

Instability analysis and forecasting using ARIMA Model in Fruit Production in India

^{1*}POOJA PATEL, ²A.J. WILLIAMS, ³G.F.AHMED, ⁴KAMINI YADAV AND ⁵POONAM DUBEY

¹Department of Farm Engineering, B.H.U., India

²BTC College of Agriculture and Research Station, IGKV, Sarkanda, Bilaspur - 495 001, India.

³College of Agriculture, Powarkheda, J.N.K.V.V., (M.P.), India

⁴College of Agriculture, Rewa, J.N.K.V.V., (M.P.), India

⁵T.R.S. P.G. College, Saharnpur, (U.P.), India

*Corresponding author email id: poojapatel14111999@gmail.com

Abstract: India is a significant agrarian nation, with 77% of India's population living in rural areas and 90% of that rural population employed directly in agriculture. The country's main fruit crops include bananas, guavas, papayas, jackfruits, pineapples, mangoes, and others. The primary aim of this study is to forecast the major types of fruit production in India by applying the Box-Jenkins Autoregressive Integrated Moving Average (ARIMA) model using time series models during 1961-2020. The findings indicate that the best ARIMA models for predicting the production of bananas, apples, grapes, mangosteens, guavas, and pineapples are ARIMA(1,1,0), ARIMA(1,1,4), and ARIMA(1,1,5). The comparison between the original and forecasted series, presented in a similar manner, indicates that the fitted model reliably forecasts fruit production in India. In the present investigation aim to analyze the instability of fruit crops in India and to estimate the decomposition analysis of fruit crops specifically in India. To achieve these objectives, secondary data spanning 60 years were analyzed using statistical tools such as the Coefficient of Variation (CV) and Cuddy Della Valle's Instability Index (CDVI).

Keywords: Fruits Productions, Forecasting, ARIMA Model, trend, instability, India

INTRODUCTION

Fruits are an essential part of the human diet. It is widely believed that a nation's per capita fruit production and consumption can be assessed by assessing its living level. The nation produces a wide range of fruits. In addition to being the world's top ten producer of apples, papayas, citrus fruits, grapes, guavas, and pineapples. India is the world's largest producer of bananas and mangoes. The Ministry of Agriculture and Farmers Welfare has published the 2nd Advance Estimates for the area and production of various horticultural crops for the year 2022-23. According to these estimates, fruit production is projected to reach 108.34 million tones in 2022-23, showing a slight increase from the 107.51 million tones recorded in 2021-22. showcasing a rise in production compared to the previous year,

with major contributors being apples, bananas, grapes, mangoes, and watermelon; the average productivity stands at approximately 15.22 tons per hectare across the cultivated fruit area of 7.06 million hectares, India's primary fruit-producing states include Uttar Pradesh, Andhra Pradesh, Bihar, Karnataka, Kerala, Tamil Nadu, Maharashtra, and West Bengal. Additionally, Gujarat, Assam, Madhya Pradesh, and Odisha are significant contributors. The per capita fruit consumption in the country is 46 grams per day. much lower than the recommended 92g/day by the ICMR and NIN. In 2024, the total area under fruit cultivation increased to 7,046ha, up from 7,024 ha in 2023.

Bananas, often referred to as the "Apple of Paradise," are among the oldest cultivated

fruits, originating from Southeast Asia. They are highly nutritious, providing approximately 89 calories per 100 grams. This tropical fruit is rich in carbohydrates, vitamins, and minerals, making it a valuable energy source. Nutritional Profile per 100 grams: Calories (89 kcal), Water (75%), Protein (1.1 grams), Carbohydrates (22.8 grams), Sugars (12.2 grams), Dietary Fiber: 2.6 grams, Fat: 0.3 grams, Vitamin B6: 0.37 mg (22% DV), Vitamin C: 8.7 mg (10% DV), Potassium: 320 mg (16% DV) and Manganese 0.36 mg (18% DV). India is the world's leading banana producer, with an estimated production of 34.9 million metric tons in the fiscal year 2022-2023. The primary banana-growing states include Tamil Nadu, Maharashtra, and Gujarat. Over the past decade, India's banana exports have increased significantly, with a compound annual growth rate (CAGR) of 5%. In 2022, India ranked 18th in banana exports, totaling \$162.8 million. The country aims to boost exports further, targeting \$1 billion within the next five years. Beyond their dietary value, bananas have various applications. They are used in traditional medicine, animal feed, and even as a source of natural fibers for textiles. Their rich nutrient content, including potassium and vitamin B6, contributes to several health benefits, such as supporting heart health and aiding digestion.

Apple cultivation constitutes approximately 2.8% of the total fruit production in India, ranking sixth in terms of output. An apple is a highly nutritious fruit, often associated with the saying, "An apple a day keeps the doctor away." Nutritional Content in a Medium-Sized Apple (100g): Calories (52 kcal), Carbohydrates (14g), Fiber (2.4g), Protein (0.3g), Fat (0.2g), Vitamin C (4.6mg), Potassium (107mg), Vitamin A (54 IU), Antioxidants: Includes quercetin, flavonoids, and polyphenols, Water Content (86%), Health Benefits of Apples, Heart Health, Digestive Health, Weight Management, Blood Sugar Control, Immune Boosting, Cancer Prevention, Bone Health, Brain Health.

Pineapple (*Ananas comosus*) is among the most significant tropical and commercial fruit crops worldwide, ranking third after bananas and citrus fruits. Renowned for its exceptional flavor and taste, it is often referred to as the "queen of

fruits." It is rich in moisture, sugars, ascorbic acid, and essential nutrients like calcium, potassium, and vitamin C, while being low in crude fiber. Pineapple is consumed fresh or as juice and supports good digestion, balanced nutrition, and ideal weight maintenance, making it a valuable supplementary fruit for health.

Grapes (*Vitis vinifera*) are a popular fruit crop worldwide. India ranks seventh in grape production. India produces 3.4 million tones of grapes annually, according to recent data. This puts India in the top ten grape-producing nations, contributing significantly to the global grape supply. Indian agriculture relies heavily on grapes (*Vitis vinifera*). They are typically cultivated in the subtropical regions of India. They can be used in raisins, jam, jelly, juice, and wine. Grapes can boost immunity. In vitro and in vivo studies show strong antioxidant activity, cancer cell proliferation inhibition, platelet aggregation suppression, and cholesterol reduction. Grapes contain antioxidants like phenolic acids, stilbenes, anthocyanins, and proanthocyanidins (Yang et al., 2013).

