Modeling ~~Forecasting~~ Long Memory Time Series for Groundnut prices in Andhra Pradesh with Autoregressive Fractionally Integrated Moving Average for Forecasting

ABSTRACT

The presence of long memory in time series is characterized by an autocorrelation function that decreases slowly or hyperbolically. The most suitable model for capturing this phenomenon is the Autoregressive Fractionally Integrated Moving Average (ARFIMA) model, which is particularly useful for modeling historical prices in financial data analysis. This research aims to assess ARFIMA modeling of long memory processes using the Geweke and Porter-Hudak (GPH) parameter estimation method. The model was applied to the monthly prices of Groundnut in Andhra Pradesh, from the period January 2002 to December 2023. The best-fitted model identified was ARFIMA(1,0.43,1), which demonstrates a strong short-term forecasting ability, closely matching actual prices with lowest AIC, MSE and RMSE values when compared to SARIMA$(1,1,3)(0,1,2)_{12}$ model. The study concluded that the ARFIMA model forecasted better than the SARIMA model indicates the model captures the existence of long memory present in the price series. Pl. chk the sentence

Formatted: Font color: Red

Formatted: Font color: Red

Keywords: Autoregressive fractionally integrated moving average, Geweke and Porter Hudak method, Groundnut, Long memory, Prices.

INTRODUCTION

The Groundnut (Arachis hypogaea L.) is a leguminous plant extensively cultivated in tropical and subtropical regions between 40°N and 40°S latitudes (ref?). Groundnut renowned for its high oil content and edible seeds, it ranks as the fourth most important source of edible oil and the third most significant source of vegetable protein globally (Ref?). In India, it is not only a vital oilseed crop but also a significant agricultural export commodity. Globally, groundnut is cultivated over 327 lakh hectares, ~~production~~ producing 539 lakh tonnes with productivity of 1648 kg per hectare (FAOSTAT, 2021). India, with an annual all season coverage of 54.2 lakh hectares, ranks first in groundnut cultivation area and is the second largest producer, achieving 101 lakh tonnes with a productivity of 1863 ~~kg~~ kilogram per hectare in 2021-22 (agricoop.nic.in) In India, Andhra Pradesh contributes an area of 5.94 lakh hectares, production 6.01 lakh tonnes and productivity 1012 kg/ha of during 2022-23 of groundnut. (des.ap.gov.in). Area and production Scenario of groundnut in India over several decades were depicted in figure.1.

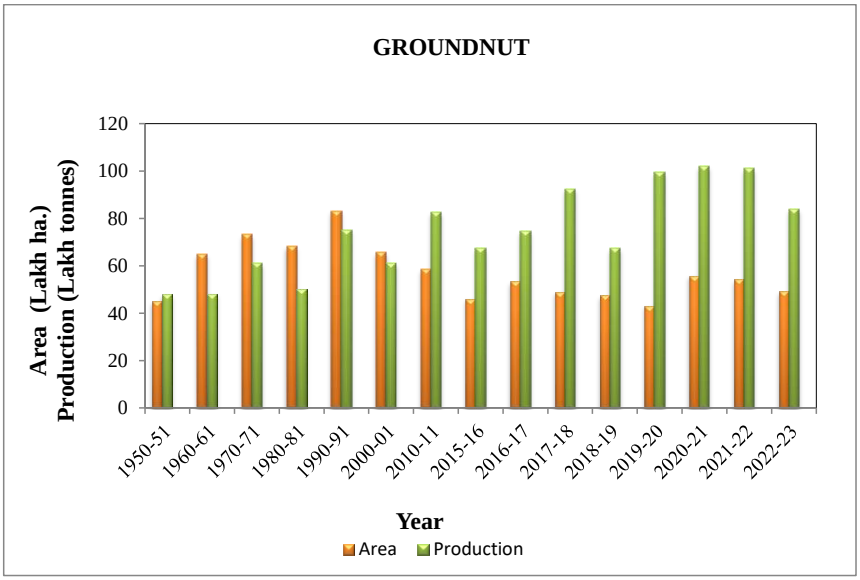


Fig.1. Area, Production ~~and Yield~~ of Groundnut in India

Formatted: Font color: Red, Strikethrough

Formatted: Font color: Red

Formatted: Highlight

Along with production, the prices of groundnut were also fluctuates rapidly and prices were also unpredictable and largely influenced by different factors that cause a significant risk and uncertainty Sentence pl. Hence, forecasting agricultural commodity prices is a crucial aspect in agricultural sector as it allows farmers, traders, and policymakers to make informed decisions regarding production, marketing, and policy implementation. The concept of long memory processes has evolved to provide substantial evidence for describing phenomena in time series data, particularly in the fields of finance and macroeconomics. The presence of long memory ~~is~~ can be identified empirically by examining the persistent autocorrelations within observed time series data. This persistence is indicated by the stationarity of the data over time, characterized by autocorrelations that decrease slowly or hyperbolically, often associated with a class of autoregressive moving average (ARMA) models.

