Indian Journal of Agricultural Sciences **94** (9): 1013–1018, September 2024/Article
https://doi.org/10.56093/ijas.v94i9.150524

Forecasting cotton (*Gossypium* spp.) prices in major Haryana markets: A time series and ARIMA approach

AJAY KUMAR[1], VINAY KUMAR[2]*, CHETNA[2], SUMAN GHALAWAT[2], JASPREET KAUR[3], KHUSHBU KUMARI[3], HIMANSHU SAHARAN[3], SHUBHI CHHABRA[3] and BASANT RAI[3]

Krishi Vigyan Kendra (CCS Haryana Agricultural University, Hisar, Haryana), Jhajjar, Haryana 124 104, India

Received: 10 January 2024; Accepted: 16 July 2024

ABSTRACT

Economic outputs are an attractive prospect in any field and hence agriculture also relies heavily on economic stability. The costs associated with cotton farming are increasing and profitability is taking a hit in cotton cultivation. Timely and accurate forecast of the price helps the farmers switch between the alternative nearby markets to sale their produce and getting good prices. Present study was carried out during 2022 to 2023 in Haryana to provide some insights into the possible future prices of cotton (*Gossypium* spp.) with the help of data collected from AGMARKNET and various major cotton markets (Adampur, Sirsa and Fatehabad) of Haryana. The Autoregressive Integrated Moving Average (ARIMA) models have been employed in order to forecast the prices of cotton crops for the years 2022–23 to 2027–28. Through a meticulous exploration of various combinations of lagged moving average and autoregressive components, the ARIMA (1,1,1) model was selected as the most suitable for the price forecasting in these districts. The results of this analysis demonstrate that the coefficient of determination (R^2) for the forecasted cotton crop prices in comparison to the real-time prices falls within acceptable ranges. This finding underscores the efficacy of the ARIMA (1,1,1) model as a reliable tool for generating short-term price estimates. This model offers valuable insights and predictive accuracy, aiding decision-makers and stakeholders in the cotton industry of Adampur, Sirsa and Fatehabad markets to make informed choices and plan effectively for the coming years. Cotton prices vary according to the season and the region, hence a valuable insight on future price assumptions will help the agriculture community.

Keywords: Autocorrelation, Coefficient of determination, Differencing, Partial autocorrelation function, Price forecast

Cotton (*Gossypium* spp.) stands as a pivotal cash crop within the realm of major commercial crops cultivated in both India and the state of Haryana. This fibrous commodity, widely acclaimed as 'White Gold,' assumes a commanding position within the domain of cash crops in India. On a global scale, India proudly occupies the 3rd rank in cotton production, trailing only behind China and the United States. India's cotton cultivation spans approximately 25% of the world's acreage dedicated to this crop; however, its contribution to global cotton production stands at a comparatively modest 14%. In 2021–22 the area, production and productivity of cotton in India was 123.71 lakh ha, 311.17 lakh bales and 428 kg/ha, respectively. Whereas in 2022–23 area, production and productivity have seen a rise to 130.61 lakh ha, 343.47 lakh bales and 447 kg/ha (COCPC 2023).

Today, cotton is a vital cash crop, particularly in developing countries, impacting both local and national economies (Gudeta and Egziabher 2019). India, with its agriculture-driven economy, leads the world in cotton cultivation, providing essential raw materials for the cotton textile industry. As we know agriculture is impacted the most because of the climatic variability, crop production and stock levels at various production and consumption centres fluctuate. This variability impacts pricing policies and trade opportunities for various agricultural commodities. Hence, reliable market predictions, including both short-term and long-term price forecasts for agricultural commodities are essential for the development of the farming community and time series forecasting models can provide important insight in this matter.

