International Journal of Statistics and Applied Mathematics 2023; 8(4): 08-17

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
Maths 2023; 8(4): 08-17

https://www.mathsjournal.com
Received: 07-02-2023
Accepted: 13-03-2023

Borsha Neog
Assistant Professor, Department of Agril. Statistics, Assam Agricultural University, Assam, India

Bipin Gogoi
Professor, Department of Statistics, Dibrugarh University, Assam, India

AN Patowary
Assistant Professor, College of Fisheries (AAU), Raha, Assam, India

Corresponding Author:
Borsha Neog
Assistant Professor, Department of Agril. Statistics, Assam Agricultural University, Assam, India

Development of hybrid model for improvement of forecast Agricultural commodity price

Borsha Neog, Bipin Gogoi and AN Patowary

Abstract
Auto regressive fractional integrated moving average (ARFIMA) is widely applied for time series forecasting in long memory for divergent domain from several decades. The major limitation of this model is presumption of linearity. In real world, most of the long memory time series data are not purely linear, therefore hybrid model is evolved to enhance the prediction ability of ARFIMA models by fusing with other non-linear models. With this reasoning, this present study attempts to predict the price of Onion using ARFIMA-ANN and ARFIMA-SVM models. Results of this experiment justified the results of hybrid model gives better results with comparison to linear and non-linear models.

Keywords: ARFIMA, ANN, forecasting, price, time series

Introduction

Price instability and uncertainty pose a restriction on decision and policy makers. Price forecasting of agricultural commodity plays a vital role for both production and market strategy. Price forecasting allows a Mach between the supply and demand of the commodity. It is a herculean task because it depends on many factors which cannot be accurately predicted. Nonlinear and non-stationary behaviour are the pivotal problems in agricultural price data. Prices of agricultural commodity are also volatile in nature due to seasonality, inelastic demand, production uncertainty and also because many agricultural commodities are destructive in nature.

The presence of long memory in price of agricultural commodity has been an important subject for both theoretical and empirical studies. Since the data points are dependent over time and to make business decisions presence of long memory is a crucial information. Long memory can be measured by Hurst exponent. Many methods have been proposed for Hurst parameter estimation such as rescaled range (R/S), aggregated variance method, periodogram method, absolute value method, Higuchi's method and Wavelet based method.

Traditional time series models namely ARIMA models cannot describe long memory phenomenon. Therefore, to overcome such limitations different models has been established, among which most widely used model is autoregressive fractionally integrated moving average (ARFIMA) model given by Granger and Joyeux (1980) [2]. The properties of the process are widely discussed by Hosking (1981) [4]. (Magsood, A. and Aqil, S.M. (2014) [5]. Romalingam (2010) made overall review of long-memory independently. Long memory studies using hybrid ARFIMA and feed forward neural network carried out by Aladag *et al*. (2012) [1], Wiri *et al*. (2022) [13] used ARFIMA modelling in Nigeria exchange rate. Not much research work has been done in agriculture sector using long memory time series.

In general, the estimators of fractional order of integration d can be classified into two groups; parametric and semi parametric methods. In parametric approach, parameters are estimated simultaneously and in semi parametric methods, parameters are estimated in two steps: In the first step, d is estimated and in second step remaining parameters are estimated (Reisen, V, Abraham, B. and Loper, S. 2001) [8]. Mostly used semi parametric methods were proposed by Geweke and Porter- Hudak (1983) [3], Kuensch (1987) [4], Robinson *et al*. (1995) [9] & Reisen and Lopes (1999) [7], while the exact maximum likelihood (EML), proposed by Sowell (1992) [11] and modified profile likelihood (MPL) are the most common parametric methods.

So, the main aim of this paper is to develop a suitable model for forecasting of monthly price of vegetable (onion) in Assam. Data collected from Directorate of Economics & Statistics, Assam during the period January, 2010 to December, 2020.

