The Pharma Innovation Journal 2022; SP-11(3): 1260-1267

www.ThePharmaJournal.com

The Pharma Innovation

ISSN (E): 2277- 7695
ISSN (P): 2349-8242
NAAS Rating: 5.23
TPI 2022; SP-11(3): 1260-1267
© 2022 TPI
www.thepharmajournal.com
Received: 21-01-2022
Accepted: 10-03-2022

Rama Shankar Yadav
Department of Agricultural Statistics, College of Agriculture, Acharya Narendra Deva University of Agriculture and Technology (ANDUAT), Kumarganj, Ayodhya, Uttar Pradesh, India

Vishal Mehta
Department of Agricultural Statistics, College of Agriculture, Acharya Narendra Deva University of Agriculture and Technology (ANDUAT), Kumarganj, Ayodhya, Uttar Pradesh, India

Ashish Tiwari
Department of Agricultural Statistics, College of Agriculture, Acharya Narendra Deva University of Agriculture and Technology (ANDUAT), Kumarganj, Ayodhya, Uttar Pradesh, India

Corresponding Author
Vishal Mehta
Department of Agricultural Statistics, College of Agriculture, Acharya Narendra Deva University of Agriculture and Technology (ANDUAT), Kumarganj, Ayodhya, Uttar Pradesh, India

An application of time series ARIMA forecasting model for predicting nutri cereals area in India

Rama Shankar Yadav, Vishal Mehta and Ashish Tiwari

Abstract
This study employed a time series modeling method (Box-Jenkins' ARIMA model) to anticipate nutri cereals area in India. The optimal ARIMA model arrangement was shown to be (0,1,1). Furthermore, by applying the ARIMA (0,1,1) model to our time series data, we attempted to anticipate the future nutri cereals Area as accurately as possible for a period up to ten years. The yearly nutri cereals Area, according to the anticipated results. It will continue to fall at a pace of about 1% each year on average.

Keywords: Forecasting, time series modeling, ARIMA, nutri cereals area

1. Introduction
India, known as the original home of nutri cereals, Minor millets (Nutri cereals) are the groups of small seeded cereals and belong to the family of *Poaceae,* which are highly drought tolerant and other extreme weather conditions and are grown with low inputs of chemical such as fertilizers and pesticides. In small millets there are up to Thirty five species of grasses from twenty genera. Millets are one of the oldest cultivated foods, grown and consumed from the past more than 5000 years. Which are highly nutritious? They are provided nearly all of the essential nutrients for normal functioning of human body. Millets are good for people who are gluten-intolerant because millets are Gluten free and non- allergenic.

Millets are a kind of cereal grain. Millets such as Sorghum (Jowar), Pearl Millet (Bajra), Finger Millet (Ragi/Mandua), Minor Millets such as Foxtail Millet (Kanngani/kakun), Kodo Millet (Kodo), Proso Millet (Cheena), Little Millet (Kutki), Barnyard Millet (Sawa/Sanwa/Jhangora), Brown top millet.

In this paper, an effort is made to forecast nutri cereals Area for the 10 leading years. The model developed for forecasting is an Autoregressive Integrated Moving Average (ARIMA) model. This model was introduced by Box and Jenkins in 1960 and hence this model is also known as Box-Jenkins Model which is used to forecast a single variable. The main reason of choosing ARIMA model in this study for the forecasting is because this model assumes and takes into account the non-zero autocorrelation between the successive values of the time series data.

For this investigation, the open source statistical programme 'R' (build 4.0.5) was utilised, as well as numerous statistical and time series packages such as 'tseries' version: 0.10-48, 'predict', and 'TTR', among others.

Table 1: Yearly data of nutri cereals area Area - Million Hectares

Year	Area	Year	Area	Year	Area	Year	Area
1951	37.67	1969	46.24	1987	39.74	2005	29.03
1952	38.88	1970	47.24	1988	36.55	2006	29.06
1953	42.45	1971	45.95	1989	38.68	2007	28.71
1954	45.37	1972	43.57	1990	37.69	2008	28.48
1955	43.92	1973	42.21	1991	36.32	2009	27.45
1956	43.45	1974	46.24	1992	33.42	2010	27.68
1957	42.02	1975	43.15	1993	34.42	2011	28.34
1958	42.91	1976	43.8	1994	32.82	2012	26.42
1959	44.66	1977	41.94	1995	32.17	2013	24.76
1960	43.79	1978	42.28	1996	30.88	2014	25.22
1961	44.96	1979	42.23	1997	31.81	2015	25.17
1962	44.73	1980	41.36	1998	31.05	2016	24.39
1963	44.29	1981	41.78	1999	29.34	2017	25.01
1964	43.93	1982	42.45	2000	29.34	2018	24.29
1965	44.35	1983	40.43	2001	30.26	2019	22.15
1966	44.34	1984	41.71	2002	29.52	2020	24.02
1967	45.09	1985	39.21	2003	26.99		
1968	47.34	1986	39.47	2004	30.8		

Source of Data: Directorate of Economics & Statistics, DAC&FW, New Delhi

2. Review of Literature

Mandal (2005) [6] ARIMA modeling of yearly sugarcane output data for the years 1950-51 to 2002-03 in India projected. Autocorrelation and partial autocorrelation functions were used to examine the data. Box-Jenkins autoregressive integrated moving average model was used to fit the data. Standard statistical approaches were used to assess the model's validity. The power of the autoregressive integrated moving average model was used to anticipate sugarcane production for three years ahead.

