Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF
AGRICULTURAL STATISTICS 72(1) 2018 49–60

Forecasting Time Series Allowing for Long Memory and Structural Break

Dipankar Mitra, Ranjit Kumar Paul, A.K. Paul and L.M. Bhar
ICAR-Indian Agricultural Statistics Research Institute, New Delhi

Received 24 June 2017; Revised 10 October 2017; Accepted 23 February 2018

SUMMARY

Long range dependency or long run persistence is a common issue in agricultural price data. These type of phenomena in time-series process can be modeled with the help of Autoregressive fractionally integrated moving average (ARFIMA) model. The feature often arises when working with real time-series data which might exhibit long memory is the possible presence of structural break in mean or in long memory parameter. In this study, the statistical tests for testing presence of long memory and structural break have been discussed. The joint test (Gil-Alana, 2002) for testing degree of fractional integration and possible presence of structural break at known time epoch is also discussed. Two stage forecasting (TSF) algorithm by Papailias and Dias (2015)is used to obtain the forecasts of a long memory process in presence of structural break. In the present investigation, TSF approach is considered for forecasting daily wholesale price of pigeon pea in Bhopal market of Madhya Pradesh, India. A comparative study of predictive performances has also been carried out among the existing forecasting methodology of a long memory time-series subjected to structural break viz. AR approximation method and AR truncation method. It is concluded that TSF approach outperforms the other methods as far as forecasting is concerned for the series under consideration.

Keywords: ARFIMA, Long memory, Pigeon pea, Structural break, Two stage forecasting.

1. INTRODUCTION

Time-series analysis is an important statistical technique used as a basis for manual and automatic planning in many application domains (Gooijer and Hyndman, 2006). Forecasting or prediction is a challenging area in scientific research. Price forecast are largely done by using time-series approaches. Much effort has been devoted over last few decades to develop and improve several time-series forecasting models. Most of the research works in time-series analysis assume that the observation separated by long time lags are independent of each other or nearly so. But in many agricultural data, particularly in daily commodity price data it is seen that the distant observations are dependent that means the data set have characteristic feature of long memory or long range dependency. Long range dependence or long memory of a time-series is a phenomenon that states statistical dependence between the observations separated at distant lags. A time-series process is called as long memory process if the autocorrelation function decays very slowly towards zero unlike the exponential decay in usual Autoregressive integrated moving average (ARIMA) model. The autocorrelation function of a long memory process exhibits persistency structure which is neither an I(1) process nor an I(0) process. The time-series process having long memory in the mean equation can be modeled by using Autoregressive fractionally integrated moving average (ARFIMA) model by allowing non-integer or fractional differencing parameter (Granger and Joyeux, 1980). Paul (2014) and Paul *et al.* (2014a, 2015) have applied ARFIMA model for forecasting of agricultural commodity prices. The good performance of model has demonstrated in terms of variability explanation and prediction performance. A time-series process may sometimes exhibits long memory due to presence of structural break in mean or a shift in long memory parameter. Literally, structural change

Corresponding author: Ranjit Kumar Paul
E-mail address: ranjit.paul@icar.gov.in

can be described as fundamental shift in the structure of the series under consideration which may be due to economic growth, policy decisions, revolution etc. Ignoring the presence of breaks can lead to seriously biased estimates and poor forecasts. Hence detection of structural break before forecasting is very important to get more accurate forecasts. In this paper, the joint test (Gil-Alana, 2002) for testing simultaneous presence of fractional integration and structural break at known period of time is studied along with its application to the agricultural commodity price data. In literatures, there are several existing methodologies to obtain forecasts of a long memory process subjected to structural break. Among these, the most well-known approaches are Autoregressive (AR) approximation method (Wang *et al.*,2013), a truncated version of the infinite autoregressive representation of the model (Peiris, 1987) and Two stage forecasting (TSF) algorithm by Papailias and Dias (2015). The TSF approach has been successfully applied for modelling maximum temperature series in India (Paul and Anjoy, 2017). But in that study, joint test has not been applied to identify simultaneous presence of long memory and structural break. In this present study the TSF approach to obtain the predictions of a long memory process having structural break in mean or long memory parameter along with the joint test for testing presence of both the property has been discussed in detail. To demonstrate the predictive ability of TSF approach of forecasting pigeon pea wholesale price data of Bhopal market, Madhya Pradesh, India is considered and comparison has been made with respect to simple ARFIMA (without considering structural break), AR approximation method (using AIC criteria) and AR truncation method.

2. LONG MEMORY PROCESS

Most of the research works in time-series analysis assume that the observation separated by long time span are independent of each other or nearly so. But in many practical situations it is seen that many empirical economic series show that the distant observations are dependent, though the correlation is small but not negligible. Diebold and Inoue (2001) studied first the joint analysis of long memory and structural break and proved the way of misinterpretation of structural break as long memory. The statistical dependence of any time-series data is generally measured by plotting the ACF of the dataset. Let X_t; (t=1,2,...) be a stationary time-series process and the autocorrelation function of the time-series with a time lag of k is given as

$$\rho_k = cov(x_t, x_{t-1})/\mathrm{var}(x_t) \tag{1}$$

The series X_t; (t=1,2,...) is said to have short memory if the autocorrelation coefficient at lag k approaches to zero as k tends to infinity, i.e. $\lim_{k\to\infty} \rho_k = 0$.The autocorrelation functions of most of stationary and invertible (ARMA) time-series process decay very rapidly at an exponential rate, so that $\rho_k \approx |m|k$, where $|m|<1$.

For long memory processes, decaying of autocorrelations functions occur at much slower rate (hyperbolic rate) which is consistent with $\rho_k \approx Ck^{2d-1}$, as k increases indefinitely, where C is a constant and d is the long memory parameter. The autocorrelation function of a long memory process exhibits persistency structure which is neither consistent with an I(1) process nor an I(0) process.

In other words a short memory process is defined as

$$\sum_{k=0}^{\infty} |\rho_k| < \infty \tag{2}$$

And a long memory process is defined as

$$\sum_{k=0}^{\infty} |\rho_k| = \infty \tag{3}$$

where ρ_k is the coefficient of autocorrelation with lag of k.

