Application of ARIMA Model for Forecasting Maize Production in Telangana State, India

ABSTRACT

The study utilized the Box-Jenkins approach for forecasting maize production in Telangana state. It involved the analysis of 55 years of empirical annual observations of maize production. Autocorrelation (ACF) and partial autocorrelationfunctions (PACF) were calculated for the data. Appropriate Box-Jenkin's ARIMA model was fitted. Validity of the model was tested using standard statistical techniques. The forecasting power of autoregressive integrated moving average model was used to forecast maize production for three leading years.

KEYWORDS: Maize, Forecasting, Autocorrelation function (ACF), Partial autocorrelation function (PACF), Autoregressive integrated moving average (ARIMA), Autoregressive moving average (ARMA).

INTRODUCTION:

Maize (*Zea mays L.*) referred to as the Queen of Cereals, is a vital crop in India, standing as the third cash crop after wheat and rice. With 16 million Indian farmers engaged in maize cultivation, states like Karnataka, Rajasthan, Madhya Pradesh, and Telangana contribute significantly to the country's maize production. To ensure remunerative prices for farmers, India must plan production by enhancing productivity and reorienting the value chain (IIMR, 2023-24).

In Telangana, maize ranks third among all crops, covering an extensive area of 12.74 lakh acres. The maize production in Telangana reached 28.65 lakh tonnes during 2022-23 (DES, 2022-23). Major maize-growing districts in Telangana include Warangal Rural, Khammam, Nirmal, Siddipet, Kamareddy, Mahabubabad, Nizamabad, Warangal Urban, Jagityal and Karimnagar. Over the last decade, both the area and production of maize have witnessed significant growth in the state (TS agriculture, 2022-23).

MATERIAL AND METHODS:

Data source:

Telangana State's maize production data, in lakh tonnes from the year 1966-67 to 2021-22 sourced from Directorate of Economics and Statistics www.ecostat.telangana.gov.in, Telangana state was used for model development. The SAS 9.3 statistical software package has been used for data analysis.

Stationarity

The noise (or residual) series for an ARMA model must be stationary, which means that both the expected values of the series and its auto covariance function are independent of time. The standard way to check for non-stationarity is to plot the series and its autocorrelation function. We can visually examine a graph of the series over time to see if it has a visible trend or if its variability changes noticeably over time. If the series is non stationary, its autocorrelation function will usually decay slowly. Another way of checking for stationarity is to use the stationarity tests. Most time series are non-stationary and must be transformed to a stationary series before the ARIMA modeling process can proceed. If the series has a trend over time, seasonality, or some other non-stationary pattern, the usual solution is to take the difference of the series from one period to the next and then analyze this differenced series. Sometimes a series may need to be differenced more than once or differenced at lags greater than one period. If the trend or seasonal effects are very regular, the introduction of explanatory variables may be an appropriate alternative to differencing.

Differencing

A deterministic seasonal pattern will also cause the series to be non-stationary, since the expected value of the series will not be the same for all time periods but will be higher or lower depending on the season. When the series has a seasonal pattern, you may want to difference the series at a lag corresponding to the length of the cycle of seasons. To take a second difference, add another differencing period to the list. There is no limit to the order of differencing and the degree of lagging for each difference.

ARIMA Process

Auto Regressive Integrated Moving Average (ARIMA) Model (Process):This process is an amalgamation to ARMA process when the time series {Yt} is Non-Stationary or “Integrated”. It is obvious that to develop the ARMA model in this situation, the series has to be “differenced” to make it stationary and this differenced series, which is now stationary has to be subjected to fitting of ARMA model. This process is referred as ARIMA (p, d, q), where p and q refer to the number of AR and MA terms and d refers to the order of differencing required for making the series a Stationary. The characteristics of the time series models, i.e., the parameters (p, d, q) and thereafter the estimation of the relevant model can be carried out in a planned approach outlined by Box and Jenkins methodology.

