Analysing and Forecasting Cashew Production in Kerala using Time Series and Regression Models

Divya P R[1], Wasma Nawaal N[2]

[1]Assistant Professor, Department of Statistics, Vimala College (Autonomous), Thrissur
[2]Student, Department of Statistics, Vimala College (Autonomous), Thrissur, University of Calicut, Kerala, India

Abstract
This paper deals with time series modelling approach (ARIMA Model) to forecast Cashew nut production in Kerala, India. The study assesses models related to area under cashew cultivation, production levels and productivity within Kerala. ARIMA (2, 0, 1) model was identified as the most suitable time series model for modelling the area of cashew cultivation in the state. Further, efforts were made to forecast, as accurate as possible, the Cashew nut production for a period up to ten years by fitting ARIMA(2,0,1) model to our time series data. The study was also explained different regression model techniques such as linear and multiple regression models, have been used to forecast cashew nut production. This paper also analyzes the combinations of linear and non linear regression models. For predicting cashew production with independent variable area, cubic model found to be best with highest R^2 (0.96). For predicting cashew production with independent variable productivity, cubic model found to be best with highest R^2 (0.59). Therefore Cubic model is the most appropriate model for predicting cashew production of the present study.

Keywords: ARIMA model, Linear and Multiple regression analysis, Cubic Model,Trend analysis

1. Introduction

Cashew (Anacardium occidentale L), popularly known as the "miracle nut," is one of the most valuable processed nuts sold on global commodities markets, as well as a significant cash crop. It has the potential to be a source of income for cashew producers, to empower rural women in the processing industry, to create jobs, and to earn foreign revenue through exports. Cashew is a very lucrative and nutrient-dense crop. The cashew tree is said to have originated in Brazil and has since spread around the world, mostly for soil conservation, afforestation, and wasteland development. Cashews are mostly produced in India's coastal states. It is grown on the west coast in Kerala, Karnataka, Goa, and Maharashtra, and on the east coast in Tamil Nadu, Andhra Pradesh, Orissa, and West Bengal. Kannur, Kasaragod, Malappuram, Palakkad, and Kollam are the major districts in Kerala where cashews are grown. In 2018, Kerala's total cashew area and production were 82,000 hectares and 88,000 metric tonnes, respectively, with a productivity of 1.06 metric tonne per Hectare. Kannur district ranked highest in cashew output, accounting for 67.6percent of total production, followed by Kasaragod and Kollam. Despite the fact that Kerala is fifth in cashew production in India, it is first in processing and exporting, followed by Tamilnadu, Karnataka, and Andhra Pradesh. However, from the year 2013, both the area and production of cashews in Kerala have been steadily declining. In this case, the study's goal is to analyse the trend in cashew area, production, and productivity in Kerala in order to determine the cashew's growth rate and, if necessary, to make recommendations for improvement.

Time series forecasting is the process of analysing time series data using statistics and modelling to make predictions and inform strategic decision-making. This study focuses on modelling and trend analysis of cashew nut production. The aim of this project is to develop a pattern of cashew nut production generated using time series analysis. The purpose of this project is to build on ARIMA model that fits in an appropriate way the evolution of the cashew nut production's time series.

The main objectives of this project are:

- To show the pattern of Production of cashew nut over the period (2000-2022)
- To forecast trends in area, production and productivity of cashew nut in Kerala using ARIMA model.
- To analyze the liner trend between production of cashew nut (Dependent variable) and independent variables such as area & productivity.
- To analyze the combinations of linear and non-linear models which are one of the most popular and widely used hybrid models for improving the forecasting accuracy.
- To study the influence of Area & Productivity on Production of cashew nut, multiple regression analysis is used.