The ARIMA model, also known as the Box-Jenkins model, is widely employed for forecasting time series data. As the several studies have employed the ARIMA model for price forecasting. Verma *et al.* (2016) applied ARIMA modelling to forecast coriander prices in Rajasthan and

[1]Krishi Vigyan Kendra (CCS Haryana Agricultural University, Hisar, Haryana), Jhajjar, Haryana; [2]Chaudhary Charan Singh Haryana Agricultural University, Hisar, Haryana; [3]Haryana Space Applications Centre (CCS Haryana Agricultural University, Hisar, Haryana), Hisar, Haryana. *Corresponding author email: vinay.luhach4@gmail.com

determined that the ARIMA (0, 1, 1) model was the most appropriate. Darekar and Reddy (2017) used ARIMA for wheat price forecasting and found that the ARIMA (0, 1, 1) (0, 1, 1) model was the best fit. Since agriculture is the main source of income for majority of the people in the state and cotton being a major cash crop, this article forecasts prices in key cotton markets of Haryana. Although price forecasting carries inherent risks due to unpredictable factors that can invalidate forecasts, forecasting is crucial for making timely decisions in an uncertain future (Jadhav *et al.* 2017). Therefore, a study was carried out to forecast cotton prices in major Haryana markets utilizing the ARIMA model.

MATERIALS AND METHODS

Present study was carried out during 2022–23 in Haryana. Time series data spanning 17 years, from September 2005–May 2022, were collected from the AGMARKNET website various major cotton markets (Adampur, Sirsa and Fatehabad) of Haryana. The primary analytical tool employed in this study was the Autoregressive Integrated Moving Average (ARIMA) model. ARIMA models, initially popularized by Box and Jenkins (1976), are utilized to analyse and predict time series data. Time series data comprises a collection of values that exhibit variation over time, where the intervals between observations may vary. Nonetheless, the range of these intervals should remain consistent throughout the observed period. In empirical time series analysis, it is generally assumed that the time series is stationary. A stochastic process, representing the collection of a variable over time, can be either stationary or non-stationary. An autoregressive model entails regressing the dependent variable on one or more lagged periods of itself. When only one lagged period is included, it constitutes a first-order autoregressive stochastic process, denoted as AR(1). Extending to include p lagged periods results in a pth-order autoregressive process, denoted as AR(p). Stationarity can manifest in various forms, with weak stationarity (or second-order stationarity) being commonly employed in empirical analyses. A stochastic process is considered weakly stationary if it maintains a constant mean and variance, with the covariance being time-invariant, indicating that statistical properties do not fluctuate over time. A white noise process is a special case of a stationary stochastic process, characterized by a mean of zero, constant variance, and serially uncorrelated observations. In forecasting time series data, it is assumed that the underlying time series is stationary. Under this assumption, various forecasting models can be constructed, including moving average processes, autoregressive processes, autoregressive and moving average processes, and autoregressive integrated moving average processes (Gujarati and Porter 2008).

The study was divided into four main stages, viz. (1) Identification stage: The stationary nature of the time series data was initially assessed, revealing non-stationarity; To make the non-stationary data stationary, first-order differencing was applied; Candidate ARIMA models were developed based on the differenced data; Initial values for non-seasonal parameters "p" and "q" were determined by examining significant spikes in autocorrelation and partial autocorrelation functions; One or more tentative ARIMA models were selected to adequately represent the data; Precise parameter estimates for the chosen model(s) were obtained using least squares estimation; (2) Estimation stage: ARIMA models were used to estimate the results of the data; Accuracy of model was checked using diagnostic statistics; (3) Diagnostic checking: AIC (Akaike Information Criteria) score that is low; The Schwarz Bayesian Criteria (SBC) were occasionally taken into account; The accuracy of the model was evaluated using the Mean Absolute Percent Error (MAPE); As an accuracy statistic, the Root Mean Squared Error (RMSE) was used; To evaluate the model's precision, the coefficient of determination (R^2) was calculated; and (4) Forecasting stage: Using the chosen ARIMA model, future values of the time series were predicted.

Statistical analysis: The statistical tools like SPSS and R were used for the analysis and model creation. Numerous scenarios requiring time series analysis and dynamic systems have successfully used ARIMA models. Notably, Gwilym Jenkins and George Box invented the detailed study of ARIMA models in 1968. Since then, the terms "ARIMA processes" and "time series analysis and forecasting" have come to mean the same thing.

RESULTS AND DISCUSSION

The initial phase of this study involves the identification of suitable ARIMA model orders, denoted as p and q, for the purpose of modelling time series data. This determination is based on the rigorous analysis of ACFs and PACFs of the stationary data series. Specifically, the ACFs offer insights into the temporal correlations between data points and their respective lags while the PACFs are utilized to discern direct relationships between observations at varying lags, while mitigating the influence of shorter lags.