In frequency domain a long memory is defined in terms of rates of explosion of low frequency spectra as

$$f_x(\omega) = \mathrm{g}\,\omega^{-2\mathrm{d}} \text{ as } \omega \to 0^+ \tag{4}$$

In general the low-frequency spectral definition of long memory is simply as

$$f_x(\omega) = \infty \text{ as } \omega \to 0^+ \tag{5}$$

Berran (1995a) have discussed some properties of a stationary long memory process.

3. ARFIMA MODEL

The Autoregressive fractionally integrated moving-average (ARFIMA) model (Granger and Joyeux, 1980) is used for modeling time-series in presence of long memory. Fractional integration is a generalization of integer integration. Here, time-series is usually presumed to be integrated of order zero or one. For example, an autoregressive moving-average process integrated of order d [denoted ARFIMA (p, d, q)] can be represented as

$$\phi(B)y_t == (1-B)^{-d}\,\theta(B)u_t \qquad (6)$$

where, μ_t is an independently and identically distributed (*i.i.d.*) random variable having zero mean and constant variance, B denotes the lag operator; $\phi(B)$ and $\theta(B)$ denote finite AR and MA polynomials in the lag operator having roots outside the unit circle. For $d = 0$, the process is stationary, for $-0.5 < d < 0.5$ the process y_t is stationary and invertible, for $d \in \left(0, \frac{1}{2}\right)$ the process is said to have long memory. For any value of d we have

$$(1-B)^d = 1 - dB + \frac{B^2 d(d-1)}{2} + \ldots + \sum_{j=0}^{\infty}\binom{d}{j}(-1)^j B^j \qquad (7)$$

with binomial coefficients

$$\binom{d}{j} = \frac{d!}{j!(d-j)!} = \frac{r(d+1)}{r(j+1)\Gamma(d-j+1)} \qquad (8)$$

where $\Gamma(.)$ represents the gamma function.

4. TESTING FOR PRESENCE OF LONG MEMORY AND STRUCTURAL BREAK

Following approaches are generally used for estimating long memory parameter

(i) Parametric method - Maximum Likelihood method of estimation (MLE)(Berran, 1995b).

(ii) Semi-parametric method – Whittle (Robinson, 1994), GPH (Geweke and Porter-Hudak, 1983) etc.

(iii) Heuristic method - R/S statistic (*Hurst, 1951)*, ACF plot, variance plot, log-log correlogram and least square regression in spectral domain.

(iv) Nonparametric method – Wavelet methodology (Jensen, 1999).

Literally, structural change can be described as fundamental shift in the structure of the series under consideration which may be due to economic growth, policy decisions, revolution etc. If the presence of breaks is completely ignored, forecasts may become poor. Therefore detection of structural break is prime importance prior to the analysis of time-series. For detection of structural break cumulative sum (CUSUM)(Ploberge and Krame, 1992), Chow test etc. can be used. A detailed description of these tests along their application for detection of structural break in mean temperature in different agro-climatic zones of India is found in Paul *et al.* (2014b).

5. JOINT TEST OF FRACTIONAL INTEGRATION AND STRUCTURAL BREAK AT KNOWN TIME PERIOD

Gil-Alana (2002) has developed a joint test for simultaneous testing of fractional integration and a need of a structural break at a known period of time. The joint test is a modified version of the test proposed by Robinson (1995) for testing a wide range of null hypothesis, mainly for non stationary null hypothesis particularly in the form of unit root.

Robinson (1995) considered the following model

$$y_t = \beta' z_t + x_t,\; t = 1, 2, \ldots \qquad (9)$$

where y_t is the observed time-series process; β is a (p×1) vector of unknown parameters; z_t is a (p×1) known vector of deterministic regressors.

He considered testing of the following hypothesis

$$H_0\colon d = d_0 \qquad (10)$$

Under the null hypothesis the least squares estimate of β and residuals are given as

$$\beta = \left(\sum_{t=1}^{T} w_t w_t'\right)^{-1} \sum_{t=1}^{T} w_t (1-L)^{d_0} y_t;\; w_t = (1-L)^{d_0} z_t \qquad (11)$$

$$\hat{u}_t = w_t(1-L)^{d_0} y_t - \hat{\beta}' w_t$$

The Lagrange multiplier (LM) test statistic for testing μ_0 is given as

$$\hat{r} = \left(\frac{T}{\hat{A}}\right)^{1/2} \frac{\hat{a}}{\hat{a}^2} \qquad (12)$$

$$\hat{a} = \frac{-2\pi}{T}\sum_{j=1}^{T-1}\psi(\lambda_j)\,g(\lambda_j;\hat{\tau})^{-1} I(\lambda_j), \quad \hat{\sigma}^2 = \sigma^2(\hat{\tau}) = \frac{2\pi}{T}\sum_{j=1}^{T-1} g(\lambda_j;\hat{\tau})^{-1} I(\lambda_j)$$

$$\psi(\lambda_j) = log\left|2\, sin\frac{\lambda_j}{2}\right|$$

$$\hat{A} = \frac{2}{T}\left(\sum_{j=1}^{T-1}\psi(\lambda_j)^2 - \sum_{j=1}^{T-1}\psi(\lambda_j)\hat{\varepsilon}(\lambda_j)' \times \left(\sum_{j=1}^{T-1}\hat{\varepsilon}(\lambda_j)\hat{\varepsilon}(\lambda_j)'\right)^{-1} \times \sum_{j=1}^{T-1}\hat{\varepsilon}(\lambda_j)\psi(\lambda_j)\right)$$

$$\hat{\varepsilon}(\lambda_j) = \frac{\partial}{\partial\tau}\log g\,(\lambda_j;\hat{\tau})$$

where $I(\lambda_i)$ is the periodogram of $\hat{u}_t$, with Fourier frequencies $\lambda_j = 2\pi_j/T$, and g is a known function derived from the spectral density of $u_t : f(\lambda;\tau) = (2\pi/\sigma^2)$, evaluated at $\hat{\tau} = \arg\min_{\tau\in'}\sigma^2(r)$.