The most notable definition of a long memory process was provided by Haslett and Raftery (1989), who stated that data exhibiting long memory are characterized by an autocorrelation function that does not decline exponentially, but rather decreases slowly or hyperbolically. The concept of long memory in time series was initially introduced by Hurst (1951) through various data sets. Subsequently, Granger and Joyeux (1980) and Hosking (1981) developed a model suited for long memory processes, known as the Autoregressive

Fractionally Integrated Moving Average (ARFIMA) model. This model effectively explains time series by incorporating both short memory and long memory components, with the differencing parameter represented as a real number. The study of long memory processes, particularly in relation to the ARFIMA model, had been extensively developed in data analysis across both time and space. One of its most compelling features was its suitability for long-term predictions and assessing the effects of shocks within conventional macroeconomic frameworks. In this study, to demonstrate the long memory process using ARFIMA models by employing the Geweke and Porter-Hudak (GPH) differencing parameter estimation method on historical data of groundnut monthly prices of Andhra Pradesh. The ARFIMA model was used in this study because its capability to directly estimate the differencing parameter, eliminating the need to initially know the order values of the autoregressive and moving average components.Ref? The study also compared the forecasting performance of SARIMA model with ARFIMA models.

Formatted: Highlight

METHODOLOGY:

Data: Groundnut monthly prices of Andhra Pradesh data was collected from Agricultural Market Intelligence Committee (AMIC), Lam farm, Guntur from the period January 2002-December 2023.The collected data was divided into training and a testing datasets. First 384 observations were used as a training dataset for model development and last 12 observations were used as testing dataset for model validation purpose.

Descriptive Statistics:

The summary statistics viz., mean median, standard deviation, skewness, kurtosis, minimum and maximum were used to study the behaviour of the monthly prices of groundnut in Andhra Pradesh.

ARIMA Model:

The Autoregressive Integrated Moving Average (ARIMA) methodology developed by Box-Jenkins is the most widely used model for analysing time series data. The Box-Jenkins model-building process is used to fit a blended ARIMA model to provided data. The basic purpose of fitting the ARIMA model is to accurately characterise and forecast the time series stochastic process (Box and Jenkins, 1970).

Initially, George Box and Gwilym Jenkins conducted substantial research on ARIMA models, and their names were frequently associated with the broad ARIMA method used in

time series analysis, forecasting, and control. The two forms of stochastic processes are stationary and non-stationary. The ARIMA model can only be used with stationary data.

Stationarity and Non-stationarity:

A process that generates data in equilibrium around a constant value and has a constant variance around the mean throughout time is referred to as "stationary." If the means shift over time and the variance is not roughly constant both mean and variance, the series is said to be non-stationary. To build the ARIMA model the series should be stationary in nature. If the original series is not stationary then it has to make stationary by differencing will be done to convert the non-stationary series into stationary series Sentence Pl.

Formatted: Highlight

Autoregressive Model of order p (AR (p)):

An autoregressive model is one in which Y_t depends only on its past values Y_{t-1}, Y_{t-2}, Y_{t-3}, etc. is called autoregressive of order *p* and abbreviated as AR (p), where ∅ is autoregressive coefficient and ε_t is white noise.

$$y_t = \emptyset_1 y_{t-1} + \emptyset_2 y_{t-2} + \cdots + \emptyset_p y_{t-p} + \varepsilon_t \quad \ldots (1)$$ -ve signs need explanation

In general, a variable y_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:

$$y_t = \sum_{i-1}^{p} \emptyset_i y_{t-i} + \varepsilon_t \quad \ldots (2)$$

Moving Average Model order q (MA (q)):

Moving Average (MA) is the one where Y_t depends on its lagged forecast errors.

$$y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} \quad \ldots. (3)$$

The MA term is represented by the order q and abbreviated as MA(q) and θ is MA coefficient.

Autoregressive Moving Average model (ARMA (p, q)):

It is often advantageous to use both autoregressive and moving average processes in order to achieve greater flexibility in fitting of time-series data. This leads to mixed autoregressive-moving average model.

$$y_t = \emptyset_1 y_{t-1} + \emptyset_2 y_{t-2} + \cdots + \emptyset_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} \quad \ldots. (4)$$

Autoregressive Integrated Moving Average Model (ARIMA (p, d, q)):

The ARIMA model allows y_t to be explained by its past, or lagged values and stochastic error terms. The models are often referred to as "mixed models." Either explain why or provide Ref. ARIMA models use a combination of autoregressive (AR), integration (I) and moving average (MA). The term integration is referred when a non-stationary series is converted into stationary series by means of differencing. Box and Jenkins propose a practical four stage procedure for finding a good model. The four-stage univariate Box Jenkins procedure is summarized schematically in Fig.2.