2. Data and methodology

Monthly wholesale price data on selected agricultural crops are obtained from the various issues of "Agricultural prices in India" published by the Directorate of economics and statistics, Government of India during Jan, 2010 Dec, 2020. Methodology part consists with detection of long memory with its different tests, tests of stationarity, ARFIMA and its parameter estimation, ANN and SVM model and hybrid ARFIMA-ANN and hybrid ARFIMA-SVM model.

2.1. Long memory Process

Long memory in time series can be defined as autocorrelation at long lags (Robinson 1995) [9]. Mathematically, if the time series x_t is said to be long memory series if the autocorrelation function ρ_t satisfies the condition:

$$\lim_{t\to\infty} \sum_{t=-n}^{n} |\rho_t| \to \infty$$

Where n is the sample size, for detection of long memory, many statistical tests are available in literature *viz*., Rescaled range (R/S), aggregated variance method, absolute value method, periodogram method, Higuchi's method and Wavelet based method.

A brief description of the test R/S test, Absolute value method are given:

2.1.1. Rescaled range statistic (R/S)

The first test for long memory was used by the hydrologist Hurst (1951) for the design of an optimal reservoir for the Nile River, of where flow regimes were persistent. Hurst gave the following formula:

$$(R/S)_n = Cn^H$$

R/S is the rescaled range statistic measured over a time index n, c is a constant and H is the Hurst exponent. The aim of the R/S statistic is to estimate the Hurst exponent which can characterize a series. Estimation of Hurst exponent can be done by transforming to

$$\text{Log } (R/S)n = \log (C) + H \log (n)$$

Rescaled range statistic (R/S) is defined as the range of partial sums of deviation of a time series from its mean, rescaled by its standard deviation: consider the sample $\{x_1, x_2,\ldots, x_k\}$ from a stationary long memory process $\{ x_n;\ t=1,2,\ldots, N\}$, and let the partial sums of x_k is $\sum_{j=1}^{k} x_j$; k=1,2,…, n.

Let $\overline{x_n}= \frac{1}{n}\sum_{j=1}^{n} x_j$ and $S_n^2= \frac{1}{n-1}\sum_{k=1}^{n}(x_k-\bar{x})^2$ be the sample mean and sample variance respectively. The rescaled range statistic (R/S) is defined as: (Magsood, A. and Aquil, S. M. 2014);

$$R/S= \frac{1}{S_n}\left[Max_{1\le k\le n}\left(\sum_{j=1}^{k}\left(x_j-\bar{x}_n\right)\right)-Min_{1\le k\le n}\left(\sum_{j=1}^{k}\left(x_j-\bar{x}_n\right)\right)\right]$$

Where n: number of observations.

S_n= the standard deviation.

In addition, the Hurst coefficient H can be used to estimate the fractional differencing parameter d by the equation:

$$d= H-0.5$$

This method is considered as semi parametric of estimating ARFIMA models.

2.2.2. Absolute value Method

The data is divided in the same way as the aggregated variance method to form aggregated series, the sum of the absolute values of the aggregated series is computed as (Sun, R., Chen, Y. and Li, Q. 2007) [12]:

$$\frac{1}{N/m}\sum_{k=1}^{m}\left|x^{(m)}(k)\right|$$

The logarithm of this statistic is plotted versus the logarithm of m. For long-range dependent time series with parameter H, the result should be a line with slope H-1.

2.2. The ARFIMA model

ARFIMA model used for long range dependent time series. ARFIMA (p, d, q) model (Granger and Joyeux (1980)) [2] is given as follows:

$$\varphi(B)(1-B)^d X_t = \theta(B)e_t, -0.5 < d < 0.5$$

Where B is the back-shift operator such that $BX_t=X_{t-1}$ and e_t is a white noise process with $E(e_t)=0$ and variance is σ_e^2. The polynomials
$\varphi(B) = \left(1-\varphi_1 B - \cdots - \varphi_p B^p\right)$ and $\theta(B) = (1-\theta_1 B - \cdots - \theta_q B^q)$ have orders p and q respectively with all their roots outside the unit circle. The process is stationary if d=0 and the effect of shock to e_t on $x_{(t+j)}$ decays geometrically as j increases. For d=1, the process is non-stationary.

