Long memory Time Series model ARFIMA based modelling of jute prices in the Samsi market of Malda district, West Bengal, India

ABSTRACT The objective of this paper is modeling and forecasting the weekly jute prices of Samsi market in the Malda district of West Bengal in the presence of long memory process. The long memory behavior of series is investigated by the ACF plot and Hurst R/S analysis. A fractionally integrated autoregressive moving-average (ARFIMA) model is fitted using 668 weekly data (January 2009-November 2022). This study shows the efficiencies of the Hurst exponent, GPH, Smoothed periodogram, Local Whittle, and Wavelet methods used to estimate the fractional difference parameter in the ARFIMA model. Furthermore, we compared the forecasting abilities of the ARFIMA and ARIMA models. The results show that long memory is present in the jute price series. The models selected according to each method are ARFIMA (3,0.348,0), ARFIMA (3,0.291,1), ARFIMA (2,0.487,0), ARFIMA (3,0.461,0), ARFIMA (2,0.311,0), and ARIMA (2,1,1) on the basis of minimum AIC and BIC using 534 in-sample data. Finally, the wavelet method based ARFIMA (2,0.311,0) model is found to be the best optimal model in terms of MAE, RMSE, and MAPE criteria using 134 out-of-sample data. A comparative study indicates that the forecasting performance of the ARFIMA model is strongly better than that of the ARIMA model in this regard.

Keywords Jute prices, long memory, fractional difference, forecasting, ARIMA model, ARFIMA model

1. Introduction

Agriculture and allied sectors have emerged as the most resilient sector of the Indian economy in the aftermath of the COVID-19 pandemic, with positive growth rates and rising output over the last two years. Jute (*Corchorus capsularis L.*), also referred to as the "Golden Fiber", is one of the most important commercial cash crops grown in India and is the second most affordable natural fiber after cotton [31]. Jute, originally intended only as a raw material for the packaging industry, has now developed into a versatile raw material for a wide range of applications, including the textile and paper industry, flooring, floor protection, furniture, and handicrafts.

India is the world's largest producer of raw jute, accounting for more than half of global jute production. Among the states, West Bengal ranks first in area and production of jute in the country with a total area of 0.52 million hectares (78.24 percent) and a total production of 7.61 million bales (79.68 percent) with a 2643 kg/hectare productivity during the year 2020–21, followed by Bihar (0.98 million bales production) and Assam (0.80 million bales production) (Directorate of Economics & Statistics, DA&FW).

Agricultural prices play an important role in the whole national economy of India. Commodity price projections and forecasts are critical for market participants making production and marketing strategies as well as for policy makers managing commodity programs and assessing the market impact of national or international events [32]. Time series modelling is one such approach that collects and analyses historical data in order to develop appropriate models that accurately capture the inherent structure and features of the series [28]. In many agricultural datasets, such as daily commodity price data, daily precipitation data, and daily temperature data, it is seen that the long-distance observations are dependent, which means that the dataset has a long memory or long-range dependency characteristic [28]. Traditional models that only describe short-term memory, such as AR(p), MA(q), ARMA(p,q), and ARIMA(p,d,q), cannot adequately explain long-term memory features [1]. A fractional-order signal processing technique, i.e., the Autoregressive Fractional Integral Moving Average Model (ARFIMA) model, has been widely used for decades to predict long-memory time series in divergent domains [14, 23].

Mandelbrot [17], Booth et al. [7], and Helms et al. [20] applied the rescaled range (R/S) method to detect the existence of long memory in the futures prices and provided evidence of long-memory behaviour in financial markets. Lo [15], Chow et al. [12], and Nawrocki et al. [19] conducted similar studies on stock markets. Erfani and Samimi [6] studied the long memory of the stock price index (TSIP) by Hurst exponent and established the ARFIMA and ARIMA models, concluding that the ARFIMA is a much better model in long memory. Paul [26] investigated the ARFIMA model along with its estimation procedure, and the study has revealed that the ARFIMA model could be used successfully for modelling as well as forecasting of daily wholesale price of pigeon pea in different markets. Again, Paul *et al.* [27] reached a similar conclusion after investigating the ARFIMA model and its various estimation procedures in both simulation and with real data. Mohamed [23] used the GPH, smoothed GPH, Local Whittle, R/S analysis and Exact Maximum Likelihood estimation methods to estimate the fractional parameter in the ARFIMA model and found that the Local Whittle method based ARFIMA model is more accurate than others for the total value of traded securities in the Arab Republic of Egypt. Similar works have been done by Liu *et al*. [14], Paul [25], Safitri *et al*. [3], and Monge and Infante [18].

2. Methodology

2.1 Data Description

Jute is the main cash crop of the Malda district and includes 3 markets, viz., Samsi, English Bazar, and Gajol markets. Here, Samsi market, where the majority of jute arrivals take place, is selected for the present study. In order to carry out our analysis, historical jute price (Rs/quintal) data has been taken from the Agricultural Marketing Information Network (https://agmarknet.gov.in) portal from January 2009 to November 2022 (668 weeks) with weekly data. In the present study, statistical analyses have been carried out using the powerful software "RStudio" version 4.2.1 (https://www.rstudio.com) using the 'forecast', 'rugarch' and 'wmtsa' packages.

2.2. Test for stationarity

Testing the stationarity of data is very important in studies where the underlying variables are time-based. A time series is said to be stationary if its underlying generating process relies on constant mean and variance and its autocorrelation function (ACF) is essentially constant over time. If these conditions are not met, the series is not stationary. The stationarity of the data was checked using the Augmented Dickie-Fuller (ADF) test and the Phillips-Perron (PP) unit root test.

2.2.1. Augmented Dickey-Fuller (ADF) test

The Augmented Dickey-Fuller Test is an extension of the Simple Dickey-Fuller Test [4]. Due to the error term, it is unlikely to be white noise. They extended the test with additional lags related to the dependent variable to eliminate autocorrelation issues. The ADF test uses a least-squares estimator to estimate the linear model. In this model, the first difference in the time series at time t is regressed on the level at time $t-1$, augmented by the lag term of the dependent variable. Stationarity is then checked based on the significance of the level term. It can be illustrated as follows:

$$\Delta y_t = \beta T_t + \psi y_{t-1} + \sum_{i=1}^{p} \phi_i \Delta y_{t-i} + \varepsilon_t$$

Where, $\sum_{i=1}^{p} \phi_i \Delta y_{t-i}$ are the corresponding lag terms up to order p; the term T_t is a vector of deterministic terms (constant, trend, etc.); and ε_t is the error term. ψ is the coefficient of interest and testing the null hypothesis (H_0): ψ = 0, corresponds to the test following the unit root process with the alternative hypothesis (H_A): $\psi < 0$, is that the time series is stationary.

2.2.2. Phillips-Perron Unit Root Tests

Phillips and Perron [22] developed the unit root test, which has become popular in financial time series analysis. Under the Phillips-Perron (PP) test, the same null and alternative hypotheses exist as under the Augmented Dickey Fuller (ADF) test. The Phillips-Perron (PP) test is generally preferred over the ADF test due to its robustness. The PP test works well with heteroscedastic errors and does not require lag lengths in the test regression. In particular, if the ADF test uses parametric autoregression to approximate the ARMA structure of the test regression's error, the PP test ignores the serial correlation of the test regression. The test regression model for this test is given by

$$\Delta y_t = \beta T_t + \psi y_{t-1} + u_t$$

Where, Δ is the first difference operator. u_t is trend stationary $I(0)$ and may be heteroskedastic.