The methodology involves the following four steps

1. Identification of the characteristics (p, d, q) for the Model

2. Estimation

3. Diagnostics Checking

4. Forecasting

1. **Identification of the characteristics (p, d, q) for the Model:** 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 stationary series. If the original series is non stationary then first of all it should be made stationary.
2. **Estimation:** On the basis of identification of the parameters (p, d, q) the series is subjected to fitting of the appropriate ARIMA (p, d, q) model. The procedure for fitting the model involves transforming the series through appropriate differencing, in case it is non-stationary, and then subjecting the differenced series to fitting. Choice of parameters is on the basis of significant ACFs and PACFs.
3. **Diagnostics:** In the model-building process, if an ARIMA (p, d, q) model is chosen based on the ACFs and PACFs, some checks on the model adequacy are required. A residual analysis is usually based on the fact that the residuals of an adequate model should be approximately white noise. Therefore, checking the significance of the residual autocorrelations and comparing with approximate two standard error bounds, *i.e.*, $\pm 2/\sqrt{n}$ are needed.

Two criteria for model selection:

Akaike's information criterion (AIC):

$$AIC = \log \hat{\sigma}^2 + 2\frac{p+q}{n}$$ Where $\hat{\sigma}^2$ is the estimated variance of e_t.

Schwarz's Bayesian Information Criterion (SC, BIC, or SBC):

$$BIC = \log \hat{\sigma}^2 + 2\frac{p+q}{n}\log(n)$$

Both criteria are likelihood-based and represent a different trade-off between "fit", as measured by the log-likelihood value, and "parsimony", as measured by the number of free parameters, p + q. If a constant is included in the model, the number of parameters is increased to p+ q+ 1.Usually, the model with the smallest AIC or BIC values are preferred. While the two criteria differ in their trade-off between fit and parsimony, the BIC criterion can be preferred because it has the property that it will almost surely select the true model.

4. Forecasting:

The model that satisfies all the diagnostic checks was considered for forecasting

RESULTS AND DISCUSSION:

As Box Jenkins models were preferred to the multiplicative time series models for forecasting purposes, these were used for the forecasting of maize production in Telangana state. The results of the analysis are presented below.

The first step in the Box Jenkins approach for selecting an autoregressive integrated moving average (ARIMA) model for forecasting maize production involved applying the Augmented Dickey Fuller (ADF) unit root test to check for stationarity or non-stationarity of the data. This test helped to choose between auto regressive moving average (ARMA) and ARIMA models for further processing and forecasting. The analysis and forecasting were based on 55 years (1966-67 to 2021-22) of empirical annual observations of maize production.

The original observed data of maize production was nonstationary and was transformed to stationary after the first difference. This was evident from Fig. 1 and 2 and was confirmed by the ADF test at a 0.05 level of significance. The model was chosen based on the lowest Akaike Information Criterion (AIC) value. For the maize production data, the ARIMA (0,1,1) model had the lowest AIC value of 291.20 and a Bayesian Information Criterion (BIC) value of 295.35 (Table 1).