Kumar *et al.* (2012) [5] modeling time series method (Box-Jenkins' ARIMA model) is used to anticipate sugarcane production in India. The best ARIMA model's order was discovered to be as follows: (2,1,0). Additionally, using our time series data, we attempted to anticipate future sugarcane output as accurately as possible for a period of up to five years. The forecasted findings suggest that annual sugarcane output would rise in 2013, and then fall sharply in 2014, before continuing to rise steadily in 2015 through 2017, with an average yearly growth rate of around 3%.

Nath *et al.* (2019) [7] was projected wheat output in India using a time series modeling technique. The best ARIMA model for this investigation was found to be the ARIMA (1,1,0) model. By fitting the ARIMA (1,1,0) model to our time series data, we were able to anticipate future wheat output as accurately as feasible for a period up to 10 years. The findings of the forecast suggest that annual wheat output will increase in 2026-27.

Athiyarath *et al.* (2020) [1] many applications, such as power consumption, cloud workload, weather and sales, cost of business items, and so on, benefit greatly from the analysis of such data. We have utilized many ways to learn and extract meaning complete information by first understanding the nature of the time series and the research purpose. The current research examines and analyses several forecasting algorithmic techniques, as well as their limits and use for various forms of time series data in diverse areas.

Kathayat and Dixit (2021) [4] The study was conducted with the goal of forecasting wholesale paddy prices for the 2020-21 agricultural years in five main states: Punjab, Uttar Pradesh, Tamil Nadu, West Bengal, and Delhi. The prices were forecasted using the ARIMA model. Wholesale prices for the agricultural year 2020-21 are expected to be in the range of Rs1810.23–Rs.2239.59 qt^{-1} in Punjab, Rs.1662.91–Rs.1674.98 qt^{-1} in Tamil Nadu, Rs.3010.00–Rs.3133.36 qt^{-1} in Delhi, Rs.1835.05–Rs.1902.22 qt^{-1} in West Bengal, and Rs.1080.90-Rs.1495.35 qt^{-1} in RMSE and MAPE were used to evaluate ARIMA models. ARIMA(4,0,12), ARIMA(0,1,6), ARIMA(0,1,12), ARIMA(0,1,3), and ARIMA(3,1,12) were the best suited ARIMA models for Punjab, Tamil Nadu, Delhi, West Bengal, and Uttar Pradesh, respectively.

Satrio *et al.* (2021) [9] this research compares the performance and accuracy of Facebook's Prophet Forecasting Model with ARIMA Forecasting Model using a dataset provided from the Kaggle website that contains verified cases, fatalities, and retrieved numbers. Both directly and indirectly, this has created significant disturbance in our everyday lives. To examine the performance of the forecast models, the last two weeks of actual data are compared. Despite being further from the real data as the number of days forecasted increases, Prophet outperforms ARIMA.

Biswas (2021) [2] Time series modeling with the ARIMA (p, d, q) model was employed in his research for individual univariate series of both area and production of kharif rice in West Bengal from 1962. Crop area estimation and crop yield predictions are important tools for auxiliary policymakers to use when making choices about land use, food security, and environmental issues. The goal of this research was to provide a full picture of West Bengal's present kharif rice output. West Bengal is the state that produces the most rice. According to India's agriculture statistics, paddy provides 15% of the country's total paddy production. Both models indicate a high level of accuracy for future area protuberance and kharif rice output in West Bengal.

Oswari *et al.* (2022) [8] To predict or forecast, the (ARIMA) technique with parameters Seasonal autoregressive integrated moving average (SARIMA) (2, 1, 2) x (0, 1, 1, 1) and the Long Short Term Memory (LSTM) algorithm with LSTM parameters 100, dropout 0.2, and 100 times will be employed. Agriculture is a vital part of the Indonesian economy, with a lot of potential. The agricultural business, particularly the small group, employs the bulk of Indonesian labour. During the COVID-19 outbreak, Indonesia's agriculture industry is claimed to have contributed to the country's economic prosperity. The findings of the forecasting analysis of the two models show that the LSTM model has more accurate prediction outcomes than the ARIMA model.

In fact, there are numerous research studies that show that a careful and exact selection of ARIMA model can be fitted to single variable time series data (with any type of pattern in the series and autocorrelations between consecutive values in the time series) to predict future values in the series with greater accuracy. This research also aims to anticipate future predicted values of nutri cereals in India using the ARIMA approach and time series data from the previous 70 years in million hectares.