2.3. Long Memory process

In time series analysis, the idea of long-term memory, or long-range dependency, is crucial. The long memory feature occurs when the autocovariances for a stationary time series decay very slowly towards zero, like a power function but more slowly than an exponential decay. In this study, the Hurst exponent method [5] has been used to test the long memory.

The time period spanned by the time series of length N is divided into m contiguous sub-periods of length n such that $m * n = N$. The range R_j, standard deviation S_j and their average ratio $(R/S)_j$ are determined for each sub-period j. The range R_j is the difference between the maximum and minimum index of accumulated deviations within the sub-period. This process is repeated until $n = N/2$ by increasing n to the next integer value; in order to avoid bias, at least two sub-periods are required. In the next step, the $(R/S)_n$ is computed by the average of the $(R/S)_j$ values for all the contiguous sub-periods with length n. A least-squares method is then applied to $(R/S)_n$ to obtain an estimate of the slope of the regression line. This slop estimate is a measure of the Hurst Index which is an indicator of market persistence. The Hurst exponent (H) ranges from 0 to 1, when $0.5 < H < 1$, indicating that long memory persistent of a time series is strong.

2.4. Autoregressive Integrated Moving Average (ARIMA) model

The ARIMA model pioneered by Box and Jenkins [8] is a widely recognized statistical forecasting model that predicts future observations of a time series on the basis of some linear function of past values and white noise terms. The stationary ARMA(p, q) process after being differenced d times is denoted by ARIMA(p, d, q):

$$\phi_p(B)(1-B)^d y_t = \mu + \theta_q(B)\ \varepsilon_t$$

Where, μ is the mean of the series, y_t is the observed time series values at time t; $\phi_p(B) = \left(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p\right)$ and $\theta_q(B) = (1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q)$ are the stationary autoregressive operator and moving average operator respectively in which $\phi_i(i = 1,2, \dots \dots, p)$ and θ_j $(j = 1,2, \dots \dots, q)$ are the Autoregressive and Moving Average coefficients respectively; p and q are integers and often referred to as orders of autoregressive and moving average respectively; ε_t is the tth white noise assumed to be independently and B is the backshift operator ($By_t = y_{t-1}$). If the series is not stationary, the first difference $\Delta y_t = y_t - y_{t-1} = y_t - By_t = (1-B)y_t$ or higher-order differences $\Delta^d y_t = (1-B)^d y_t$ will produce a stationary time series. Differencing is done until a plot of the data indicates the series varies about a fixed level, and the graph of ACF either cuts off fairly quickly or dies down fairly quickly.

2.5. Autoregressive Fractionally Integrated Moving Average (ARFIMA) model

The ARIMA model, which has been widely used to forecast social, economic, agricultural, engineering, and financial problems [8], can only capture the short-range dependence (SRD) property, but many practical agricultural data sets, principally commodity prices data, show the typical feature of a long memory process. Therefore, we need a model that has the long memory property. There is a family of models that satisfy this property by generalizing the ARIMA (p, d, q) model. Generalizing, the degree of differencing d is not restricted to integer values, but can take any real value. For reasonable values of d, especially $0 < d < 1/2$, we find that these 'fractionally differenced' processes can be modelled for long-term persistence. The ARFIMA (p, d, q) model was established by Granger and Joyeux [2] and Hosking [11] as a generalized version of the ARIMA (p, d, q) model which is given as follows:

$$\phi_p(B)(1-B)^d y_t = \mu + \theta_q(B)\varepsilon_t \qquad \text{for } 0 < d < 1/2$$

The process of establishing an ARFIMA model consists of three steps. First, we test long-term memory on time series and determine the fractional differencing parameter d. Second, impose a fractional differencing on the series and get the ARMA process. Third, determine the other two orders of the ARFIMA (p, d, q) model, namely p and q. After determining the fractional differencing d, we get the time series of the fractional differentiation as $w_t = (1 - B)^d y_t$; where w_t is the fractional differenced time series and B is the lag operator, and the fractional difference operator defined by the binomial expansion. The partial auto correlation function (PACF) is most useful for identifying the order (p) of an autoregressive model, while the autocorrelation function (ACF) is most useful for identifying the order (q) of a moving average model and can be chosen as [30]: if spikes decay towards zero in the ACF plot and spikes cutoff to zero in the PACF plot, we choose the AR(p) model. Similarly, if Spikes cutoff to zero in the ACF plot and spikes decay towards zero in the PACF plot, we choose the MA(q) model.

2.6. Long memory parameter estimation

We deal with some well-known estimation methods of long memory parameter (d) to estimate $\hat{d}$ which are the R/S Hurst exponent method, the semiparametric methods (i.e., Geweke and Porter–Hudak (GPH), Smoothed periodogram (Sperio), Local Whittle methods) and Wavelet Estimator.

2.6.1. Hurst Rescaled Range Analysis (R/S) method

The British hydrologist Hurst [5] was the first to study the long memory characteristics in a system, defining the Hurst parameter based on rescaled range (R/S) analysis. He used Hurst index (H) to depict the long memory strength of a time series. The two parameters Hurst exponent H and long memory parameter d are closely related through the simple formula [13]:

$$d = H - 0.5.$$

This method is considered as heuristic method of estimating ARFIMA models.

2.6.2. Geweke and Porter-Hudak (GPH) method

Geweke and Porter-Hudak [9] proposed the log-periodogram estimator as one of the earliest semiparametric model estimators. The GPH estimator is based on a regression model in which the spectral density of a process is replaced by the logarithm of its periodogram,

$$\log\left(I_y(\lambda_j)\right) = \alpha + d\, x_j + e_j$$

with a constant term α and an error term e_j. The x_j is explanatory variable which is given by $-\log\left[4\sin^2\left(\frac{\lambda_j}{2}\right)\right]$ with the bandwidth m; where, $m = n^{\upsilon}, 0 < \upsilon < 1$ and the Fourier frequency $\lambda_j = 2\pi j/n\,; j = 1,2, \dots\dots, m; n$ is the number of observations. The $\hat{d}_{GPH}$ is obtained by minimizing the sum of squared residuals of the above regression model with respect to the slope coefficient.

2.6.3. Reisen smoothed periodogram method

Reisen [29] proposed a modified form of the regression method based on a smoothed periodogram function. The smoothed periodogram function with the parzen lag window is used to replace the ordinates in the log periodogram regression to obtain the regression estimator $\hat{d}_{Sperio}$. The bandwidth m in the lag window generator, also referred as the truncation point, is a function of sample size, which is set to $m = n^{\beta}, 0 < \beta < 1$. The smoothness of the estimate is determined by the truncation point m. The bandwidth g(n) used in the regression equation is chosen similarly to the $\hat{d}_{GPH}$ method.

2.6.4. Local Whittle method

The local Whittle estimator is another semi-parametric estimator that is also commonly used to estimate d. Kuensch [24] proposed this estimator, which Robinson [21] modified. The local Whittle estimator $\hat{d}_{LW}$ of fractional differencing parameter d is obtained by maximizing the local Whittle log likelihood at Fourier frequencies close to zero,

$$f(\lambda) \sim G\lambda^{1-2d}\ if\ \lambda \longrightarrow 0$$

Where, G is a constant. This computation includes an additional parameter m (an integer less than $n/2$) that controls the number of frequencies included in the local likelihood.