2. Review of Literature

Kumar Manoj (2024) suggested a Time Series Model Approach (Box –Jenkins ARIMA Model) has been used in this study to forecast Sugarcane Production in India. The order of the best ARIMA model was found to be (2,1,0) in **"**An Application of Time Series ARIMA Forecasting Model for Predicting Sugarcane Production in India." Okeke Daniel Chukwujioke and Akarue Blessing Okiemute (2018) in "Forecasting of Cashew Area Harvested, Yield and Production using Trend Analysis." In this the study was conducted to examine the trend Analysis of Area, Yield and Production for cashew in Nigeria. The findings of the study are based on data from the years (1961 to 2016). Three models of trend analysis were applied. The models were linear trend model, Quadratic trend model and Cubic trend model. The most appropriate model for trend analysis of the present study was Cubic trend model with highest R^2. Ali J.Ramandhan, Tufleuddin Biswas, Soumik Ray, S.R Anjanawe, Deepa Rawat, Binita Kumari, et.al (2024) suggest "Modelling And Forecasting Of Coconut Area, Production And Productivity Using A Time Series Model." Dr.Richard Paul & Lakshmi (2015) suggest "Emerging Trend Analysis Of Coconut Production In India." N Karunakaran, (2013) suggest "Trends and Overall Growth Analysis of Cashew nut Cultivation in Kerala." Sandip Shil, Sankar Ch. Paul, G.C. Acharya, Soumen Pal, (2013) "Trend Analysis and Forecasting Coconut Production in Assam." S.U. Aparana & M.Asokan, (2020), "Impact Analysis of Coconut Producer Societies in Kerala State." Vaisakh Venu, Vipin P.R and Prajitha N.K (November 2023) "A Comparative Analytical Study of Many Regression Model Approaches, ARIMA Model and a Hybrid Model for Forecasting Area, Production and Productivity of Coconut in Kerala, India. This study is intended to provide reliable and context- specific forecasting methodologies to support sustainable agriculture planning, resource allocation and policy formulation for the coconut industry in Kerala.

3. Methodology

This study based on secondary data. This study used yearly data of cashew nut production during the period 2000 to 2022. The Data are collected from Kerala state Economics and Statistics Department. The Data analysis undertaken involves descriptive analysis, stationarity tests, model fitting and forecasting. This study mainly employs a time series approach using ARIMA model and forecast total production of cashew nut. In this project work different regression model techniques such as linear & multiple regression models have been used to forecast cashew nut production. This study also analyzes the combinations of linear & non- linear regression models.

4.1. Time Series Model

Time series model is a dynamic research area which has attracted attention of researchers' community over last few decades. The main aim of series modelling is to carefully collect and rigorously study the past observations of a time series to develop an appropriate model which describes the inherent structure of the series. This model is then used to generate future values for the series i.e., to make forecasts. In this project work ARIMA model is used to fit the model.

4.2. Regression Analysis

Regression analysis is a set of statistical methods used for the estimation of relationships between a dependent variable and one or more independent variables.

5. Forecasting Of Production of Cashew, Area Harvested and Productivity Using Time Series and Regression Analysis

This section includes a comprehensive time series analysis of the production of cashew in Kerala, aiming to uncover underlying patterns, trends over a period of 2000-2022.Table 1 represents the Cashew production of Kerla during the period 2000 to 2022.

Table 1: Cultivation of Cashew in Kerala during 2000 to 2022

Year	Area	Production	Productivity
2000-01	92122	66178	718
2001-02	89718	65867	734
2002-03	88548	66087	746
2003-04	86376	65655	760
2004-05	81547	60584	743
2005-06	78285	68262	872
2006-07	70463	61680	875
2007-08	58381	52402	898
2008-09	53007	42334	799
2009-10	48972	35818	731
2010-11	43848	34752	793
2011-12	54052	36743	680

2012-13	52086	37919	728
2013-14	49105	33375	680
2014-15	45436	29715	654
2015-16	43090	24733	574
2016-17	41661	27944	671
2017-18	39720	25629	645
2018-19	38781	15635	403
2019-20	39898	19444	487
2020-21	37923	20909	551
2021-22	32369	15861	490

The descriptive statistics is presented in Ttable 2 which delineated the properties of all the variables under study.