3. Materials and Methods

3.1 Box-Jenkins (ARIMA) Model: Basics

A time series is a collection of data collected over a period of time. ARIMA models are a type of model that may give precise predictions of a single variable based on a description of past data and can correspond to both stationary and non-stationary time series. This approach differs from previous forecasting models in that it does not assume any specific pattern in the past data of the time series to be predicted. The Box-Jenkins methodology is based on the following phases in order to construct ARIMA models: (1) Model identification; (2) parameter estimation and selection; (3) diagnostic checking (or modal validation); and (4) model application.

The orders (p, d, and q) of the AR (autoregressive) and MA (mean average) components of the model are determined during model identification. Essentially, it aims to determine whether data is stationary or non-stationary, as well as the order of differentiation (d) that causes time to be stationary.

3.2 Time Series Analysis and Building ARIMA

The given set of data in Table 1 is used to develop forecasting model. The Figure 1 below represents the line plot of Nutri Cereals Area in India.

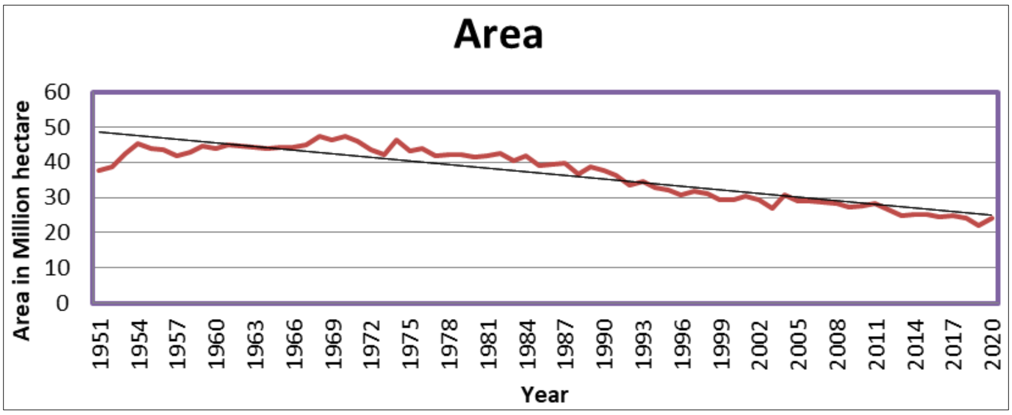

Fig 1: Nutri cereals Area (Million Hectares) in India from 1951 to 2020

Since we've already reviewed how to design an ARIMA model for variable forecasting, here are the steps to take: Before we can utilize the Model for forecasting purposes, we must first identify the model, estimate and choose parameters, and do diagnostic checking or modal validation. As a result, we'll start by trying to figure out the model's fitness.

3.3 Model Identification

The first step in creating an ARIMA model is determining if the variable being projected is stationary in time series. The term "stationary" refers to the fact that the values of a variable fluctuate around a constant mean and variance across time. Figure 1 provides a temporal plot of the Nutri Cereals Area data, which clearly illustrates that the data is not stationary (actually, it shows an decreasing trend in time series). When we make this series stationary, we may build the ARIMA model. To create an ARIMA (p,d,q) model with 'd' as the order of differencing, we must first difference the time series 'd' times to obtain a stationary sequence. Over differencing tends to increase the standard deviation rather than reduce it, hence caution should be exercised during differencing. The optimal approach is to use lowest order differencing (d=1) to construct the data and then test it for unit root issues.

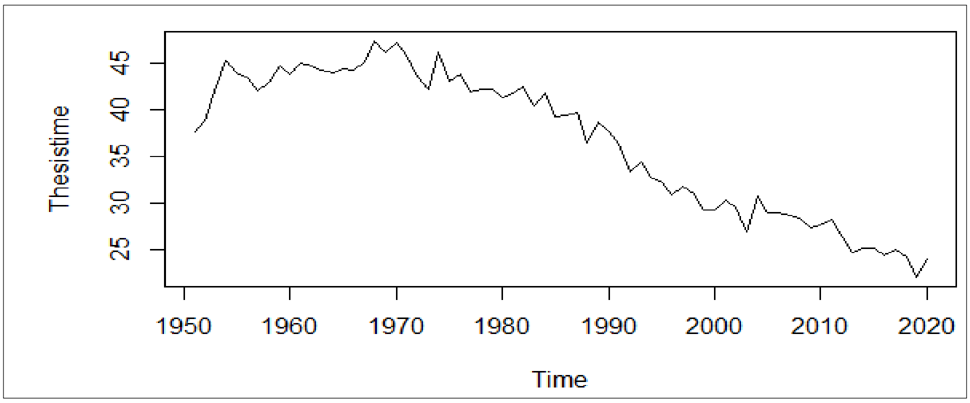

Fig 2: Line plot of Nutri cereals area data

The time series appears to be non-stationary in both its mean and variance, as can be seen in the graph above (Figure 2). But, before we go any further, we'll use the augmented Dickey-Fuller test to see if the differenced time series data is stationary (unit root problem).

