ISSN: 0974-5823 Vol. 7 No. 5 May, 2022

International Journal of Mechanical Engineering

A PRODUCTION OF GROUNDNUT IN TAMILNADU USING ARIMA AND NEURAL NETWORK ANALYSIS

M. Saranyadevi[1, 2] and A. Kachi Mohideen[2]

[1]Assistant Professor (Guest), Department of Statistics, Government Arts College (Autonomous), kumbakonam, Affiliated to Bharathidasan University, Tiruchirappalli – 620024

[2]Assistant Professor, Department of Statistics, Thanthai Periyar Government Arts College (Autonomous), Affiliated to Bharathidasan University, Tiruchirappalli – 620024

ABSTRACT

Peanut is one of the important oilseeds in Tamil Nadu. Peanuts are produced in many countries. In India, peanuts play an important role in oilseed production and India is the second largest producer of peanuts. In 2020, India's agricultural sector will account for 15.4% of India's GDP and employ 41.5% of India's workforce. Since peanut prices are volatile and unpredictable, farmers need realistic price forecasts at harvest time to choose where to grow their peanut crops. This study used time series data from Tamil Nadu province from 2003 to 2018. The main goal of the study is to find the best ARIMA model and also to predict peanut production in Tamil Nadu. A neural network approach using ARIMA and R software is used to draw certain conclusions from the data obtained about the expected yield of peanut crops.

Key words: ARIMA, Neural Network Approach, Production of Groundnut

INTRODUCTION

Groundnut is the worldwide important crop of oilseeds. In the 16th century, the groundnut cropping has been introduced in the southwestern part of India. Groundnut is the world's 13th most essential food gather. It is the world's fourth largest manufacturer of edible oil and third largest manufacturer of vegetable protein. Universally, in Groundnut production India is the second largest country throughout in Asia. Among number of oilseeds, groundnut production is higher as 25% than others. Virtually, every part of Groundnut crop is important items which are sold one. The Groundnut is used by the people in numerous ways such as, cooking with the groundnut oil, groundnut cake, groundnut kernels, groundnut shell, and groundnut straw. Contempt the truth that the production of oilseeds has been increased pointedly subsequently from 1960s, the consumption of edible oils is increased in the human. India imports the groundnut to various countries, though due to a discrepancy between domestic availability and actual consumption of edible oils.

Attention has been given to the univariate time series Auto-Regressive Integrated Moving Average (ARIMA) Models, which is principally due to World of Box and Jenkins (1970). Among the stochastic ARIMA types are robust, effective, and famous as they can correctly describe the found facts and can make forecasts with minimum forecast error. These types of models are challenging to pick out and estimate. Muhammed et al (1992) empirical modeling and prediction of time series data of rice production in Pakistan. Ibrahim Usman et al., 2013 have done a gross margin and cost-benefit analysis to study the profitability of groundnut production in Tamilnadu. A comparative analysis of groundnut crops at the state level and national level, district level in India is discussed in Madhusudhana(2013). SitaRambabu et al.,(2013) analyzed trends and Compound growth rates for the area, production, and productivity of groundnut in Andhra Pradesh over a period of 1995-96 to 2010-2011. Rahul (2014) has done an "in agriculture and allied sciences, time series models are used for prediction milk production, the yield of a certain breed of cows, the yield of a crop, prices, production, and productivity". These models can play a significant role in stock market decision-making have been discussed by Qiu and Song (2016). Hemavathi et al.,(2018) ARIMA Model for Forecasting of productivity of Rice and Its Growth Status in Thanjavur District of Tamil Nadu, India, also use the ARIMA Model.

Different time series analyses are available to study the nature of time-series data and also for prediction. A study of the Autoregressive model of various types is used as a statistical tool for analyzing a variety of types of time series data. Yule and Walker proposed the Autoregressive Moving Average (ARMA) model, and Box and Jenkins proposed the method (ARIMA) model afterward. The main objective of the article is to analyze and also forecast the production of groundnuts. Auto Regression Integrated Moving Average models and Neural Network techniques are used to observe groundnut production in Tamil Nadu. This study elaborated that the forecast results are obtained, and a detailed clarification of model choice and forecasting accuracy is provided.