2.6.5. Wavelet-Ordinary Least Squares Estimator

The maximum overlap discrete wavelet transform (MODWT) is a modified DWT that avoids the subsampling process and provides a higher level of information in the resulting scaling coefficient and wavelet coefficient compared to the DWT [10]. The MODWT calculates the scaling coefficient $V_{\phi}(m,n)$ and wavelet coefficient $W_{\psi}(m,n)$ by applying low-pass and high-pass filters, respectively, to the original dataset. In MODWT, the largest level of decomposition is commonly selected such that $J_0 \leq \log_2(N)$ in order to preclude decomposition at scales longer than the total length of the time series.

The algorithm based on wavelet is used to estimate the long memory parameter of the ARFIMA model [16]. Let y_t be a mean zero fractionally differenced $I(d)$ process with $0 < d < 0.5$. Using the autocovariance function of the $I(d)$ process, Jensen [16] found that as $j \longrightarrow 0$, the wavelet coefficients, $W_{\psi}(m,n)$; where m is scale parameter and n is translations parameter, associated with a mean zero $I(d)$ process with $0 < d < 0.5$ are distributed $N(0,\ \sigma^2 2^{-2md})$; where σ^2 is a finite constant. The wavelet coefficients from a fractionally differenced $I(d)$ process have a variance $R(m)$ that is a function of the scaling parameter m, but is independent of the translation parameter n. The correlation of the wavelet coefficients from an $I(d)$ process decay exponentially over time and scale. Hence, define $R(m)$ to be the variance of wavelet coefficients at scale m, i.e., $R(m) = \sigma^2 2^{-2md}$. Taking the logarithmic transformation of $R(m)$, we obtain the relationship $\ln R(m) = \ln \sigma^2 - d \ln 2^{2m}$. Where ln $R(m)$ is linearly related to ln 2^{-2m} by the fractional differencing parameter, d. Hence, the unknown d of a fractionally differenced series can be estimated by the ordinary least squares estimator $\hat{d}_{Wavelet}$ [27].

2.7. Information criteria and accuracy measures

Model selection has become an important focus in recent years in statistical modelling. Many tools for identifying the "best model" among a set of candidates have been suggested in the literature. In this paper, we used two widely applied criterion Akaike information criterion (AIC) and the Bayesian information criterion (BIC) to select the best model among a set of candidate models:

Akaike's Information criteria (AIC) $= n\,ln(SSE) - n\,ln(n) + 2(k+1)$

Bayesian Information criteria (BIC) $= n\,ln(SSE) - n\,ln(n) + (k+1)\,ln(n)$

Where, n is sample size and k is number of predictor terms so $(k+1)$ is total number of parameters in the model being evaluated. The model with the lowest AIC and BIC values are treated as the best model. Furthermore, the RMSE, MAE and MAPE are used as an accuracy measure to evaluate the performance of the models:

Root Mean Square Error (RMSE) $= \sqrt{\sum_{i=1}^{n} \frac{(y_i - \hat{y}_i)^2}{n}}$

Mean Absolute Error (MAE) $= \frac{\sum_{i=1}^{n} |y_i - \hat{y}_i|}{n}$

Mean Absolute Percentage Error (MAPE) $= \frac{100}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right|$

Where, y_i and $\hat{y}_i$ are the actual value and predicted value of response variable.

Table 1. List of variables used in the study

Variable	Description
y_t, y_i	Response variable (Observed jute price data in Rupees/quintal)
y_{t-i}	Response variable at time $t - i$
$\hat{y}_i$	Predicted value of response variable
T_t	Vector of deterministic terms
w_t	Fractional differenced Response variable
p	Order of an autoregressive model
q	Order of a moving average model
d	Order of differencing
x_j	Explanatory variable
λ_j	Fourier frequency
N	Length of response variable
n	Sample size of response variable
m	Contiguous sub-periods of response variable
k	Number of predictor terms
J_0	Largest level of decomposition
e_j, ε_t	White noise

3. Results and Discussions

3.1. Primary statistical analysis

The weekly series of jute prices for the Samsi market of Malda district is shown in Figure 1. The Samsi market series depicts an up-and-down pattern, with two sharp rises between 2015 and 2018, and another between 2021 and 2022. The descriptive statistics to summarize information from the weekly jute price data are listed in Table 2. As Table 2 shows, the series of jute prices in the Samsi market has a mean of 3655, a standard deviation of 1385.54 and, 37.90% coefficient of variation, suggesting that it has been volatile. In addition, skewness and kurtosis statistics show that the price series is positively skewed and leptokurtic in nature, i.e., the series is not normally distributed.

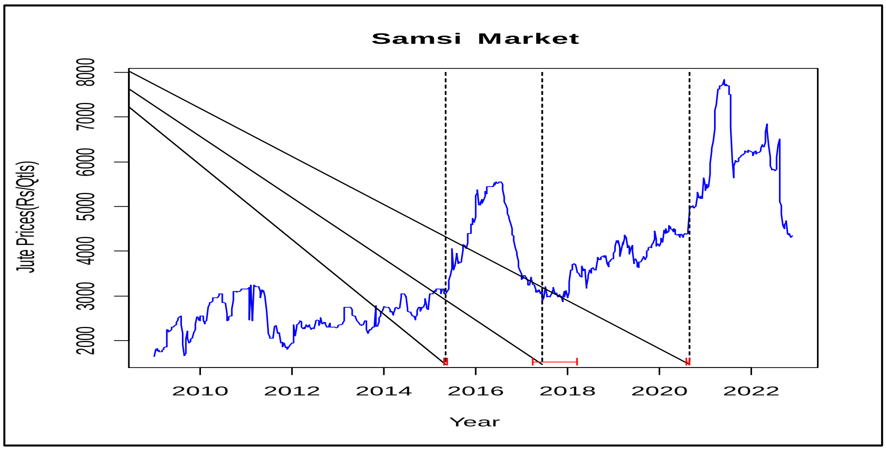

Figure 1. Weekly jute price series for the Samsi market, including all breaks and confidence intervals

Table 2. Descriptive statistics

No. of observations	Min	Max	Mean	Median	SD	CV (%)	Skewness	Kurtosis
668	1650	7836	3655	3186	1385.54	37.90	0.95	3.24

To begin with the implementation of ARIMA and ARFIMA models, the data series are divided into two sets: the training set and the testing set. First, the model is fitted using the training data set, and then it is predicted over the validation period. Out of 668 total observations, the first 534 observations (1st week, 2009 to 6th week, 2020) are used for the training set data, and the last 134 observations are used for model validation purpose (7th week, 2020 to 44th week, 2022).

3.2. Test for stationarity

The first step in applying ARIMA and ARFIMA models is to check whether the time series is stationary or not. In order to test for stationarity, we first conducted Augmented Dickey-Fuller and Phillips-Perron unit root tests on the training dataset of the series, and the results of these tests are given in Table 3. According to the results, the value of the Augmented Dickey-Fuller test statistic has been found to be -2.3166 with lag order 8, whereas the Phillips-Perron test statistic has been found to be -9.5586 with lag order 6. The p-value are 0.444 and 0.577, respectively, which indicate that the time series analyzed is clearly non-stationary. Analysis of time-series observations with unit root (non-stationary) could lead to spurious results. The study, therefore, proceeded to find the stationary series.