Mango is one of the most widely consumed fresh fruits globally, continuing to dominate the Indian fruit market. The mango is recognised as the preeminent tropical fruit of Asia, cultivated commercially in over 87 countries. Mango occupies the foremost position in the area dedicated to fruit crops, comprising 21.83%, and ranks second in total fruit production at 35.53%, following the banana crop (HSD, 2014-15). Mangoes are rich in vitamins A, C, E, and K, which bolster immune function and promote skin health. They comprise minerals including magnesium, potassium, and calcium. They play a crucial role in maintaining bone health and supporting muscle function. Mangoes contain antioxidants such as beta-carotene and zeaxanthin, which safeguard cells from damage. A serving of mango substantially fulfills daily vitamin C and vitamin B6 needs.

This paper aims to predict fruit production for the next nine years. The forecasting model constructed is an (ARIMA) model. The Box-Jenkins Model, introduced by Box and Jenkins in 1960, is utilized for forecasting a singular variable. The primary rationale for selecting

the ARIMA model in this study for forecasting is its assumption and consideration of the non-zero autocorrelation among successive values of the time series data. This study utilizes the open-source statistical software 'R' (version 3.0.1) in conjunction with various statistical and time series packages, as well as other standard packages. Instability is an intrinsic attribute of agriculture universally. The area, yield, and production of crops are significantly influenced by weather conditions and are subject to considerable annual variation. The primary aim of this study is to create an ARIMA model for forecasting the production of various fruit crops in India. The specific objectives include developing an Autoregressive Integrated Moving Average (ARIMA) model to predict the production of seasonal fruits such as banana, apple, Mango, guava, and Pineapple in India. Additionally, the study will assess the production performance of key fruit crops, explore the instability in their production, and analyze the trends and challenges associated with major fruits in India.

MATERIAL AND METHODS

Source of the Data

Research methodology outlines the scientific approach to conducting research, while the techniques and methods used are termed research methods (Sahu, 2013). These methods focus on collecting and analyzing information to address the research problem and evaluate the accuracy of results. This chapter offers a summary of the study area, outlines the nature and sources of the data, and specifies the study period for the analysis. It also details the statistical tools and techniques employed in the data analysis.

Analytical tools are vital to research methods. Selecting appropriate tools that align with the research subject is essential for achieving study objectives. These tools begin with describing and summarizing the collected information. To assess the characteristics of the data series, various descriptive statistical measures are utilized. Descriptive statistics emphasize the fundamental characteristics of the data, providing concise summaries of the sample and its key measures. When combined with graphical representations,

they serve as the foundation for quantitative analysis, setting them apart from inferential statistics.

Descriptive Statistics

Descriptive statistics offer a summary of the data, emphasizing its displayed features. On the other hand, inferential statistics make conclusions that go beyond the immediate dataset. For instance, inferential statistics enable us to deduce population characteristics from sample data and determine whether observed differences between groups are significant or merely due to chance.

Descriptive statistics simplify large datasets into clear, manageable summaries. They can be categorized into measures of central tendency, measures of dispersion, and measures of association. Common methods include the arithmetic mean, standard deviation, skewness, kurtosis, correlation, and regression, which are effective tools for accurately describing data.

Trend models

A model represents a process or system, with statistical models specifically designed to illustrate the progression of a process while highlighting its statistical characteristics. This study examines the behavior and nature of a series through various models. Among the types of statistical models—parametric, non-parametric, and semi-parametric—parametric models are particularly recognized for their useful statistical features. In this work, we analyzed several parametric models, and a brief summary of them is provided below.

Linear Model: $Y_t = b_0 + (b_1 t)$

Quadratic Model: $Y_t = b_0 + (b_1 t) + (b_2 t^2)$

Compound Model: $Y_t = b_0 (b_1^t)$ Or $\ln(Y_t) = \ln(b_0) + t \ln(b_1)$

Cubic Model: $Y_t = b_0 + (b_1 t) + (b_2 t^2) + (b_3 t^3)$

Exponential Model: $Y_t = b_0 e^{(b_1 t)}$ Or $\ln(Y_t) = \ln(b_0) + (b_1 t)$

Logarithmic Model: $Y_t = b_0 + b_1 \ln(t)$

Growth Model: $(Y_t) = b_0 + b_1 t Y_t$

Where Y_t is the value of the series at time t and b_0, b_1, b_2, b_3 are the parameters.

The best model for each series is chosen based on the highest R^2 value, lowest standard error, and significant coefficients. If multiple models meet these criteria, the one with the fewest parameters is selected. The best-fitting models are also shown graphically alongside the observed values.

Instability and its measures

Since independence, India has experienced three phases of agricultural production: a) pre-green revolution (1950-1967), b) green revolution (1968-1988), and c) post-green revolution (1989-2009). Each phase demonstrates unique levels of growth and development. Studies typically analyze these periods separately. The first phase had limited technological advancement, while the second saw the early adoption of green revolution technologies and significant policy initiatives. The third phase is marked by widespread technology dissemination and trade liberalization.

Instability in agriculture is usually measured using the coefficient of variation and variance, with standard deviation also employed for comparing import instability across countries.

$$\bar{Y} = \frac{1}{n} \sum_{t=1}^n Y_t$$

$$\text{Variance of } Y = \frac{1}{n} \sum (Y_t - \bar{Y})^2$$

Standard deviation of Y_t is the positive square root of its variance.

For measuring the instability in area, production and yield the index given by Cuddy and Della (1978) and used by Larson *et.al.*, (2004):

$$CV_t = (CV) \times \sqrt{1 - R^2} \quad \text{where,} \quad C.V. = \frac{\sigma}{\bar{X}} \times 100$$

Where σ = Standard Deviation

$$\bar{X} = \text{Mean}$$

R^2 = coefficient of determination of the linear trend model of the variable concerned.