3.4 Test for stationary: Augmented Dickey-Fuller (ADF) Test

The test's null hypothesis (H_0) is that the time series data is non-stationary, whereas the alternative hypothesis (H_a) is that it is stationary. The hypothesis is then assessed by applying the ADF tests to the differenced time series data after proper differencing of the data d_{th} in order. We construct a table of differenced data for the current and preceding one $(X_t = X_t - X_{t-1})$ instants using first order differencing (d=1). The following is an example of an ADF test result acquired ahead of time:

Dickey-Fuller = -2.2664, Lag order = 4, p-value = 0.4669

As a result, we reject the and thus infer that the alternative hypothesis (H_a) is correct, namely, that the series' mean and variance are stationary. As a result, we don't need to differentiate the time series any further, and we use d = 1 for our ARIMA (p,d,q) model. This test allows us to go further in the ARIMA model building process, especially in terms of determining optimal values for p in AR and q in MA in our model. To do so, we need to look at the stationary (first order differenced) time series' correlogram (auto-correlation function, ACF) and partial correlogram (partial auto-correlation function, PACF).

3.5 Correlogram and Partial Correlogram

The plot of a correlogram (ACF) for lags 1 to 18 of the first order differenced time series of the nutri cereals region in India is depicted in Figure 3.

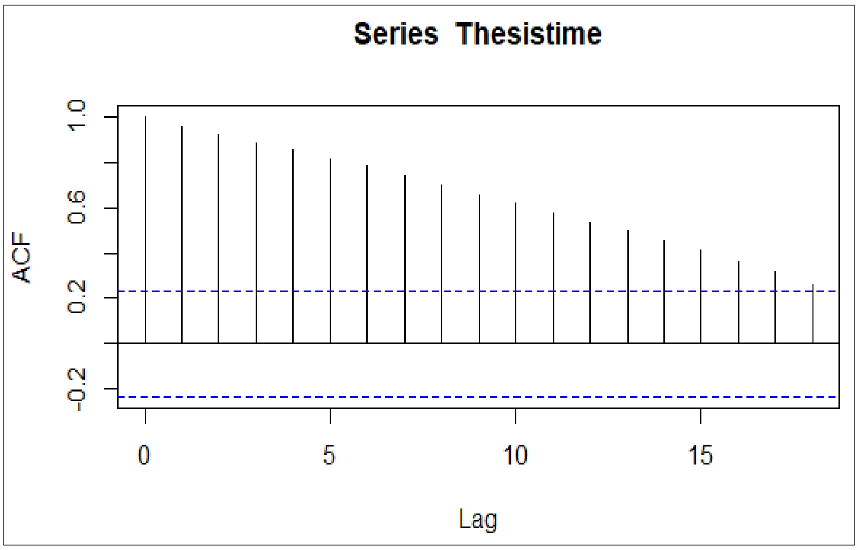

Fig 3: Autocorrelations (ACF) of first differenced series by lag

Infers that the auto-correlation at lags 1 to 18 does not surpass the significance limits, and auto-correlations tail off to zero in the above correlogram. Although the autocorrelation does not surpass the significant bounds at any latency, the remainder of the coefficients between lag 0 and 18 do not.

The partial correlogram (PACF) for lags 1 to 18 of the differenced time series is shown in Figure 4.

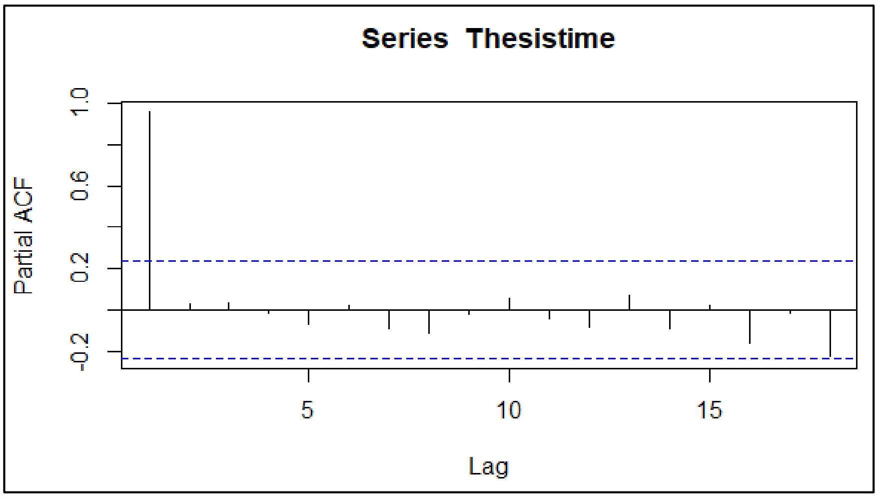

Fig 4: Partial Autocorrelations (PACF) of first differenced series by lag

The partial correlogram in Figure 4 also suggests that the partial autocorrelation coefficient does not reach meaningful limits at lag 1, and that partial autocorrelation tails down to zero at lag 2. Although there is one outlier at lag 18 (the coefficient is almost reaching the significant limits), we may infer that this is an error that occurred only by chance because all of the other PACFs from lag 2 to 18 are inside the significant limits.