Formatted: Highlight

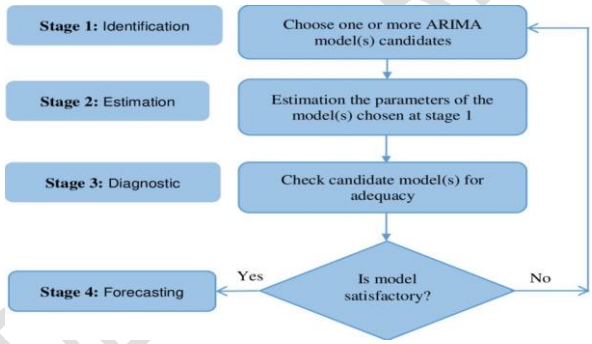


Fig.2. Flow chart of Box-Jenkins Methodology

The main stages in setting up a Box-Jenkins forecasting model are described below:

1. **Identification:**

The autocorrelation function (ACF) and partial autocorrelation function (PACF) are two graphical devices to measure the correlation between the observations within a single data series and they give an idea about the patterns and relationship in the available data. As the time series under study is a particular realization of the process, the theoretical ACF and PACF must resemble the estimated ACF and PACF of the data.

Table.1. Pattern of ACF and PACF for identification of AR, MA and ARMA process

PROCESS	ACF	PACF
AR	Decays towards zero	Cut off to zero (lag length of last spike is the order of the process)
MA	Cut off to zero (lag length of last spike is the order of the process)	Decays towards zero
ARMA	Tails off towards zero	Tails off towards zero

2. Estimation of parameters

At the estimation stage, coefficients of the identified models are estimated using method of least squares or maximum likelihood estimation methods are used to estimate the parameters. Stationarity and invertibility are checked for the coefficient obtained and at the same time diagnostic checking is done in order to know whether the model fit the data satisfactorily or not. The importance of the estimation coefficients is measured in terms of the statistical significance.

3. Diagnostic Checking

Different models can be obtained for various combinations of AR and MA individually and collectively. The best model is obtained with following diagnostics.

(a) Low Akaike Information Criteria (AIC) / Bayesian Information Criteria (BIC)Not discussed

Formatted: Highlight

AIC is given by (-2 log L + 2m) where m = p + q + P + Q ist time introduction in the article, needs elaboration and L is the likelihood function. Since -2log L is approximately equal to $\{n (1+\log 2\pi) + n \log \sigma^2\}$ where σ^2 is the model MSE. Thus, AIC can be written as AIC = $\{n (1+\log 2\pi) + n \log \sigma^2 + 2m\}$ and because first term in this equation is constant, it is usually omitted while comparing between models. The model having lowest AIC/BIC is considered as the best model.

Formatted: Highlight

Formatted: Font color: Red

(b) Plot of residual ACF

Once the appropriate model has been fitted, the goodness of fit can be examined by plotting the ACF of residuals of the fitted model. If most of the sample autocorrelation coefficients of the residuals are within the limits $\pm$ 1.96/N where N is the number of observations on which the model is based, then the residuals are white noises indicating that the model is good fit.

(c) Box-Pierce or Ljung-Box texts

Box-Pierce statistic is a test to measure the overall adequacy of the chosen model by examining a quantity Q, whose approximate distribution is Chi-square.

$$Q = n \sum_1^k r_{(j)}^2 \quad \ldots (5)$$

Where *k* as maximum lag considered, and is usually around 20, *n* = number of observations, *r(j)* is the estimated autocorrelation at lag *j*. Chi-square with *(k-m-1)* degrees of freedom where *m-1* is the number of parameters estimated in the model.

A modified Q statistics is the Ljung-box which is given by

$$\text{Chk q} = n(n+2) \sum \frac{r_{(j)}^2}{n-j} \quad \ldots (6)$$

Formatted: Highlight

The critical value of *Q* statistic is compared with Chi-square *(n-1)* degrees of freedom. Residuals should be uncorrelated and Q should be small if model is correctly specified. A significant value of test statistic indicates the chosen model is not a good fit.

4. Forecasting

The model that satisfies all the diagnostic checks is considered for forecasting. If the model is based on differencing / de-trending transformations, then the model must be represented with relevant expressions of original series. Then only, the forecasts can be made.

Seasonal ARIMA

In the time series analysis, seasonality is defined as the pattern of changes that repeats over *S* time periods, where *S* is the number of time periods between the repeats of the pattern. For quarterly data, *S* = 4 time periods per year and for monthly data *S* = 12 time periods per

year are considered. As the regular differencing was applied to the series having non-stationary nature similarly seasonal differencing will be applied to the seasonal non-stationary series. The seasonal Autoregressive (SAR) and Seasonal Moving Average (SMA) are the parameters of seasonal ARIMA. In the seasonal ARIMA model, seasonal AR and MA terms predicts the x_t often with the lags that are multiples of *S*.