An ARFIMA (p, d, q) process may be differenced a finite integral number until d lies in the interval [-0.5, 0.5] and will then be stationary and invertible (sun, R., Chen, Y. and Li, Q.2007) [12]. The ARFIMA process is stationary but not invertible if d=-0.5, the ARFIMA process has an intermediate memory or over differenced if -0.5<d<0. In that case, the inverse autocorrelation decay hyperbolically. When d=0, the ARFIMA process can be white noise and when 0<d<0.5, the ARFIMA (p, d, q) process is a stationary process with long memory and the auto covariance function exhibits hyperbolic decay (Doornik, J.A. and Ooms, M. 2004; Sun, R., Chen, Y. and Li, Q.2007) [12].

2.2.1. Estimation of ARFIMA models

There are two types of estimation methods: Parametric and Semi-Parametric methods. The methods which are implemented in their chapter are described below.

2.2.1.1 Parametric estimation methods

The parametric methods estimate all parameters of the ARFIMA process in one step. Exact maximum likelihood (EML), proposed by Sowell (1992) [11]. The EML method is chosen in this chapter for estimating the parameters of ARFIMA model.

Let y be the sample time series. The log-likelihood of the estimation is simple and it is based on the normality assumption and with a procedure to compute the autocovariances in the T x T covariance matrix $\Sigma\sigma_e^2 R$ of a T x 1 vector of observations y, the log-likelihood for the ARFIMA (p, d, q) model with k regressors is (Lildholdt, P.2000; Doornik, J.A. and Ooms, M. 2004):

$$\sigma_\epsilon^2 - \frac{1}{2} ln|R| - \frac{1}{2\sigma_e^2} ZR^{-1}Z$$

Where

Z= y-xβ

when σ_ε^2 and β are concentrated out, the resulting normal profile likelihood function becomes:

$$\log L(d, \phi, \theta) = C - \frac{1}{2} ln|R| - \frac{T}{2} \ln[\acute{\hat{z}}\, R^{-1}\, \hat{z}]$$

Where $\hat{z}$ = y- x$\hat{\beta}$

and $\hat{\beta} = (\acute{x}R^{-1}x)^{-1}\acute{x}\, R^{-1}y$

2.2.1.2. Semi - Parametric estimation methods

The most popular and widely used of all semi-parametric estimation method was proposed by Geweke and Porter – Hudak (1983) [3] (Joe, B and Sisir, R. 2014), this method is based on an approximated regression equation obtained from the logarithm of the spectral density function. The GPH estimation procedure is a two – step procedure which begins with the estimation of d (Paul, R.K. 2014) and is based on the following log – periodogram regression (Bhardwaj, G., and Swanson, N.R. (2006))

$$\text{Ln}[I(w_j)] = \beta_0 + \beta_1 \ln[4\sin^2(w_j^2)] + \upsilon_j$$

Where $w_j = \frac{2\pi j}{T}$, j=1,2,…,m

The estimate of d, say $\hat{d}_{GPH}$; is $\hat{\beta}_1$, w_j represents the m=√T. Fourier frequencies, and $I(w_j)$ denote the sample periodogram defined as:

$$I(w_j) = \frac{1}{2\pi T}\left|\sum_{t=1}^{T} y_t e^{-wjt}\right|^2$$

The second step of the GPH estimation procedure involves fitting an ARMA model to data according to box and Jenkins method, given the estimate of d. The GPH estimate is denoted as $\hat{d}_{GPH}$.Reisen and Lopes (1999) [7] modified the GPH procedure by replacing the periodogram by a “smoothed” estimate of the spectral density. Reisen and Lopes (1999) [7] proposed to use a Blackman- Tukey type estimate of the spectral density (Joe, B. and Sisir, R. 2014)

$$F_m(x) = \frac{1}{2\pi}\sum_{s=-m}^{m} k(\frac{s}{m})\hat{p}\,(s)\cos(sx)$$

Henceforth, the smoothed periodogram estimate of d is denote as $\hat{d}_{sperio}$. Both $\hat{d}_{GPH}$ and $\hat{d}_{sperio}$ although simpler to implement are inefficient in the non-stationary region i.e., $|d| > 1 = 2$.