Robinson (1995) showed that under the null hypothesis H_0 the test statistic $\hat{r}$ is asymptotically distributed as

$$\hat{r} \xrightarrow{d} N(0,1) \text{ as } T \to \infty \tag{13}$$

H_0 is rejected against the alternative hypothesis $H_1 : d > d$ ($d < d_0$) if $\hat{r} > Z_\alpha$ ($\hat{r} > -Z_\alpha$), where $-Z_\alpha$ is the critical value of a standard normal variate at 100α% level.

To test the long memory parameter in presence of structural break at a point Gil-Alana (2002) considers the following model

$$y_t = \beta SB(T_b)_t + x_t;\ (1-L)^d x_t = u_t, t = 1,2,... \tag{14}$$

where T_b denotes the known break point and $SB(T_b)_t = \begin{cases}1, & t \ge T_b\\ 0, & t \le T_b\end{cases}$.

The test based on testing the following hypothesis

$$H_0 : d = d_0 \text{ and } \beta = 0 \tag{15}$$

against the alternative

$$H_1 : d \ne d_0 \text{ or } \beta \ne 0$$

The test statistic is given as

$$\hat{S} = \hat{r}^2 + \sum_{t-1}^{T}\tilde{u}_t w_t^* \times \left(\sum_{t-1}^{T} w_t^{*2}\right) \times \sum_{t=1}^{T} w_t^* \tilde{u}_t \tag{16}$$

where $w_t^* = (1-L)^{d_0} SB(T_b)_t$ and $\tilde{u}_t = (1-L)^{d_0} y_t$; and $\hat{r}$ is as defined earlier.

H_0 is rejected against the alternative hypothesis H_1 if $\hat{S} > \chi^2_{2,\alpha}$, where $\chi^2_{2,\alpha}$ is the upper 100α% limit of a chi-square distribution with 2 degrees of freedom.

6. FORECASTING LONG MEMORY PROCESS IN PRESENCE OF STRUCTURAL BREAK

Sometimes real time-series data may exhibit long memory pattern due to possible presence of structural change. This phenomenon is commonly called as spurious long memory. There are mainly two situations –

(a) The structure of the time-series process might be mistaken as long memory due to presence of structural break and

(b) The co-existence of long memory and structural break in the given data set. There are different approaches to obtain the forecasts of a long memory process subjected to structural break (Ngene and Lambert, 2015). Only three methods are discussed below.

6.1 AR truncation method

An ARFIMA (p, d, q) process subject to a mean shift and a change in the long memory parameter can be well approximated by an autoregressive (AR) modelof infinite order(Peiris,1987)and the AR(∞) representation of $\{y_t\}$, $t = 1,2,...$ is given as

$$y_t = \sum_{j=1}^{\infty}\beta_j x_{t-j} + \epsilon_t \tag{17}$$

where t_t is a white noise process with mean zero and constant variance σ^2 and $\beta_i = \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(d)}$ with $\Gamma(.)$ being gamma function.

Using information available information up to time T and equation (17) the s-step ahead forecast can be obtained as

$$\hat{x}_{T+s} = \sum_{j=1}^{\infty}\hat{\beta}_j x_{T+S-j} \tag{18}$$

Using a truncated version of equation (18) the s-step ahead forecast is given by

$$\hat{x}_{T+s} = \sum_{j=1}^{\hat{p}}\hat{\beta}_j x_{T+S-j} \tag{19}$$

The optimal lag order $\hat{p}$ can be chosen as $\hat{p} = [(\ln T)^v]$, $\Gamma(.)$ denotes floor function and T is total number of data points.

6.2 AR approximation method

Wang *et al.* (2013) suggested to use two selection criteria namely AIC or Mallows' C_p to choose the order of the approximation of anAR(∞) model. This method avoids the issue of estimation inaccuracy of the long memory parameter and the issue of spurious breaks in finite sample.

6.3 Two-Stage Forecast (TSF) Algorithm

Papailias and Dias (2015) have proposed a forecasting methodology known as two-stage forecast (TSF) to obtain the forecasts of a long memory process subjected to structural break in the mean or in the long memory parameter. These forecasts are more accurate and robust forecasts. The TSF algorithm avoids the loss of information which might be a case in the truncated version of an infinite AR model. It is an intuitive, simple and commonly used methodology to obtain the forecasts. The methodology begins with the estimation of long memory parameter. In the next step the fractional differencing operator is applied to the underlying process in order to obtain the weakly dependent series. The multi-step ahead forecasts of the later series are computed and finally obtain the corresponding forecasts of the original series by employing the fractional cumulation operator. Therefore, the TSF algorithm has mainly two steps. In first step forecasts of the underlying weakly dependent time-series are computed and in the second step corresponding forecasts of the original series are computed by applying the fractional cumulation operator. TSF procedure can be envisaged in the following steps:

(i) The methodology starts with estimation of long memory parameter of the long memory process (x_t, (t=1,2,...) using any consistent estimator and obtain $\hat{d}$.

(ii) In the next step the differencing operator is applied to the original long memory process and obtain weakly dependent process y_t as

$$y_t = (1-L)^{\hat{d}} x_t,\ t = 1,2,... \tag{20}$$

(iii) Fit an AR(P) model and compute the one step ahead forecast

$$\hat{y}_{T+1} = \sum_{i=1}^{\hat{p}} \hat{\pi}_i y_{T+1-i} \tag{21}$$

The $\hat{p}$ can be chosen based on (i) minimum AIC value or (ii) sample size as $\hat{p} = [(\ln T)^2]$, (Γ.) denotes integer part).

(iv) Write the weakly dependent series incorporating the one step ahead forecast, such that

$$\tilde{y}_T = (y_1, ..., y_T, y_{T+1}) \tag{22}$$

(v) Apply fractional cumulation operator to the $\tilde{y}_T$ series to go back to the original series and obtain the one step ahead forecast for original long memory series

$$\tilde{x}_t = (1-L)^{-\hat{d}} \tilde{y}_{T'},\ t = 1,2,...,T,T+1 \tag{23}$$

The one-step ahead forecast is given as $\tilde{x}_{t+1}$.