Seasonal ARIMA model is denoted by ARIMA (p, d, q) $(P, D, Q)_S$, where p represents the number of autoregressive terms, q represents the number of moving average terms and d denotes order of differencing to induce stationarity, P represents the number of seasonal Autoregressive components, Q represents the number of seasonal moving average terms and D represents the number of seasonal differences required to make the series stationarity. The seasonal ARIMA model expressed as follows;

$$\emptyset(B)\Phi(B)\nabla^d\nabla_s^D r_t = \theta(B)\Theta(B)\varepsilon_t \quad \ldots. (7)$$

$$w_t = \nabla^d\nabla_s^D r_t \quad \ldots. (8)$$

$\nabla^d = (1-B)^d$ denotes the number regular differences and $\nabla_s^D = (1-B^s)^D$ denotes number of seasonal differences.

Where, $\emptyset(B)$ is stationary Autoregressive operator, $\theta(B)$ is a stationary moving average operator, ε_t is a white noise (Brockwell and Davis, 1996).

Long memory model:

Long memory in time-series can be defined as autocorrelation at long lags. According to Jin and Frechette (Yr?) memory means that observations are not independent (each observation is affected by the events that preceded it). The ACF of a time-series x_t is defined as

$$\rho_k = \text{cov}(x_t, x_{t-k})/\text{var}(x_t) \quad \ldots..(37)$$ Chk equation number from this one and onwards

Formatted: Highlight

for integer lag k. A covariance stationary time-series process is expected to have autocorrelations such that $\lim_{k\to\infty} \rho_k = 0$. Most of the well-known class of stationary and invertible time-series processes have autocorrelation at decay at the relatively fast exponential rate, so that ρ_k $|m|^k$ where, $m < 1$ and this property is true, for example, for the well-known stationary and invertible ARMA (p, q) process.

For long memory processes, the autocorrelations decay at an hyperbolic rate which is consistent with $\rho_k \approx$, $\rho_k \approx ck^{2d-1}$, k increases without limit, where C is a constant and d is the long memory parameter. ACF is considered to be the time domain analogue of spectral density. In terms of the auto covariance sequence (ACVS), i.e. $\{S_{x,k}\}$ for $\{x_t\}$, $\{x_t\}$ is a stationary long memory process if there exist $\lim_{k\to 0} S_{x,k}/(\mathrm{CsK}^b)$ constants b and C_S satisfying $-1 < b < 0$ and $C_S > 0$ such that the standard time-series models such as stationary autoregressive processes have ACVS s such that $S_{x,k} \approx c\phi^k$ forlarge k, where $C \geq 0$ and $|\phi| < 1$ Sentence pl. For a long memory process, $S_{x,k} \approx C_S k^b$ for large k. In both the cases $S_{x,k} \to 0$ as $k \to \infty$, but the rate of decay toward zero is much slower for a long memory process, implying that the observations that are widely separated in time can still have a non-negligible covariance i.e. the current observations retain some 'memory' of the distant past. An alternative definition can be stated as if $\{x_t\}$ is a stationary process with the spectral density function (SDF) denoted by $S_x(.)$, then $\{x_t\}$ is a stationary long memory process if there exist constants a and C_S satisfying $1 < a < 0$ and $C_S > 0$ such that

$$\lim_{f\to 0} S_x(f)/(C_s|f|^a) = 1 \qquad \ldots\ldots(38)$$

Where, a = -b-1. In other words, a stationary long memory process has an SDF $S_x(.)$ such that $S_x(.) \approx (C_s|f|^a)$, with the approximation improving as f approaches zero.

GPH estimator:

The approximated regression equation is used in this method, which was calculated by logarithmic transformation of the spectral density function (SDF). This method is based on least squares regression in the spectral domain (Geweke and Hudak, 1983), utilize the sample form of the pole of the spectral density at the origin, $f(\eta) \sim \eta - 2d$, $as\ \eta \to 0$. The SDF of a stationary model yt, t=1,…, can be expressed as,

$$f(\eta) = [4\sin(\eta 2) - d] f\varepsilon(\eta) \qquad \ldots(39)$$

where $f\varepsilon(\eta)$ is the spectral density of εt, assumed to be a finite and continuous function on the interval $[-\pi,\pi]$. After logarithmic transformation of the SDF, the log-spectral density can be written as,

$$log(f_y(\eta)) = log(f_\varepsilon(0)) - dlog[4sin^2(\frac{n}{2})] + log\frac{f\varepsilon(\eta)}{f\varepsilon(0)} \qquad \ldots(40)$$

Let, $I(\eta i)$ be the periodogram obtained at the Fourier frequencies, $\eta_l = 2\pi lk$, l=1,2,…,t; k is the total number of observations and t is the number of considered Fourier frequencies,

that is the number of periodogram series which is utilized in regression,

$$log(I_y(\eta_l))=log(f_\varepsilon(0))-dlog[4sin2(\frac{n}{2})]+log\frac{f\varepsilon(\eta)}{f\varepsilon(0)}+\frac{Iy(\eta l)}{fy(\eta l))}) \quad \text{....(41)}$$

where, $log(f\varepsilon(0))$ is being constant, $log[4sin^2(\frac{n}{2})]$ is the exogenous variable and $log\frac{Iy(\eta l)}{fy(\eta l))}$ is the unforeseen term. The GPH estimate is based on two assumptions which are reason behind the asymptotic behaviour of the equation. These are,