The investigation of ACF plots yielded a discernible pattern of gradual decline, indicative of non-stationarity within the data. In response, a first-order differencing (d=1) operation was applied to all data series under consideration to establish stationarity, an essential prerequisite for the accurate application of time series models. Moreover, the examination of PACF plots revealed a conspicuous peak at lag one, suggesting the potential presence of an autoregressive (AR) component of order one.

Following these preliminary steps, tentative ARIMA models were formulated and fitted to the data. These included ARIMA (1,1,1), ARIMA (0,1,1) and ARIMA (1,1,0) models. Based on a thorough examination of the temporal aspects of the data, these model formulations sought to predict district-level cotton yield for the markets of Adampur, Sirsa, and Fatehabad. As a result, it can help in making well-informed choices. Autocorrelations and Partial autocorrelation of cotton price for Adampur, Sirsa and Fatehabad markets are presented in Fig.1 and 2. Identification entails figuring out the correct p and q values for the AR and MA polynomials as well as their suitable

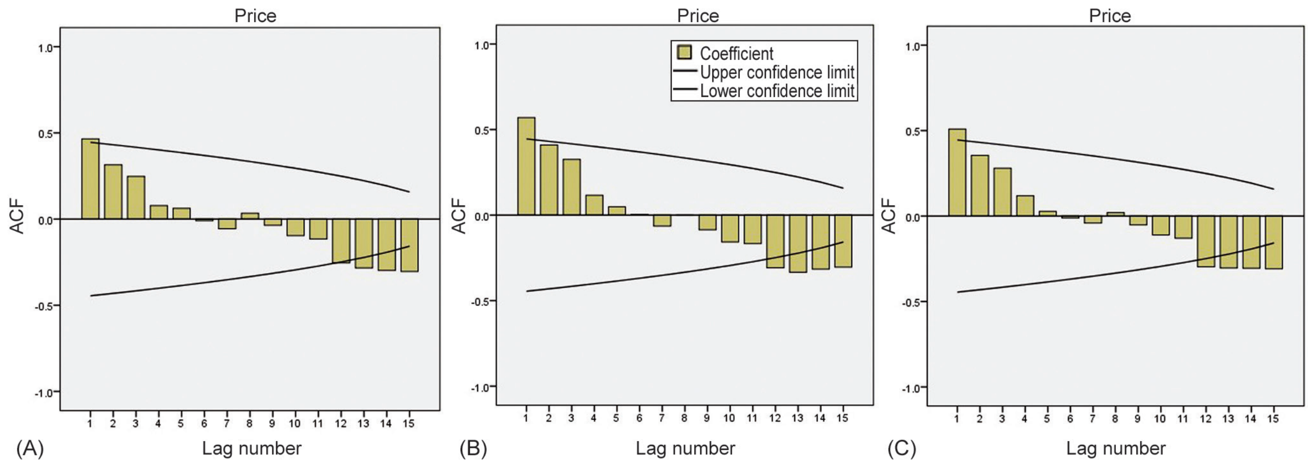

Fig. 1 Auto correlation of cotton prices for (A) Adampur; (B) Sirsa; (C) Fatehabad market.
ACF, Autocorrelation of function.

orders. The stationary series' autocorrelation and partial autocorrelation functions (ACFs and PACFs) were used to calculate the ordering. According to Fig. 1 charting of the acfs for all the districts under consideration, ACFs declines suggest non-stationarity. Order one differences were sufficient to produce an adequate stationary series for each market. The series may have one order of AR component, as seen in Fig. 2, because the PACFs revealed the presence of one substantial spike at lag one. The first differencing of the original data series converted the non-stationary data series of all the districts into stationary series (Jadhav *et al.* 2017).

Parameter estimation: We took into account three alternative ARIMA models: ARIMA (1,1,1), ARIMA (0,1,1), and ARIMA (1,1,0) throughout the identification phase of our investigation. In order to produce estimates that the sum of squared residuals, we performed parameter estimation using a non-linear least squares (NLS) method, reflecting our desire for more accurate forecasts with smaller error variance. It's worth noting that linear least squares can only be employed to estimate pure AR models, while all other models require the use of non-linear least squares.