(vi) Iterating over the previous steps recursively provide the h step ahead forecast for the underlying process $\tilde{x}_{T+h}$.

7. EMPIRICAL ILLUSTRATION

7.1 Data set

For the present investigation, daily pigeon pea wholesale price data is collected from AGMARKNET (www.agmarknet.nic.in) website for the period 1stJanuary, 2010 to 3rdMay, 2017 considering Bhopal market of Madhya Pradesh, India. The data is composed of maximum price, minimum price and model price of pigeon pea. The dataset is partitioned into two sets namely training set and test or holdout dataset. The training dataset comprises 1078 observations which are used for model estimating purpose. The test dataset includes last 50observations which are used for forecast evaluation and model validation.

The first and foremost step in time-series analysis is to plot the data and visualize the presence of several time-series components. The time plots of the series under consideration are given in Fig. 1. A perusal of figure 1 indicates that maximum, minimum and modal prices of pigeon pea in Bhopal market has increased suddenly in February 2015 and it stayed high up to end of 2016. It indicates that the variability in price increases during this period in comparison to the previous period. One of the reasons for almost constant price before 2015 is due to very less differences in production and consumption of pigeon

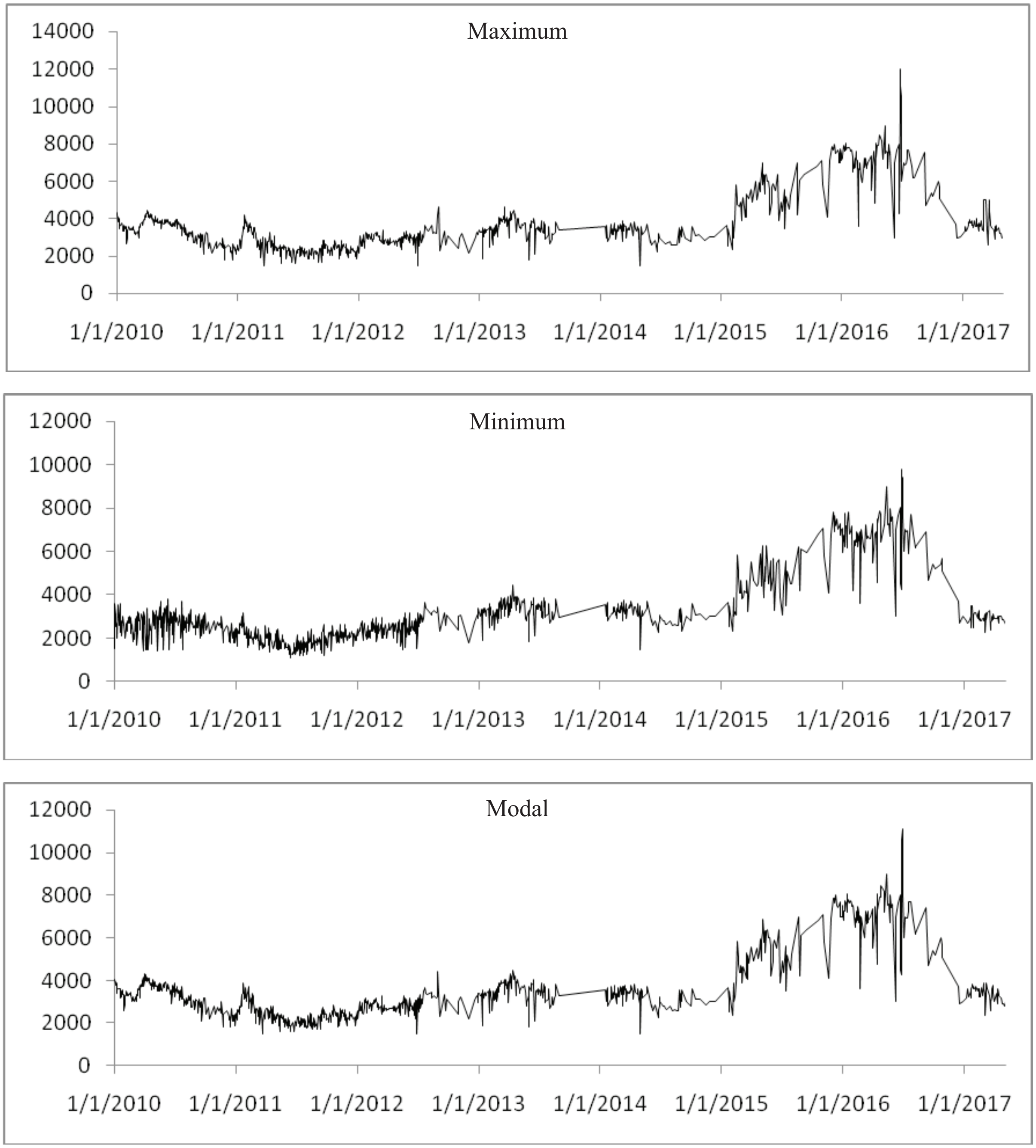

Fig. 1. Time plot of pigeon pea price data

pea. The increasing mismatch between production and consumption of pulses has resulted in larger imports of pulses in recent year resulting increase in price.

7.2 Descriptive statistics

Table 1 reports the descriptive statistics of all the series under consideration. A perusal of table 1 indicates that the average maximum pigeon pea price is higher than the minimum and the modal price of pigeon pea of the selected market. Higher value of coefficient of variation in minimum price data indicates the higher variability as compared to other price data. Higher deviations between maximum and minimum value for each of the series indicate more dispersive nature of the data set. All the series are consideration are positively skewed and leptokurtic.