3.6 Selecting the candidate model for forecasting

We can have only following nine tentative ARIMA (p,d,q) models:

Table 2: Tentative ARIMA models

ARIMA (2,1,2) with drift	260.9009
ARIMA (0,1,0) with drift	262.2882
ARIMA (1,1,0) with drift	259.1303
ARIMA (0,1,1) with drift	258.8167
ARIMA (0,1,0) with drift	261.3716
ARIMA (1,1,1) with drift	260.8028
ARIMA (0,1,2) with drift	260.7998
ARIMA (1,1,2) with drift	260.0632
ARIMA (0,1,1) with drift	259.2810

Best model: ARIMA (0,1,1) with drift

Table 3: Coefficients

	ma1	Drift
	-0.2904	- 0.2157
s.e.	0.1182	0.1301

Out of the nine models listed above, we will choose the one with the lowest BIC (Bayesian Information Criterion) and AIC (Akaike Information Criterion) values as the best model for forecasting. The output of each fitted ARIMA model in our time series (of nutri cereals area data) is summarized in Table 4:

To select as the best suitable model for forecasting out of nine above, we will choose the one with lowest BIC and AIC values. Following Table 4 summarizes the output of each of the fitted ARIMA model in our time series (of nutri cereals area data):

Table 4: AIC and BIC value of fitted ARIMA Model

ARIMA model	Coefficients		σ^2 (Estimated)	Log likelihood	AIC	BIC	AICc
	MA	Drift					
(0,1,1)	-0.2904	-0.2157	2.35	-126.41	258.82	265.52	259.19

The ARIMA(0,1,1) model with (p=0, d=1 and q=1) has the lowest AIC and BIC values, as shown in the table above, and hence this model may be the best predictive model for forecasting future values of our time series data.

4. Result and Discussion

4.1 Forecasting using selected ARIMA model

The ARIMA (0,1,1) model that we are applying to our time series data signifies that we are fitting an Autoregressive moving average ARMA(0,1) model of first order difference to our data. Also, because q is 1 in MA, the ARMA (0,1) model, which has two parameters, may be expressed as an AR model of order 0, or AR(0) model. As a result, this model may be written as:

$$X_t = \mu + (\beta_1 \times (Z_{t-1} - \mu)) + (\beta_2 \times (Z_{t-2} - \mu)) + \varepsilon_t$$

Where X_t is the stationary time series we're looking at, μ is its mean, β_1 and β_2 are parameters that need to be estimated, and ε_t is white noise with a mean that isn't constant variance? One caveat: the mean (μ) should be equal to or extremely close to zero as a norm for a stationary differenced time series. If μ is not zero, the mean value in the above equation is used to anticipate future values.

We'll now fit the specified ARIMA (0,1,1) model to our time series to anticipate future values. The projection for the next ten years is shown in Table -5, with 95 percent (low and high) prediction intervals:

Table 5: 10 year forecasted nutria cereals area in million hectare

Point	Forecast	Lo 95	Hi 95
2021	23.36824	20.36388	26.37260
2022	23.15249	19.46866	26.83632
2023	22.93675	18.68057	27.19292
2024	22.72100	17.96080	27.48120
2025	22.50525	17.28952	27.72099
2026	22.28951	16.65494	27.92407
2027	22.07376	16.04941	28.09811
2028	21.85801	15.46761	28.24841
2029	21.64227	14.90568	28.37886
2030	21.42652	14.36068	28.49236

Figures 5 and 6 provide the plot for a ten-year projection of the nutri cereals area obtained by fitting the ARIMA (0,1,1) model to our time series data:

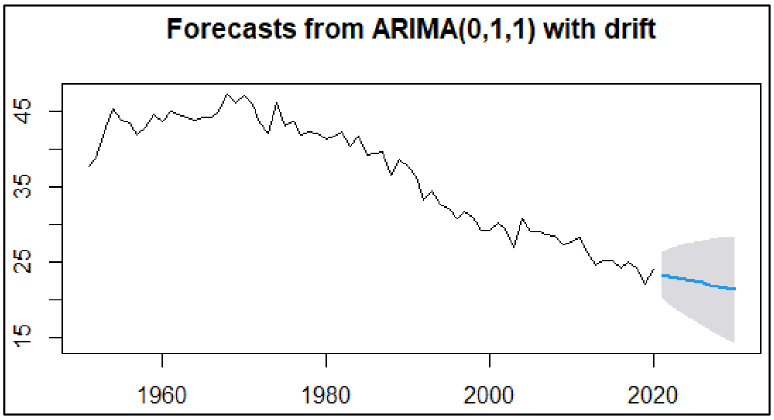

Fig 5: Forecasts from ARIMA (0,1,1)

The one-shaded forecast zones in the image above indicate the 95 percent (lower and upper side) projection of prediction intervals.

