Original Research Article

Forecasting Paddy Prices in Ambikapur Market of Chhattisgarh State using SARIMA Approach

Chowa Ram Sahu[1], M. L. Lakhera[1] and S. S. Walke[2]*

[1]*Department of Agricultural Statistics & Social Science (L.), IGKV,*
Raipur - 492012 (C.G.), India
[2]*College of Agriculture and Allied Science, Baramati MPKV Rahuri (MS), India*

*Corresponding author

A B S T R A C T

The present study was carried out to forecast prices of paddy through univariate Seasonal Auto Regressive Integrated Moving Average (SARIMA) model. The secondary data (Time Series data) pertaining to monthly prices (Rs/q) of paddy were collected the website of Chhattisgarh State Marketing Board (Mandi) for the period from April 2009 to February 2020 have been used for the study. The SARIMA (2,1,2) $(1,1,1)_{12}$ were found as the most suitable models to forecasts of paddy prices. With these models there were made forecast for 13 months, which are from March 2020 to March 2021. The best models were chosen based on highest R^2-value and least Mean Absolute Percentage Error (MAPE) value.

Introduction

The importance of cereals in the world of today is cereal grains have been considered as the principal component of human diet for thousands of years and have played a major role in shaping human civilization. Among cereals, paddy undoubtedly constitutes the large stand most important role in food consumption. Paddy is one of the most important food crops of India and is second in importance throughout the world. It feeds more than 50 percent of the world's population. It is the staple food of most of the people in South-East Asia. Asia accounts for about 90 percent of the world's paddy cultivation and production.

Among the paddy growing countries, India has the largest area under cultivation, though in terms of volume of output, it is second to China. In India, rice is the most preferred staple food for about 65 percent of the population. In India, Paddy occupies nearly 43 with 110.15 million tonnes of the production with an average yield of 2550 kg

per hectare (*www.agricoop.nic.in*). Among the states West Bengal ranks first in paddy production (15.37 million tonnes) and productivity (2788 kg/ha) with nearly 5.51 million hectares of area during 2016-17 followed by Uttar Pradesh (14.63 million tonnes), Andhra Pradesh (13.03 million tonnes), Punjab, Tamil Nadu, Bihar and Chhattisgarh.

Chhattisgarh is an important state as it contributed about 2.97 percent of the total annual food grains production in India during 2012-13 (Department of Agriculture statistic & cooperation Directorate economics & agriculture). Chhattisgarh state has total geographical area of 137.90 lakh hectare with net sown area 47.75 lakh hectares.

The cropping intensity of Chhattisgarh is 138 percent. Chhattisgarh, one of the largest contributors of rice in India, Chhattisgarh is agricultural chief land & due to large production of rice Chhattisgarh is known as the "bowl of rice". Chhattisgarh used to produce over seventy percent of the total paddy production in the state. Paddy is one of the most important crop grown in Chhattisgarh with an area (38.80 lakh hectares), production (57.49 lakh tonnes) and productivity (1482 kg/ha) (*https://agriportal.cg.nic.in*) during 2017-18.

For this purpose, forecasting acts as a key tool to alarm about the need of nation in advance. The time series approach to forecasting is one such approach which relies on the past patterns in a time series to forecast prices in the future.

Among the various methods for analyzing a time series, one of the most simple and bench mark method is that of Box and Jenkins which is popularly known as ARIMA methodology (Box, G.E.P., Jenkins, and GM.1976). Previously attempts were made to forecast tomato prices (Kumuda ad Mohan 2013); onion prices using SARIMA models (Mohan et. al 2014); ARIMA models were used to obtain seasonal forecast of paddy in Tamil Nadu and food grains in India (Balasubramanian and Dhanavantham, 2002) and Forecasting maize prices in Andhra Pradesh using Seasonal ARIMA model (Mohan Naidu 2015).

Materials and Methods

Data

The secondary data pertaining to monthly prices (Rs/q) of paddy were collected the website of Chhattisgarh State Marketing Board (Mandi) http://agriportal.cg.nic.in/ agrimandi/ for the period from April 2009 to December 2017.

Description of the model

In this section, a brief description of statistical tools employed was presented. Box-Jenkins (ARIMA model) model for forecasting of prices was employed.

