Turkish Online Journal of Qualitative Inquiry (TOJQI)
Volume 12, Issue 9, August 2021: 6462-6471

Forecasting Groundnut Production With Arima and A Neural Network Approach

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

[1]Department of Statistics, Government Arts College (Autonomous), kumbakonam-612001, Tamilnadu, India.

Affiliated to Bharathidasan University, Tiruchirappalli – 620024

[2] Department of Statistics, Periyar EVR College (Autonomous), Tiruchirappalli – 620024, Tamilnadu, India.

Affiliated to Bharathidasan University, Tiruchirappalli – 620024

Email: statsara04@gmail.com

Abstract

The king of oilseeds is groundnut. Groundnut is the world's most widely produced oilseed, and India's second-largest oilseed. In 2020, the Indian agriculture sector will account for 15.4 percent of India's GDP and employ 41.49 percent of the country's workers. Because groundnut prices are prone to fluctuating swiftly and unpredictably, farmers require a realistic forecasting of harvest season prices in order to choose the amount of groundnut acreage to plant. The study employed time series data from Tamilnadu that spanned the years 2003 to 2018. The goal of the research is to find the best ARIMA model for fitting and forecasting groundnut production in Tamilnadu. Using ARIMA and Neural Network Approach using R software, certain inferences about the anticipated production of groundnut crop are drawn based on the obtained data.

Keywords: ARIMA, Neural Network Approach, Groundnut Production

INTRODUCTION

In the 16th century, the Portuguese introduced groundnut from Brazil to West Africa and ultimately to southwestern India. Groundnut is the world's 13th most significant food crop. It is the world's fourth-largest producer of edible oil and third-largest producer of vegetable protein. After China, India is the world's second-largest producer of groundnuts, accounting for over 20% of global production. It is known as the "king" of oilseeds, accounting for over 25% of the country's total oilseed production. Almost every component of the peanut can be sold. Groundnuts are now used in a variety of ways, including groundnut oil, groundnut cake, groundnut kernels, groundnut shell, and groundnut straw. Despite the fact that oilseed output has increased significantly since the 1960s, demand for oilseeds continues to rise due to rising population growth rates and per capita edible oil consumption. However, due to a discrepancy between domestic availability and actual consumption of edible oils, India must rely on imports.

Concentration has been given on the uni-variate time series Auto Regressive Integrated Moving Average (ARIMA) Models, which primarily due to World of Box and Jenkins (1970). Among the stochastic ARIMA types which are powerful, effective and popular as they can correctly describe the found facts and can make forecast with minimum forecast error. These types of models are very difficult to pick out and estimate. Muhammed et al (1992) conducted an empirical of modelling and forecasting time series data of rice production in Pakistan. Ibrahim Usman et al., 2013 have done gross margin and cost benefit analysis to study the profitability of groundnut production in Nigeria. Madhusudhana, 2013 has done a comparative analysis for groundnut crop at national level, state level, and Anantapuram district level in India. SitaRambabu et al., 2013 analyzed trends and Compound growth rates for area, production and productivity of groundnut in Andhra Pradesh over a period of 1995-96 to 2010-2011. Parmar, 2013 observed Flowering and Peg initiation stage was the most critical phase for moisture requirement of groundnut using regression. M.K Debnath et al. (2013) for forecasting Area, production, and Yield of Cotton in India using ARIMA Model ,. Tripathi Rahul (2014) has done an In agriculture and allied sciences, time series models are used for forecasting milk production, milk yield of certain breeds of cows, 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 M. and Song Y. (2016). M. Hemavathi et al.(2018) ARIMA Model for Forecasting of Area, Production and productivity of Rice and Its Growth Status in Thanjavur District of Tamil Nadu, India ,also use the ARIMA Model.

Autoregressive models of various types are used as a statistical tool for analyzing various types of time series data. Yule and Walker proposed the Autoregressive Moving Average (ARMA) model, and Box and Jenkins proposed the methodology of the Autoregressive Integrated Moving Average (ARIMA) model later on. The primary goal of this paper is to forecast groundnut production. The ARIMA models and the Neural Network Approach are used to examine groundnut production in Tamil Nadu. The forecast results are obtained, and a detailed explanation of model selection and forecasting accuracy is provided.

SOURCE OF DATA

The study relied on secondary data sources. The statistical year book of India 2018 published by MOSPI, govt. of India, provided the time series data on yearly totals production of groundnut crop from 2003-2004 to 2017-2018, which was required for the study. The model was validated using 15 years of groundnut production data. The data sets were obtained for statistical reporting purposes.