Table 1. Descriptive statistics of pigeon pea price data

Statistics	Maximum	Minimum	Modal
Observations	1128	1128	1128
Mean (Rs/quintal)	3644.73	3137.26	3535.87
Median	3310.50	2800.00	3201.50
Minimum	1500.00	1100.00	1500.00
Maximum	12000.00	9800.00	11100.00
Standard deviation	1480.24	1490.01	1487.02
Coefficient of variation (%)	40.61	47.49	42.05
Skewness	1.82	1.72	1.80
Kurtosis	6.43	5.62	6.16

7.3 Test for stationarity, presence of long memory and structural break

The first step in analyzing the time series data is to check for presence of unit root. Several tests are available for this purpose. In the present investigation, PP test (Phillips and Perron, 1988) is employed to see the presence of unit root in the data set as it is more robust. The results of the test are reported in table 2. It is clear from the results of PP test that the null hypothesis of unit root test is rejected at 5% level of significance indicating stationarity of the series.

Table 2. PP test for stationarity

Price series	PP test	
	Test statistic	p-value
Maximum	-5.736	<0.001
Minimum	-7.524	<0.001
Modal	-5.249	<0.001

The autocorrelation function (ACF) and partial autocorrelation function (PACF) of the original price series are investigated. The ACF plots of the maximum, minimum and modal price series are shown in Fig. 2 and it is clear that the autocorrelation functions are decaying very slowly towards zero indicating the possible presence of long memory. The dotted lines in this figure represent 95% confidence intervals. The pattern is in contradiction to the usual Box Jenkins ARIMA methodology. In general, the stationary series will have rapid decay of autocorrelation function towards zero. But as depicted in Fig. 2, the autocorrelation function is significant even beyond 200 lags which clearly indicates the possible presence of long memory property.

After investigating the ACF plot, the long memory test is conducted to the data set and the test results are provided in table 3. Since the calculated Z-value greater than 1.96 for all the series under consideration, the test is found to be significant. It establishes the significance presence of long range dependency in price series. Hence we can use long memory time-series models to get forecasts of the under lying series.

Table 3. Long memory parameter estimate of pigeon pea price data

	Maximum	Minimum	Modal
d	0.221	0.225	0.222
S.E.	0.013	0.013	0.013
Z	17.000	17.307	17.076

S.E.: Standard Error

Empirical fluctuation process according to OLS based CUSUM test is applied to the price series and the results are presented in figure 3. Structural changes in the price series of the selected market are shown here. A perusal of figure 3 indicates that there is a structural break in the price series during the year 2015-2016. To confirm this OLS based CUSUM test, Chow test is applied to the price series and the results are reported in table 4. The tests for significance of tentatively selected structural break are highly significant for both the tests in all the series. According to these tests there is a break point in the data set which at 11th February, 2015 and it is same for all the series.The F-statistic values of the Chow test are plotted in Fig. 4. From the plot it is clear that break is present at 0.8 which is euivaltent to the serially numbered 902 observation for all the sries. The 902 numbered observation refers to the observation corresponding to 11th February, 2015.

Table 4. Test for detection of structural break

Series	OLS-based CUSUM test		Chow test	
	Test statistic	p-value	Test statistic	p-value
Maximum	10.23	<0.001	1544.60	<0.001
Minimum	10.62	<0.001	1461.30	<0.001
Modal	10.16	<0.001	1497.60	<0.001

7.4 Joint test and Forecasting using TSF

To detect presence of spurious long memory, the joint test for testing fractional integration and structural break is carried out by the procedure described in Section 5. The test results are highly significant at 5% level of significance indicating