Table 3. Testing for stationarity

ADF test			PP test		
Test statistic	Lag order	p-value	Test statistic	Lag order	p-value
-2.3166	8	0.4440	-9.5586	6	0.5770

Notes: p-values < 0.05: Significant at 5% level and p-values > 0.05: non-significant at 5% level.

3.3. Test for long memory and estimation

The presence of long memory in a time series (training set) was confirmed by investigating the autocorrelation function (ACF) plot of the data series and using the Hurst rescaled range (R/S) analysis. The autocorrelation function plot (Figure 2) up to 200 lags shows that the correlations decay very slowly towards zero (they look closer to hyperbolic than exponential), indicating the presence of long memory processes. Accordingly, the presence of long memory is tested as discussed in Section 2.3, and it is found that the R/S Hurst value ($H = 0.848$) is higher than 0.5, which firmly concludes the existence of the long memory characteristic of the jute prices.

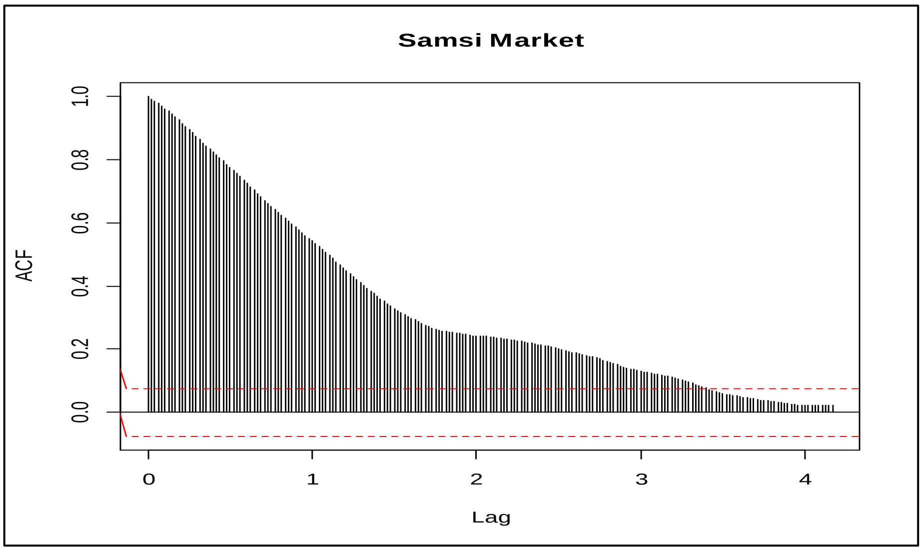

Figure 2. ACF plot of the weekly series of jute prices

The models that consider the long memory property are very sensitive to estimation of the long-memory parameter d (i.e., the fractional differencing parameter), and for this reason, in this study, it has been estimated by using the five estimation procedures: R/S Hurst exponent analysis ($d_{R/S}$), GPH method (d_{GPH}), smoothed periodogram method (d_{Sperio}), local whittle method (d_{LW}), and wavelet-based ordinary least squares estimator ($d_{Wavelet}$). The results are reported in Table 4.

Table 4. Estimate of long memory parameter by different methods

$\hat{d}_{R/S}$	$\hat{d}_{GPH}$	$\hat{d}_{Sperio}$	$\hat{d}_{LW}$	$\hat{d}_{Wavelet}$
0.348	0.291	0.487	0.461	0.311

After determining the fractional differencing parameters d_i, we obtained the fractional and first-order differencing time series shown in Figure 3. Table 5 shows the stationary test results for the fractional difference series (Z_t) and first order difference series (Z_t^1). The p-values of the ADF and PP tests are less than 5%, which reveal the series has become stationary after computing for the fractional difference and first-order difference, which is also confirmed by Figure 3.

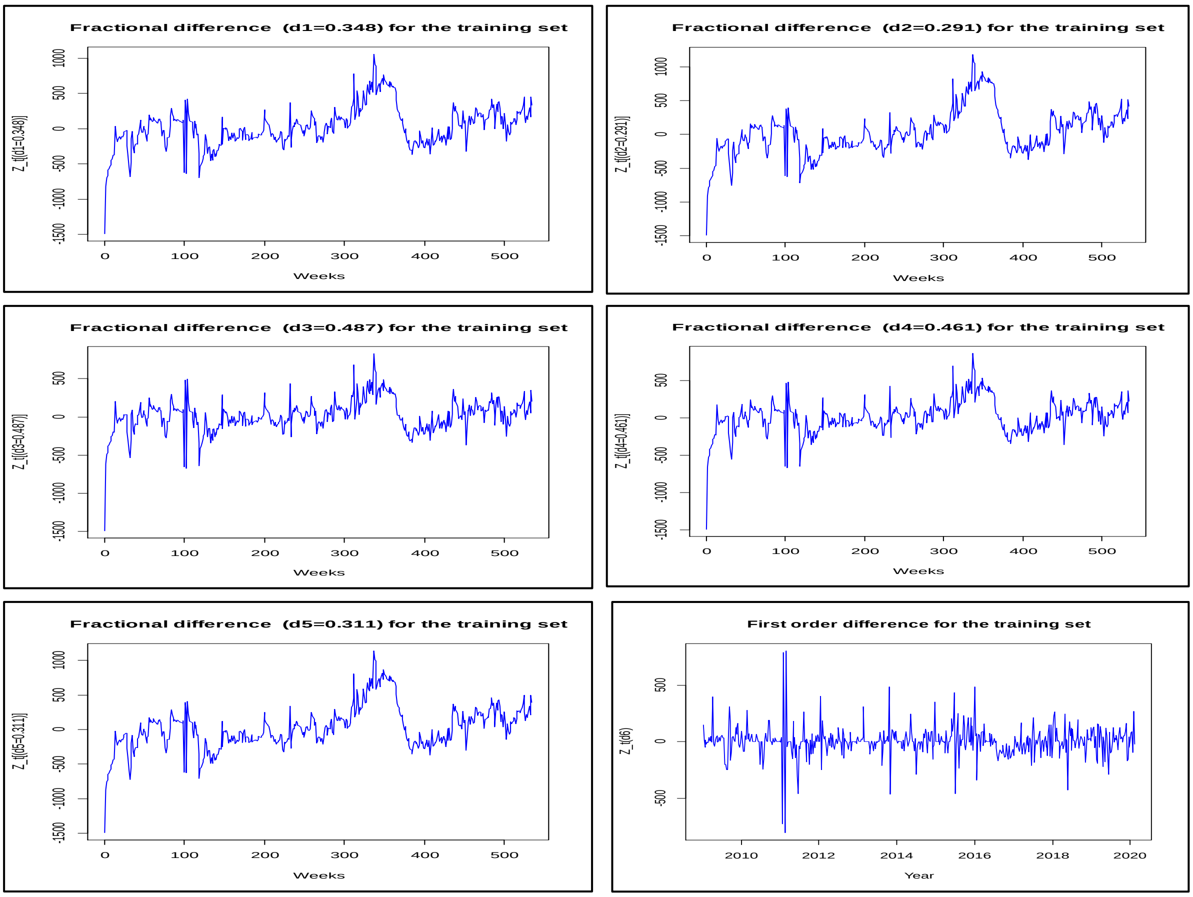

Figure 3. Fractional difference series and first order difference series

Table 5. Stationary test for the fractional difference series and first order difference series

Method	Test for Z_t and (Z_t^1) series	Test statistic	Lag order	p-value
$\hat{d}_{R/S}$ (0.348)	ADF test	-3.50	7	0.042
	PP test	-89.61	6	0.010
$\hat{d}_{GPH}$ (0.291)	ADF test	-3.45	5	0.047
	PP test	-59.64	6	0.010
$\hat{d}_{Sperio}$ (0.487)	ADF test	-3.95	8	0.011
	PP test	-216.47	6	0.010
$\hat{d}_{LW}$ (0.461)	ADF test	-3.83	8	0.017
	PP test	-186.99	6	0.010
$\hat{d}_{wavelet}$ (0.311)	ADF test	-3.56	5	0.037
	PP test	-68.875	6	0.010
$d(1)$	ADF test	-3.56	5	0.037
	PP test	-588.72	6	0.010

Notes: p-values < 0.05: Significant at 5% level and p-values > 0.05: non-significant at 5% level.