CV_t = CV around trend

The ordinary coefficient of variation (CV) is widely used but often inadequate for addressing trend components in time series data (Hasan *et al.*, 2008). In response, the method proposed by Cuddy and Della (1978) offers a more robust alternative, focusing on the coefficient of variation around the trend, or CV_t. Although the Cuddy and Della method is beneficial, its linear trend assumption can limit its effectiveness, especially in long-duration time series where pure linear trends are uncommon. To enhance our analysis, we incorporate nonlinearity into trend models and use the coefficient of determination (R^2) from the best-fitting models to calculate CV_t values. This modified approach provides a more accurate reflection of trends, leading to potential differences in R^2 compared to the original Cuddy and Della method. However, the variation in nonlinear trend models across different series can complicate comparisons of CV_t values. This challenge highlights the significance of our approach in delivering deeper insights into time series data.

Autoregressive Integrated Moving Average model (ARIMA)

ARIMA models are a comprehensive set of models used for forecasting time series data, which can be made stationary through transformation or differencing. These models are essentially advanced versions of random-walk and random-trend models, incorporating lags of the differenced series and/or forecast errors into the prediction equation to eliminate any remaining autocorrelation in the forecast errors.

The acronym ARIMA stands for "AutoRegressive Integrated Moving Average." In this model, lags of the differenced series are referred to as "auto-regressive" terms, while lags of forecast errors are called "moving average" terms. A time series that requires differencing for stationarity is considered an "integrated" version of a stationary series.

A nonseasonal ARIMA model is typically denoted as "ARIMA (p, d, q)," where:

- **p** is the number of autoregressive terms,
- **d** is the number of nonseasonal differences, and

- q is the number of lagged forecast errors included in the prediction equation.

To find the right ARIMA model for a time series, first determine the necessary differencing order to achieve stationarity and remove seasonal features, possibly using transformations like logging. If you predict a constant differenced series, you have a random walk or random trend model. The random walk model predicts a constant first difference, while the seasonal random walk predicts a constant seasonal difference. Even the best models may show autocorrelated errors, suggesting the need for additional factors. The process involves four key steps: Identification, Estimation, Diagnostic Checking, and Forecasting.

Identification

The objective is to determine suitable values for p , d , and q in the ARIMA model by examining the autocorrelation function (ACF), partial autocorrelation function (PACF), and their respective correlograms. By analyzing the ACF and PACF patterns for ARMA processes such as AR(1), AR(2), MA(1), MA(2), and ARMA(1,1), the model can be identified if it matches one of these patterns. Finally, diagnostic tests should be conducted to confirm that the chosen ARIMA model appropriately fits the data.

Estimation

After determining p and q , estimate the parameters of the autoregressive and moving average components. This can usually be done with least squares, though non-linear methods may be needed in some cases.

Model Formulation

The whole period under consideration (1961-2020) has been divided into two parts.

- The model formulation period (1961-2020).
- Forecasting period up to (2021-2029).

Model Evaluation

The best Box-Jenkins model is selected based on maximum R^2 , minimum RMSE, MAPE, MaxAPE, MaxAE, and normalized BIC. The chosen model

meets the most criteria. This section defines the relevant goodness-of-fit measures.

RESULTS AND DISCUSSION

This descriptive analysis provides insights into the range, variability, and distribution shape of the data, highlighting potential differences in the patterns across these fruits. Bananas range from 2,257 to 31,504, with a mean of 12,282.65. The standard deviation is high (10,155.63), indicating a substantial spread in the data. The variance (103,136,734.16) further supports this. The skewness is 0.784, suggesting a moderately right-skewed distribution. The kurtosis is -0.960, indicating a flatter distribution than a normal curve. Apple data ranges from 90 to 2,891, with a mean of 1,153.46. The standard deviation is 751.97, reflecting a moderate spread. The variance is 565,464.21. The skewness (0.440) suggests slight right skewness, while kurtosis (-0.621) indicates a slightly flatter-than-normal distribution. Grapes have values ranging from 70 to 3,125, with a mean of 889.95. A standard deviation 928.92 reflects a moderate to high spread, with a variance of 862,890.14. The skewness (1.110) shows significant right skewness, while the kurtosis (0.023) is close to normal. Mangoes, Mangosteens, and Guavas have values from 6,988 to 25,631, with a mean of 11,091.62. The standard deviation (4,863.78) and variance (23,656,356.14) indicate substantial spread. The skewness (1.633) points to a highly right-skewed distribution, and the kurtosis (2.093) suggests a sharper peak compared to a normal distribution. Pineapples range from 76 to 1,984, with a mean of 904.23. The standard deviation is 515.11, indicating moderate variability. The variance is 265,336.38. The skewness (0.353) suggests minimal skewness, and the kurtosis (-0.901) shows a flatter-than-normal distribution (Table 1).

Bananas and mangoes/mangosteens/guavas show the greatest variability and positive skewness, indicating greater spread and more frequent higher values. Grapes and mangoes/mangosteens/guavas exhibit the most pronounced skewness and kurtosis, suggesting non-normal distributions. Apples and pineapples have the least variability and near-symmetric

Table 1: Descriptive Statistics of major fruits in India

	Minimum	Maximum	Mean	Std. Deviation	Skewness	Kurtosis
Bananas	2257.000	31504.000	12282.653	10155.626	0.784	-0.960
Apples	90.000	2891.000	1153.465	751.974	0.440	-0.621
Grape	70.000	3125.000	889.954	928.919	1.110	0.023
Mangoes, mangosteens, guavas	6988.000	25631.000	11091.620	4863.780	1.633	2.093
Pineapples	76.000	1984.000	904.227	515.108	0.353	-0.901

distributions, with flatter shapes than normal. These statistics provide useful information about the spread, central tendencies, and shapes of the distributions, which can guide further analysis or decision-making.

Test of outliers and randomness for production of fruits in India

Having gained an understanding of the production of major fruit crops in India, the next objective is to examine the growth patterns of all the parameters under study. Before identifying trends in each series, it is essential to assess whether the series show any patterns or exhibit random behavior. Prior to conducting the test for randomness, the series are first tested for outliers, following the method outlined in the materials and methods section. The results of both the randomness and outlier tests are summarized in a table 2.