SOURCE OF DATA

Studies depend on secondary data sources. Data taken from India's MOSPI Statistical Yearbook. A book containing a time series of annual peanut production since 2003-2004 to 2017-2018 required for the study. The model was validated using 15 years of peanut production data. Data sets are obtained for statistical reporting purposes.

This periodical describes the sources of information taken for the study and the analytical methods and characteristics used in the study.

Autoregressive Integrated Moving Average Model (ARIMA) Model

Auto Regressive Integrated Moving Average is an assorted model which be uses the most common process which is called Box Jenkins process. As a result of fitting the ARIMA model, the basic goal is to identify the stochastic process that motivates the time series and exactly a variety of situations involving the construction of models for discrete time series and dynamic systems have also made use of these methodologies. As early as 1926, Yule introduced autoregressive (AR) models. Slutsky, which introduce Moving Average (MA) methods in 1937, added to these. As Wold (1938) established, ARMA processes can be used to represent all stationary time series as long as the required order of p and q for AR and MA components is maintained. This article presents some basic ideas of linear time series analysis, such as stationarity and seasonality, and a transitional focus on the most popular forms of time series forecasting process.

Stationarity and non-stationarity

Stationary time series data that reflects regular mean and variance, and an autocorrelation function (ACF) that remainder largely constant over time. The Dickey and Fuller test also occupies an important position in testing the stationary of the data.

$$\Delta Y_{t-1} = \gamma Y_{t-1} + \varepsilon_t \qquad \ldots (1)$$

Where γ = Φ-1, Then, null hypothesis of H_0: γ =0 against the alternative hypothesis H_1:γ<0. v If the series is stationary, the null hypothesis should be accepted. Differences are usually drawn until the ACF exhibits a pattern that can be understood only with a few basic autocorrelations.

Seasonality

It is not uncommon for stationary series to exhibit seasonal behavior outside of the trend currently under consideration. Seasonal patterns can also show consistent changes over time. SD is applied to seasonal non-stationarity just as regular differencing was applied to the overall trending series, and as well as auto regressive and moving average tools are available for the overall series, they are also available for seasonal phenomena using seasonal autoregressive parameters (SAR) and seasonal moving average parameters (SMA). (1996, Brockwell et al.).

Autocorrelation Function (ACF)

The most important tool for studying dependencies is the sample autocorrelation function. The correlation coefficient between any two random variables X, Y, which measures the strength of the linear relationship between X, Y, always takes a value between 1 and 1. Assuming stationary, the autocorrelation function ρ_k for a set of lags K = 1,2, ... is estimated by simply calculating the sample correlation coefficient between the temporally separated pairs in k units. The correlation coefficient between Y_t and Y_{t-k} is called the lag-k autocorrelation or serial correlation coefficient of Y_t and expressed by the symbol ρ_k, which under the assumption of weak stationarity, defined as:

$$\rho_k = \frac{\sum_{t-k+1}^{T}(Y_t - \overline{Y})(Y_{t-k} - \overline{Y})}{\sum_{t-1}^{T}(Y_t - \overline{Y})^2} = \frac{\gamma_k}{\gamma_0}\ ;\ for\ k = 1, 2, \ldots\ldots where, \gamma_k = \mathrm{cov}(Y_t, Y_{t-k}) \qquad \ldots (2)$$

It ranges from -1 to +1. Box and Jenkins has suggested that maximum number of useful ρ_k are roughly N/4 where N is the number of periods upon which information on y_t is available.