In this publication, the study's data sources and nature, as well as the analytical techniques used in the investigation, are described.

Methodology

Autoregressive Integrated Moving Average Model (ARIMA) Model

ARIMA (Auto Regressive Integrated Moving Average) is a mixed model that is fitted to a collection of data using the Box-Jenkins process. As a result of fitting the ARIMA model, the fundamental goal is to identify the stochastic process that underlies the time series and precisely 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 auto-regressive (AR) models. Slutsky, who introduced Moving Average (MA) methods in 1937, added to these. As Wold (1938) demonstrated, ARMA processes may 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. A few basic ideas of linear time series analysis, such as stationary and seasonality, and a brief reference to the most popular forms of time series forecasting processes are presented before moving on to the ARIMA model-building

Stationarity and non-stationarity

Stationary Time Series have a constant mean and variance, and an autocorrelation function (ACF) that remains largely constant across time. Other than that, it's referred to Dickey and Fuller have presented a statistical test for stationarity (1979).

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

Where $\gamma = \Phi$-1, Then, null hypothesis of H_0: γ =0 against the alternative hypothesis H_1: γ<0. If the series is stationary, the null hypothesis must be accepted Differencing is usually done until the ACF exhibits a pattern that can be understood with only a few major autocorrelations.

Seasonality

It is not uncommon for stationary series to display seasonal behaviour in addition to the trend, which has now been addressed. Additionally, the seasonal pattern may show a steady change 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 tools for study dependence is the sample autocorrelation function. The correlation coefficient between any two random variables X, Y, which measures the strength of linear dependence between X, Y, always takes values between -1 and 1. If stationarity is assumed and autocorrelation function ρ_k for a set of lags K = 1,2, ... is estimated by simply computing the sample correlation coefficient between the pairs, k units apart in time. The correlation coefficient between Y_t and Y_{t-k} is called the lag-k autocorrelation or serial correlation coefficient of Y_t and denoted by the symbol ρ_k , which under the assumption of weak stationarity, defined as:

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

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 called (PACF) at lag k and denoted by ϕ_{kk}

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

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

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

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 process (AR)

The autoregressive model is a stochastic model that can be extremely useful in representing certain practically occurring series. The current value of the process is expressed in this model as a finite, linear aggregate of previous values of the process and a shock ε_t.

A model written on the form

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

is called 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 represented 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 appear to be very similar to the AR model, the underlying concept is quite different. Moving average parameters only relate what happens in period t to random errors that happened in previous time periods. A series $\{r_t\}$ is called moving average of order q and abbreviated as MA (q), expressed in following form of equation:

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

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

The eq(7) can be written as:

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

Autoregressive Integrated Moving Average process (ARIMA)

ARIMA is a well-known method for analysing non-stationary time series. Unlike regression models, the ARIMA model allows rt to be explained by its past, or lagged, values as well as stochastic error terms. These are commonly referred to as "mixed models." Although this complicates the forecasting method, the structure may indeed better simulate the series and produce a more accurate forecastPure models imply that the structure is made up of only AR or MA parameters, not both This approach's models are commonly referred to as ARIMA models because they employ a combination of autoregressive (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 autoregressive integrated moving average is expressed in the form:

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

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

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:

$$\phi(B) w^t = \theta_0 + \theta(B)\varepsilon_t \qquad \text{… (9)}$$

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:

$$\phi(B)\Phi(B)\nabla^d \nabla_s^D r_t = \theta(B)\Theta(B)\varepsilon_t \qquad \text{… (10)}$$

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 the seasonal ARIMA model, p represents 1the number of autoregressive components, q represents the number of moving average terms, and d denotes the number of times that a series must be differenced to achieve stationarity. Seasonal autoregressive components are represented by P, seasonal moving average terms are denoted by Q, and seasonal differences are designated by D. (1994), Brockwell and Davis (1996).

ACF AND PACF PLOTS

The Autocorrelation function (ACF) and Partial Autocorrelation function (PACF) are the primary tools for identifying the relationships that exist between time series at different lags.