Another semi-parametric estimator, the local whittle estimator, is also often used to estimate d. This estimator was proposed by Kuench (1987) and modified by Robinson (1995) [10]. The local whittle estimator of d; say $\hat{d}_w$ is obtained by maximizing the local whittle log likelihood at Fourier frequencies close to zero, given by (Bhardwaj, G. and Swanson, N.R. 2006):

$$\Gamma(d) = -\frac{1}{2\pi m}\sum_{j=1}^{m}\frac{I(w_j)}{f(w_j;d)} - \frac{1}{2\pi m}\sum_{j=1}^{m} f(w_j;d)$$

2.2.2. Ljung-Box test

The Ljung-Box test, named after statisticians Greta M. Ljung and George E.P. Box, is a statistical test that checks if autocorrelation exists in a time series. It is sometimes called Box-Pierce test. The test identifies whether errors are iid (i.e., white noise). The null hypothesis of Ljung-Box test is

H_0: The residuals are independently distributed and the alternative hypothesis is

H_1: The residuals are not independently distributed; they exhibit serial correlation.

The test statistic for the Ljung-Box test is as follows:

$$Q = n(n+2)\,\Sigma p_k^2 / (n-k)$$

Where

n = sample size

p_k = sample autocorrelation at lag *k*

The test statistic *Q* follows a chi-square distribution with *h* degrees of freedom; that is,

$$Q \sim \chi^2(h).$$

We reject the null hypothesis and say that the residuals of the model are not independently distributed if $Q > \chi^2_{1-\alpha,\, h}$

2.2.3. Brock- Dechert- Scheinkman (BDS) Test

This test was proposed by Brock *et al.* (1991) called the Brock- Dechert- Scheinkman (BDS) statistic, for residual analysis. This test can be applied for estimation of residuals of time series model and used as a model selection tool. In the recent work, BDS test is used in detecting deterministic nonlinear dynamics and chaos theory. Now it is not only used for detecting deterministic chaos, but also used for testing residuals. Under the null hypothesis of independent and identical distribution (IID), the BDS test statistic have its ability in distinguishing random time series from the time series generated by low dimensional chaotic or nonlinear stochastic processes.

Under the null hypothesis, the BDS statistic for m>1 is defined as

$$\text{BDS}(m, M, r) = \frac{\sqrt{M}}{\sigma}[C(m,r) - C^m(1,r)]$$

It has a limiting standard normal distribution as M→∞ and obtain its critical values using the standard normal distribution.

2.3 Artificial Neural Network (ANN) model

ANN(s) models are set of nonlinear models that can capture different nonlinear structures present in the data set. The specification of ANN model does not require any prior assumption of the data generating process, instead it is largely depended on characteristics of the data known as data-driven approach. Single hidden layer feed forward network is the most widely used model for time series modelling and forecasting. This model is constructed by a network of three layers of simple processing units, and thus termed as multilayer ANNs. The first layer is input layer, the middle layer is the hidden layer and the last layer is output layer.