3.4. Model Identification

To establish ARIMA(p, d, q) and ARFIMA(p, d, q) models, the values of p, q and d must be determined. In the above section, we have identified the value of d using different methods. Now in this section, we are going to find the optimal value of p and q which are order of autoregressive and moving average terms. We used the training set as in-sample data for the determination of the parameters p and q of the ARIMA and ARFIMA models. First, we computed the values of autocorrelation and partial autocorrelation for fractionally differenced series and first-order differenced time series, as illustrated in Figures 4-9. On computation of ACF and PACF for each estimated difference parameter, it is observed that the decay rate of ACF has improved as compared to the decay of ACF in the actual training set (Figs. 4-8). The orders of non-seasonal parameters p and q are obtained by looking for significant spikes in autocorrelation and partial autocorrelation functions. Whereas Figures 4-8 show that ACF drops slowly, this shows that the AR model will be used as an estimate in this case.

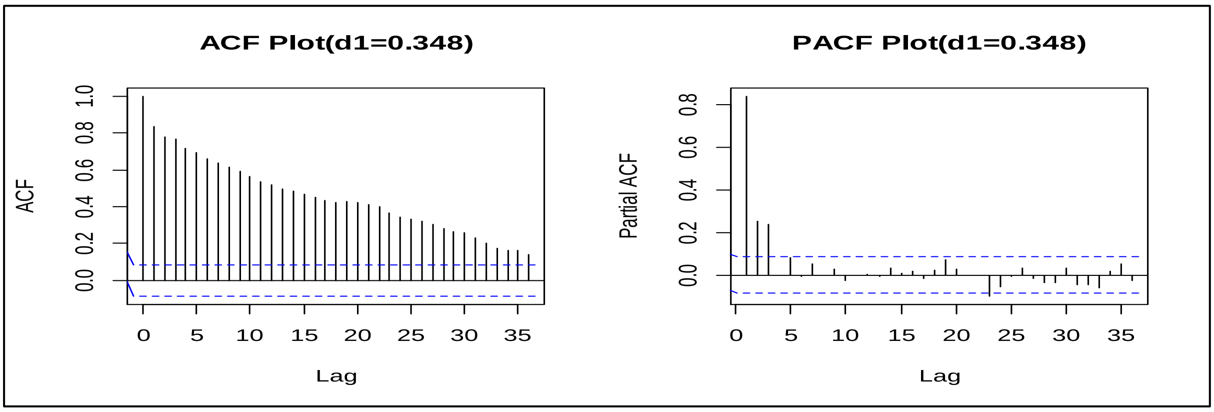

Figure 4. ACF and PACF plot of R/S Hurst method based fractional difference series

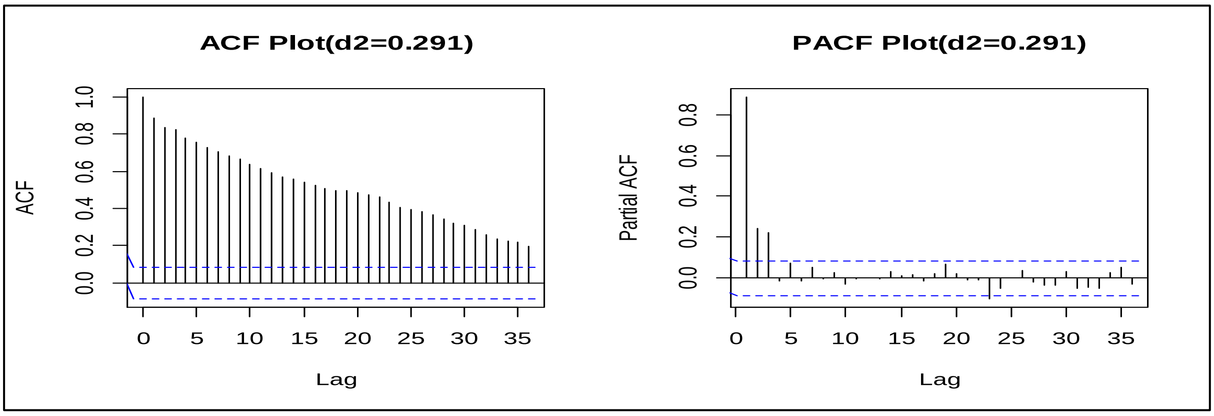

Figure 5. ACF and PACF plot of GPH method based fractional difference series

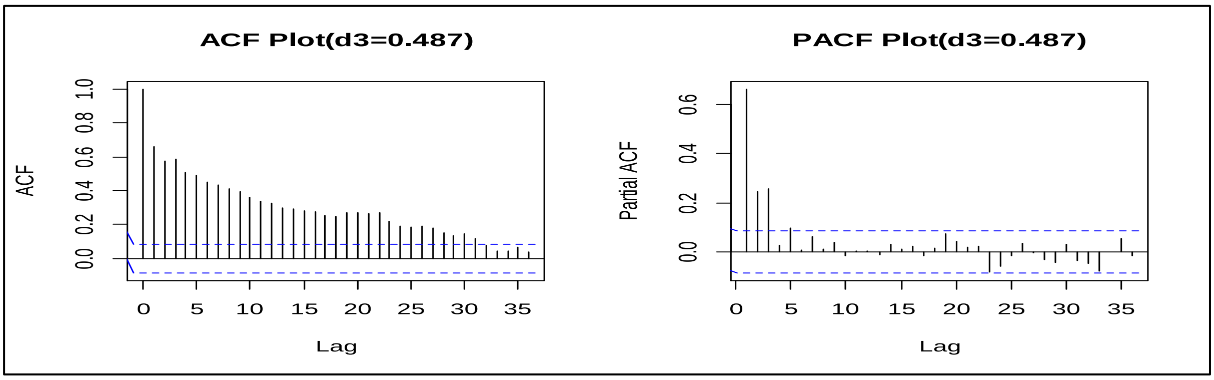

Figure 6. ACF and PACF plot of smoothed periodogram method based fractional difference series