Table 2: Test of randomness of major fruits in India

	Banana	Apple	Grape	Mango	Pineapple
No. of Observation	60	60	60	60	60
No. of Point (p)	23	35	22	21	26
E (P)	38.67	38.67	38.67	38.67	38.67
V(P)	10.70	10.34	10.34	10.34	10.34
t_{cal}	-4.79	-1.14	-5.18	-5.49	-3.94
Inference	Trend	Random	Trend	Trend	Trend
outliers Test	NO	NO	No	NO	No

The t-statistic of banana (**tc_{al} = -4.79, Trend**) suggests a significant deviation from the expected value, leading to the inference of a trend. There are no outliers in the data. The t-statistic of **APPLE (tc_{al} = -1.14, Random)** is within the critical limits, indicating no significant deviation from randomness. The t-statistic of **GRAPE (tc_{al} = -5.18, Trend)** exceeds the critical value, suggesting a significant trend. A t-statistic of **mango (tc_{al} = -5.49, Trend)** also indicates a

significant trend. The t-statistic of **PINEAPPLE (tc_{al} = -3.94, Trend)** suggests a significant trend as well. The “No” under the outliers test means no extreme values were found in the data for any product. This implies that the results (trend or randomness) are not influenced by unusually large or small observations.

Trend analysis

Based on the overall performance outlined above, the trajectory of the series was analyzed using parametric trend models. To examine the trends in area, production, and yield, various parametric models such as polynomial, logarithmic, compound, growth, and exponential models, as described in the Materials and Methods section, were applied. Among the competing models, the best model was selected based on the highest R² value, the significance of the model, and the relevance of its coefficients

In all cases, non-linear patterns are revealed, in all the series cubic models fitted well. This table summarizes a **trend analysis** for different fruits (Banana, Apple, Grapes, Mango, and Pineapple) using a best fitted cubic equation to model the relationship between the dependent variable crop production and its independent variables time. **Equation** indicates that the trend is modeled using a **cubic equation**, which includes terms up to the third power. **R Square:** Coefficient of determination, showing how well the cubic model fits the data. A value closer to **1.0** indicates a very good fit (e.g., Banana: 0.972 or 97.2% of the variability is explained by the model). F-statistic from the regression analysis, indicating the overall significance of the model. Larger values signify stronger predictive power. **df1** and **df2:** Degrees of freedom for the regression and error terms, respectively. Significance value (p-value). A value of **.000** means the model is highly significant and the relationship is

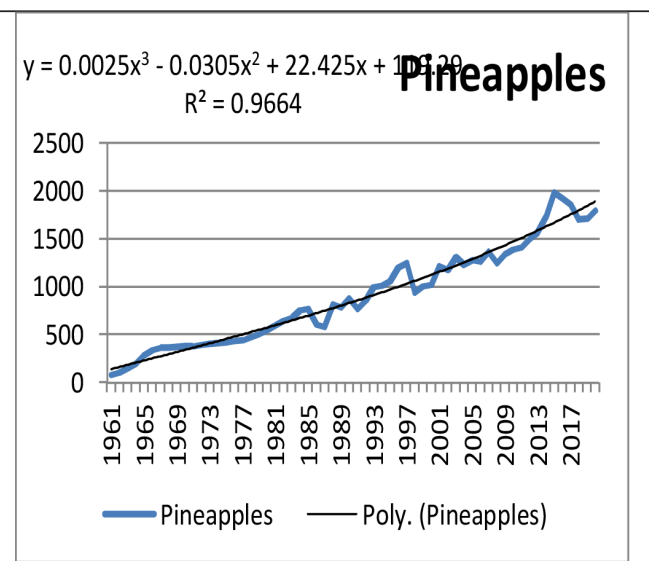
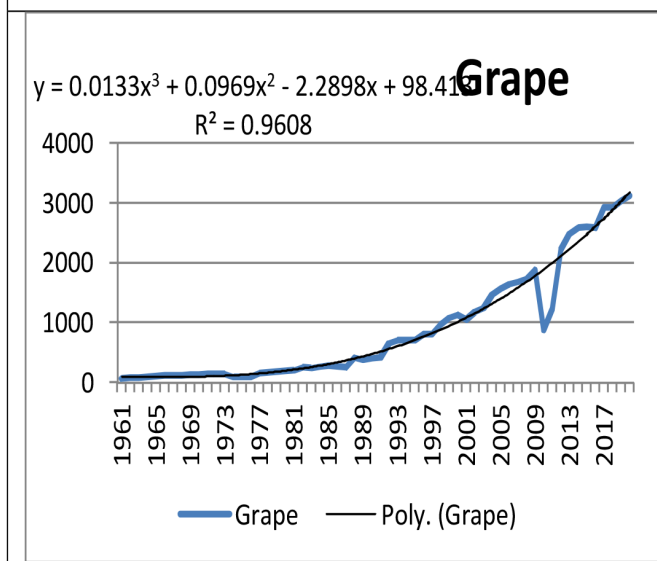
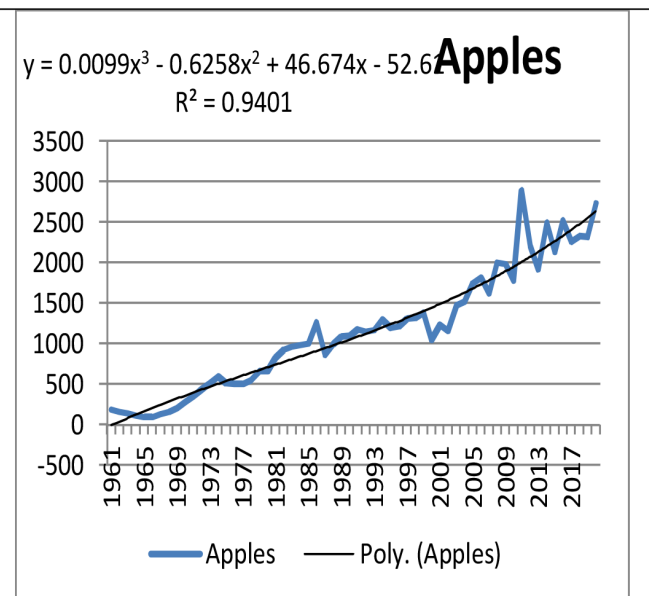
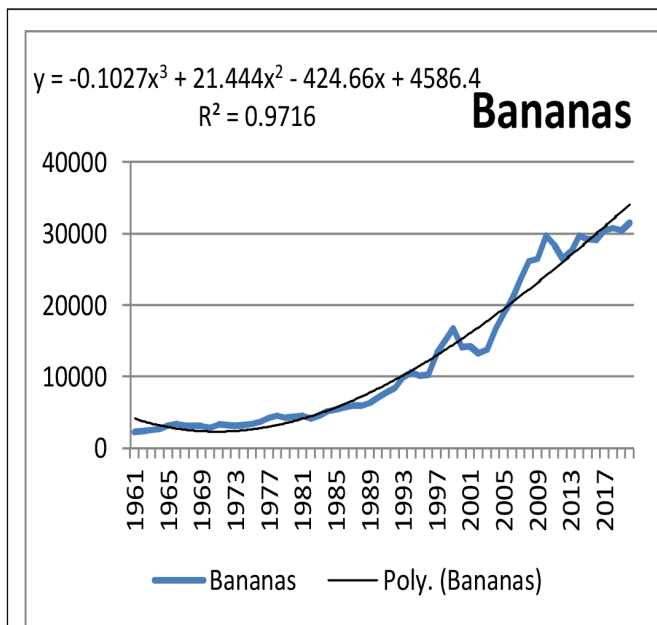
Table 3: Trends in production of major fruits of India