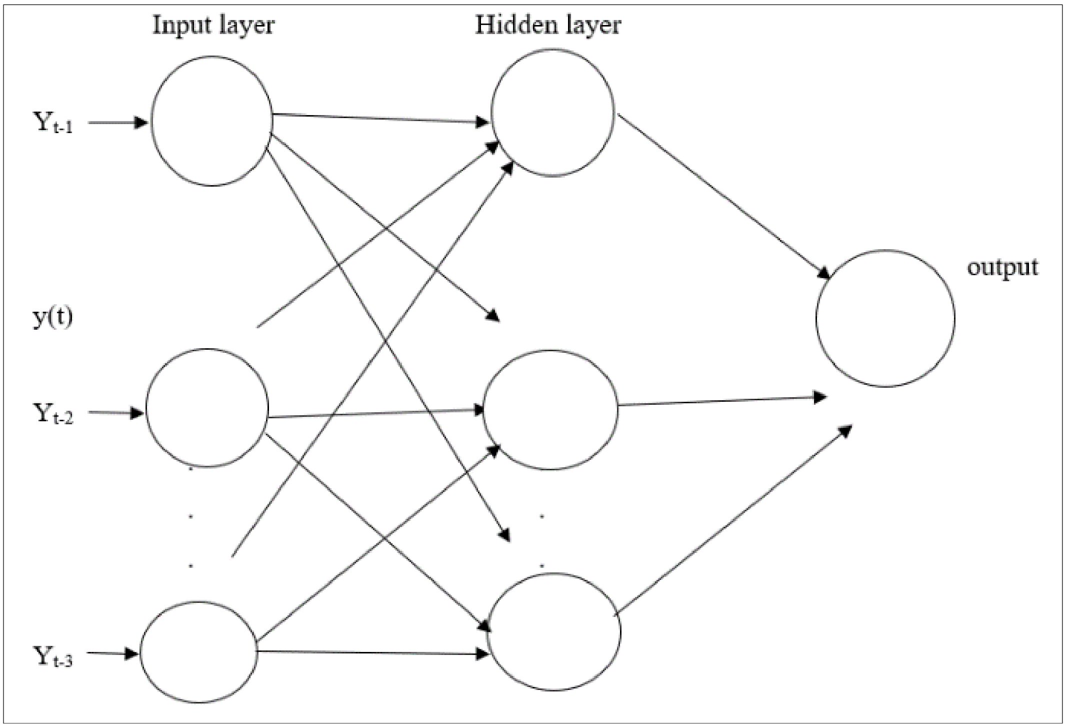

Fig 1: Neural Network architecture

The relationship between the output (y_t) and the inputs (y_{t-1},y_{t-2}, …, y_{t-p}) can be mathematically represented as follows:

$$Y_t = f\left(\sum_{j=0}^{q} w_j g\left(\sum_{i==0}^{q} w_{ij}\ y_{t-i}\right)\right) \quad (1)$$

Where w_j(j=0,1,2,…,q) and w_{ij}(i=0,1,2,…,p; j=0,1,2,…,q) are the model parameters often called the connection weights; p is the number of input nodes and q is the number of hidden nodes, g and f denote the activation function at hidden and output layer respectively.

Recently, support vector machine (SVM) method developed by Vapnik *et al*. (2000) has wide range of applications such as data mining, classification, regression and time series forecasting. In time series forecasting, SVM becomes successful by solving nonlinear regression estimation problems.