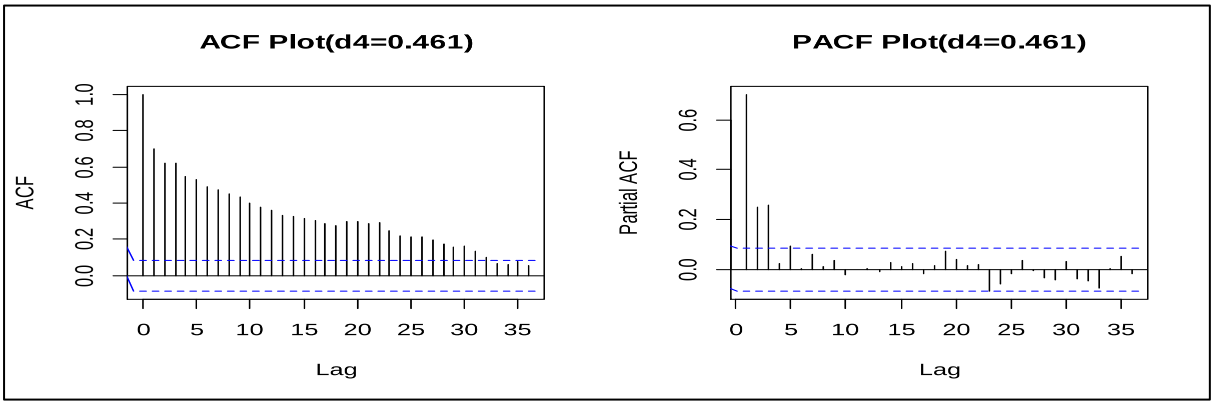

Figure 7. ACF and PACF plot of Local Whittle method based fractional difference series

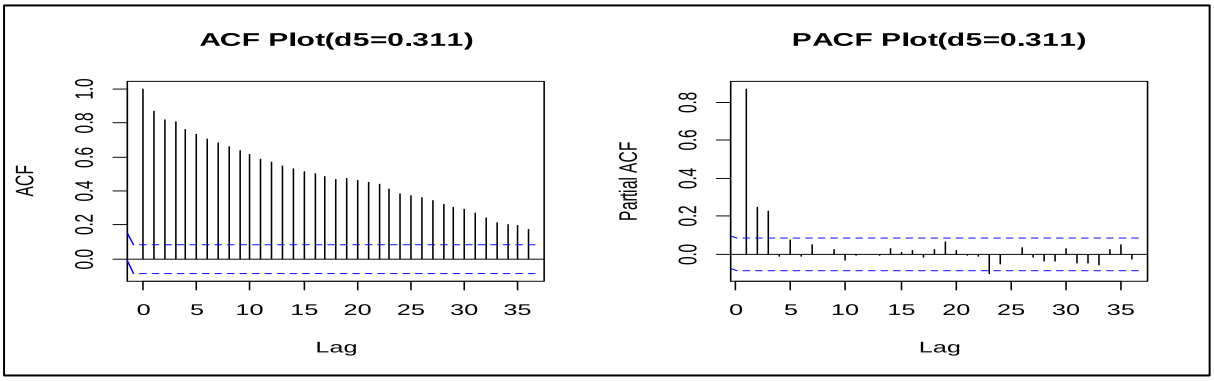

Figure 8. ACF and PACF plot of Wavelet method based fractional difference series

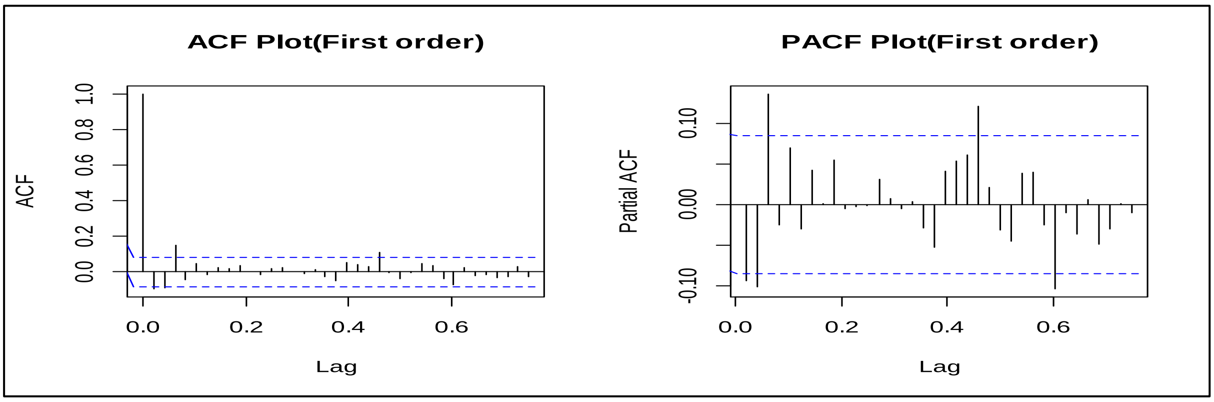

Figure 9. ACF and PACF plot of first order difference series

In the identification stage, we estimated different ARIMA and ARFIMA specifications with different combinations of p (AR terms) and q (MA terms) that we have chosen, as well as compared the fractional difference parameters d_i which are listed in Table 6, and selected the appropriate model from each method as having the minimum values of AIC and BIC. As expected, the AR based models are more accurate than the other models. Thus, the models selected for the training period according to each method are ARFIMA (3,0.348,0), ARFIMA (3,0.291,1), ARFIMA (2,0.487,0), ARFIMA (3,0.461,0), ARFIMA (2,0.311,0), and ARIMA (2,1,1).

Table 6. AIC and BIC values of the ARFIMA and ARIMA models

$\hat{d}_{R/S}$ = 0.348 $ARFIMA\ (p, 0.348, q)$			$\hat{d}_{GPH}$ = 0.291 $ARFIMA\ (p, 0.291, q)$			$\hat{d}_{Sperio}$ = 0.487 $ARFIMA\ (p, 0.487, q)$		
Model	**AIC**	**BIC**	**Model**	**AIC**	**BIC**	**Model**	**AIC**	**BIC**
(1,0.348,0)	6775.283	6792.404	(1,0.291,0)	6763.799	6780.921	(0,0.487,2)	6896.845	6918.247
(1,0.348,1)	6700.135	6721.537	(1,0.291,1)	6699.999	6721.401	(1,0.487,0)	6784.282	6801.404
(2,0.348,0)	6731.282	6752.683	(1,0.291,2)	6694.09	6719.773	**(2,0.487,0)**	**6743.28**	**6764.682**
(2,0.348,1)	6696.381	6722.063	(2,0.291,0)	6723.924	6745.326	(2,0.487,1)	6785.865	6811.547
(3,0.348,0)	**6690.227**	**6715.909**	(2,0.291,1)	6696.447	6722.13	(0,0.487,3)	6816.652	6842.335
			(3,0.291,1)	**6690.21**	**6720.172**			
$\hat{d}_{LW}$ = 0.461 $ARFIMA\ (p, 0.461, q)$			$\hat{d}_{wavelet}$ = 0.311 $ARFIMA\ (p, 0.311, q)$			$ARIMA\ (p, 1, q)$		
Model	**AIC**	**BIC**	**Model**	**AIC**	**BIC**	**Model**	**AIC**	**BIC**
(1,0.461,0)	6785.087	6802.209	(0,0.311,1)	7234.237	7251.359	(1,1,1)	6684.1	6696.93
(1,0.461,1)	6699.315	6720.717	(0,0.311,2)	7119.326	7140.728	**(2,1,1)**	**6669.65**	**6686.76**
(1,0.461,2)	6694.701	6720.383	(1,0.311,0)	6768.178	6785.3	(1,1,2)	6671.72	6688.84
(2,0.461,0)	6742.012	6763.414	**(2,0.311,0)**	**6726.569**	**6747.971**	(2,1,2)	6671.3	6692.69
(3,0.461,0)	**6693.78**	**6719.462**	(0,0.311,3)	6991.438	7017.12	(3,1,2)	6669.89	6695.56
(2,0.461,1)	6696.128	6721.81				(2,1,0)	6680.19	6693.02

3.5. Validation and Diagnostic checking

After appropriate ARFIMA and ARIMA models have been obtained, the next step is to see their ability to forecast the data. The model verification process is concerned with examining residuals obtained from fitted models to see if they contain any systematic pattern that could still be removed to improve the chosen models. This has been done through the Ljung-Box diagnostic test, and it is found that the p-value of the Ljung-Box test is more than 5% except for ARFIMA (2,0.487,0) (Table 7), which means that the model residual meets the assumption of white noise residuals. The evaluation of forecasting performance has been done for the test set as an out of-sample period of 134 observations (i.e., 134 weeks). Table 6 represents the results of the models based on the three different accuracy performance measures: RMSE, MAE, and MAPE.