Partial Autocorrelation Function (PACF)

The correlation coefficient between two random variables Y_t and Y_{t-k} after removing the impact of the intervening Y_{t-1}, Y_{t-2} ,..., Y_{t-k+1} is

$$\phi_{00} = 1 \quad \phi_{11} = p_1$$

$$\phi_{kk} = \frac{p_k - \sum_{j=1}^{k-1}\phi_{k-1,j}\, p_{k-j}}{1 - \sum_{j=1}^{k-1}\phi_{k-1,j}\, p_j}, k = 2,3,\ldots\ldots where\, \phi_{k,j} = \phi_{k-1,j} - \phi_{k,k}\phi_{k-1,k-1} \qquad \ldots (3)$$

Autocorrelation function (ACF) and partial autocorrelation function (PACF)

Autoregressive process (AR)

It is possible to calculate the theoretical ACFs and PACFs (autocorrelations versus lags) for the various models considered (Pankratz, 1983) for various orders of autoregressive and moving average components (i.e. p and q. In order to discover a reasonable match between the TS data and the theoretical ACF/PACF, one or more ARIMA models might be tentatively selected by comparing the correlogram (plot of sample ACFs versus lags) generated from the TS data. The following are the general characteristics of theoretical ACFs and PACFs: (here, ‘spike' represents the line in the plot at various lags with length equal to magnitude of autocorrelations).

Autoregressive models are probabilistic models that can be very useful for representing series that actually occur. The current value of a process is expressed in this model as a finite linear set of previous values and jumps of the process ε_t.

A model written on the form

$$r_t = \phi_1 r_{t-1} + \phi_2 r_{t-2} + \ldots + \phi_p r_{t-p} + \varepsilon_t \qquad \ldots (4)$$

is known as autoregressive of order p and abbreviated as AR (p), where ϕ is autoregressive coefficient and ε_t is white noise .

In general, a variable r_t is said to be autoregressive of order p [AR (p)], if it is a function of its p past values and can be denoted as:

$$r_t = \sum_{i-1}^{P} \phi_i r_{t-i} + \varepsilon_t \qquad \ldots (5)$$

Moving Average process (MA)

A moving average model is another type of Box Jenkins model. Although these models look very similar to AR models, the underlying concepts are completely different. The moving average parameter relates events that occur in period t to random errors that occur in the previous period. The series $\{r_t\}$is called the moving average of orders q and MA(q) and is expressed in the form of the following equation:

$$r_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots\ldots - \theta_q \varepsilon_{t-q} \qquad \ldots (6)$$

Where, θ is moving average coefficient and ε_t is white noise

The eq(7) can be written as:

$$r_t = \varepsilon_t - \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} \qquad \ldots (7)$$

Autoregressive Integrated Moving Average process (ARIMA)

ARIMA is a well-known method for analyzing unsteady time series. Unlike regression models, ARIMA models allow rt to be described in terms of stochastic error terms as well as historical or lag values. Commonly referred to as "mixed models". This complicates the prediction method, but the structure can actually better mimic the series and produce more accurate predictions. A pure model means that the structure consists of only AR or MA parameters and not both. This approach's models are commonly referred to as ARIMA models because they employ a combination of auto regressive (AR), integration (I) - referring to the reverse process of differencing to produce the forecast - and moving average (MA) operations. An ARIMA model is usually stated as ARIMA (p, d, q). An auto regressive integrated moving average is expressed in the form:

If $w_t = \nabla^d r_t = (1 - B)^d r_t$ then

$$w_t = \phi_1 w_{t-1} + \phi_2 w_{t-2} + \ldots. + \phi_p w_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots.. - \theta_q \varepsilon_{t-q} \qquad \ldots (8)$$

If $\{W_t\}$ follows the ARMA (p, q) model, and $\{r_t\}$ is an ARIMA (p, d, q) process. For practical purposes, we can take is usually d = 1 or 2 at most. Above equation is also written as:

$$\phi(B) w^t = \theta_0 + \theta(B)\varepsilon_t \qquad \ldots (9)$$

Where $\phi(B)$ is a stationary autoregressive operator, $\theta(B)$ is a stationary moving average operator, and ε_t is white noise and θ_0 is a constant. In the case of the pattern of seasonal time series ARIMA model is written as follows:

$$\phi(B)\Phi(B)\nabla^d \nabla_s^D r_t = \theta(B)\Theta(B)\varepsilon_t \qquad \ldots (10)$$

Where;

$$w_t = \nabla^d \nabla_s^D r_t$$

$\nabla^d = (1 - B)^d$ is number of regular differences.