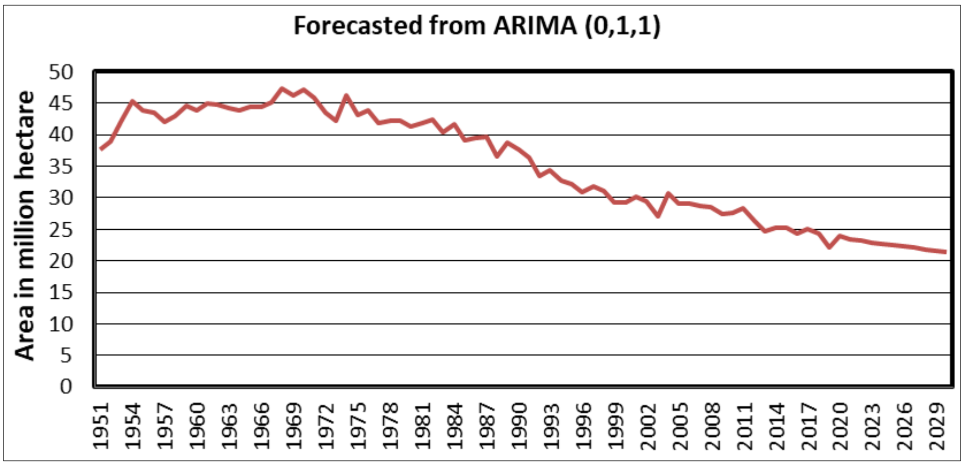

Fig 6(a): Forecast fitted with ARIMA (0,1,1)

The fitted ARIMA (0,1,1), as well as the upper and lower control limits of the forecast, are shown in Figure 6(a).

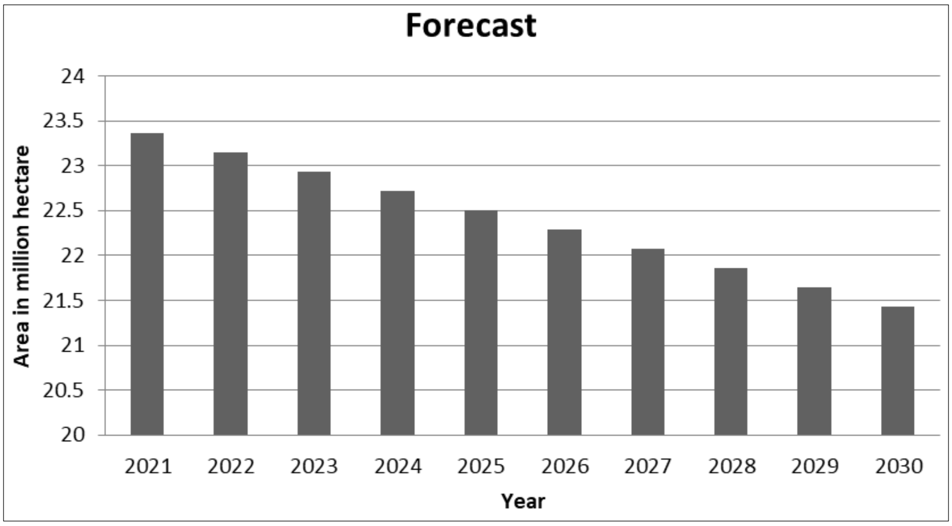

Fig 6(b): Forecast area with ARIMA (0,1,1)

Following that, we'll look at (1) if our ARIMA (0,1,1) model's forecast errors are normally distributed with mean zero and constant variance; (2) whether there are any correlations between subsequent forecast errors; and (3) whether residuals are white noise.

The standard residuals will be plotted to evaluate the distribution of forecasting mistakes. The plots and histograms of standard residuals of the fitted ARIMA (0,1,1) model are shown in Figures 7(a), and 7(b).