Time series model

For analysis of time series data, a model is essential. Generally two broad approaches are resorted too. One is a multiplicative model and the other is an additive model. In this present study multiplicative model has been employed, since many of agricultural data admit such a model as a more appropriate one.

The following two structures are considered for basic decomposition models:

1. Additive: Y_t = Trend (T_t) + Seasonal (S_t) + Cyclical (C_t) + Irregular (I_t)

2. Multiplicative: Y_t = Trend (T_t) x Seasonal (S_t) x Cyclical (C_t) x Irregular (I_t)

How to choose between additive and multiplicative decompositions

The additive model is useful when the seasonal variation is relatively constant over time.

The multiplicative model is useful when the seasonal variation increases over time.

Box-Jenkins model

ARIMA models, also called Box-Jenkins models, are models that may possibly include autoregressive terms, moving average terms, and differencing operations. ARIMA model is an extrapolation method for forecasting and like any other such method, it requires only the historical time series data on the variables under forecasting. It is robust to handle any data pattern. The prices of paddy were forecasted in the Chhattisgarh State.

The seasonal part of an ARIMA model has the same structure as the non-seasonal part: it may have an AR factor, an MA factor, and/or an order of differencing. In the seasonal part of the model, all of these factors operate across multiples of lag s (the number of periods in a season).

A seasonal ARIMA model is classified as an ARIMA (p,d,q)x(P,D,Q) model, where P=number of seasonal autoregressive (SAR) terms, D=number of seasonal differences, Q=number of seasonal moving average (SMA) terms.

Identification

In identifying a seasonal model, the firststep is to determine whether or not a seasonal *difference* is needed, in addition to or perhaps instead of a non-seasonal difference. You should look at time series plots and ACF and PACF plots for all possible combinations of 0 or 1 non-seasonal difference and 0 or 1

seasonal difference. If the series has a strong and consistent seasonal pattern, then you should use an order of seasonal differencing-- but never use more than one order of seasonal differencing or more than 2 orders of total differencing (seasonal+non-seasonal).

Estimation

At the identification stage one or more models are tentatively chosen that seem to provide statistically adequate representations of the available data. Then we attempt to obtain precise estimates of parameters of the model by least squares as advocated by Box and Jenkins. Standard computer packages like R program, SAS, SPSS etc. are available for finding the estimates of relevant parameters using iterative procedures.

Diagnostic checking

In this step, model must be checked for adequacy by considering the properties of the residuals whether the residual from an ARIMA model must has the normal distribution and should be random. An overall check of model adequacy is provided by the Ljung-Box Q statistic.

If the p-value associated with the Q statistic is small (p-value<), the model is considered inadequate. The analyst should considered a new or modified model and continue the analysis until a satisfactory model has been determined.

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

Model selection

Many researchers were used R^2 and MAPE as the criterion of model selection are given below (Naidu 2015):

R-Square

R-squared is a statistical measure of how close the data are to the fitted regression line. It is also known as the coefficient of determination, or the coefficient of multiple determination for multiple regression.

$$R^2 = 1 - \frac{SS_{res}}{SS_{total}}$$

Where, SS_{res} = residual sum of squares and SS_{total} = total sum of squares,ingeneral, the higher the R-squared, the better the model fits.

Percentage errors

The percentage error is given by $P_i = (e_i/y_i)$x100.Percentage errors have the advantage of being scale-independent, and so are frequently used to compare forecast performance between different data sets. The most commonly used measure is:

Mean absolute percentage error: MAPE = mean $(|P_i|)$.

Results and Discussion

For forecasting prices of paddy, ARIMA model was estimated only after transforming the variable under forecasting into a stationary series. The stationary series is the one whose values vary over time only around a constant mean and constant variance. There are several ways to ascertain this. The most common method is to check stationary by examining the graph or time plot of the data. Non-stationary in mean is connected through appropriate differencing. If there is both trend and seasonality, apply a seasonal difference to the data and then re-evaluate the trend. If a trend remains, then take first differences.

Identification of the model

The stationary check of time series data was performed, which revealed that the prices of paddy were shows an upward trend and seasonality. With trend and monthly seasonality we will likely need both a 1st and a 12th difference and best fit ARIMA models were developed using the data from April 2009 to February 2020. A seasonal $ARIMA\,(p,d,q)\,x\,(P,D,Q)_{12}$ model was identified by finding the initial values for the orders of non-seasonal parameters p and q and seasonal parameters P and Q. They were obtained by looking for significant spikes in autocorrelation and partial autocorrelation functions. The result ACF and PACF of prices are graphically presented in the Figure 1.