$\nabla_s^D = (1 - B^s)^D$ is number of seasonal differences.

In a seasonal ARIMA model, p denotes the number of 1 autoregressive components, q denotes the number of moving average terms, and d denotes the number of times the series must differ in order to achieve normality. The components of seasonal autoregression are P, the seasonal moving average term is Q, and the seasonal difference is D. (1994), Brockwell and Davis (1996).

ACF AND PACF PLOTS

The autocorrelation function (ACF) and the partial autocorrelation function (PACF) are the main tools for determining the relationship that exists between time series with different lags.

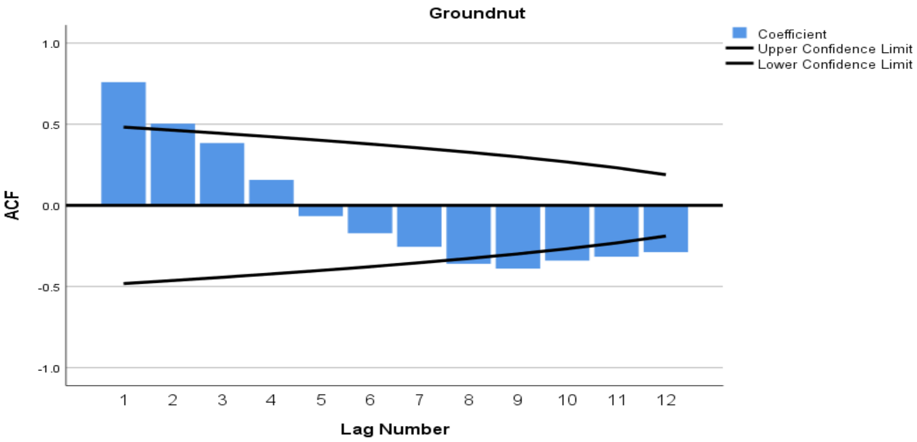

Fig 1

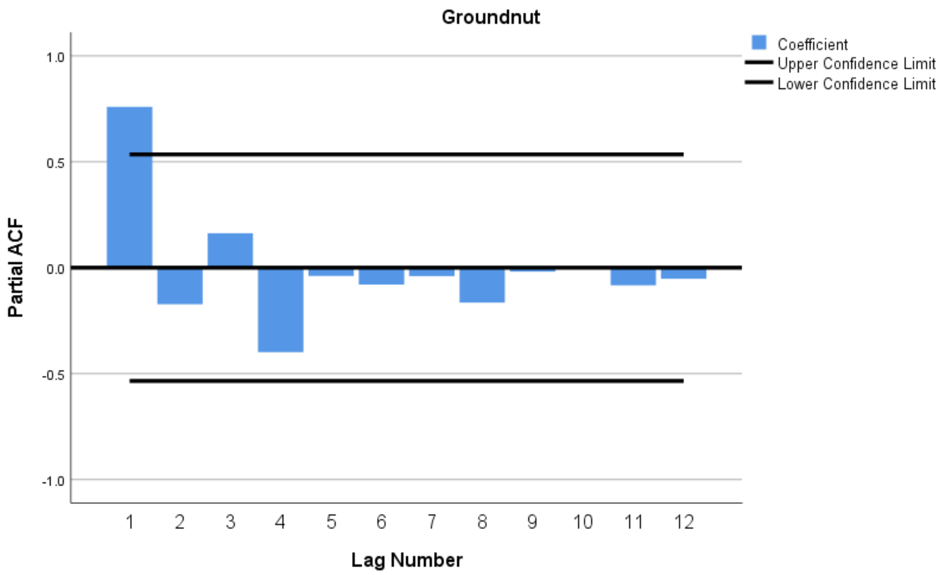

Fig 2

Figures 1 and 2 show the correlation between time series observations using ACF and PACF plots. The series shows changes over time with constant mean and variance, indicating that it is not stationary.

MODELING OF TIME SERIES DATA

A fitted autoregressive model was used to analyze the peanut production data in this section. ARIMA and neural network approaches are two autoregressive models used in data analysis.