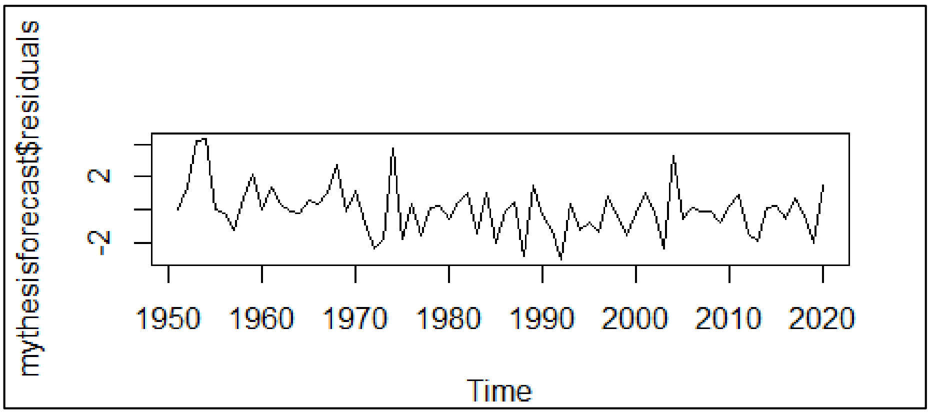

Fig 7(a): Plot of Residuals

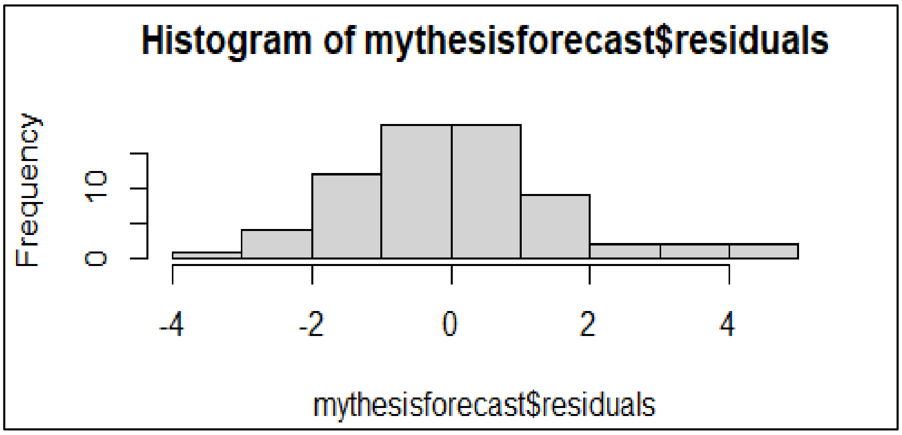

Fig 7(b): Histogram of residual of forecasted data by ARIMA (0,1,1)

The fitted model's line plots and standard residuals (shown above in Pictures) indicate that standard errors are roughly consistent in their mean and variance across time (although there seems to be some higher variance towards the end of the time series i.e. in the most recent decade). The histograms of the residuals (Picture 7c) also confirm this. The histograms (showing the errors' distribution) above indicate that the errors are normally distributed, with a mean of zero.

We will plot the correlogram (ACF) and partial correlogram (PACF) of the forecast errors to see whether there are any connections between successive forecast mistakes. The ACF and PACF of forecast errors are represented in the following Figures 8(a) and 8(b):

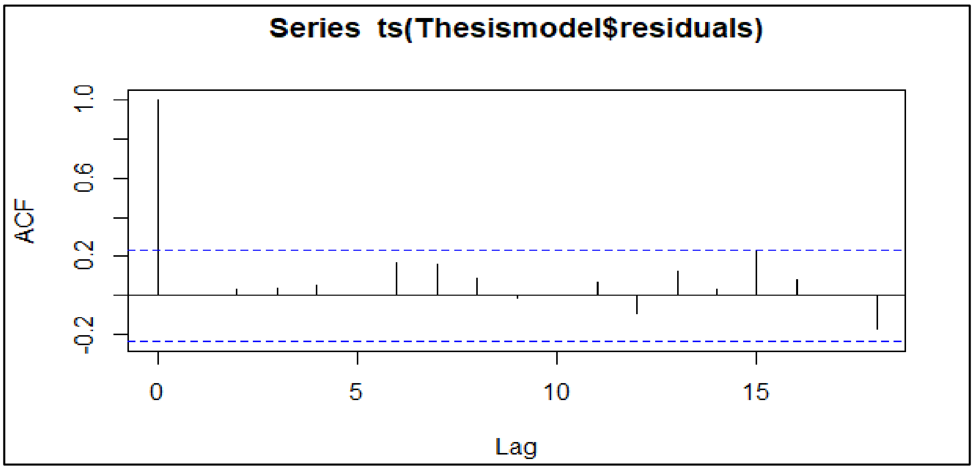

Fig 8(a): Estimated ACF of Residuals – ARIMA (0,1,1)

The autocorrelation coefficients between lags 1 and 18 do not violate the significant limits, as seen in the ACF figure above, and all values of the ACF are well inside the significant boundaries.