Among the various NLS methods available, Marquardt's compromise algorithm (1963), is commonly used to estimate ARIMA models. This algorithm initially selects preliminary parameter estimates and then iteratively refines them to minimize the sum of squared residuals. Our criteria for selecting the appropriate ARIMA model included Log Likelihood, Akaike's Information Criterion (AIC, 1969), Schwarz's Bayesian Criterion (SBC, 1978), as well as residual variance.

Based on our analysis and parameter estimation, we have determined that the ARIMA models ARIMA (1,1,1), ARIMA (0,1,1), and ARIMA (1,1,0) all have significant parameters (Table 1). However, after assessing the models using criteria such as R^2, MAPE, MAE, and BIC, ARIMA (1,1,1) model was chosen for Adampur, Sirsa, and Fatehabad districts due to its superior performance in terms of these metrics. Similar results and methodology was used by Kathayat and Dixit (2021).

Diagnostic checking: The model verification process aimed to assess the residuals of the chosen ARIMA models for any systematic patterns that could be eliminated to enhance model accuracy. To accomplish this, we computed approximate t-values for the autocorrelation coefficients of the residuals, employing Bartlett's approximation to estimate the standard error of these autocorrelations. In

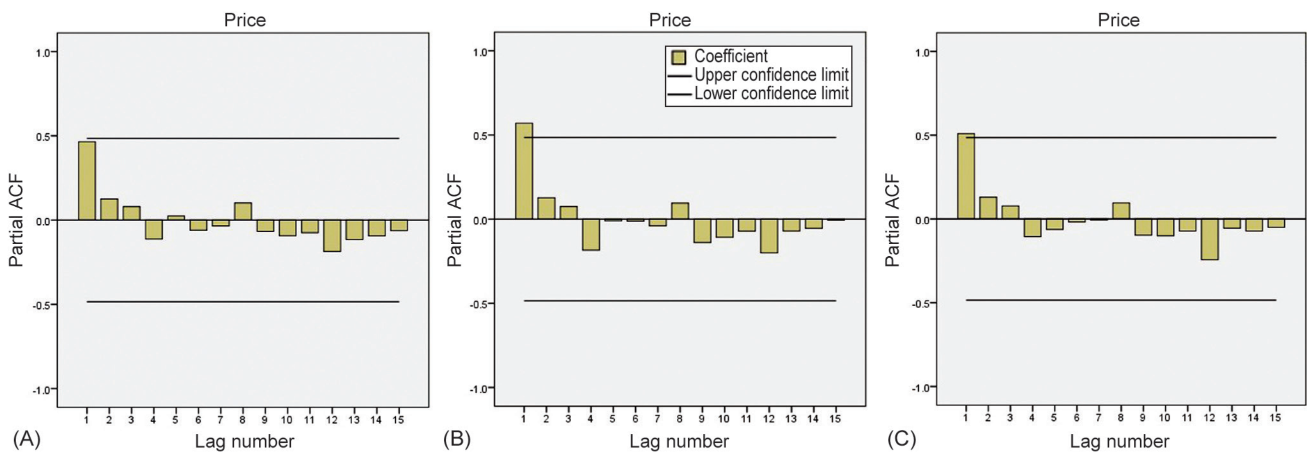

Fig. 2 Partial auto correlation of cotton price for (A) Adampur; (B) Sirsa; and (C) Fatehabad market.
ACF, Autocorrelation of function.

Table 1 Criteria for choosing ARIMA models in major markets of Haryana

District	Model	Model fit statistic(s)			
		R^2	MAPE	MAE	BIC
Adampur	ARIMA(1,1,1)	0.66	12.71	669.83	14.65
	ARIMA(0,1,1)	0.62	12.97	670.87	14.53
	ARIMA(1,1,0)	0.61	13.54	694.55	14.54
Sirsa	ARIMA(1,1,1)	0.70	11.20	599.60	14.39
	ARIMA(0,1,1)	0.66	11.10	589.69	14.22
	ARIMA(1,1,0)	0.65	11.38	590.92	14.26
Fatehabad	ARIMA(1,1,1)	0.74	10.31	510.91	13.89
	ARIMA(0,1,1)	0.74	11.41	558.98	13.67
	ARIMA(1,1,0)	0.68	11.63	568.69	13.86

ARIMA, Autoregressive integrated moving average model; R^2, Coefficient of determination; MAPE, Mean absolute percent error; MAE, Mean absolute error; BIC, Bayesian information criterion.

graphical representation it was discovered that none of the residual autocorrelation functions (ACFs) in any of the districts exhibited statistically significant departures from zero at a reasonable significance level. Additionally, the residuals appeared to approximate a normal distribution, further supporting the absence of systematic patterns in the residuals (Darekar and Reddy 2017).