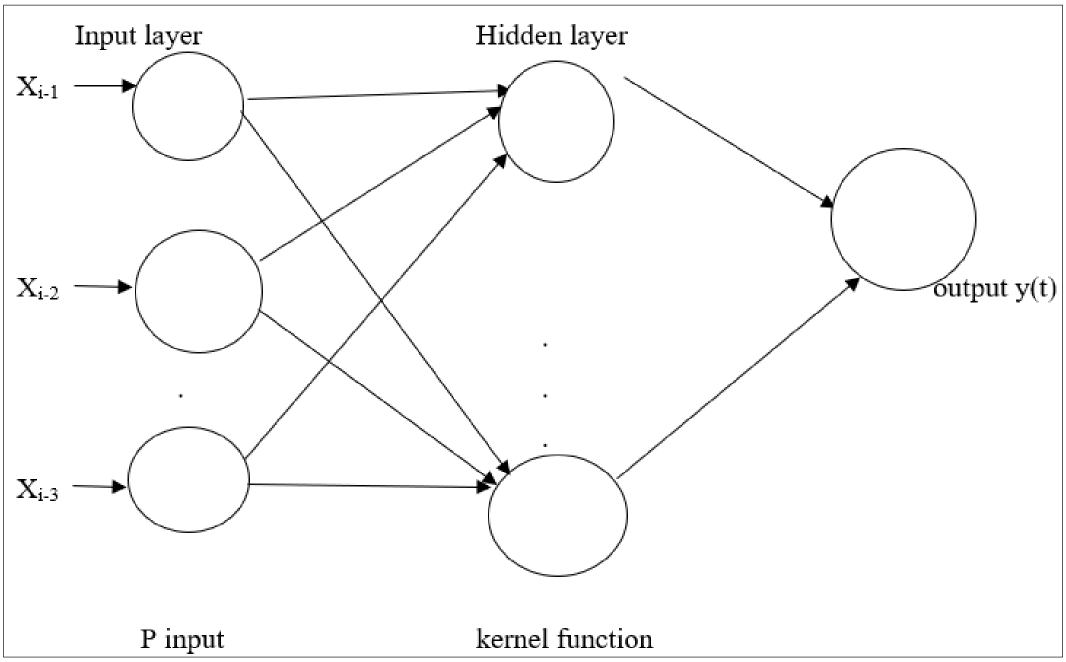

Fig 2: SVM architecture

2.4. Hybrid approach

Hybrid model is a combination of two or more models. Use of hybrid models becomes necessary because a single model may not handle the inherent data patterns like nonlinearity and non-stationarity simultaneously. However, both statistical and AI based time series models have their prerequisite assumptions. These limitations demand the necessity for combining two or more models for forecasting. The proposed approach considered time series (y_t) as a function of linear and non-linear components. Hence y_t= f (L_t, N_t) where y_t is a time series data; L_t and N_t represents the linear and nonlinear component respectively. This approach follows the Zhang's

(2003) hybrid approach, accordingly the relationship between linear and nonlinear components can be written as following

$$Y_t = L_t + N_t$$

The main strategy of this approach is to model the linear and nonlinear components separately by different model. The methodology consists of three steps. Firstly, ARFIMA model is applied to the data series to fit the linear part. Let the prediction series provided by ARFIMA model denoted as $\widehat{L_t}$. In the second step, instead of predicting the linear component, the residuals denoted as e_t which are nonlinear in nature are predicted. The residuals can be obtained by subtracting the predicted value $\widehat{L_t}$ from actual value of the considered time series y_t.

$$e_t = y_t - \widehat{L_t}$$

Now the residuals are predicted employing an ANN and SVM model. Let the prediction series provided by ANN/SVM model denoted as $\widehat{N_t}$. Finally, the predicted linear and nonlinear components are combined to generate aggregate prediction.

$$\hat{y}_t = \widehat{L_t} + \widehat{N_t}$$

BDS test is used to test for non-linearity in this study.
The graphical representation of proposed approach is expressed in the figure 6.2.3 & 6.2.4