Fig 4

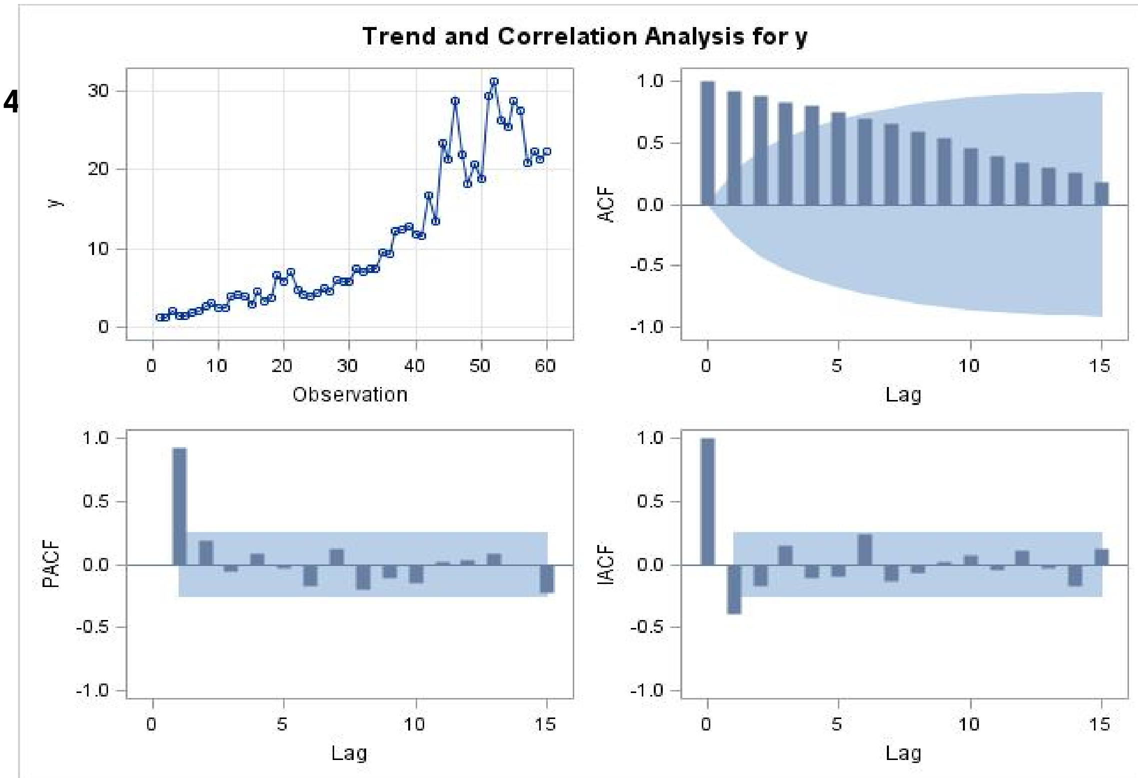

Fig 1: Line plot of the original series of maize production data

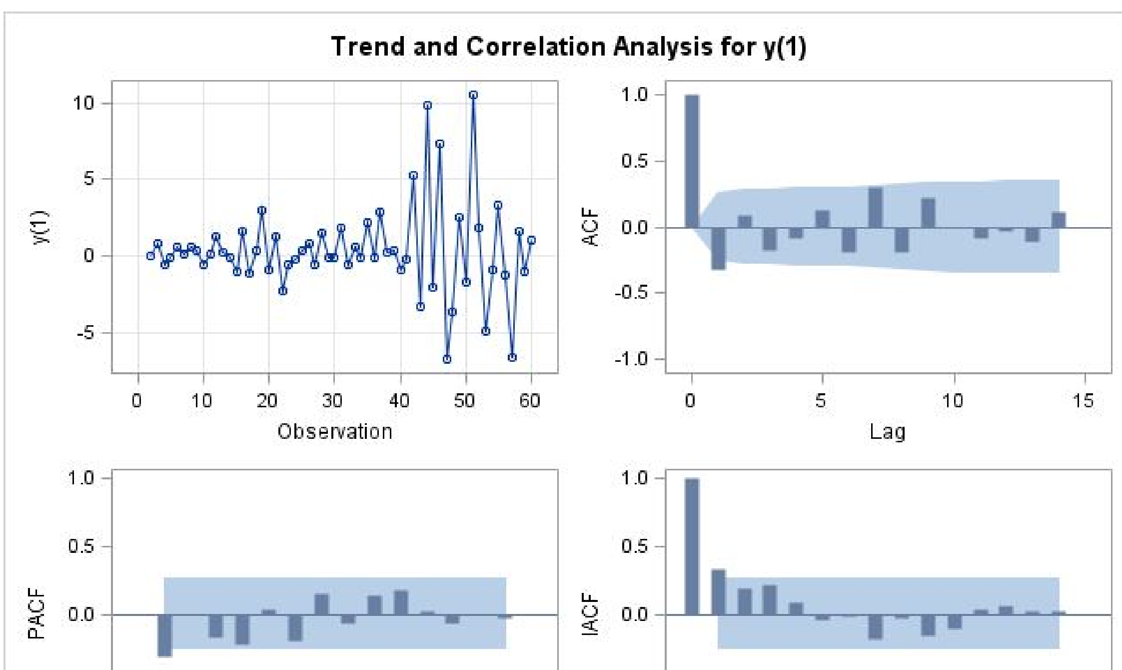

Fig 2: Line plot of the first order differenced maize production data

The next step in the Box-Jenkins approach was to perform diagnostic checks on the fitted model. The residuals of the estimated model were examined to test the randomness of the series and determine the adequacy of the estimated models. The autocorrelation and partial autocorrelation plots of the residuals (Fig. 3) showed that they were random in all cases, as none of the coefficients were significantly different from zero. This was supported by the significant ADF test of residuals of maize production at a 0.05 level of significance. These statistics proved that the tentatively identified and estimated models were appropriate for predicting maize production in Telangana.

The Box-Jenkins method was used to obtain ex-post values for maize production, as shown in Figure 4. The ARIMA model was fitted based on 55 years of data from 1966-67 to 2021-22, and the forecast was done for three years from 2022-23 to 2024-25. From Figure 4 and Table 2, it can be seen that maize production in Telangana is forecasted to be 23.10 LT in 2022-23, 23.46 LT in 2023-24, and 23.82 LT in 2024-25.