During the parameter estimation stage, we experimented with various lags for both the moving average and autoregressive processes. Ultimately, the ARIMA models (0,1,1) and (1,1,0) were chosen. However, during the diagnostic checking stage, we found that the ARIMA (0,1,1) model was the most suitable fit based on non-significant Ljung-Box Q statistics (1978), indicating that the residuals resembled white noise. This model was selected to estimate cotton prices in the Adampur, Sirsa, and Fatehabad districts (Table 2).

For a more detailed description of the ARIMA (1,1,1) model fitted to these districts, the following can be elaborated:

The ARIMA (1,1,1) model incorporates three components: The three components are, differencing (d=1), autoregressive (AR) order of 1, and moving average (MA) order of 1. The differencing (d=1) suggests that we applied a first-order differencing operation to the time series data to achieve stationarity. The autoregressive component (AR=1) indicates that the current value of the time series is regressed on its previous value, with a lag of 1 time period. The moving average component (MA=1) implies that the current value of the time series is influenced by the previous error term, also with a lag of 1 time period. This combination of differencing, autoregressive, and moving average components captures the underlying patterns in the data, making it a suitable model for forecasting cotton prices in the Adampur, Sirsa, and Fatehabad districts (Biswal and Sahoo 2020).

Regenerate

(1-B) Yt= (1-θ1B) at

Yt-BYt= at- θ1B at

Yt = Yt-1 - θ1 at-1+ at……..(1)

The appropriate forecast equation is equation 1. The observed-to-predicted value plot (Fig. 3) demonstrated that all fitted models accurately capture the underlying mechanism.

The ARIMA (1,1,1) model was employed to analyze cotton price data spanning from the periods of 2005–06 to 2021–22, and to make forecasts for the subsequent periods from 2022–23 to 2027–28 (Table 3). A comparative evaluation was conducted between ARIMA-based yield estimates and estimates provided by the Department of Agriculture (DOA) in terms of key statistical metrics, namely, R^2, MAPE, MAE, and BIC. The results clearly favoured the utilization of ARIMA models for obtaining short-term forecast estimates.

It can be inferred from the examination of the cotton yield time series data that the Box-Jenkins approach

Table 2 Parameter estimates for cotton price models in major markets of Haryana

District	Model		Estimate Std.	Std. Error	t-value	*P*-value
Adampur	ARIMA(1,1,1)	AR	0.535	0.865	0.618	0.547
		MA	1.000	269.33	0.004	0.997
	ARIMA(0,1,1)	MA	0.286	0.558	0.512	0.617
	ARIMA(1,1,0)	AR	-0.613	0.487	-0.334	0.743
Sirsa	ARIMA(1,1,1)	AR	0.419	0.710	0.589	0.566
		MA	0.999	56.101	0.018	0.986
	ARIMA(0,1,1)	MA	0.505	0.401	1.261	0.228
	ARIMA(1,1,0)	AR	-0.192	0.465	-0.414	0.685
Fatehabad	ARIMA(1,1,1)	AR	0.282	0.514	0.549	0.592
		MA	1.000	234.164	0.004	0.997
	ARIMA(0,1,1)	MA	1.000	165.397	0.006	0.995
	ARIMA(1,1,0)	AR	-0.349	0.338	-1.033	0.319

ARIMA, Autoregressive integrated moving average model.