FRUIT	Equation	Model Summary					Parameter Estimates			
crop		R Square	F	df1	df2	Sig.	Constant	b1	b2	b3
Banana	Cubic	.972	637.801	3	56	.000	4586.414	-424.657	21.444	-.103
Apple	Cubic	.940	293.075	3	56	.000	-52.610	46.674	-.626	.010
Grapes	Cubic	.961	458.070	3	56	.000	98.413	-2.290	.097	.013
Mango	Cubic	.962	473.738	3	56	.000	5458.234	367.885	-16.469	.263
Pineapple	Cubic	.966	536.498	3	56	.000	119.288	22.425	-.031	.002

unlikely due to random chance. **Parameter Estimates, Constant (b0):** The intercept of the equation, representing the predicted value of the dependent variable when all predictors are zero. **b1, b2, b3:** Coefficients for the linear, quadratic, and cubic terms, respectively. These

coefficients indicate the strength and direction of the relationship for each term.

In this table clearly visible that R Square value of Banana is 0.972 (97.2% of the variability in the data is explained by the model). F: 637.801, showing a strong model fit. Coefficients:



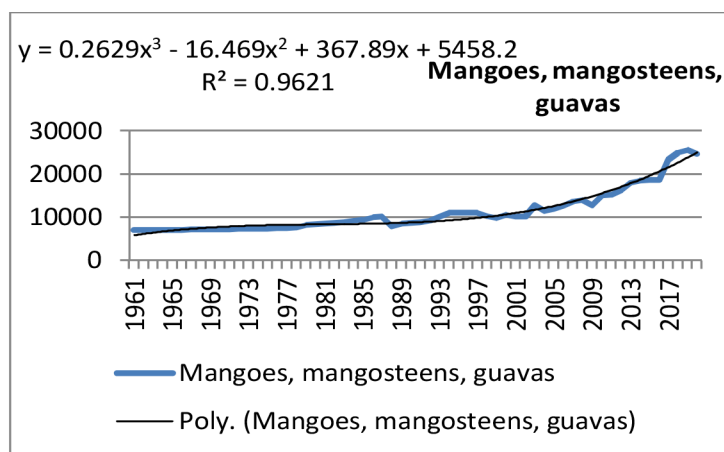


Figure 1: Observed and expected trends of major fruits of India

Constant: 4586.414 (baseline value). b_1 (linear): -424.657 (a negative relationship with the linear term). b_2 (quadratic): 21.444 (a positive quadratic effect). b_3 (cubic): -0.103 (a small negative cubic effect). **R Square** value of **Apple** is 0.940 (94% of the variability explained). **F** is 293.075, indicating a significant fit but weaker than Banana. **Coefficients** of banana are Constant: -52.610 (baseline **Pineapple** is negative). b_1 : 46.674 (positive linear effect). b_2 : -0.626 (small negative quadratic effect). b_3 : 0.010 (minimal positive cubic effect). **R Square** value of **Grapes** is 0.961 (96.1% of variability explained). **F** is 458.070. **Coefficients**: Constant: 98.413. b_1 : -2.290 (small negative linear effect). b_2 : 0.097 (positive quadratic effect). b_3 : 0.013 (minimal positive cubic effect). **R Square** value of **Mango** is 0.962 (96.2% variability explained). **F** is 473.738. **Coefficients**: Constant: 5458.234. b_1 : 367.885 (strong positive linear effect). b_2 : -16.469 (moderate negative quadratic effect). b_3 : 0.263 (positive cubic effect). **Similarly R Square** value 0.966 (96.6% variability explained). **F**: 536.498. **Coefficients**: Constant: 119.288. b_1 : 22.425 (positive linear effect). b_2 : -0.031 (small negative quadratic effect). b_3 : 0.002 (minimal positive cubic effect).

Instability analysis

Instability analysis of fruit crop production over two periods (Period-1 and Period-2) and for the whole dataset. The purpose is to evaluate the variability and consistency of crop production trends for each fruit across the specified periods. Different statistics of this table

shows **R²** Coefficient of determination, which measures the goodness-of-fit of the regression model. A value closer to **1.0** suggests a stronger and more stable model. **(CV)** Coefficient of Variation (%), a measure of relative variability in crop production. Higher values indicate greater instability. **(CVt)** Trend Coefficient of Variation (%), which reflects the variability of production over time due to trends. Higher values suggest more instability caused by trends.

Table 4: Instability analysis in production of major fruits of India

fruit crop production	statistics	period-1	period-2	whole
banana	R ²	0.962	0.953	0.986
	CV	31.368	40.359	80.450
	CVt	6.115	8.750	9.519
apple	R ²	0.948	0.843	0.940
	CV	64.867	30.900	62.484
	CVt	14.792	12.243	15.305
grape	R ²	0.927	0.908	0.961
	CV	54.901	52.192	101.526
	CVt	14.833	15.831	20.050
mango	R ²	0.809	0.966	0.962
	CV	12.029	35.037	43.498
	CVt	5.257	6.460	8.479
pineapple	R ²	0.929	0.867	0.966
	CV	44.576	24.388	53.676
	CVt	11.878	8.894	9.897

All fruits exhibit good model fit ($R^2 > 0.8$) across periods, with the whole dataset showing the best fit in most cases. Mango and Apple have slightly lower R^2 in specific periods, indicating less stable production trends. **Banana and Grapes** have high overall variability (CV), indicating

more unstable production. **Apple** shows high variability in Period-1 but stabilizes in Period-2. **Mango and Pineapple** have lower variability and better stability, especially in recent periods. Instability due to trends (CV_t) is generally lower than total variability (CV), but it increases over the whole period for most fruits, especially Grapes. This analysis helps in identifying crops that require more focused management to reduce production instability and optimize trends.