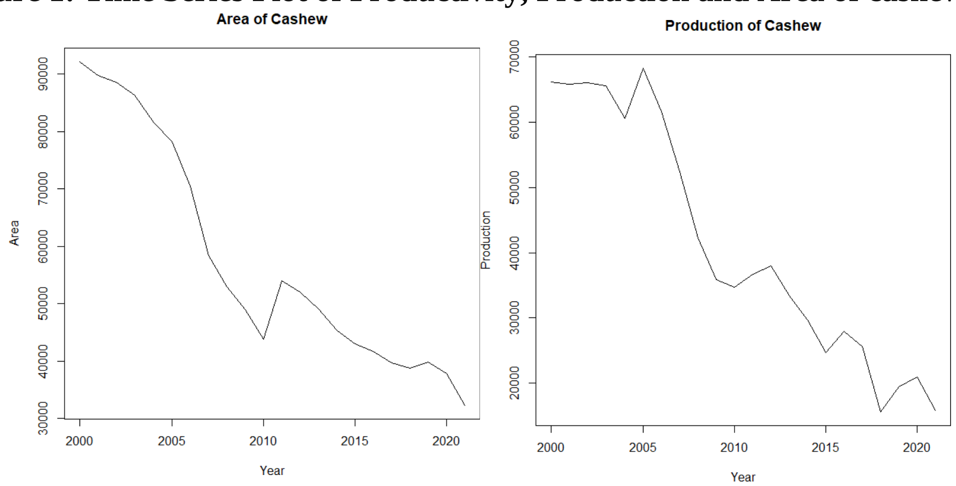
Table 2: Descriptive Statistics of Total Production of Cashew in Kerala

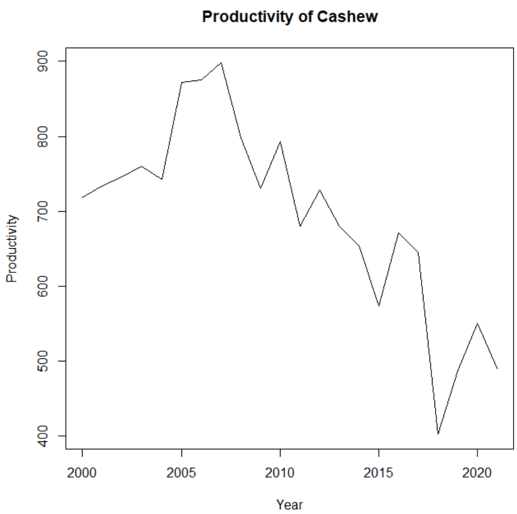
Variables	**Std. Error**	**Mean**	**Median**	**Skewness**	**Kurtosis**
Productivity	129.0823	692.4	723	-0.5038	2.7468
Production	18604.07	41251	36281	0.2412	1.5756
Area	19760.21	57518	50596	0.6304	1.8713

Summary statistics of cashew Production are presented in Table 2. Results indicate that Production and Area are positively skewed. Productivity is negatively skewed, which means that it has more values on higher side.

From the above table it is very clear that, kurtosis coefficients which are less than 3 (Productivity, Production and Area) indicate that the distribution of cashew production are platykurtic.

Figure 1: Time Series Plot of Productivity, Production and Area of cashew





From the Figure 1, it is very clear that the time series plots of Productivity, Production and Area of Cashew Nut shows a decreasing trend.

5.1. Identification Of ARIMA (Auto Regressive Integrated Moving Average) Model

A statistical model that forecasts future values in a time series based on:

- AR (Auto regression): Past values
- I (Integration): Differences between past values
- MA (Moving Average): Past errors

It helps capture patterns, trends and seasonality in time series data.

$Yt=\phi1Yt-1+\phi2Yt-2+\theta1\epsilon t-1+\epsilon t$

This study based on time series analysis and a mathematical model has been established, to predict it with a reasonable prediction method.

Before the model is estimated, stationarity tests are performed using Augmented Dicky-Fuller (ADF) test. The Augmented Dickey-Fuller Test (ADF) is a statistical test used to determine if a unit root is present in a given time series sample. The null hypothesis of the test is that a unit root is present, whereas the alternative hypothesis can be either stationary or trend stationary, depending on the version of the test used. The ADF test is an augmented version of Dickey Fuller test and is designed to handle a wider range of time series models. The ADF statistic, which is negative, is used in the test. The more negative the statistic, the stronger the evidence against the null hypothesis. Table 3 represents ADF Test result. The results of the ADF (Augmented Dickey- Fuller) test for the presented time series (Production) are as follows.