The ARIMA (0,1,1) model was selected based on the lowest AIC value and was found to be appropriate for predicting maize production in Telangana. The residuals of the estimated models were found to be random in all cases, which was supported by the significant ADF test of residuals of maize production at 0.05 level of significance. These findings suggest that the Box-Jenkins approach and ARIMA models can be useful in forecasting agricultural production in similar contexts. The similar results were also obtained by Mandal (2005) for sugarcane production in India.

Table 1: AIC and BIC values for different ARIMA models for Maize production in Telangana State

Model	AIC	BIC
ARIMA(000)	436.41	438.50
ARIMA(111)	292.29	298.52
ARIMA(011)	291.20	295.35
ARIMA(110)	291.65	295.81
ARIMA(210)	293.65	299.88
ARIMA(012)	292.90	299.13
ARIMA(211)	295.63	303.94
ARIMA(112)	295.17	303.48
ARIMA(212)	296.48	306.87

Table 2: Forecasts of Maize production in Telangana up to 2025

Forecasts for maize production				
Observations	**Forecast**	**Std Error**	**95% Confidence Limits**	
2022-23	23.1097	3.3140	16.6144	29.6051
2023-24	23.4689	3.7529	16.1134	30.8244
2024-25	23.8281	4.1456	15.7029	31.9533