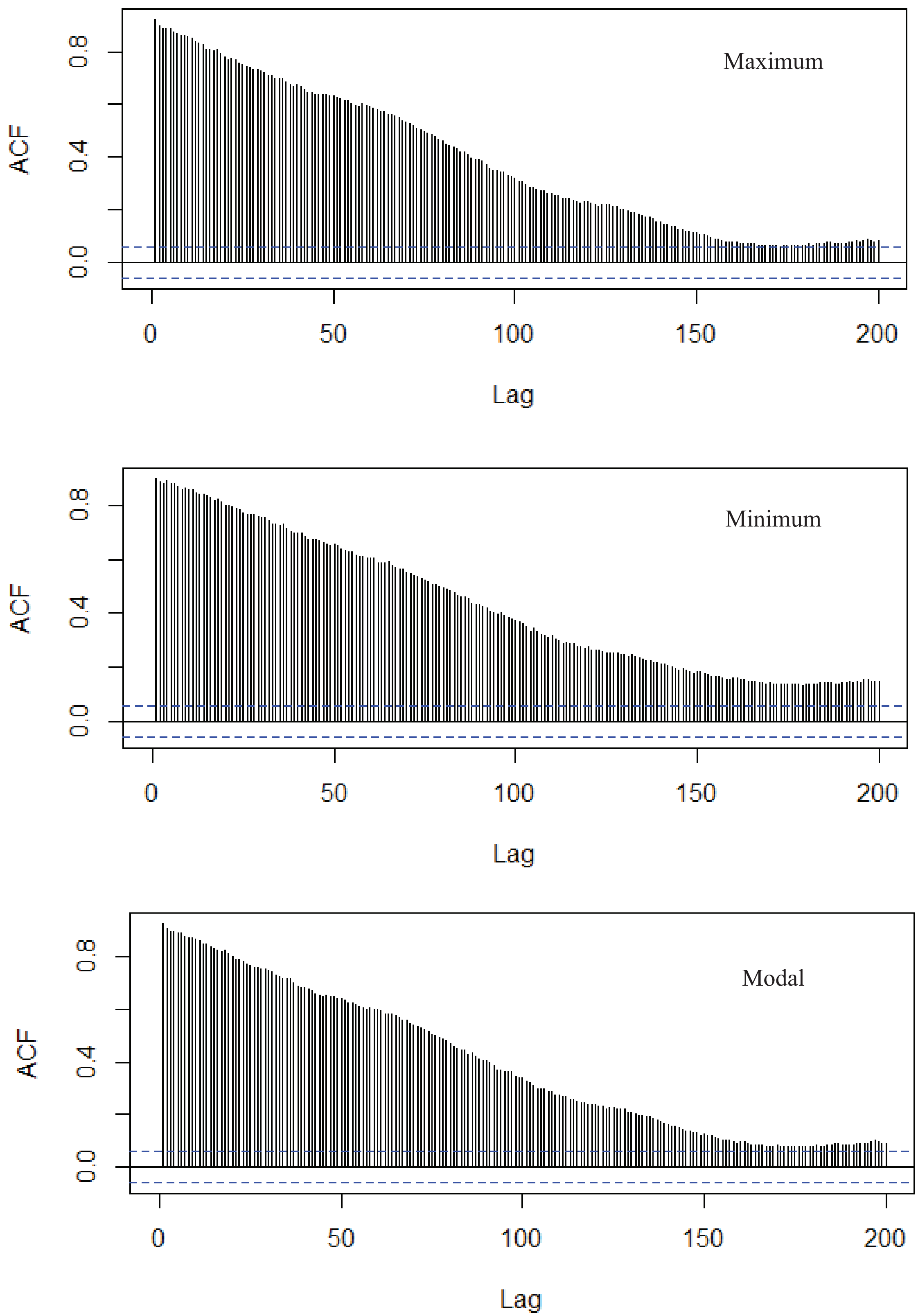

Fig. 2. ACF plot of pigeon pea price data

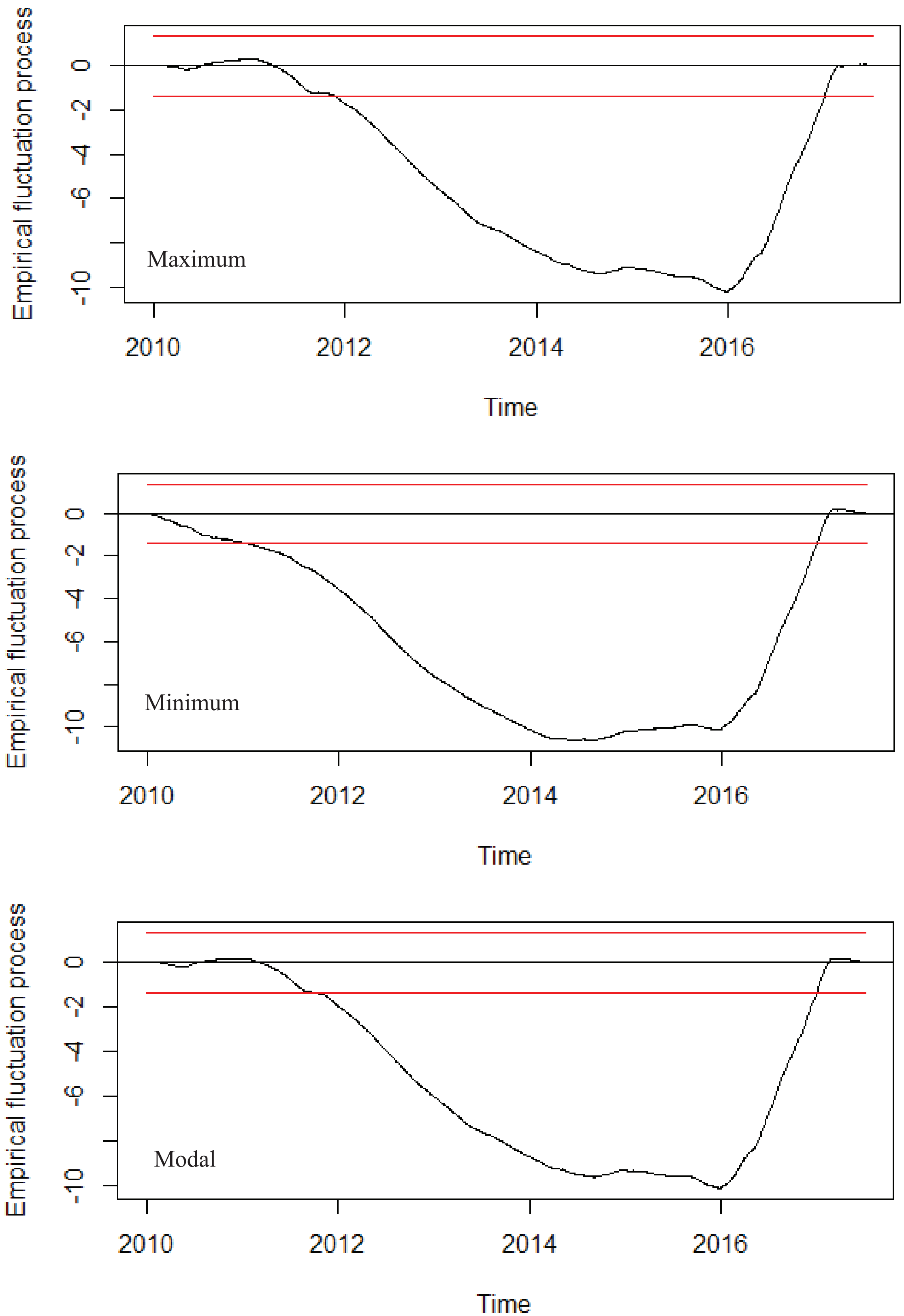

Fig. 3. OLS based CUSUM test for structural break in price series

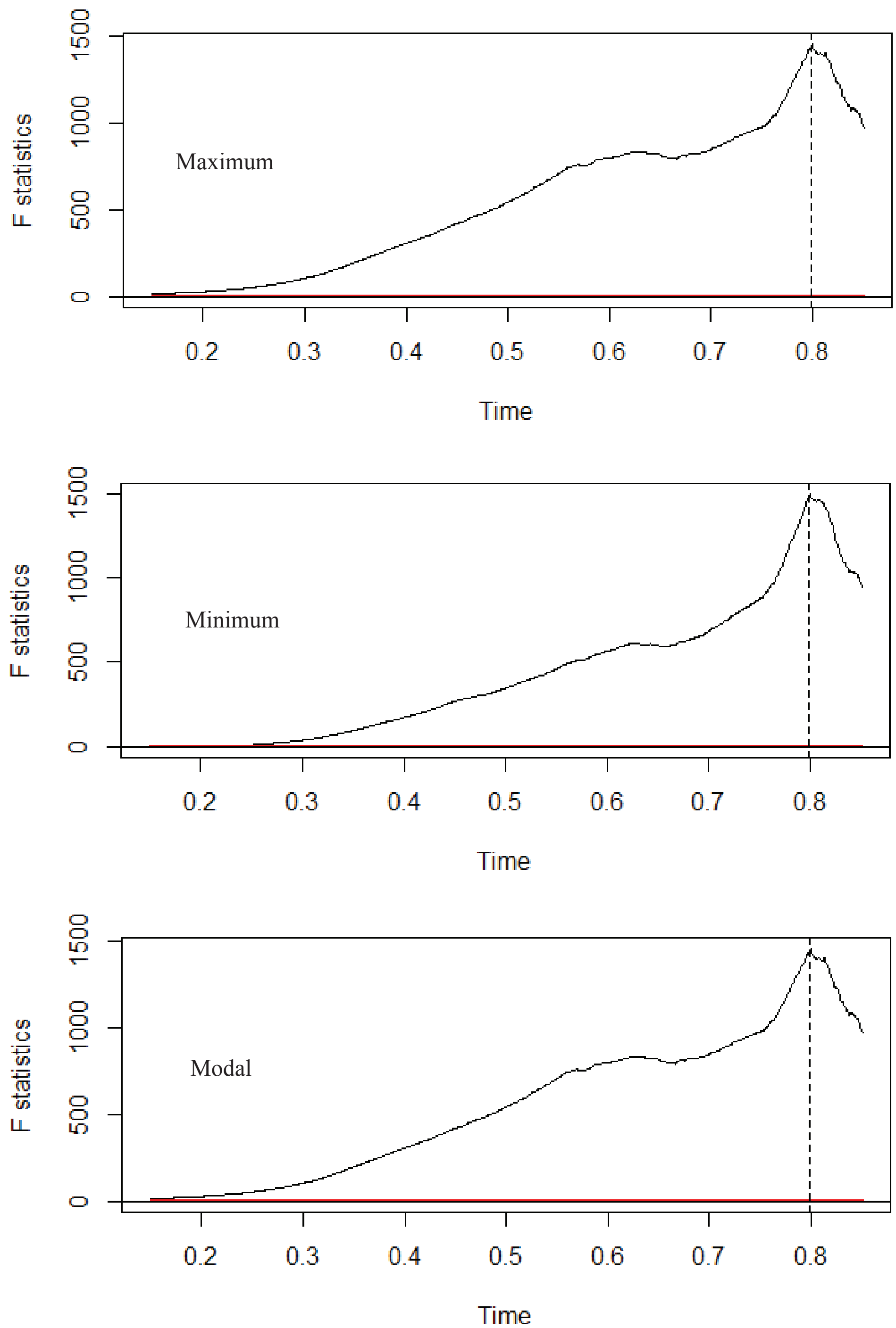

Fig. 4. Plot of F-statistics of Chow test for structural break

simultaneous presence of long memory and structural change.The critical value at 5% level of significance is 5.99 and the calculated test statistics for all three series are much higher than the critical value. Accordingly, TSF approach of forecasting is applied to the data set under consideration. The forecasting methodology starts with estimation of long memory parameter and differencing the price series to obtain the weakly dependent series. Then a suitable ARMA model is fitted to the weakly dependent series based on minimum AIC or SBC value and obtain one step ahead forecast of the underlying weakly dependent series using best chosen AR model. Finally, by applying cummulation differencing operator the one-step ahead forecast of the original price series is computed. The previous steps are repeated 50 times to compute 50-steps ahead forecasts for the dataset.

7.5 Forecasting using AR truncation method and AR approximation method

Truncated AR model of order 49 is fitted to the each of price series and the one-step ahead forecast of 50 observations is obtained using equation (19) for each of the series.

An approximate AR model with order chosen based on minimum AIC value is fitted to each of the price data. The orders of approximated AR models along with corresponding AIC values are provided in table5. After fitting best selected AR model to the price data one-step ahead forecast is calculated for last 50 observations for each of the price series.

Table 5. Approximated AR model

Series	Order of AR model	AIC value
Maximum	19	17124.89
Minimum	18	17408.18
Modal	18	17057.94

7.6 Validation of results