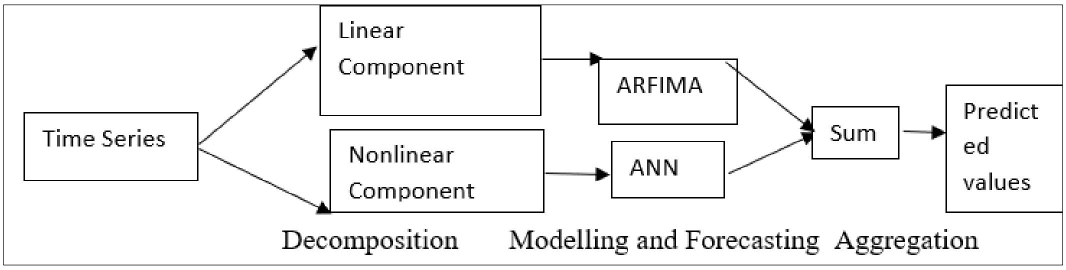

Fig 3: Schematic representation of ARFIMA-ANN hybrid methodology

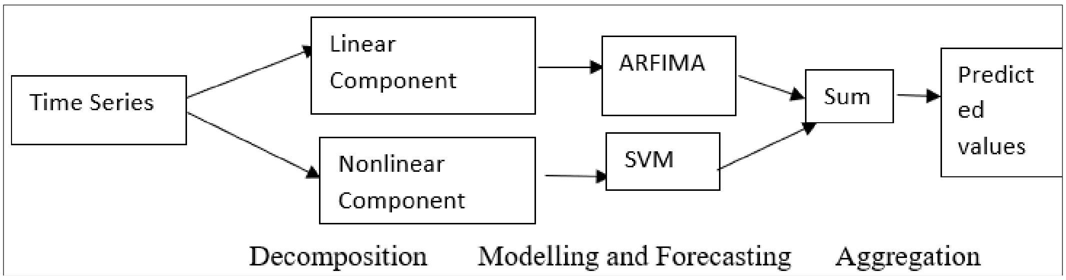

Fig 4: Schematic representation of ARFIMA-SVM hybrid methodology

For long memory time series ARFIMA have been used. However, it is obvious that ARFIMA is not adequate for non-linear part. It can be applied only for linear part. In real situation, it is difficult to completely know the characteristics of data. Hybrid methods have the capability to modelling for linear and non-linear dataset. To model long memory time series, any hybrid approaches using nonlinear techniques SVM has not been proposed in any literature. In this chapter, a new hybrid approach has been proposed ARFIMA-ANN & ARFIMA-SVM for forecasting of price of agricultural commodity.

2.5. Forecasting Performance

Forecasting Performance of the model has been adjusted by computing mean absolute error (MAE). The model with minimum values of MAE for training and testing data set is preferred for forecasting purpose. The MAE is computed as

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|y_t - \hat{y}_t|$$

Where n is the total number of forecast values. Y_t is the actual value at period t and $\hat{y}_t$ is the corresponding forecast value. The model with better forecasting power has lower values of MAE compared with other models.
The statistical software *viz.*, R and Excel were used for modelling and forecasting of price of vegetables in Assam. R software package arfima and TSA package was used for ARFIMA, 'Forecast' was used for modelling and forecasting using NN and package 'e 1071' was used for modelling and forecasting using SVM.

3. Results and Discussion

For the present study, Monthly wholesale price data on selected agricultural crops are obtained from the various issues of "Agricultural prices in India" published by the Directorate of economics and statistics, Government of India during the period Jan,2010 to Dec., 2020. Out of 120 observations, 84 observations have been used for model estimation and remaining 36 observations are used for validation. Summary statistics of price of onion is given in the Table 6.1.

Table 1: Summary statistics of price of Onion

Statistic	Series	Statistic	Series
Observation	120	Standard Deviation	1396.09
Mean	2205.54	Kurtosis	4.908
Median	1750.00	Skewness	1.991
Mode	1600.00	Coefficient of Variation (CV)	63.29

Fig 1: Trend of Prices of Onion in Assam

As Table 6.1 shows, the price series of onion has the mean of 2205.54 and the standard deviation of 1396.09 in the sample period. To validate the stationarity of the time series, two tests namely Augmented Dickey- Fuller test and Philips-Perron test are used. Results of the Stationarity tests are reported in table 6.3.2. The results indicate that price of onion time series is stationary.