ARMA Model

The ARMA model used a two-dimensional time series of peanut production data. The ARMA model is applied to the original series in Tables 1 and 2 below, which shows the prediction of peanut production and the definition of the ARIMA model.

Table. 1 Groundnut production in Tamil Nadu is being forecasted.

Year	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027
Production Predictive	454.96	604.58	607.48	505.72	529.07	488.13	416.94	391.19	390.83	347.72

Table 2. ARIMA (p,d,q) model identification for groundnut production in Tamil nadu

Model	Production ARIMA	Coefficient	S.E	σ^2	Log likelihood	AIC
(1,0,1)	AR1	-0.2185	0.4489	158261	-109.31	228.27
	MA1	-0.0745	0.4351			
(2,2,2)	AR1	-0.8915	0.2054	183012	-100.02	206.71
	AR2	-0.5149	0.1865			
	MA1	-1.0000	0.2657			
(1,0,0)	AR1	-0.2913	0.2126	158234	-109.51	225.34
(2,0,2)	AR1	-0.5147	0.4702	126932	-108.84	230.24
	AR2	-0.7519	0.3674			
	MA1	0.2061	0.4416			
	MA2	0.8272	0.7786			
(1,2,1)	AR1	-0.5562	0.1968	325395	-101.19	210.87
	MA1	-1.000	0.2054			
(1,1,1)	AR1	-0.3790	0.2519	185182	-104.56	214.82
	MA1	-0.7512	0.2026			
(0,0,1)	MA1	-0.2720	0.1953	160759	-110.42	226.75

In the present study, the ARIMA (2,2,2) model is the best fitted model based on the lowest AIC value, and it is then used to predict the production of groundnut in ceded districts for up to ten years using 15 years of time series data, i.e. from 20032004 to 20172018. ARIMA (2,2,1) is widely used due to its ability to predict time series data with any type of pattern and with auto correlated successive values of the time series. The study was also validated and statistically tested to ensure that the successive residuals in the fitted ARIMA (2,2,1) are not correlated and that the residuals appear to be normally distributed with a mean of zero and a constant variance. As a result, it can be a reliable predictor of groundnut yield in various districts of Tamil Nadu from 2018 to 2027. The ARIMA (2, 2, 1) models predicted an increase in production from 2018 to 2027. The projection for 2027 is about 347,7251,000 tonnes. ARIMA models, like other forecasting models, have limited predictive accuracy, but are widely used to predict future values in time series. This has been observed in the peanut production of Tamil Nadu.

Table. 3 Comparison of ARIMA and Feed-forward neural networks during full,

training, and testing sets of groundnut production

Models	Criteria	Full model	Training	Testing
ARIMA (2,1,1)	RMSE	1.152	1.003	1.617
	MAPE	14.317	12.984	21.850
	R-Square	0.512	0.567	0.214
ARIMA (2,2,2)	RMSE	1.021	1.002	1.674
	MAPE	13.843	12.156	20.547
	R-Square	0.527	0.547	0.053
ANN 3-2-1	RMSE	0.891	0.843	1.005
	MAPE	11.297	10.790	12.817
	R-Square	0.682	0.687	0.630

Neural networks outperform ARIMA models when it comes to capturing complex nonlinearities in data series. The resulting model baseline is the average of 10 networks. According to the results in Table 6, the entire data set is approximated using a feedforward network (321). Compared to the ARIMA model, the neural network model with 3 inputs, 2 hidden units and 1 output unit accurately predicted the production data. When fitted to the full dataset and training set, the ARIMA (2, 2, 2) model showed the best criterion of accuracy, but when fitted to the test set, the accuracy dropped sharply (e.g. 0.053). Despite being consistent with the test data set, ARIMA (2, 1, 1) out performed ARIMA (2, 2, 2) with an r- square value of 26.6%. In terms of RMSE (Full-0.891, Training-0.843, Testing-1.005), MAPE (Full-11.297, Training- 10.790, Testing-12.817), and MAPE (Full-11.297, Training- 10.790, Testing-12.817), NN outperformed ARIMA models (Table 3 When fitted to all data series except testing in predicting production by Vijay Shankaret.al, the model with parameters p-2, d-1-2, and q -1-2 produced better r-square and other accuracy criteria (2019).