Last 50 observations of the data set were kept for the validation of the models. After confirming for the structural break and long memory in the series, TSF approach was used to obtain forecasts of the price series. A comparative performance of forecasting ability of TSF approach with usual ARFIMA model (without considering structural break), AR truncation method and AR approximation method has been carried out in terms of Mean absolute percentage error (RMAPE) and Root mean square error (RMSE) value. The computed RMAPE and RMSE values for each of the forecast horizons are reported in tables 6, 7 and 8 for maximum, minimum and modal price series respectively. Forecast evaluation is carried out for five moving windows (10-step, 20-step, 30-step, 40-step, 50-step ahead). For each series, the final columns of tables 6-8labeled as "Average" show average RMAPE and RMSE, across all the forecast horizons. A perusal of tables 7-9 indicate that the forecast performance of TSF approach is better than the other approaches. To test for the significance difference in forecast

Table 6. RMAPE and RMSE values for maximum price dataof pigeon pea

Forecasting methods	Forecast horizon					
	10	20	30	40	50	Average
RMAPE (%)						
TSF	8.44	6.03	7.75	19.71	15.95	11.57
ARFIMA	57.33	49.14	46.53	44.57	40.90	47.69
AR(P)	24.32	23.39	22.42	21.71	20.84	22.53
AR(AIC)	8.46	9.54	13.01	15.79	16.67	12.70
RMSE						
TSF	577.85	730.78	665.81	717.65	661.98	670.81
ARFIMA	1880.46	1726.52	1668.21	1613.09	1509.74	1679.60
AR(P)	817.51	832.27	866.23	880.02	865.35	852.27
AR(AIC)	679.60	749.92	709.05	669.33	647.12	691.00

Table 7. RMAPE and RMSE values for minimum price dataof pigeon pea

Forecasting methods	Forecast horizon					
	10	20	30	40	50	Average
RMAPE (%)						
TSF	28.25	8.28	5.54	12.20	4.64	11.76
ARFIMA	25.60	25.48	27.35	27.01	29.24	26.94
AR(P)	10.11	11.84	13.11	13.60	17.60	16.25
AR(AIC)	16.65	17.15	14.92	12.58	13.38	14.93
RMSE						
TSF	169.38	317.82	288.76	441.61	178.87	279.28
ARFIMA	735.35	746.44	798.81	798.95	880.69	792.07
AR(P)	336.59	314.50	324.16	457.89	687.97	424.22
AR(AIC)	160.78	220.99	235.88	252.79	610.78	296.24