Modeling and forecasting

ARIMA (Auto-Regressive Integrated Moving Average) models used to analyze and forecast the production of major fruit crops in India. The metrics evaluate the performance and accuracy

of each ARIMA model for Banana, Apple, Grape, Mango, and Pineapple.

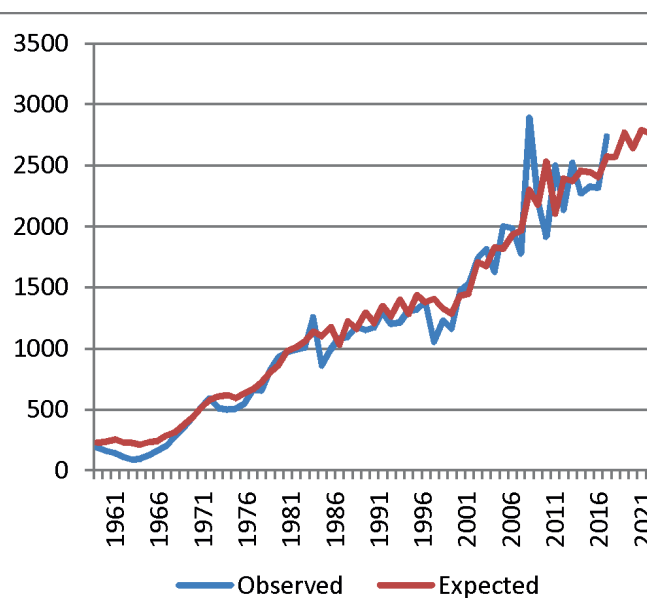
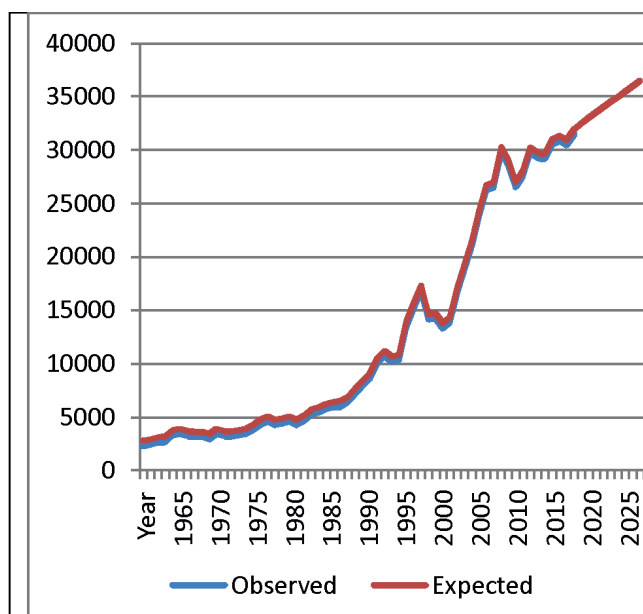
Pineapple shows **Best Performing Model** Strongest overall with low RMSE, MAPE, and MaxAE, indicating reliable forecasting. **Mango** Also performs well with the lowest MAPE but higher RMSE suggests more variability in certain cases. **Banana is High Variability Crops** Despite excellent fit, high RMSE and MaxAE suggest significant production variability. and **Grape** is Strong fit but highest MAPE indicates challenges in accurate forecasting. **Extreme Outliers** are shown in **Apple** with High MaxAPE reflects extreme variability in some instances. Models with higher complexity (e.g. ARIMA (1,1,5)) generally fit better but may still face issues with variability.

Table 5: ARIMA model of production major fruit crop in India

	PRI	R-square	RMSE	MAPE	MAE	MaxAPE	MaxAE	Normalized BIC
banana	ARIMA(0,1,0)	.990	1119.173	8.191	792.864	26.630	3168.412	14.110
apple	ARIMA(1,1,4)	.942	188.103	19.833	122.346	155.469	926.921	10.889
grape	ARIMA(1,1,,5)	.963	190.136	21.359	93.983	122.098	1075.319	10.979
mango	ARIMA(1,1,,5)	.969	905.411	4.966	558.803	29.231	3104.367	14.101
pineapple	ARIMA(1,1,,5)	.975	85.399	7.023	54.835	32.276	282.488	9.378

Conclusion of this table is ARIMA analysis identifies **Pineapple** as the most stable and accurate crop for production forecasting, while

Banana and Grape show higher variability, requiring further refinement in modeling or additional data preprocessing.