H0: The time series data is not stationary

H1: The time series data is not stationary

Table 1: ADF Test Result

Variables	**Statistic**	**P value**
Productivity	-2.2286	0.4853
Production	-2.1691	0.508
Area	-1.9717	0.5832

Here the p values of variables are greater than 0.05, which means it is in non-stationary. Hence by differerencing these variables, we can make the non-stationary to stationary. ADF Test output is presented in Table 4.

Table 2: ADF Test output

Variables	**Statistic**	**P value**
Productivity	-4.4411	0.01
Production	-4.0667	0.02
Area	-4.6686	0.01

Here, the p value of Productivity, Production and Area is less than 0.05. Reject null hypothesis. Now the time series data is stationary. Correlogram errors are used to match the ARIMA model specification.Figure2,3 and 4 shows ACF and PACF plots of productivity and production.

Figure 2: ACF and PACF Plots of Productivity

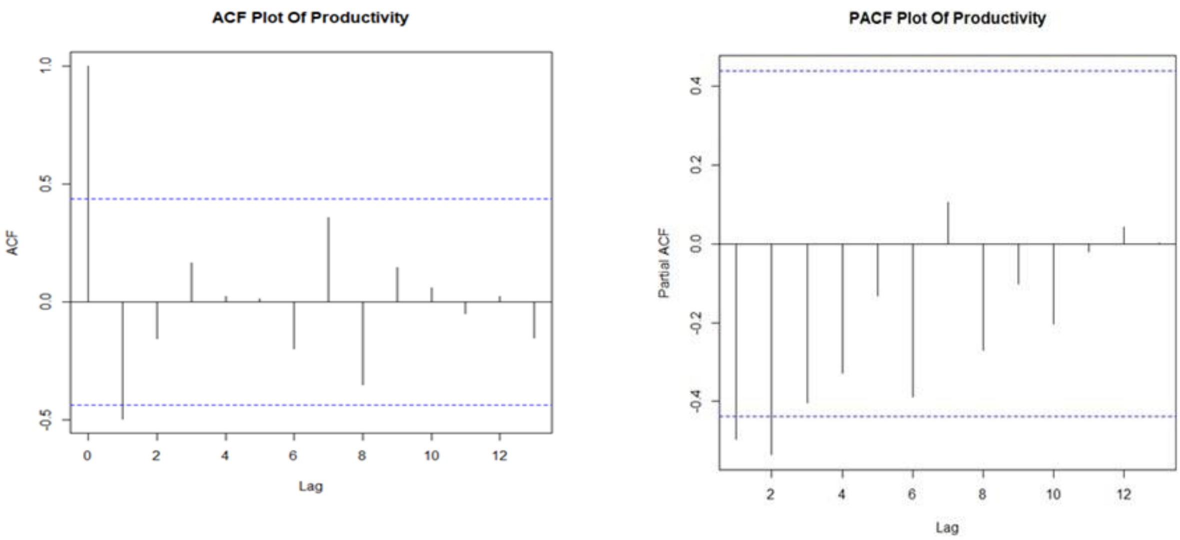


Figure 3: ACF and PACF plots of Production

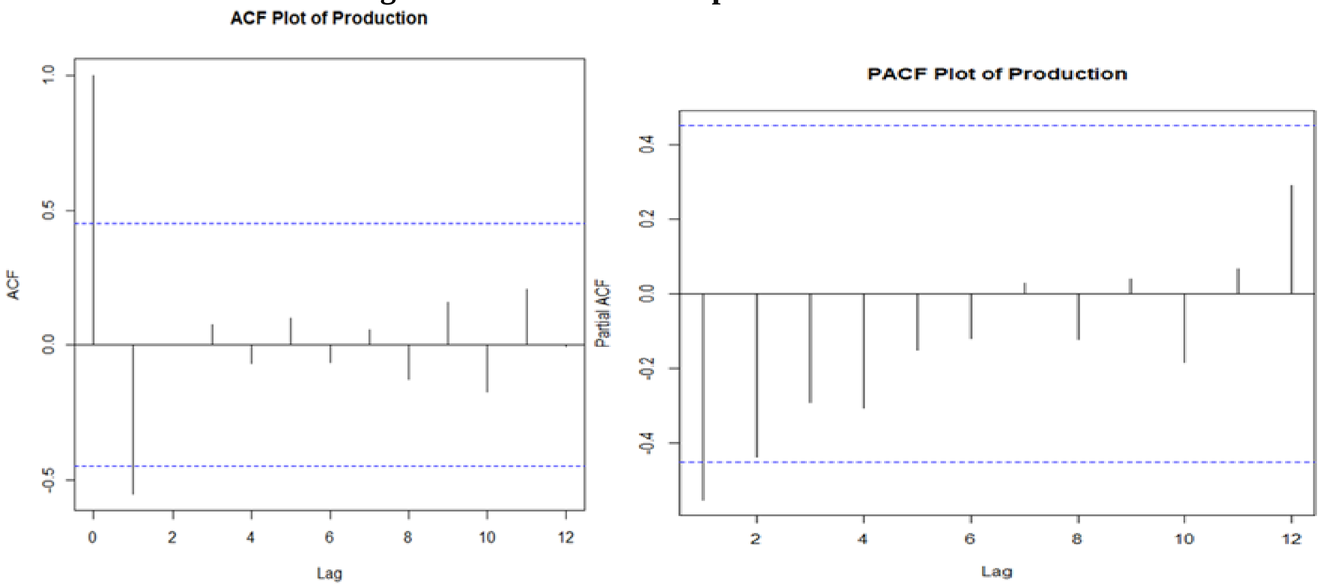
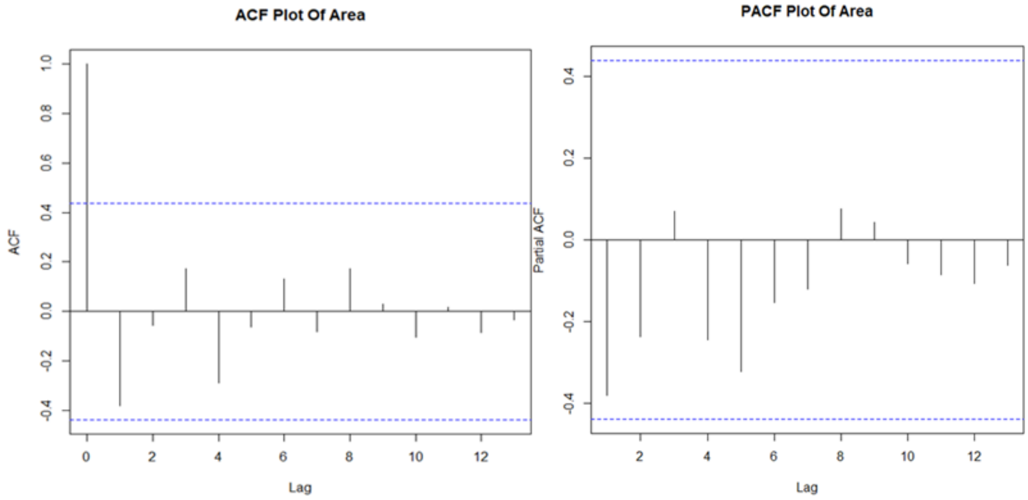


Figure 4: ACF and PACF plots of Area



5.2: Establishing Time Series Model of Total Production of Cashew

The recognition and order of the ARIMA (p, d, q) model can be obtained through the observation of the autocorrelation and partial autocorrelation function of the sample. Using R software, the best model for total production of cashew nut is ARIMA (2,0,1). Table 5 shows predicted values of Cashew production of Kerala from 2023 to 2032. The results of the model estimate of Productivity, Production and Area are as follows.