At the identification stage, one or more models were tentatively chosen which seem to provide statistically adequate representations of the available data. Then precise estimates of parameters of the model were obtained by least squares. Finally the ARIMA (2,1,1) (1,0,0)$_{12}$ were selected as the most suitable models to forecasts of paddy prices based on least Mean Absolute Percentage Error (MAPE) and highest R^2 value that are presented in Table 1.

Estimating the parameters

The parameters of best fitted model of prices were presented in Table 2, the non-seasonal specification of AR, differencing, and MA, and then the seasonal specification of seasonal AR, seasonal differencing, seasonal MA, and period for the seasonality.

Diagnostic checking

The time series plot of the standardized residuals mostly indicates that there's no trend in the residuals, no outliers, and in

general, no changing variance across time. The ACF of the residuals shows no significant autocorrelations that is a good result. The bottom plot gives p-values for the Ljung-Box-Pierce statistics for each lag up to 36. These statistics consider the accumulated residual autocorrelation from lag 1 up to and including the lag on the horizontal axis. The dashed blue line is at .05. All p-values are above it. That's a good result. Diagnostic checking for paddy prices are graphically presented in the Figure 2.

Table.1 Seasonal ARIMA Models for the prices of paddy in Ambikapur Market of Chhattisgarh

Type	Model	R-squared	MAPE
Prices	(2,1,1)(1,0,0)	0.97	2.26

Table.2 Parameters of best fitted model of prices of paddy in Ambikapur Market of Chhattisgarh

Type	Model	Term	Coefficient	SE
		AR1	-0.8604	0.1907
Prices	ARIMA (2,1,1) (1,0,0)$_{12}$	AR2	-0.4832	0.089
		MA1	0.3919	0.2103
		Seasonal AR1	0.4432	0.0877

Table.3 Predicted values for prices (Rs/Qtls) of paddy in Ambikapur Market of Chhattisgarh

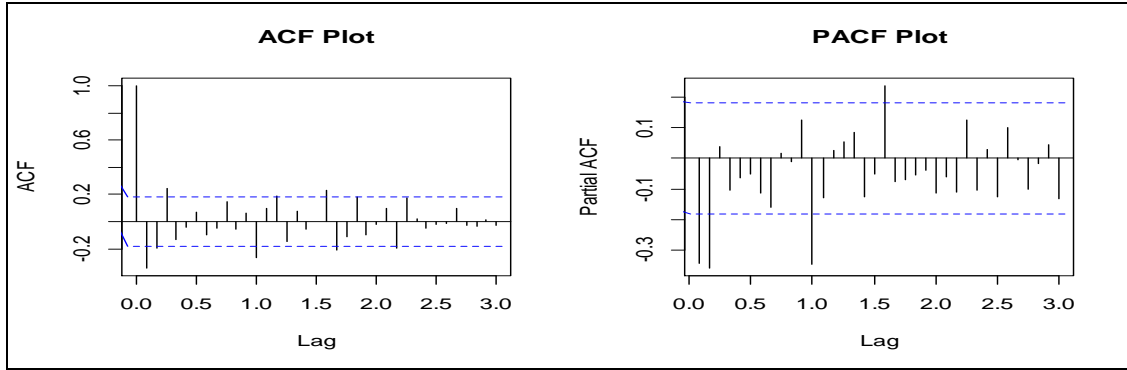| Year	Months	Predicted Prices
2020	March	1935.76
	April	1915.23
	May	1928.15
	June	1932.94
	July	*1876.92*
	August	1912.03
	September	1958.78
	October	1877.56
	November	1994.69
2021	December	1977.69
	January	1974.92
	February	*2028.49*
	March	2000.99