CONCULISION

Such predictive models will be very useful to the agricultural community in terms of better planning of agricultural production. These practical ideas will help policy makers make important changes in key areas of production. Most importantly, crop-based marketing management can significantly reduce price volatility in the market. According to the above results, feedforward neural networks converge faster to local minima and have the ability to analyze complex data structures as discussed by Qiu M. and Song Y. (2016). A time series model is the best predictive model when using the full data set, but splitting the data into training and testing reduces the accuracy of the model. Because ANN models are designed for complex nonlinear data sets, they consistently predict when the data set is split into a training set and a test set. If the data is linear, we can use R- studio can be used to choose the appropriate p, d, and q parameters, and the BoxJenkin model performs better.

Reference

1. Vijay Shankar Pandey, and Abhishek Bajpai. (2019). Predictive Efficiency of ARIMA and ANN Models: A Case Analysis of Nifty Fifty in Indian Stock Market. International Journal of Applied Engineering Research. ISSN 0973-4562. Volume 14(2), 232–244.
2. Qiu M. and Song Y. (2016). Predicting the Direction of Stock Market Index Movement Using an Optimized Artificial Neural Network Model. PLoS ONE. 11(5), e0155133.
3. Tripathi Rahul, Nayak A. K, Raja R. Shahid, Mohammad Kumar, AnjaniMohanty, Sangita Panda, Bipin Lal, B. and Gautam Priyanka. (2014). Forecasting Rice Productivity and Production of Odisha, India, Using Autoregressive Integrated Moving Average Models. Advances in Agriculture. 10, 1155/2014/621313.
4. Muhammed (1992) conducted “An empirical of modelling and forecasting time series data of rice production in Pakistan”, 2.27-3lp.
5. Madhusudana, B (2013) for ”A survey on area, production and productivity of groundnut crop in India”, IOSR Journal of Economics and Finance, volume 1, PP 01-07.
6. M.K Debnath , (2013) for “ Forecasting Area, production, and Yield of Cotton in India using ARIMA Model”, Journal of Space Science & Technology, Volume 2, Issue 1.
7. Hemavathi, M. and Prabakaran, K. (2018). ARIMA Model for Forecasting of Area, Production and Productivity of Rice and Its Growth Status in Thanjavur District of Tamil Nadu, India. Int.J.Curr.Microbiol.App.Sci. 7(2), 149–156.
8. Qiu M. and Song Y. (2016). Predicting the Direction of Stock Market Index Movement Using an Optimized Artificial Neural Network Model. PLoS ONE. 11(5), e0155133.
9. E Sakthivel, K Senthamarai Kannan (2014) Fuzzy Time Series Forecasting on Intuitionistic Fuzzy Sets, International Journal of Advanced Research in Computer Science and Software Engineering, vol.4, Issue 2.
10. Box G.E.P., Jenkins G. (1970). “Time Series Analysis, Forecasting and Control”, Holden-Day, San Francisco, CA, 1970.
11. Ibrahim Usman, et.al (2013). "Socio Economic Factors Affecting Groundnut Production In Sabongari Local Government Of Kaduna State," Nigeria, International Journal Of Food And Agricultural Economics, vol. 1, pp. 41-48, 2013.
12. Muhammad, F., M. Siddique, M. Bashir and S. Ahmad (1992). Forecasting rice production in Pakistan using ARIMA Models. J.Anim.Plant.Sci. 2: 27– 31.
13. Sitarambabu V., Chowdary N. A. and SolmonRaju Paul K, "Trends, Growth and Variability of Bengal Gram in Andhra Pradesh,"International Journal of Development Research, vol. 4, Issue 12, pp. 2560-2562, 2013.
14. Parmar R. S., "Influence of Rainfall Distribution on the Productivity of Groundnut in Bhavnagar District of Gujarat State,"International Journal of Agricultural Sciences and Veterinary Medecine, vol. 1, 2013.