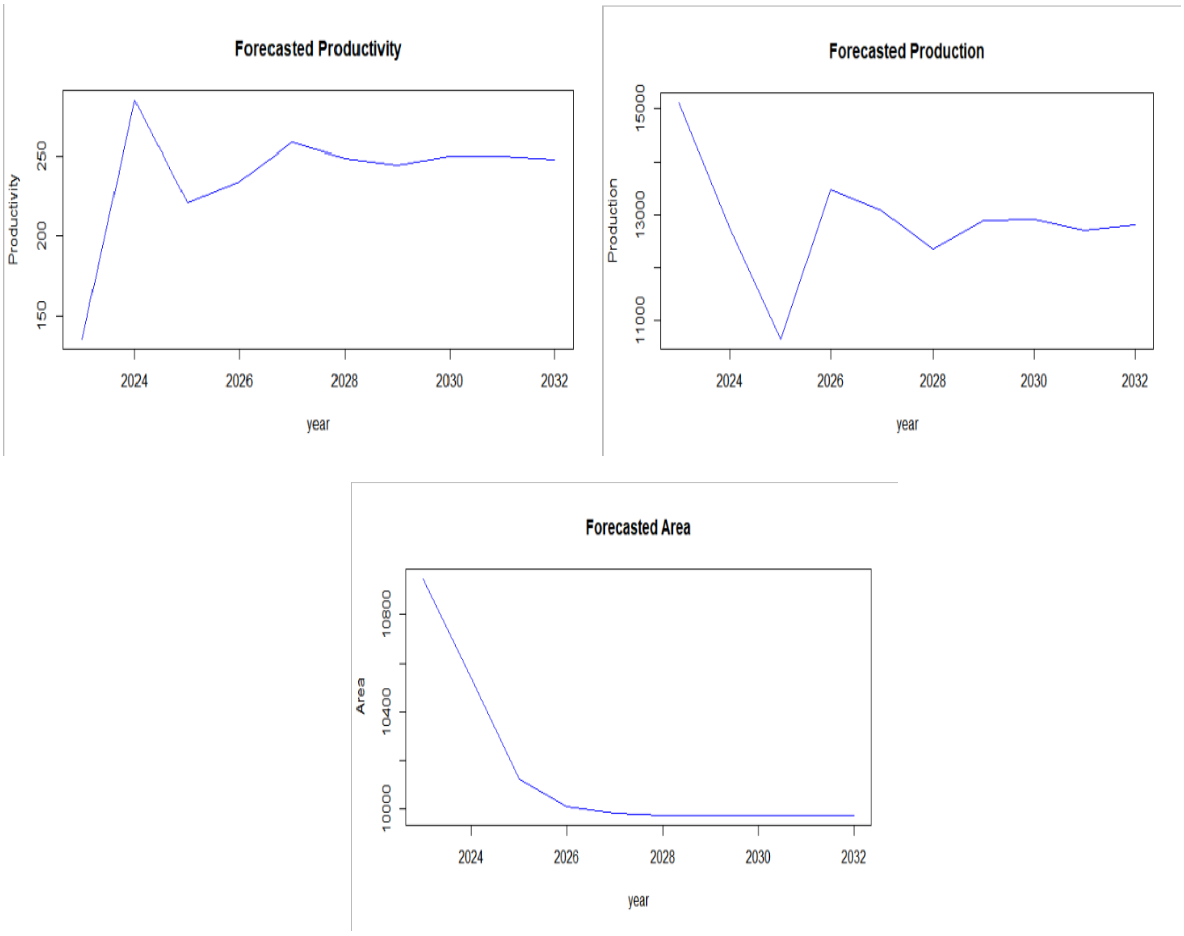
Yt (Productivity)= -0.3791Yt-1 - 0.3763Yt-2 - 0.9958ϵt-1 + ϵt
Yt (Production)= -0.4996Yt-1+ (-0.3815) Yt-2+ (-71.11533) ϵt-1 + ϵt
Yt (Area) = 0.2265Yt-1 +(0.0023) Yt-2 + 111.5163ϵt-1 + ϵt

Table 3: Predicted Values of Cashew production of Kerala from 2023 to 2032

Year	Productivity	Production	Area
2023	135.2140	15114.39	10947.206
2024	284.4952	12753.86	10541.342
2025	220.8593	10656.79	10124.668
2026	233.5834	13467.22	10010.745
2027	258.2326	13085.12	9982.999
2028	248.4159	12357.23	9976.399
2029	244.2770	12902.79	9974.838
2030	249.8701	12922.89	9974.470
2031	249.5790	12712.97	9974.382
2032	247.6254	12811.84	9974.362

Figure 5 represents the forecasted Area, Production and productivity during the period 2023 to 2032.