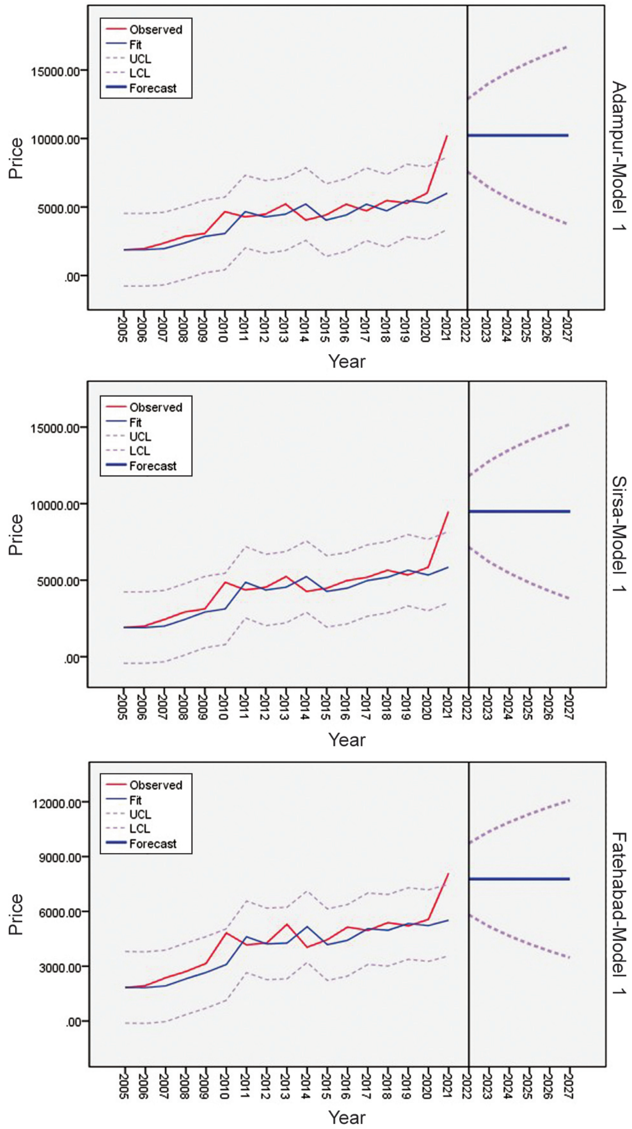

Fig. 3 Comparison of actual and predicted cotton prices using ARIMA models in major markets of Haryana.
UCL, Upper confidence limit; LCL, Lower confidence limit.

produced forecast numbers that were remarkably accurate. These forecasted yields closely aligned with the actual observed yields. It is important to emphasize, however, that certain aspects of the modelling process, such as the selection of the order of differencing and the determination of auto-regressive and moving average components, exhibit high sensitivity and can significantly impact the model outcomes. Therefore, a meticulous approach is imperative when identifying and generating these parameters for analysis, as incorrect choices may lead to misleading conclusions for decision-makers. Nonetheless, it is recommended that this technique be primarily employed for short-term forecasting purposes.

With the help from price forecasting study Adampur and Sirsa market showed seeing almost similar price increase pattern, whereas Fatehabad market on other hand showed a different price pattern with slight increase (Table 3). These findings will help in attaining better cropping pattern

Table 3 Cotton price forecast for markets of Haryana using the best-fitting ARIMA models

Year	Adampur (₹/q)	Sirsa (₹/q)	Fatehabad (₹/q)
2022–23	10751.16	9961.37	6866.70
2023–24	11272.89	10435.45	7147.19
2024–25	11794.61	10909.53	7427.68
2025–26	12316.34	11383.61	7708.16
2026–27	12838.071	11857.69	7988.65
2027–28	13359.79	12331.77	8269.14

strategies for the farmers and also will give a hint about the future prices that can be there in the markets. Kumar and Baishya (2020) reported that prices of potato will increase Madhya Pradesh, Tripura and Punjab market.

Time series analysis and forecasting is an active research area over the last few decades. The accuracy of time series forecasting is fundamental to many decision processes and hence the research for improving the effectiveness of forecasting models has never stopped. In this paper, we propose to take an approach to time series forecasting. The linear ARIMA model is used, aiming to capture different forms of relationship in the time series data. Time series is an important dimension to forecast in which past values of the variables are considered in order to develop a model. The model is then used to apply time series data into future. The existing approach is only used when lesser information is available on the data generating process or when there is no satisfactory explanatory model that relates the prediction variable to other explanatory variables. A lot of effort has been applied over the past several time to develop and improve time series forecasting models. The complex problems that have both linear and nonlinear correlation structures, the combination method can be an effective way to improve forecasting performance. The empirical results with three real data sets clearly suggest that the hybrid model is able to outperform each component model used in isolation.