***H*1**: For low frequencies, we suppose that $logf(\eta)/f_\varepsilon(0)$ is negligible

***H*2**: The random variable $logI(\eta_l)/f_y(\eta_l)$, l=1,2,…,t are asymptotically iid

Under the hypotheses $H1$ and $H2$, we can write the linear regression,

$$lo(I_y(\eta_l))= \alpha-dlog[4sin^2(\frac{n}{2})+e_l \quad \text{.....(42)}$$

where, $e_l \sim iid(-c,\pi^2/6)$. Considering the equation (42) as a regression equation of $lo(I_y(\eta_l))$ on α and yl , where y_l=log $[4sin2(\frac{n}{2})]$. The OLS estimate of d is obtained as,

$$\hat{d}GPH=\sum_{i=1}^{t}(yi-\bar{y})log(I_y(\eta_l))/=\sum_{i=1}^{t}(yi-\bar{y})^2$$

List 1 : According to the value of d long memory process can be sub-divided into 4 groups and these are:

Value of d	Names
$d\in(-1/2,0)$	Intermediate Memory and Anti-persistence
d=0	White noise(Short-Memory)
$d\in(0,1/2)$	Stationary and Persistence Long Memory
$d\in[12,1)$	Non-stationary and Persistence Long Memory

ARFIMA model:

Fractional integration is a generalization of integer integration, under which time-series are usually presumed to be integrated of order zero or one. Hosking and Reinsen described autoregressive fractionally integrated moving average (ARFIMA) process in details. x_t is called an ARFIMA(p, d, q) process with degree of differencing as d, if it satisfies

$$(1 - L)d \text{ chk } \Phi (L)\, xt = \Theta (L)\, \varepsilon t \quad \text{.... (39)}$$

where, ε_t is an independently and identically distributed (i.i.d.) random variable with zero mean and constant variance, L denotes the lag operator; and $\Phi(L) = 1\ \varphi_1 L \dots\ \varphi_p L^p$ and $\Theta(L) = 1\ \theta_1 L \dots\ \theta_q L^q$ denote finite polynomials in the lag operator with roots outside the unit circle For d = 0, the process is stationary, and the effect of a shock to ε_t on $x(t+j)$ decays geometrically as j increases. For d = 1, the process is said to have a unit root, and the effect of a shock to ε_t on $x(t + j)$ persists into the infinite future. In contrast, fractional integration defines the function $(1 - L)^d$ for non-integer values of the fractional differencing

Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Highlight

parameter d. For $0.5 < d < 0.5$ chk the process $x(t)$ is stationary and invertible. For such processes, the effect of a shock ε_t on x $(t+j)$ decays as j increases, but the rate of decay is much slower than for a process integrated of order zero. More precisely, the ACF for zero-integrated processes decays geometrically, whereas the ACF for a fractionally integrated process decays hyperbolically, with the sign of the autocorrelations being the same as the sign of *d*. In this sense, fractional integration captures long memory dynamics more parsimoniously than non-integrated ARMA processes. In the use of ARFIMA (p, d, q) models, correct specification of p and q is important. According to Robinson(Yr), under-specification of p or q leads to inconsistent estimation of AR and MA coefficients, but also of long memory parameter *d*, as does over-specification of both, due to a loss of identifiability.

RESULTS AND DISCUSSIONS:

Secondary data on monthly price series of groundnut in Andhra Pradesh were collected from Agricultural Market Intelligence Centre (AMIC) Lam, Guntur from January 2002 to December 2023. There are 264 observations, first 252 observations were used for training data set, used for development of model and last 12 observations were used for validation (testing data set) repletion of same information as provided in 1st para of methodology. The actual prices scenario of monthly prices of groundnut in Andhra Pradesh was plotted and depicted in figure.3.