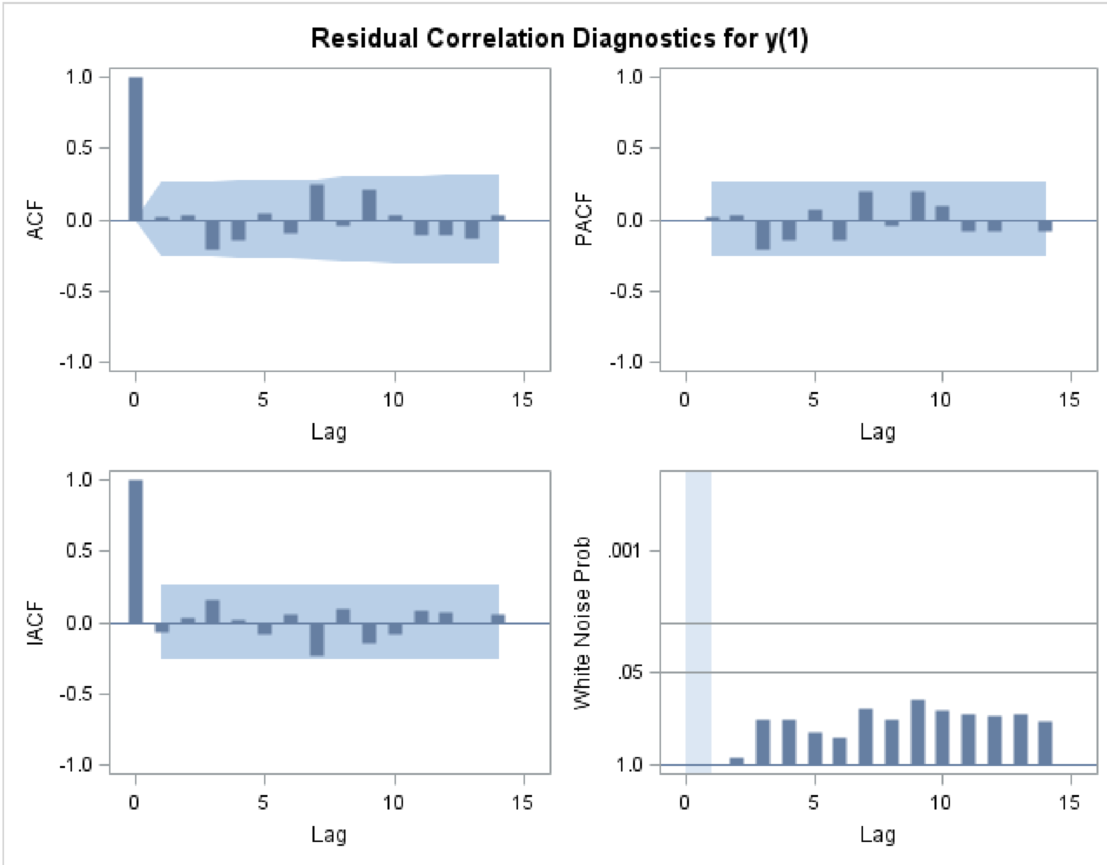

Fig 3: ACF and PACF of the residuals fitted in the ARIMA model

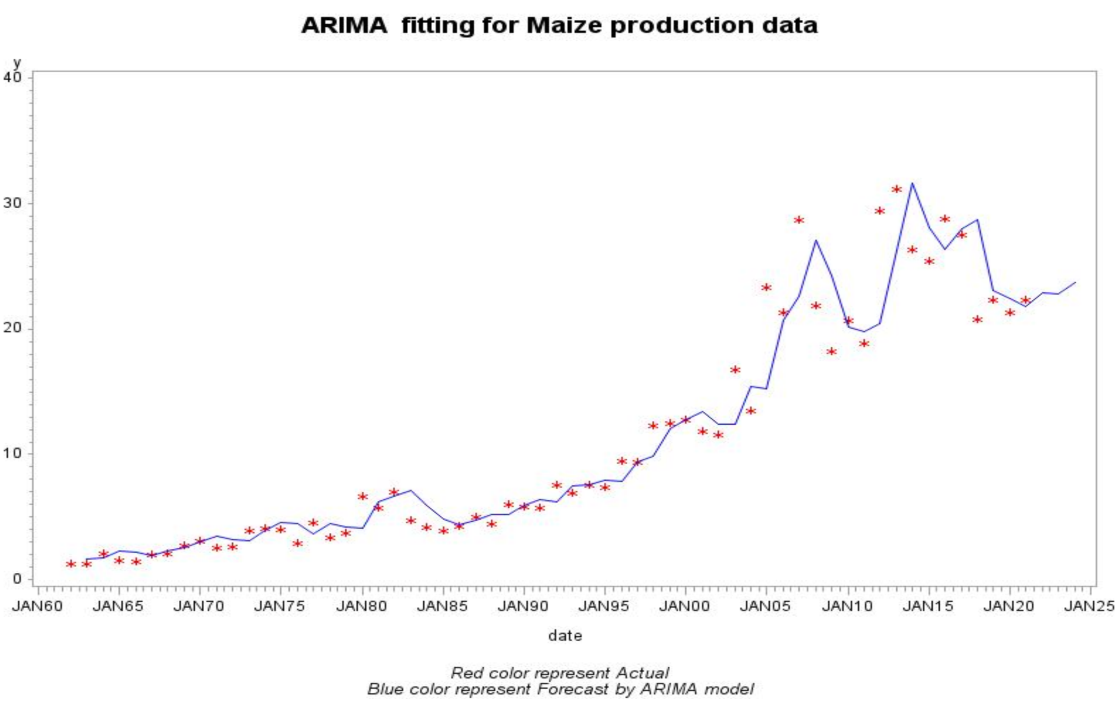

Fig. 4: Maize production in Telangana actual and forecasted data for the year 1960 to 2024