Figure 5: Predicted value of Productivity, Production and Area



6. Regression Analysis of Total Production of Cashew with Independent Variables Area and Productivity

This section examines the trend analysis of area, production and productivity for Cashew in Kerala. The findings of the study are based on data from the years (2000 to 2022) and were taken from the database of Kerala State Economics and Statistics Department. Three Models of trend analysis were applied. The models were Linear Trend Model, Quadratic Trend Model, and cubic trend Model. The most appropriate Model for trend analysis of the present study was Cubic Trend Model based on the highest R^2. Table 6 represents different Linear and non Linear models.

Table 4: Linear and Non-Linear Models

SI.no.	Model	Equation	Description
1.	**Linear**	$Yt= a + bt$	Y and t are Production of Cashew nut and the independent variables Area and Productivity. a and b are constants to be estimated
2.	**Quadratic**	$Yt = b0+ b1ti + b2 t i^2$	Y and ti 's are Production of Cashew nut and the independent variables Area and Productivity. b0 and bi 's are constants to be estimated.
3.	**Cubic**	$Yt=b0+b1ti+b^2 ti^2+b3ti^3$	Y and ti's Production of Cashew nut and the independent variables Area and Productivity. b0 and bi's are constants to be estimated.

6.1 Prediction Model of Production of Cashew in Kerala

6.1.1: Linear Model

$Yt = b0 + b1 X + ex$

Yt = Area harvested, Yield and production in year t

b0 = Constant

b1 = Regression coefficient

X = Time period (years)

ex = Error

6.1.2: Quadratic Model

The quadratic trend model is known to account for simple curvature in the data and it is presented as:

$Yt = b0 + b1 *X + b2 X^2 + ex$

6.1.3: Cubic Model

The cubic trend model is presented as follows:

$Yt = b0 + b1 *X + b2 X^2 + b3X^3 + ex$

The linear and various nonlinear models were used to predict production with area and productivity are given in Table 7. The best model was selected based on the highest R^2 value. Cubic model was found to be best with highest R^2 values (0.96).

Table 5: Comparison of models for predicting Cashew Production with independent variables Area and Productivity

Model	Area		Productivity	
	Coefficient	R^2	Coefficient	R^2
Linear	0.905	**0.92**	111.78	**0.5816**
Quadratic	82031.3	**0.959**	66123.6	**0.56**
	-16540.3		-727.8	
Cubic	82031.3	**0.96**	66123.6	**0.59**
	-16540.3		-727.8	
	-6389.7		-18463	

6.2: Multiple Linear Regression Analysis

The multiple linear regression equation is, Y = α + $\beta_1 X_1$+ $\beta_2 X_2$ +...... + $\beta_p X_p$+ ε, where, Y is the dependent variable and X_i's are the independent variables with β as the partial regression coefficients of Y on X_i's where i=1, 2,…, p.

In present Y is taken as production of cashew and X_i's are productivity and area. Multiple linear regression fitted by using cashew production as dependent variable and area and productivity as independent variables and the results are given in Table 8. The fitted model was found to be significant (5%) with R^2 value of (0.9939) and out of two independent variables selected for the study.

Table 6: Regression Output

Variables	**B Coefficients**	**Standard Error**	**t**	**P Value**
Constant	-32860	1736	-18.93	$8.64e^{-14}$
Area	0.7278	0.01984	36.68	$2e^{-16}$
Productivity	46.57	3.038	15.33	$3.75e^{-12}$

(Significant at 5%)

Model Summary[b]

Model	**R square**	**Standard Error**
1	0.9939	1457

a. Predictors: (Constant), Area, Productivity

b. Dependent Variable: Production

Here the fitted model is **Y=-32860+0.7278X1+46.57X2**. The Table 9 provides data that relate to the goodness of fit of the regression equation. The R^2 of 0.9939 indicates that 99% of the variation in production of cashew is explained by the regression variables X1 and X2.