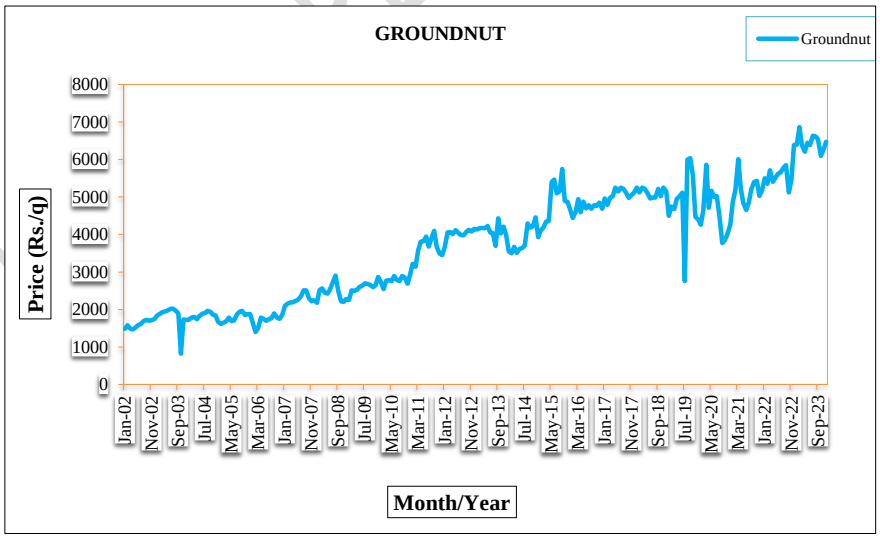


Fig.3. Groundnut Actual Prices Scenario in Andhra Pradesh during 2002-2023

Formatted: Highlight

Formatted: Highlight

Formatted: Superscript

Descriptive Statistics:

Descriptive Statistics were conducted to examine the behaviour of groundnut monthly prices of Andhra Pradesh. The findings were depicted in Table.2, provided valuable insights in to the characteristics of the data. It was observed that the prices of groundnut during the study period had varied from Rs.822/q to Rs.6864/q with an average of Rs.3679.19/q. Standard Deviation was recorded as 1491.03, which indicates that the prices were dispersed highly over the months. It was also revealed that the data was positively skewed and platykurtic in nature. Further lists the summary statistical measures which were self-explanatory. Not clear The price series were also verified for the presence of outliers by Grubb's test. It was confirmed that there were no outliers detected from the Grubb's test during the study period.

BDS (Brock - Dechert- Scheinkman) test for non-linearity: Not discussed in Methodology section

To test the linearity characteristics of the price series BDS test was conducted and results of the test was presented in Table.3. Here, the embedding dimensions were set to 2 and 3 what is the basis?. The probability value for both dimensions was less than 0.05 (at 5 % LOS) which shows that the data under consideration was nonlinear in nature.

Table.2. Descriptive statistics of Groundnut prices of Andhra Pradesh

Statistic	Groundnut
No of observations	264.00
Mean	3679.19
Median	3959.96
Standard Deviation	1491.03
Minimum	822.00
Maximum	6864.00
Skewness	0.06
Kurtosis	-1.24
Outliers detected (Grubbs test)	No

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Table.3. BDS test for non-linearity in Grondnut prices of Andhra Pradesh

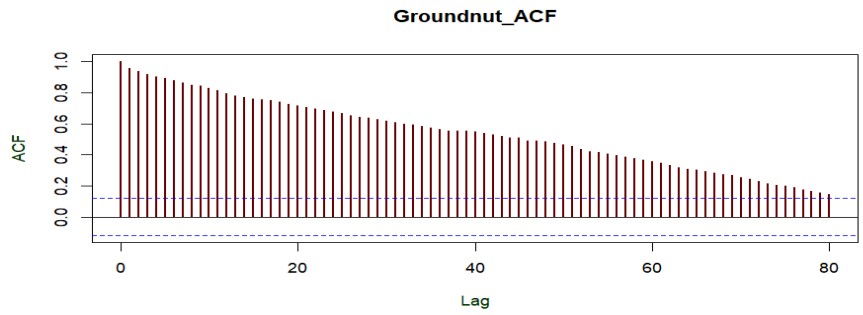
Sample	Dimension	Groundnut	
		Statistics	Probability
eps (1)	*m*=2	255.82	p<0.0001
	m=3	463.02	p<0.0001
eps (2)	*m*=2	246.57	p<0.0001
	m=3	320.99	p<0.0001
eps (3)	*m*=2	86.06	p<0.0001
	m=3	96.80	p<0.0001
eps (4)	*m*=2	50.62	p<0.0001
	m=3	51.18	p<0.0001

Autocorrelation (ACF) and Partial Autocorrelation (PACF) plots for Groundnut prices of Andhra Pradesh:

The Autocorrelation (ACF) and Partial Autocorrelation Function (PACF) plots of groundnut price series were depicted in the below figure.4. The figure shown that the prices were autocorrelated, which was supported by Box-Jung test statistic as the probability value was less than 0.05. It indicates that data under consideration was autocorrelated in nature. Once the price series were autocorrelated, the ARIMA model was built for the series. Further, the groundnut prices contains seasonal component which was confirmed by Q-S test Not discussed in Methodology section

Formatted: Highlight

as the probability value obtained was 0.02 ($p<0.05$) which indicates the presence of seasonality in the dataset. So, SARIMA model was built for the price series.