Table 8. RMAPE and RMSE values for modal price data of pigeon pea

Forecasting methods	Forecast horizon					
	10	20	30	40	50	Average
RMAPE (%)						
TSF	9.00	5.13	8.93	9.06	8.13	8.05
ARFIMA	8.92	8.75	9.11	7.84	8.20	8.56
AR(P)	18.99	16.83	20.10	20.68	20.34	19.38
AR(AIC)	8.99	9.13	9.85	10.08	10.15	9.70
RMSE						
TSF	320.46	344.03	369.92	341.57	354.01	345.99
ARFIMA	322.00	348.24	365.81	357.35	362.89	351.26
AR(P)	625.29	582.90	740.21	791.97	807.24	709.52
AR(AIC)	364.28	395.26	422.48	412.26	465.19	411.89

performance of TSF approach with others, Diebold-Mariano test has been performed based on five moving windows for last 50 observations and it is revealed that the predictive performance of TSF approach is significantly different from other approaches for the data under consideration.

8. CONCLUSIONS

In this paper, the forecasting of long memory time-series focusing on spurious long memory is discussed. CUSUM test, Chow test have been applied to identify the structural break and Joint test is also applied in order to detect simultaneous presence of structural break as well as long memory in the series under consideration. In all the series it is found that there is presence of structural break as well as long memory. Single break was found in the series and it was observed that the break appears in February, 2015. Accordingly, TSF approach of forecasting along with the other methods e.g. AR approximation and AR truncation method are investigated for forecasting the maximum, minimum and modal price of pigeon pea in Bhopal market of Madhya Pradesh. Simple ARFIMA model was also used in order to see the effect of ignorance of structural break on forecasting performance. A comparative study is carried out among the existing forecasting methodologies. Tables 6to8 clearly indicate the outperformance of TSF approach over usual ARFIMA model, AR approximation and AR truncation method in terms of RMAPE and RMSE value.

ACKNOWLEDGEMENT

We would like to express our sincere thanks and gratitude to the editor and anonymous reviewer for their valuable suggestions that helped us a lot in improving this manuscript.

REFERENCES

Beran, J. (1995a). *Statistics for Long-Memory Processes*, Chapman and Hall Publishing Inc., New York.

Beran, J. (1995b). Maximum Likelihood Estimation of the Differencing Parameter for Invertible Short and Long Memory Autoregressive Integrated Moving Average Models, *J. Roy. Statist. Soc.*, Series B (Methodological), **57**(**4**), 659-672.

Diebold, F. X. and Inoue, A. (2001). Long memory and regime switching, *J. Economet.*, **105**, 131-159.

Geweke, J. and Porter-Hudak, S. (1983). The estimation and application of long- memory time series models, *J. Time Series Analy.*, **4**, 221-238.

Granger, C. W. J. and Joyeux, R. (1980). An introduction to long-memory time series models and fractional differencing, *J. Time Series Analy.,* **4**, 221–238.

Gil-Alana, A. L. (2002). A joint test of fractional integration and structural break at a known period of time, *J. Time Series Analy.*, **25**, 691-700.

Gooijer, J. G. D. and Hyndman, R. J. (2006). 25 years of time-series forecasting, *Int. J. Forecasting*, **22**(**3**), 443-73.

Hurst, H. E. (1951). Long term storage capacity of reservoirs, *Transactions of the American Society of Agricultural Engineers*, **116**, 770–799.

Jensen, M.J. (1999). Using wavelets to obtain a consistent ordinary least squares estimator of the long-memory parameter, *J. Forecasting*, **18**, 17-32.

Ngene, G. M. and Lambert, C. A. (2015). Testing long memory in the presence of structural breaks: an application to regional and national housing markets, *The Journal of Real Estate Finance and Economics*, **50**(**4**), 465-483.

Papailias, F. and Dias, G. F. (2015). Forecasting long memory series subject to structural change: A two-stage approach, *Int. J. Forecasting*, **31**, 1056-1066.

Paul, R. K. (2014). Forecasting wholesale price of pigeon pea using long memory time-series models, *Agril. Eco. Res. Rev.*, **27**(**2**), 167-176.

Paul, R. K., Birthal, P. S. and Khokhar, A. (2014b). Structural breaks in mean temperature over agro-climatic zones in India, *The Scientific World Journal*, dx.doi.org/10.1155/2014/434325.

Paul, R. K., Gurung, B. and Paul, A. K. (2014a).Modelling and forecasting of retail price of arhar dal in Karnal, Haryana, *I. J. Agril. Sci.*, **85**(**1**), 69–72.

Paul, R. K., Saxena, R., Chaurasia, S., Zeeshan and Rana,S. (2015). Examining export volatility, structural breaks in price volatility and linkages between domestic & export prices of onion in India. *Agril. Eco. Res. Rev.*, **28**, 101-116.

Paul, R K. and Anjoy, P. (2017). Modeling fractionally integrated maximum temperature series in India in presence of structural break. *Theoretical App. Climatology*. https://doi.org/10.1007/s00704-017-2271-x

Peiris, M. S. (1987). A note on the predictors of difference sequences. *Austr. J. Statist.*, **29**, 42-48.

Ploberge, W. and Krame, W. (1992). The CUSUM test with OLS residuals, *Econometrica*, **60**(**2**), 271-285.

Robinson, P. M. (1994). Semi-parametric analysis of long-memory time series, *Annals Statist.*, **22**, 515-539.

Robinson, P. M. (1995). Long-periodogram regression of time-series with long-range dependence, *Annals Statist.*, **23**, 1048-1072.

Wang, C. S. H., Bauwens, L. and Hsiao, C. (2013). Forecasting a long memory process subject to structural breaks, *J. Economet.*, **177**, 171-184.