Table 7: Regression Model Values and Corresponding Residual values

Year	**Regression model values**	**Residual(Error)**
2000	67623.7	1445.65
2001	66619.1	752.14
2002	66326.5	239.454
2003	65397.7	-257.347
2004	61091.4	507.417
2005	64724.9	-3537.14
2006	59171.7	-2508.28
2007	51449.6	-952.448
2008	42927.9	593.925
2009	36824.5	1006.49

2010	35982.6	1230.58
2011	38146.6	1403.65
2012	38951.2	1032.15
2013	34546.2	1171.22
2014	30665.1	950.101
2015	25232.1	499.082
2016	28709.3	765.346
2017	26085.9	456.866
2018	14132.5	-502.48
2019	18857.4	-586.646
2020	20400.4	-508.571
2021	13517.5	-2343.54

7. Conclusion

This study mainly employs a time series approach using ARIMA model and forecast total production of cashew nut. In this project work different regression model techniques such as linear and multiple regression models have been used to forecast cashew nut production. This study evaluates the models for area of cashew nut cultivation, cashew nut production, and cashew nut productivity in Kerala. There is a decrease in trend of area, production and productivity over time. All variables in the data (Production, productivity, area) are stationary state. In forecasted values of area, there exist a decreasing trend. In the case of production and productivity, both are fluctuated over time. The ARIMA (2, 0, 1) model is preferred for modelling area of cashew nut production in Kerala. Linear regression analysis was conducted, focusing on area and productivity as the independent variables and production as the response variable. In multiple linear regression analysis Y is taken as production of cashew and Xi's are productivity and area. Multiple linear regression fitted by using cashew production as dependent variable and area and productivity as independent variables. The fitted model was found to be significant (5%) with R^2 value of (0.9939). For predicting cashew production with independent variable area, cubic model found to be best with highest R^2 (0.96). For predicting cashew production with independent variable productivity, cubic model found to be best with highest R^2 (0.59).

References

1. Ali J. Ramandhan ,Tufleuddin Biswas, Soumik Ray, S.R. Anjanawe, Deepa Rawat, Binita Kumari, et.al (2024), **"Modeling And Forecasting Of Coconut Area, Production And Productivity Using A Time Series Model."** BIO Web of Conferences. (Vol 97. P. 00113).
2. Dr.Richard Paul & Lakshmi, (2015), **"Emerging Trend Analysis of Coconut Production in India."**
3. Kumar Manoj, Anand Madhu (April 2024): **"An Application of Time Series ARIMA Forecasting Model for Predicting Sugarcane Production in India."** Studies in Business and Economics, 9(1), 81-94.
4. N Karunakaran, (2013): **"Trends and Overall Growth Analysis of Cashew nut Cultivation in Kerala."** Artha Journal of Social Sciences, 12(4), 31-48.
5. Okeke Daniel Chukwujioke, Akarue Blessing Okiemute (July 2018): **"Forecasting of Cashew Area Harvested, Yield and Production Using Trend Analysis."** International Journal of Agricultural Economics, 3(4), 65.

6. Sandip Shil, Sankar Ch. Paul, G.C. Acharya, Soumen Pal, (2013): **"Trend Analysis and Forecasting Coconut Production in Assam."** Journal of Plantation Crops, 41(2), 238- 241.
7. S.U. Aparana & M.Asokan, (2020):**"Impact Analysis of Coconut Producer Societies In Kerala State."** Int. J. Curr. Microb. And Appl. Sci, 9(4), 2319-7706.
8. Vaisakh Venu , Vipin P.R and Prajitha N.K (November 2023):**"A Comparative Analytical Study of Many Regression Model Approaches, Arima Model and a Hybrid Model For Forecasting Area, Production and Productivity of Coconut in Kerala, India."** Current Journal of Applied Science and Technology, 42(40), 37-49.