Fig.1 The ACF and PACF of the detrended seasonally differenced prices data of paddy in Ambikapur Market of Chhattisgarh State
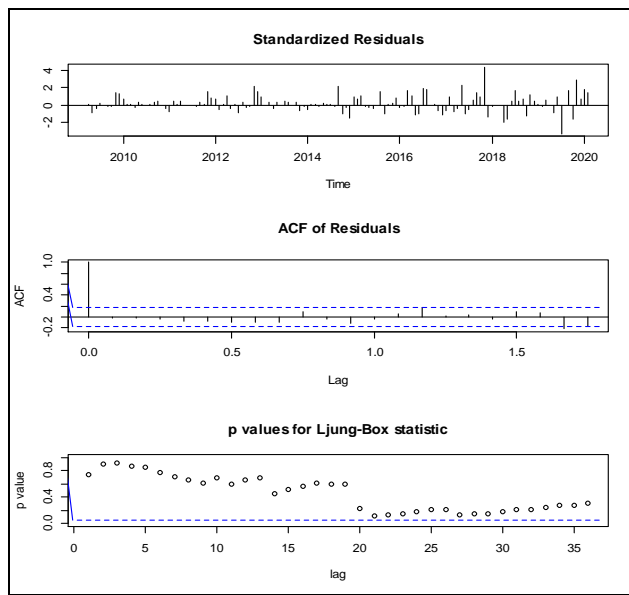

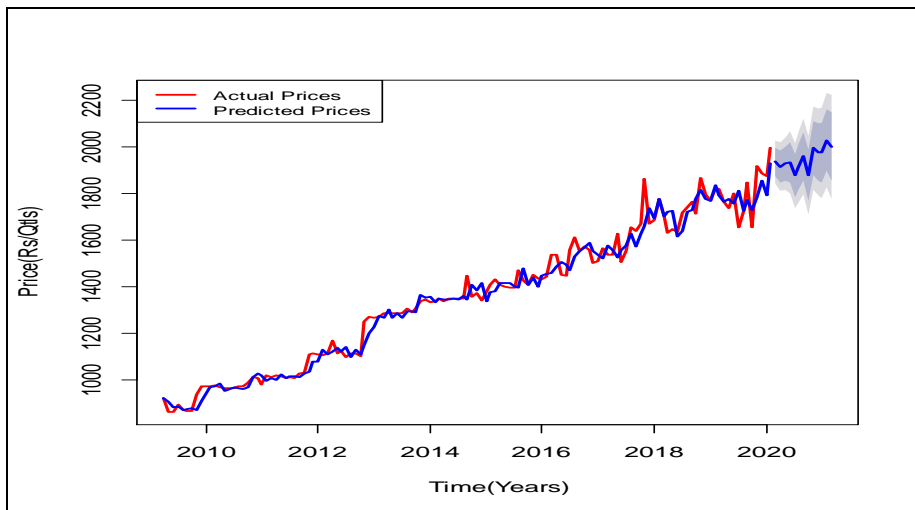**Fig.2** Diagnostic checking for prices of paddy in Ambikapur Market of Chhattisgarh

Fig.3 Actual and Predicted values for prices of paddy in Ambikapur Market of Chhattisgarh

Forecasting

After identification of the model and its adequacy check, it is used to forecast the prices of paddy in the coming periods. Hence, we used the identified SARIMA model to forecast the prices of paddy in the Ambikapur Market of Chhattisgarh for the period of March 2020 to March 2021 and the results of forecasted prices were presented in Table 3 and illustrated in Figure 3. From the forecast values obtained by the developed model, it can be said that the highest paddy prices in Ambikapur market will be 2028.49 per quintal in February, 2021 while least Rs.1876.92 in July, 2020 (Table 3).

References

Balasubramanian P and Dhanavanthan P. 2002. Seasonal modelling and forecasting of crop production. *Statistics and Applications* 4(2): 107-118.

Box, G.E. and Jenkins, G.M. (1976). Time Series analysis: Forecasting and Control. Holden Day, San Francisco.

Mohan Naidu, G. (2015). Forecasting maize prices in Andhra Pradesh using seasonal ARIMA Model. Research Journal of Agricultural Sciences 6(1): 149-152.

Rajur, B.C., Patil, B.I., Kunnall, B and Basavaraj, H. 2009. Market arrivals and prices of chilli in Karnataka – *An Economic Analysis. Agricultural Situation in India*. 65(11): 671-674.

Virender Kumar, Sharma H.R and Kamlesh Singh. 2005. Behaviour of market arrivals and prices of selected vegetable crops: A study of four metropolitan markets. *Agricultural Economics Research Review*. 18(2): 271-290.