The Indian economy has transitioned towards a free market system, wherein commodity prices are primarily determined by market forces driven by demand and supply dynamics. Commercial crop prices are now vulnerable to changes in both domestic supply and demand as well as changes in global supply and demand as a result of this change. In the aftermath of liberalization, price fluctuations in crops like cotton have become more pronounced, creating uncertainty for farmers regarding the profitability of their cultivation.

In this context, the importance of forecasting future prices, specifically harvest month prices, cannot be overstated. Providing farmers with such forecasts empowers them to make informed decisions when choosing which crops to cultivate. This proactive approach can significantly impact their choices and ultimately benefit their agricultural endeavours. To achieve this, the paper employs an ARIMA (Auto Regressive Integrated Moving Average) model, which leverages historical monthly cotton prices from major cotton-producing states to make predictions as stated by Box and Jenkins (1976).

It is crucial to acknowledge that, like any forecasting method, the ARIMA model does not guarantee perfect forecasts. Nevertheless, it has proven to be a valuable tool in forecasting future prices. The model operates as an extrapolation technique, relying solely on historical time series data of the variable under consideration. The forecasted prices generated by the ARIMA model reveal an upward trend in cotton prices in the coming years, accompanied by an anticipated increase in demand for this crop Borkar (2022).

As a result, this forecasting information can guide decisions regarding the expansion of cotton production areas and the timing of sales in suitable markets. It is crucial to keep in mind, that this projection is based on estimates from the model and past data. Due to unforeseen economic, environmental, or other circumstances, actual market prices may differ from the values predicted. Therefore, while this forecasting approach is valuable, it should be used as a tool for informed decision-making rather than an absolute predictor of future market prices. These findings are in accordance with Pardhi *et al.* (2018).

REFERENCES

Akaike H. 1969. Fitting autoregressive models for prediction. *Annals of Institute of Statistical Mathematics* **21**: 243–47.

Biswal K S and Sahoo A. 2020. Agricultural product price forecasting using ARIMA model. *International Journal of Recent Technology and Engineering* **8**(5): 5203–07.

Borkar P. 2022. Time series analysis of major cotton production states in India using Box-Jenkins Approach. *Special Proceedings 24th Annual Conference* **1**(2): 115–27.

Box G E P and Jenkins G M. 1976. *Time Series Analysis: Forecasting and Control*, pp. 575. Holden Day, San Francisco.

Committee on Cotton Production and Consumption Meeting (COCPC). 2023. *Cotton Corporation of India*.

Darekar A and Reddy A A. 2017. Cotton price forecasting in major producing states. *Economic Affairs* **62**(3): 373–78.

Gudeta B and Egziabher A G. 2019. Cotton production potential areas, production trends, research status, gaps and future directions of cotton improvement in Ethiopia. *Greener Journal of Agricultural Sciences* **9**(2): 163–70.

Gujarati N D and Porter D C. 2008. Basic Econometric, 5th edn. New York, USA.

Jadhav V, Reddy C B V and Gaddi G M. 2017. Application of ARIMA model for forecasting agricultural prices. *Journal of Agricultural Science and Technology* **19**: 981–92.

Kathayat B and Dixit A K. 2021. Paddy price forecasting in India using ARIMA model. *Journal of Crop and Weed* **17**(1): 48–55.

Kumar R R and Baishya M. 2020. Forecasting of potato prices in India: An application of ARIMA model. *Economic Affairs* **65**(4): 473–79.

Ljung G M and Box G E P. 1978. On a measure of lack of fit in time series models. *Biometrika* **65**: 297–303.

Marquardt D W. 1963. An algorithm for least-squares estimation of non-linear parameters. *Society for Industrial and Applied Mathematics* **2**: 431–41.

Pardhi R, Singh R and Pual R K. 2018. Price forecasting of mango in Varanasi market of Uttar Pradesh. *Current Agriculture Research Journal* **6**(2): 218–24.

Schwarz G. 1978. Estimating the dimension of a model. *The Annals of Statistics* **62**: 461–64.

Verma V, Kumar P, Singh S and Singh H. 2016. Use of ARIMA modelling in forecasting coriander prices for Rajasthan. *International Journal of Seeds and Spices* **6**(2): 40–45.