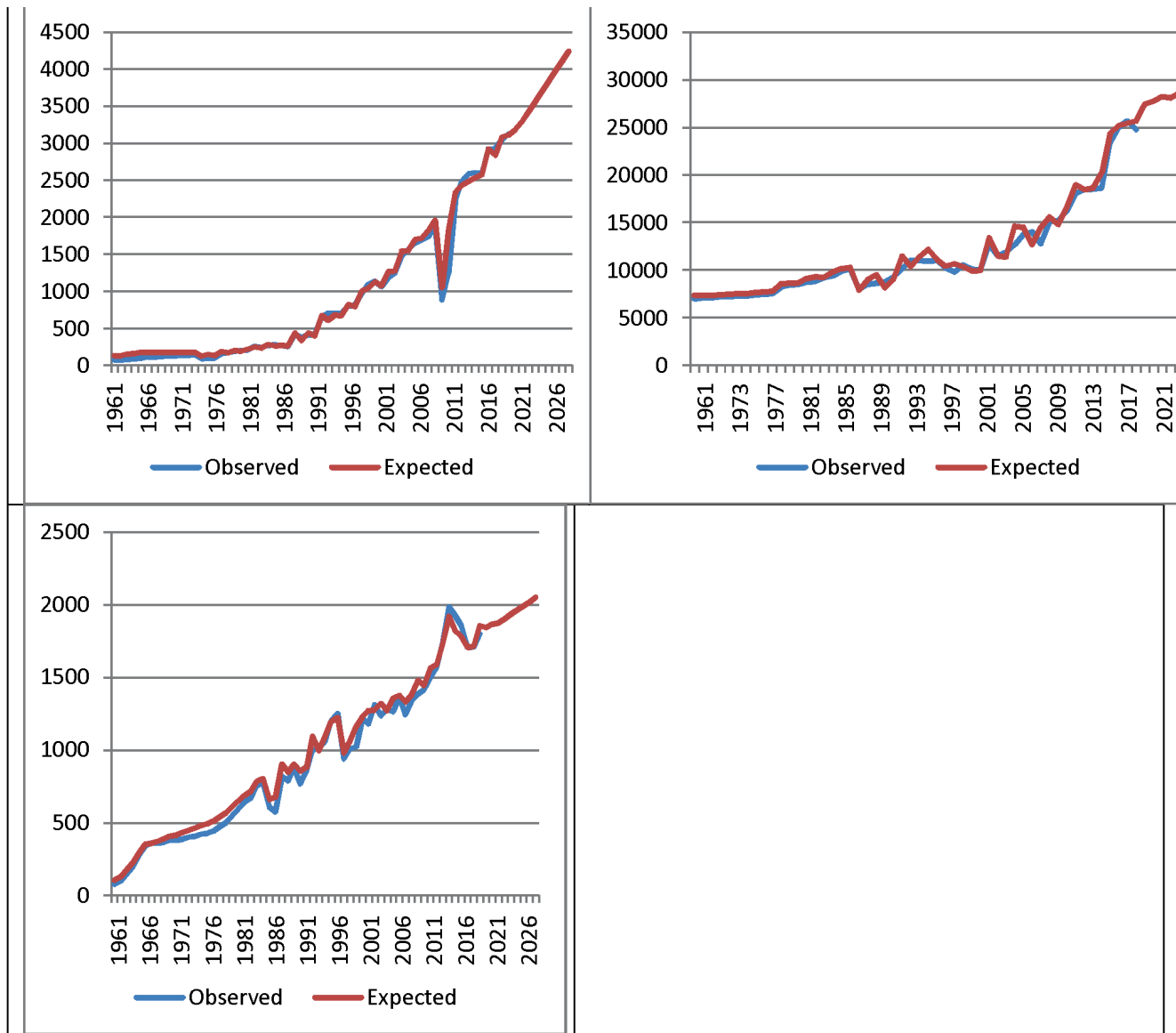
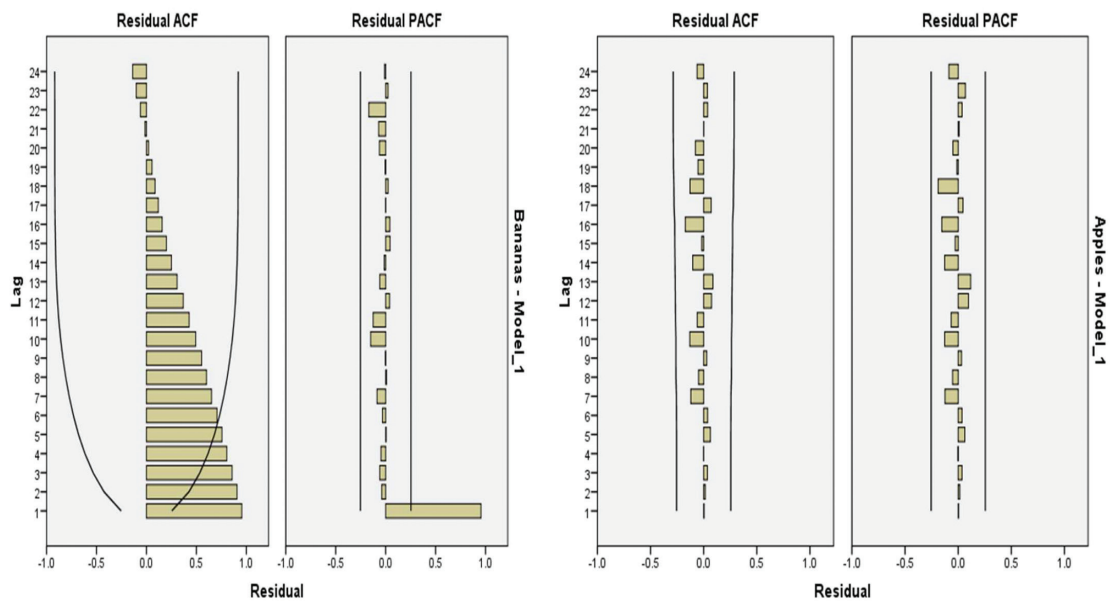