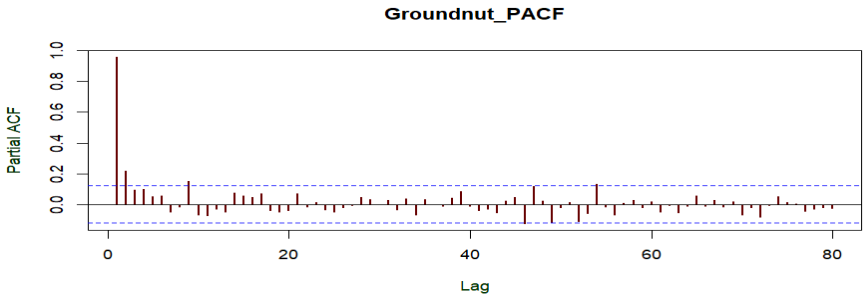


Fig.4. ACF and PACF plots for Groundnut prices of Andhra Pradesh

To develop ARIMA modelling, Augmented Dickey-Fuller test (ADF) Not discussed in Methodology section was used to check the stationarity of the data and the results were presented in Table.4. The groundnut price series shows the probability value was 0.18 ($p>0.05$), confirmed that the data under consideration was non- stationary and became stationary at first difference as probability value was 0.01 ($p<0.05$). Before model estimation to ensure that the data for model consideration was autocorrelated by applying Box- Pierce non-correlation test and it was found significant as the probability value was 0.05 (at 5% LOS) that the data was autocorrelated in nature. All the possible combinations of SARIMA models were developed and best performed SARIMA models were presented in Table.5. Out of the best performed models, the final SARIMA model order i.e., SARIMA $(1,1,3)(0,1,2)_{12}$ model was selected based on least RMSE, MAE, MAPE and AIC values.

Formatted: Highlight

The results of the final selected SARIMA $(1,1,3)(0,1,2)_{12}$ model parameter specification viz., AR, MA, SAR and SMA were presented in Table.6. After determining the SARIMA model order, the model parameters were estimated using maximum likelihood method. After fitting of the model, the diagnostic checking of the residuals by Box- Pierce non-correlation test and it was showed that the residuals were non-autocorrelated in nature as probability value was 0.90 ($p>0.05$). The performance criteria of RMSE, MSE values for both training and testing data sets were illustrated in in Table.8 and Table.9 respectively.

Table. 4. ADF test for stationarity in Groundnut prices of Andhra Pradesh

Groundnut	Data type	ADF statistic	P-value	Decision
	ADF at level	-2.96	0.18	Non-Stationary
	ADF at 1st Difference	-8.04	0.01	Stationary

Table.5. SARIMA models and Error values for Groundnut prices of Andhra Pradesh

SARIMA	RMSE	MAE	MAPE	AIC
SARIMA$(1,1,1)(1,0,1)_{12}$	341.85	198.56	6.04	3685.21
SARIMA$(1,1,1)(1,0,2)_{12}$	340.93	197.23	6.02	3645.20
SARIMA $(1,1,1)(1,0,3)_{12}$	336.78	195.63	5.98	3642.12
SARIMA $(2,1,1)(1,1,1)_{12}$	336.23	195.23	5.97	3643.23
SARIMA$(2,1,1)(2,0,1)_{12}$	336.27	196.67	5.94	3646.45
SARIMA $(2,1,2)(0,1,1)_{12}$	335.98	195.84	5.92	3642.89
SARIMA $(1,1,3)(0,1,2)_{12}$	**335.94**	**194.27**	**5.79**	**3642.02**
SARIMA $(1,1,3)(0,1,1)_{12}$	336.39	195.98	5.84	3643.25
SARIMA $(1,1,2)(1,0,2)_{12}$	338.25	196.49	5.94	3643.89
SARIMA$(2,1,3)(1,1,0)_{12}$	338.25	196.49	5.85	3642.23
SARIMA$(2,1,3)(0,1,1)_{12}$	336.78	194.63	5.98	3642.69
SARIMA$(3,1,3)(0,1,2)_{12}$	336.55	196.67	5.99	3642.96

Table.6.Parameter estimation of SARIMA model for Groundnut prices of Andhra Pradesh

Model	Parameters	Estimation	S.E.	Z-value	Probability	Model fitting	
SARIMA $(1,1,3)(0,1,2)_{12}$	AR1	-0.97	0.08	-21.66	p<0.0001	log likelihood	-1830.1
	MA1	-0.65	0.08	-8.44	p<0.0001		
	MA2	-0.20	0.07	-2.51	p<0.05		
	MA3	-0.15	0.06	-2.11	P<0.05	AIC	3642.02
	SAR1	0.16	0.05	2.75	p<0.0001		
	SMA1	-0.80	0.05	-16.07	p<0.0001		
Box-Pierce test for non-correlations of			Actual series		**χ^2 = 242.15, p<0.0001(<0.05)**		
			Residuals		**χ^2 =0.02, p=0.90 (>0.05)**		

To develop ARFIMA model, first step was to ensure that the data exhibiting long memory by employing Geweke-Porter-Hudak (GPH) test and the differencing parameter (d) value obtained was 0.43 (d<0.5) indicated a strong evidence of existence of long-memory. It was also confirmed by observing the autocorrelations function plot does not fall exponentially but decreases slowly or hyperbolically as shown in figure 3. Before model estimation, to ensure that the data under consideration was autocorrelated by applying Box- Pierce non-correlation

test found significant as the probability value was p< 0.0001 (<0.05) and concluded that the data was autocorrelated in nature. The differenced series was used to fit the ARFIMA (1, 0, 1) model and the parameter specification of AR and MA parameters were presented in Table.7. After model development the diagnostic checking of the residuals by Box- Pierce non-correlation test found non-significant as the probability value was 0.35(>0.05). The modelling and forecasting performance of the training and testing data sets were given in Table.8 and Table.9 respectively.