Table 7. Validation of estimated models

Method	MAE	RMSE	MAPE	Ljung-Box test
$\hat{d}_{R/S} = \mathbf{0.348}$ (0.161)* $\boldsymbol{ARFIMA}\ (\mathbf{3}, \mathbf{0.348}, \mathbf{0})$	114.593	208.671	2.010	0.027 [0.87]
$\hat{d}_{GPH} = \mathbf{0.291}$ (0.094)** $\boldsymbol{ARFIMA}\ (\mathbf{3}, \mathbf{0.291}, \mathbf{1})$	113.204	208.316	1.980	0.068 [0.79]
$\hat{d}_{Sperio} = \mathbf{0.487}$ (0.049)** $\boldsymbol{ARFIMA}\ (\mathbf{2}, \mathbf{0.487}, \mathbf{0})$	110.703	200.741	1.940	5.846 [0.02]
$\hat{d}_{LW} = \mathbf{0.461}$ (0.121)** $\boldsymbol{ARFIMA}\ (\mathbf{3}, \mathbf{0.461}, \mathbf{0})$	113.421	207.922	1.990	0.170 [0.68]
$\hat{d}_{wavelet} = \mathbf{0.311}$ (0.119)** $\boldsymbol{ARFIMA}\ (\mathbf{2}, \mathbf{0.311}, \mathbf{0})$	108.941	201.623	1.920	3.874 [0.05]
$\boldsymbol{ARIMA}\ (\mathbf{2}, \mathbf{1}, \mathbf{1})$	115.564	211.113	2.036	0.004 [0.95]

Notes: * Significant at 5%; ** significant at 1%; the value in the parenthesis is the SE of coefficient: p-values of the Ljung & Box statistics are reported between square brackets. The p-values < 0.05: Significant at 5% level and p-values > 0.05: non-significant at 5% level.

As shown in Table 7, comparing the validation results of all six models, it is observed that the smoothed periodogram method based ARFIMA (2,0.487,0) model presents the lowest RMSE (200.741) value while wavelet method based ARFIMA (2,0.311,0) model presents the lowest MAE (108.941), and MAPE (1.920) values. It can be concluded that the wavelet method based ARFIMA (2,0.311,0) model is the most accurate compared to other models, where predictions indicate that there are narrow variations between the actual and predicted values of jute prices (Figure 10). The strength of the ARIMA model in forecasting jute prices in the Samsi market is considerable, but Table 7 shows that the ARIMA does not perform well. That the most accurate model is conclude to forecast the weekly jute prices in the Samsi market of Malda district is the ARFIMA (2,0.311,0) model, The parameter estimates of ARFIMA models for Samsi market along with their standard errors, t-values and p-values are given in Table 8. The ARFIMA (2,0.311,0) model with parameters as shown in Table 8 is written as:

$$(1 - 0.677B - 0.276B^2)(1 - B)^{0.311} y_t = 1684.069 + \varepsilon_t$$

Table 8. Parameter estimates of ARFIMA (2,0.311,0) model

Parameter	Estimate	Std. Error	t-values	p-values
Constant	1684.069	114.689	14.684	0.000
$\hat{d}_{wavelet}$	0.311	0.119	2.622	0.009
AR(1)	0.677	0.070	9.628	0.000
AR(2)	0.276	0.029	9.468	0.000

Notes: p-values < 0.05: Significant at 5% level and p-values > 0.05: non-significant at 5% level.