Figure 2: Observed and forecasted production under major fruit crop of India



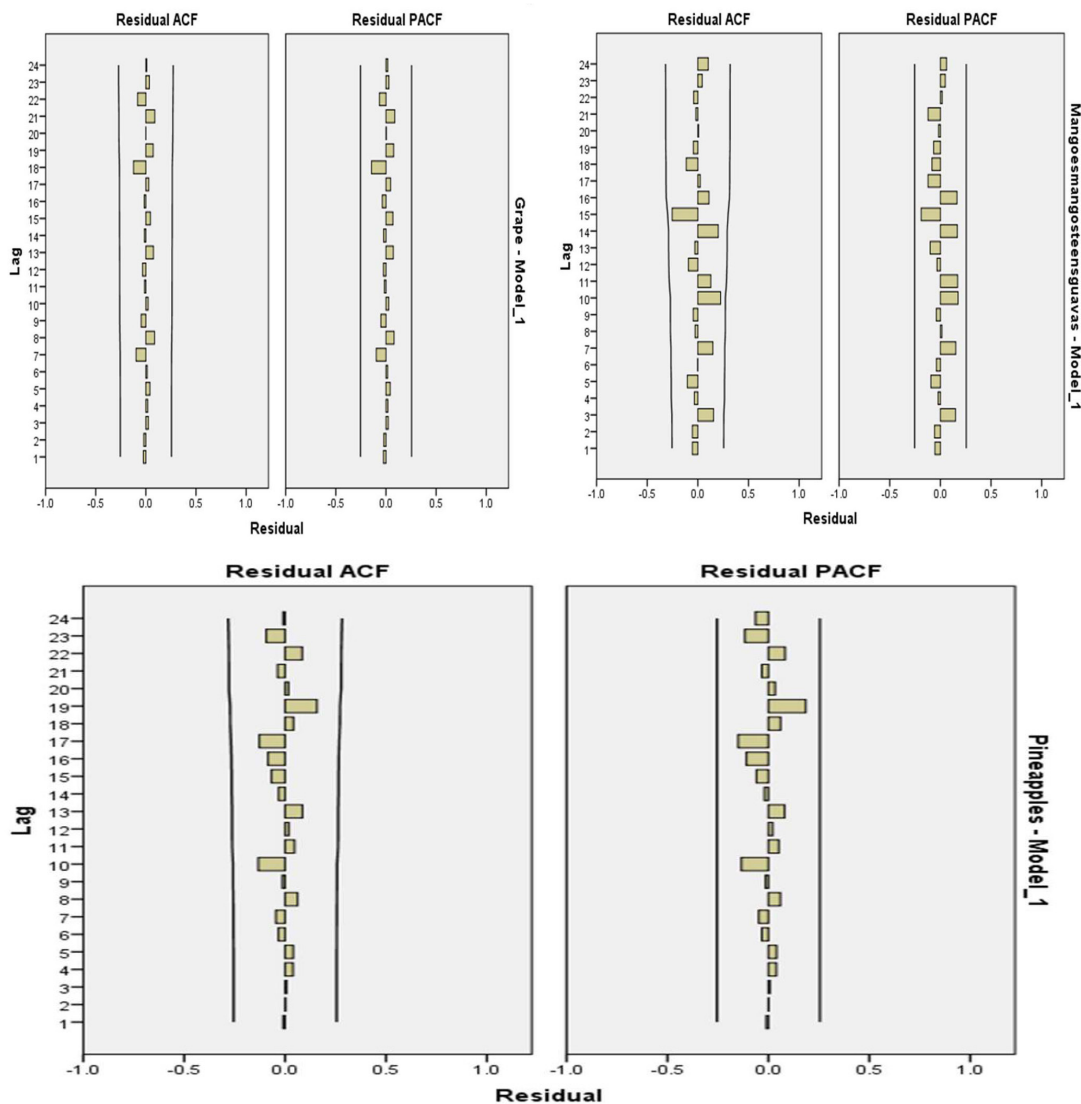


Figure 3: ACF and PACF graph

This image presents two diagnostic plots for residuals from a statistical model:

Residual ACF (Autocorrelation Function)
 The left plot shows the autocorrelations of residuals at different lags. The bars represent the strength of the autocorrelation at each lag, and the curved lines indicate the confidence bounds (typically 95% confidence intervals). If most bars fall within the confidence bounds, it suggests that the residuals are uncorrelated, this is ideal for a good model. If bars fall outside the bounds, there may be autocorrelation (dependence) in the residuals, which can indicate a model issue.

Residual PACF (Partial Autocorrelation Function)
 The right plot shows the partial autocorrelations of residuals at different lags. Partial autocorrelation removes the influence of intermediate lags, isolating the direct relationship

between the residuals at specific lags. If the bars fall within the confidence bounds, it suggests no direct correlation at specific lags. Large bars outside the bounds may indicate problems like model under fitting or a need for additional predictors.

CONCLUSION AND RECOMMENDATIONS

India’s total fruit production was estimated at approximately 107 million metric tons in the fiscal year 2023. The country remains a leading producer of various fruits, including bananas, mangoes, papayas, and apples, with significant contributions to both domestic consumption and the export market. To forecast future agricultural production, including fruit yields, the ARIMA (Auto-Regressive Integrated Moving Average) model is commonly employed. This statistical

approach analyzes time-series data to predict future trends by accounting for past values and the relationships between them. In the context of Indian agriculture, ARIMA models can be utilized to anticipate future fruit production levels by analyzing historical data, identifying patterns, and accounting for seasonal variations. This method aids policymakers and farmers in making informed decisions regarding crop planning, resource allocation, and market strategies. In summary, while India's fruit production growth has slowed in recent years due to various challenges, the application of predictive models like ARIMA offers valuable insights for future planning and development in the agricultural sector.

REFERENCE

- Azhar, B.A., Chaudhary, G. M. and Shafique, M. (1972). "A Model for Forecasting Wheat Production in the Punjab". *Pakistan Development Review*, **11**(2): 407-415.
- Borthakur, S. and Bhattacharya, B.K. (1998). "Trend analysis of area, production and productivity of potato in Assam". *Economic Affairs*, **43**(4): 221-226.
- Box, G.E.P. and Jenkins, G.M. (1976). *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco.
- Brockwell, P. J. and Davis, R. A. (2002). *Introduction to time series and forecasting*. Springer. New York.
- Chand, R. and Raju, S.S. (2009). "Instability in Indian Agriculture during Different Phases of Technology and Policy". *Indian Journal of Agricultural Economics*, **21**(2): 283-288.
- Chand, R. and Tewari, S.C. (1991). "Growth and instability of Indian exports and imports of agricultural commodities". *Indian Journal of Agricultural Economics*, **46**(2): 159-165.
- Cuddy, J.D.A. and Della, V.P. A. (1978). "Measuring the Instability of Time Series Data". *Oxford Bulletin of Economics and Statistics*, **40**(1):79-85.
- Gupta, G.S. (1993). "ARIMA model for forecasts on tea production in India". *Indian Economic Journal*, **41**(2): 88-110.
- Hamjah, M. A. (2014). Forecasting major fruit crops productions in Bangladesh using Box-Jenkins ARIMA model. *J. Econ. Sustain. Dev*, 5(7).
- Iqbal, N. B. K., Asif, M. and Ahmad, S. A. (2005). "Use of the ARIMA Model for Forecasting Wheat Area and Production in Pakistan" *Journal of Agriculture and Social Sciences*, **1**(2): 120-122.
- Mahendradev, S. (1987). "Growth and Instability in Food grains Production: An Interstate Analysis". *Economic and Political Weekly*, **22** (39): A82-A92.
- Mehra, S. (1981). "Instability in Indian agriculture in the context of new technology". International Food Policy Research Institute, Research Paper, 25 July, Washington, DC/
- Pankratz, A. (1983). "Forecasting with univariate Box-Jenkins models: concepts and cases". *John wiley, New York*.