SUMMARY AND CONCLUSIONS:

The study utilized the Box-Jenkins approach for forecasting maize production in Telangana state. It involved the analysis of 55 years of empirical annual observations of maize production. Initially, the observed data for maize production was nonstationary, but it was transformed to stationary after the first difference, as confirmed by the Augmented Dickey Fuller (ADF) test.

The ARIMA (0,1,1) model was chosen based on the lowest Akaike Information Criterion (AIC) value, indicating its suitability for forecasting maize production in Telangana. Diagnostic checks were performed on the fitted model, including examining the residuals to test for randomness. The autocorrelation and partial autocorrelation plots of residuals showed randomness, and the ADF test of residuals supported the adequacy of the model. Using the ARIMA model, maize production in Telangana was forecasted to be 23.10 LT in 2022-23, 23.46 LT in 2023-24 and 23.82 LT in 2024-25.

For maize production forecasting in Telangana, the ARIMA (0,1,1) model emerged as the most suitable, with thorough diagnostic checks ensuring the randomness of residuals and the adequacy of the model. The forecasts for the upcoming years will provide valuable insights for policymakers and stakeholders in the agricultural sector.

REFFERNCES

Badmus, M.A. and Ariyo, O.S. 2011. Forecasting cultivated areas and production of Maize in Nigeria using ARIMA Model. *Asian Journal of Agricultural Sciences*. 3(3): 171-176.

Bourdon, M.H. 2011. Agricultural commodity price volatility: An overview. *OECD Food,AgricultureandFisheries Papers. OECD Publishing.*

Box, G.E.P. and Jenkins, G. 1976. Time series analysis, forecasting and control. *Holden-Day. San Francisco. CA.*

Dareakar, A. and Reddy, A.A. 2017. Price forecasting of maize in major states. *Maize Journal*. 6(1&2): 1-5.

Dareakar, A. and Reddy, A.A. 2017. Cotton price forecasting in major producing states. *Economic Affairs.* 62(3): 373-378.

Dareakar, A. and Reddy, A.A. 2017. Forecasting of Common Paddy Prices in India. *Journal of Rice Research.* 10(1): 71-75.

DES. Directorate of Economics and Statistics. Government of Telangana 2021-22. https://des.telangana.gov.in/

DES. Directorate of Economics and Statistics. Government of Telangana 2022-23. https://des.telangana.gov.in/

INDIASTAT. 2021-22. www.indiastat.com.

IIMR. Indian Institute of Maize Research. Government of India. 2021-22.

https://iimr.icar.gov.in/?page_id=51#:~:text=The%20average%20productivity%20during%20the,many%20lead%20maize%20producing%20countries.

Jadhav, V., Chinnappa Reddy, B.V. and Gaddi, G.M. 2017. Application of ARIMA Model for forecasting agricultural prices. *Journal of Agricultural Science & Technology.* 19: 981-992.

Kumari, R.V., Ramakrishna, G., Panasa, V. and Sreenivas, A. 2019. Price movements ofredgrammajormarketsinIndiabyusingcointegrationanalysis.*InternationalResearchJournal ofAgricultural Economicsand Statistics.*2(10): 234-239.

Kumari, R.V., Ramakrishna, G., Panasa, V. and Sreenivas, A. 2019. An application of ARIMAX model for forecasting of castor production in India. *Journal of Oilseeds Research.* 36(4): 244-249.

Lama, A., Jha, G.K., Paula, R.K. and Gurunga, V. 2015. Modelling and forecasting of price volatility: An application of GARCH and EGARCH Models. *Agricultural Economics Research Review.* 28: 73-82.