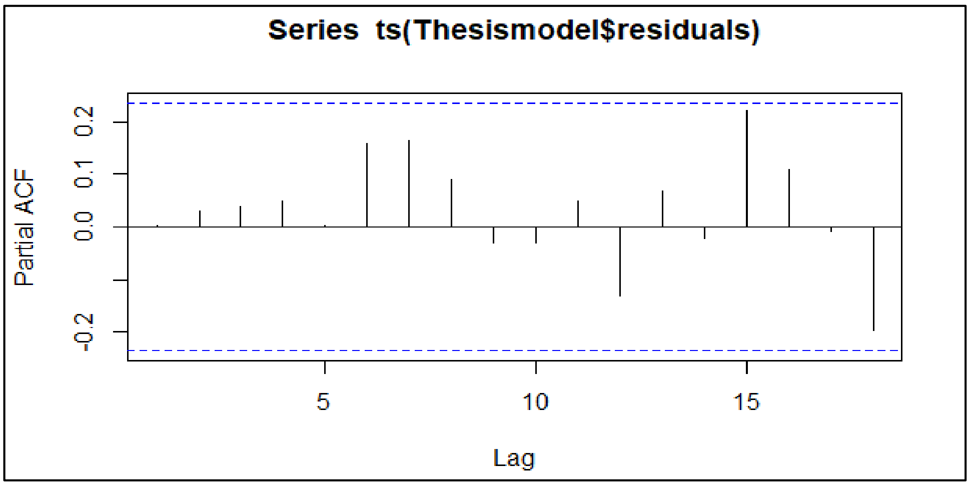

Fig 8(b): Estimated PACF of Residuals – ARIMA (0,1,1)

Similarly, for lag 1 to lag 18, all ACFs, PACFs, and partial autocorrelation coefficients of residuals of fitted ARIMA are within the significant bounds. This indicates that under the fitted ARIMA (0,1,1) model, ACF and PACF found no non-zero autocorrelations in the forecast residuals (or standard errors) at lags 1 to 18. The Box-Ljung and Box-Pierce test statistics for the fitted model are provided in Table 6 below:

Table 6: Box-Ljung Test Statistics

Lag	X^2	Degree of freedom	p-value
4	0.37164	4	0.9847
12	6.391	12	0.8951
20	17	20	0.6529

The statistics and high p-values in both tests indicate that we should accept the null hypothesis that all autocorrelation functions in lag 1 to 18 are zero. In other words, we may infer that in our fitted model, there is no (or practically no) evidence for non-zero autocorrelations in forecast errors at delays 1 to 18.

5. Summary and Conclusions

The ARIMA (0,1,1) model was chosen as the best candidate model for producing forecasts for up to ten years for the nutri cereals area in India utilizing a 70-year time series data in this study. ARIMA was chosen because of its ability to anticipate utilizing time series data with autocorrelations between successive values in the time series and with any type of pattern. The study also found that the consecutive residuals (prediction mistakes) in the fitted ARIMA time series were not associated, and that the residuals were not normally distributed. As a result, we may infer that the ARIMA (0,1,1) model chosen appears to be an appropriate forecasting model for the nutri cereals sector in India.

The ARIMA (0,1,1) model predicted a decline in area for 2021 and every year after that until 2030, with an overall loss in area (Table 2). The projection for 2021 is about 23.37 million hectare (at confidence interval 95 percent), 22.51 million hectare (at confidence interval 95 percent) for 2025, and 21.43 million hectare (at confidence interval 95 percent) for 2030. (at confidence interval 95 percent).

Despite the fact that ARIMA, like any other predictive model in forecasting, has limits in terms of prediction accuracy, it is nevertheless commonly employed for projecting future consecutive values in time series.

6. References

1. Athiyarath S, Paul M, Krishnaswamy S. A Comparative Study and Analysis of Time Series Forecasting Techniques. SN Computer Science. 2020;1(175):1-7.
2. Biswas R. Study on ARIMA Modeling to Forecast Area and Production of kharif Rice in West Bengal. Cutting-Edge Research in Agricultural Sciences. 2021;12:142-151.
3. India. Ministry of Agriculture and Farmers Welfare, Department of Agriculture. Agricultural Statistics at a Glance. New Delhi, Government of India, 2021, 71p.
4. Kathayat B, Dixit AK. Paddy price forecasting in India using ARIMA model. Journal of Crop and Weed. 2021;17(1):48-55.
5. Kumar M, Anand M. An Application of Time Series ARIMA Forecasting Model For Predicting Sugarcane Production In India. Studies in Business and Economics. 2012;8:81-94.
6. Mandal BN. Forecasting Sugarcane Production in India with ARIMA Model. A Journal of Theoretical and Applied Statistics. 2005;1:1-13.
7. Nath B, Dhakre D, Bhattacharya D. Forecasting wheat production in India: An ARIMA modeling approach. Journal of Pharmacognosy and Phytochemistry. 2019;8(1):2158-2165.
8. Oswari T, Yusnitasari T, Kusumawati RD, Setiawan I. Prediction Analysis of Food Crop Farmer Index Price during Covid-19 Pandemic using ARIMA and LSTM. Journal of Management Information & Decision Sciences. 2022;25:1-10.
9. Satrioa CB, Darmawana W, Nadiaa BU, Hanafiah N. Time series analysis and forecasting of corona virus disease in Indonesia using ARIMA model and PROPHET. Procedia Computer Science. 2021;179:524-532.