponential smoothing (SES), Double Exponential smoothing (DES), Winters Multiplicative Method (WMM) and ANN, with those of real time prices for chillies in selected markets in order to find out the closeness of the forecasted values of different models with real time prices.

The results obtained through trend analysis were relatively closer in respect of Guntur market (Table 3). WMM and ANN models were better than the other models in respect of Khammam (Table 4). Najafi *et al.*(2007) reported that ANN had the lowest error in prediction of prices of different commodities.

CONCLUSION

From the results it can be observed that there was not much variation between actual and forecasted prices. Hence marketing information is an important tool for decision making at all the stages right from farm production to ultimate consumption for all the participants in marketing channel. Marketing information is essential for producers in market led production and is equally important for other market participants for trading and also for consumers. Therefore marketing information is to be provided to all the stakeholders at right time.

REFERENCES

Ajjan, N., Raveendran, N., Raghu Ram, P., Bhavani Devi, I., Srikala, M., Ranganath, P.D and Balaji, T. 2012. Establishment and networking of agricultural market intelligence centres in India, Commodity report (Red Chillies).

Box, G.E. and Jenkins, G.M. 1970, “Time series analysis: Forecasting and control”, Holden Day, San Francisco

Kumuda Keerthi, P and Mohan Naidu, G. 2013. Forecasting monthly prices of tomato in Madanapalli market of Chittoor district. *BIOINFOLET* 10(1B):201-203.

Mohan Naidu, G., Meena Kumari, V., and Srikala, M. 2014. Behaiour of market arrivals and prices of red chillies (*Capsicum annuum*). *Plant Archives*. 14(1): 511-519

Najafi, B., Zibaei, M.; Sheikhi, M. H. and Tarazkar, M.H. 2007. Forecasting price of some Crop products in Pars province: application of artificial neural network.[Persian]. *Journal of Science and Technology of Agriculture and Natural Resources.* 11: 1(B), 501-512.