Table 7. ARFIMA model parameter estimation for Groundnut prices of Andhra Pradesh

Model	Parameters	Estimation	S.E.	Z value	Probability	Model fitting	
ARFIMA (1,0,1)	d	0.43	0.14	3.07	p<0.05	Log-likelihood	-1489.50
	Φ (AR)	0.65	0.12	5.33	p<0.0001		
	Θ (MA)	0.46	0.06	7.74	p<0.0001	AIC	2988.33
Box-Pierce test for non-correlations: χ^2= 0.88 , p=0.35 (>0.05)							

Table 8. Performance metrics of Training set for Groundnut prices of Andhra Pradesh

Training set	SARIMA	ARFIMA
AIC	3642.02	**2988.33**
RMSE	335.94	**186.22**
MSE	112855.68	**34677.89**

Table.9 Performance metrics of Testing set for Groundnut prices of Andhra Pradesh

Testing set	ACTUAL	SARIMA	ARFIMA
Jan-23	**6391**	6045.906	6211.29
Feb-23	**6864**	5945.865	6088.19
Mar-23	**6380**	5847.585	5946.46
Apr-23	**6207**	5873.262	5846.46
May-23	**6443**	5882.055	5799.52
Jun-23	**6384**	5907.222	5884.42
Jul-23	**6634**	5916.509	5871.78
Aug-23	**6613**	5941.196	5965.28
Sep-23	**6540**	5950.949	5962.53
Oct-23	**6099**	5975.184	5964.57
Nov-23	**6255**	5985.375	5978.96
Dec-23	**6470**	6009.186	6098.59
RMSE		**541.01**	**516.27**
MSE		**292691.82**	**266534.71**

CONCLUSION:

No future valu have been forecasted, only the data have been modelled using ARFIMA and SARIMA followed by comparison.

The present study modelled ~~forecasted~~ long memory time series data for the monthly prices of groundnut in Andhra Pradesh using the Autoregressive Fractionally Integrated Moving Average (ARFIMA) model. This model was denoted as ARFIMA (p,d,q), incorporates a fractional differencing operator with *d* as a real number in the range $0<d<1/2$. The ARFIMA (p,d,q) model demonstrated superior performance compared to the SARIMA model, based on performance metrics of both training and testing data sets. This model was particularly effective for short-term predictions of long memory time series, where the forecasting results closely match the actual data.

REFERENCES: Missing ref need to be incorporated and to follow the journal pattern

Box, G.E.P and Jenkins, G.M. 2015. *Time Series Analysis: Forecasting and Control*. Second Edition, Holden Day.

Divianto, D., Maiyastri and Damayanti, S. 2015. Forecasting long memory time series for stock price with Autoregressive Fractoinally Integrated Moving Average. *International Journal of Applied Mathematics and Statistics.* 53(5): 87-95.

Granger WL and Joyeux R. 1980. An Introduction to Long Memory Time Series Models and Fractional Differencing. *Journal of Time Series Analysis* 1:15-29.

Haslett J and Raftery AE. 1989. Space Time Modeling with Long Memory Dependence: Assessing Ireland's Wind Power Resource. *Applied Statistics*. 38(1):1-50.

Hosking JRM. 1981. Fractional Differencing. *Biometrika*. 68:165-176.

Hurst, H.E. 1951. Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers*. 116: 770-799.

Mithiya, D., Mandal, K and Datta, L., 2019. Forecasting of potato prices of hooghly in West Bengal: Time series analysis using SARIMA model. *International Journal of Agricultural Economics*. *4*(3):101-180.

Palma W. *Long-Memory Time Series.* John Wiley and Sons, Inc. New Jersey. 2007.

Paul, R. K., Gurung, B., Paul, A. K and Samanta, S. 2015. Monte Carlo simulation for comparison of different estimators of long memory parameter: An application of ARFIMA model for forecasting commodity price. *Model Assisted Statistics and*

Applications. 10:117–128.

Paul, R. K., Gurung, B., Paul, A. K and Samanta, S. 2016. Long memory in conditional variance. *Journal of the Indian Society of Agricultural Statistics*. *70*(3):243-254.

Rathod, S., Singh, K.N., Paul, R.K., Meher, S.K., Mishra, G.C., Gurung, B., Ray, M and Sinha, K. 2017. An improved ARFIMA Model using Maximum Overlap Discrete Wavelet Transform (MODWT) and ANN for forecasting agricultural commodity price. *Journal of the Indian Society of Agricultural* Statistics. 71(2):103-111.