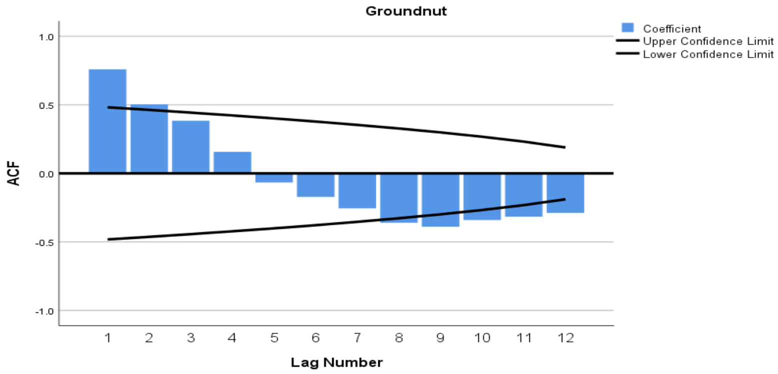

Fig 1

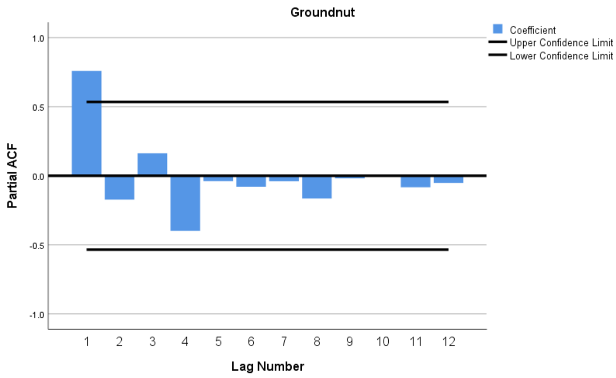

Fig 2

Figures 1 and 2 depict the correlation between time series observations using ACF and PACF plots. The series has shown variation with time over a constant mean and variance, indicating that it is not stationary.

MODELING OF TIME SERIES DATA

The fitted autoregressive models are used to analyse Groundnut production data in this section. ARIMA and the Neural Network Approach are two autoregressive models used for data analysis.

ARMA Model

A bivariate time series of groundnut production data was used in the ARMA model. The ARMA model is applied to the original series in tables 1 and 2 below, which show forecasting groundnut production and identifying ARIMA models.

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 current 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 groundnut production in ceded districts for up to ten years using 15 years of time series data, i.e. from 2003-2004 to 2017-2018. ARIMA (2,2,1) is 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 forecast for 2027 is approximately 347.7251'000 tonnes. The ARIMA model, like any other predictive model for forecasting, has limitations in prediction accuracy, but it is widely used for forecasting future values for time series. It has been observed in Tamilnadu groundnut production.

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

When it comes to capturing complex nonlinearity in a data series, neural networks outperform ARIMA models. The model criteria obtained are the mean of ten networks. According to the results in Table 6, the full data set is fitted using a feed-forward 3-2-1 network. When compared to ARIMA models, a neural network model with three inputs, two hidden units, and one output unit accurately predicted production data. When fitted to the full data set and training set, the model ARIMA (2, 2, 2) exhibited better accuracy criteria, but when fitted to the testing set, its accuracy dropped dramatically, i.e., it predicted testing data series with low r-square (0.053). While fitted to the testing data set, ARIMA (2, 1, 1) performed better than ARIMA (2, 2, 2) with a 26.6 percent r-square value. 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).

Conclusion

These prediction models will be extremely beneficial to the farming community in terms of better agricultural production planning. These real insights will assist policymakers in making significant changes in leading areas of production. Most importantly, price volatility in the markets can be reduced to a greater extent through marketing management based on crop production. According to the above results, a feed forward neural network converges faster to local minima and has the ability to analyse complex data structures, as discussed by Qiu M. and Song Y. (2016). When full data sets are used, time series models are better predictive models, but model accuracy decreases when the data is split into training and testing. Because ANN models are designed for complex nonlinear data sets, they consistently predict when the data set is divided into training and testing sets. When the data is linear, the auto.arima function of R-studio can be used to select appropriate p, d, and q parameters, and Box-Jenkin models perform better.

References

1. Ayres, L. (2008). Thematic coding and analysis. In L. M. Given (Ed.), The SAGE Encyclopedia of Qualitative Research Methods (pp. 867–868). Thousand Oaks, CA: Sage.
2. Denzin, N. K., & Lincoln, Y. S. (2005). Introduction: The discipline and practice of qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), Handbook of qualitative research (3rd ed., pp. 1-32). Thousand Oaks, CA: Sage.
3. 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.
4. 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.
5. Tripathi Rahul, Nayak A. K, Raja R. Shahid, Mohammad Kumar, Anjani Mohanty, 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.
6. Muhammed (1992) conducted “An empirical of modelling and forecasting time series data of rice production in Pakistan”, 2.27-3lp.
7. 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.
8. 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.
9. 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.
10. 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.

11. Box G.E.P., Jenkins G. (1970). “Time Series Analysis, Forecasting and Control”, Holden-Day, San Francisco, CA, 1970.
12. 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.
13. 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.
14. 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.
15. 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.