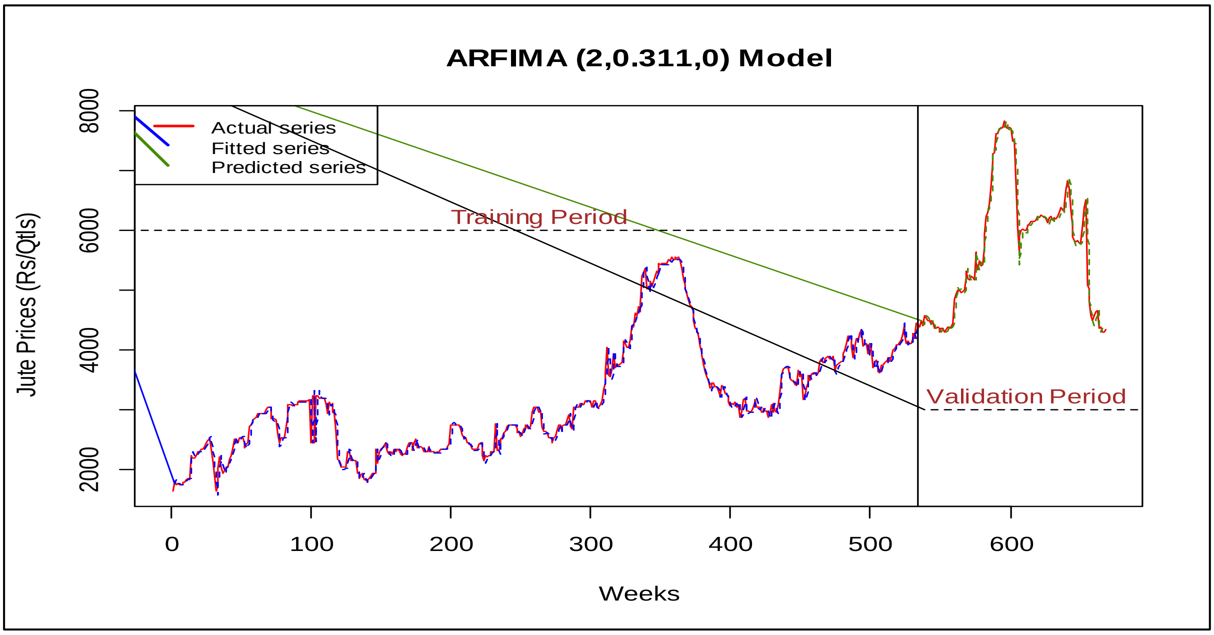

Figure 10. Plot of ARFIMA (2,0.311,0) with training and validation periods

4. Conclusion and implications

The aim of this paper was to introduce an appropriate model for modeling and forecasting the weekly jute prices in the Samsi market of Malda district, for this purpose, an ACF plot and Hurst rescaled range (R/S) analysis are employed to identify the long memory behavior of series. The presence of long memory found in jute price series indicates that it would be better to develop and employ ARFIMA models. We considered the Hurst exponent, semiparametric, and wavelet methods for estimating the fractional difference parameter d. ARIMA and ARFIMA models are fitted to the jute price series, and the models selected according to each method are ARFIMA (3,0.348,0), ARFIMA (3,0.291,1), ARFIMA (2,0.487,0), ARFIMA (3,0.461,0), ARFIMA (2,0.311,0), and ARIMA (2,1,1) on the basis of the minimum AIC and BIC value. A comparative study has been made among the performances of different estimation procedures of the fractional difference parameter d along with forecasting performances between the ARIMA and ARFIMA models, and it is found that the wavelet method based ARFIMA (2,0.311,0) model outperforms the other methods and best fitted ARIMA model in terms of the MAE, RMSE, and MAPE criteria. Hence, it is evident that long memory plays an important and dominant role in describing and modeling the jute prices. The results show that the wavelet method for estimating d is more accurate than the other methods, similar result has been found by Paul et al. [27] when carried out long memory studies in both simulation as well as in real data set. Validation results show that the forecasting performance of the ARFIMA model is strongly better than that of the model that is also concluded by Erfani and Samimi [6], Mohamed [23]. Finally, the ARFIMA (2,0.311,0) model is found to be the best optimal model to forecast the jute prices for the Samsi market.

The model demonstrated good performance in terms of explained variability and predicting power. The comparative study between the ARFIMA model and the ARIMA model revealed that the ARFIMA model is a much better model, and it was concluded that the ARFIMA model could be used successfully for modelling as well as forecasting, especially for data with the long memory property. Again, the information obtained from this study can be utilized for agriculture planning with regard to the jute crop in Malda district.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Author contributions SB as supervisor conceived the research design and the main approaches used. DSG read the first draft, enriched and helped to improve the content of the paper. CRS performed the collection and analysis of the secondary data. SB and CRS wrote the paper.

References

[1] Omekara CO, Okereke OE, Ukaegeu LU. Forecasting liquidity ratio of commercial banks in Nigeria. Microeconomics and Macroeconomics. 2016; 4(1): 28-36.

[2] Granger CWJ, Joyeux R. An introduction to long memory time-series models and fractional differencing. J. Time-Series Anal. 1980; 4: 221-238.

[3] Safitri D, Mustafid, Ispriyanti D, Sugito H. Gold price modeling in Indonesia using ARFIMA method. Journal of Physics: Conference Series. 2019; 1217: 1-11

[4] Dickey, Fuller W. Distribution of the estimators for autoregressive time series with a unit root. Journal of the American Statistical Association. 1979; 74: 427–431.
[5] Hurst E. Long-Term Storage Capacity of Reservoirs. Transactions of the American Society of Civil Engineers. (1951); 116(1): 770-799.
[6] Erfani, Samimi AJ. Long Memory Forecasting of Stock Price Index Using a Fractionally Differenced Arma Model. Journal of Applied Sciences Research. 2009; 5(10): 1721-1731.
[7] Booth G, Kaen FR, Koveos PE. R/S analysis of foreign exchange rates under two international monetary regimes. Journal of Monetary Economics. 1982; 10(3): 407-415.
[8] Box GE, Jenkins G M. Time Series analysis: Forecasting and Control. Holden Day, San Francisco. 1976.
[9] Geweke, Porter HS, The estimation and application of long memory time series models. Journal of time series analysis. 1983; 4(4):221–238.
[10] Ghaemi, Rezaie-Balf M, Adamowski J, Kisi O, Quilty J. On the applicability of maximum overlap discrete wavelet transform integrated with MARS and M5 model tree for monthly pan evaporation prediction. Agricultural and Forest Meteorology. 2019; 278: 1-18.
[11] Hosking JRM., Fractional differencing. Biometrika. 1981; 68: 165–176.
[12] Chow KV, Denning KC, Ferris S, Noronha G. Long-term and short-term price memory in the stock market. Economics Letters. 1995; 49: 287-295.
[13] Shang L. A Comparison of Hurst Exponent Estimators in Long-range Dependent Curve Time Series. Journal of Time Series Econometrics. 2020; 12: 1-39.
[14] Liu, Chen YQ, Zhang Xi. An Evaluation of ARFIMA (Autoregressive Fractional Integral Moving Average) Programs†. Axioms. 2017; 6: 1-16.
[15] Lo. Long-term memory in stock market prices. Econometrica.1991; 59: 1279-1313.
[16] Jensen MJ. Using wavelets to obtain a consistent ordinary least squares estimator of the long-memory parameter. Journal of Forecasting. 1999; 18: 17–32.
[17] Mandelbrot. Statistical methodology for nonperiodic cycles from covariance to R/S analysis. Annals of Economic and Social Measurement. 1972; 1(3): 259–290.
[18] Monge, Infante J. A Fractional ARIMA (ARFIMA) Model in the Analysis of Historical Crude Oil Prices. Energy RESEARCH LETTERS 2022; 3: 1-3.
[19] Nawrocki. R/S Analysis and Long Term Dependence in Stock Market Indices. Managerial Finance. 1995; 21:78-91.
[20] Helms P, Kaen FR, Rosenman RE, Memory in commodity futures contracts. The Journal of Futures Markets. 1984; 4(4): 559-567.
[21] Robinson PM. Gaussian Semiparametric Estimation of Long Range Dependence. The Annals of Statistics. 1995; 23(5): 1630-1661.
[22] Phillips PCB, Perron P. Testing for unit roots in time series regression. Biometrika. 1988; 75: 335-346.
[23] Mohamed RAH. Using Arfima Models in Forecasting the total value of Traded Securities on the Arab Republic of Egypt. IJRRAS. 2016; 27(1): 26-34.
[24] Kuensch R. Statistical aspects of self-similar processes. Paper presented at First World Congress of the Bernoulli Society, Tashkent, Uzbekistan, September 8–14. 1987; 67–74.
[25] Paul R. Modelling long memory in maximum and minimum temperature series in India. Mausam. 2017; 68: 317-326.

[26] Paul RK. Forecasting wholesale price of pigeon pea using long memory time-series models. Agric. Econ. Res. Rev. 2014; 27(2): 167-176.
[27] Paul RK, Samanta S, Gurung B. Monte Carlo simulation for comparison of different estimators of long memory parameter: An application of ARFIMA model for forecasting commodity price. Model Assist. Statist. Appl. 2015; 10(2): 117-128.
[28] Paul RK, Sarkar S, Mitra D, Panwar S, Paul AK, Bhar LM. Wavelets based estimation of trend in sub-divisional rainfall in India. Mausam. 2020; 71(1): 69-78.
[29] Reisen VA. Estimation of the fractional difference parameter in the ARFIMA(p,d,q) model using the smoothed periodogram. Journal Time Series Analysis. 1994; 15(1): 335–350.
[30] Chen Z, Peng D, Zheng F, Zheng Y, Chen X. Research on forecast method of dynamic angle measurement. In Seventh International Symposium on Precision Engineering Measurements and Instrumentation 2011.
[31] Rajpoot S, Saxena S, Sehgal S, Dubey SK, Choudhary K, Gavli AS, Verma A, Ray SS. Jute crop production estimation in major states of India: a comparative study of last 6 years' fasal and DES estimates. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. 2019; 42: 129-136.
[32] Sharma H. Applicability of ARIMA models in wholesale wheat market of Rajasthan: An investigation. Economic Affairs. 2015; 60: 687-691.