Table 2: Testing for stationarity of Onion

ADF test statistic	P	PP test statistic	p
-4.731	0.001	-5.034	0.001

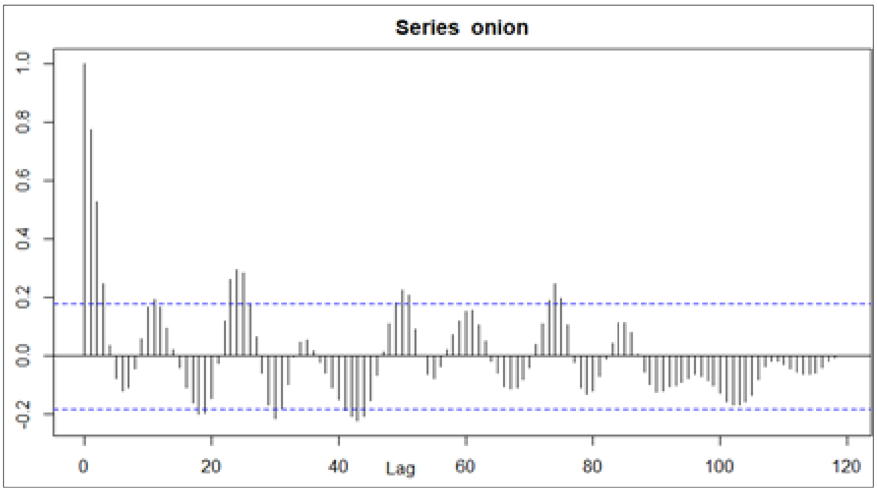

Fig 2: Plot of ACF of Onion

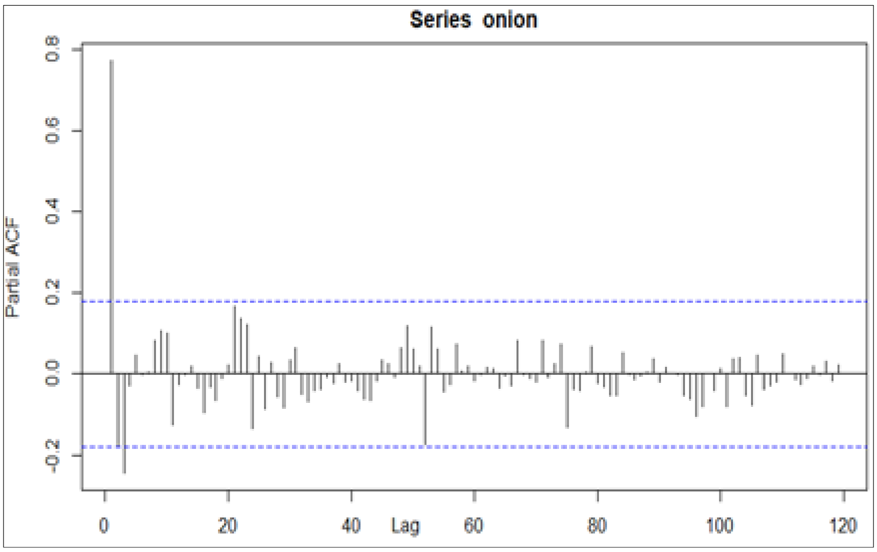

Fig 3: Plot of PACF of Onion

The auto correlation function (ACF) and Partial autocorrelation function (PACF) for the actual price series were investigated and it has been found that though the stationarity tests validated that the series is stationary but plot of ACF shows a slow decay towards zero. This indicating the possible presence of long memory in the series. The long memory for the series was tested using R/S method and Absolute value method. The results of the above tests are given in Table 6.3.

Table 3: Test for long memory of Onion

R/S method	Absolute value method
H=.897	H=.862

The estimated Hurst exponent 0.5<H<1, so this result provides more evidence for the existence of the long memory property in the price series. In this study, common parameter estimation exact maximum likelihood (EML) proposed by Sowell (1992) [11] were used for estimation of parameters. Based on the smallest value of AIC, the best ARFIMA model was chosen. The value of all parameters are given in table 6.4.

Table 4: Parameter estimates of ARFIMA model of Onion

Parameters	Estimates	Probability	Log-Likelihood	AIC
d	0.153	0.001	-977.8	1967.625
AR1	1.414	0.000		
AR2	-0.670	0.000		
MA1	0.708	0.001		
MA2	-0.233	0.000		

Table 5: Forecasted monthly mustard prices using ARFIMA (2, 0.15, 2) of Onion

Months	Point forecast	Lowest confidence interval	Upper confidence interval
Jan	3261.299	1630.04753	4892.551
Feb	2481.293	330.11818	4632.469
Mar	1967.229	-514.49189	4448.950
Apr	1755.510	-874.48199	4385.501
May	1794.404	-872.85967	4461.667
June	1986.230	-682.21598	4654.676
July	2227.263	-445.74475	4900.272
Aug	2436.055	-248.46373	5120.574
Sep	2566.752	-127.06983	5260.574
Oct.	2608.959	-87.95643	5305.875
Nov	2578.620	-118.33043	5275.570
Dec	2505.193	-193.09381	5203.481

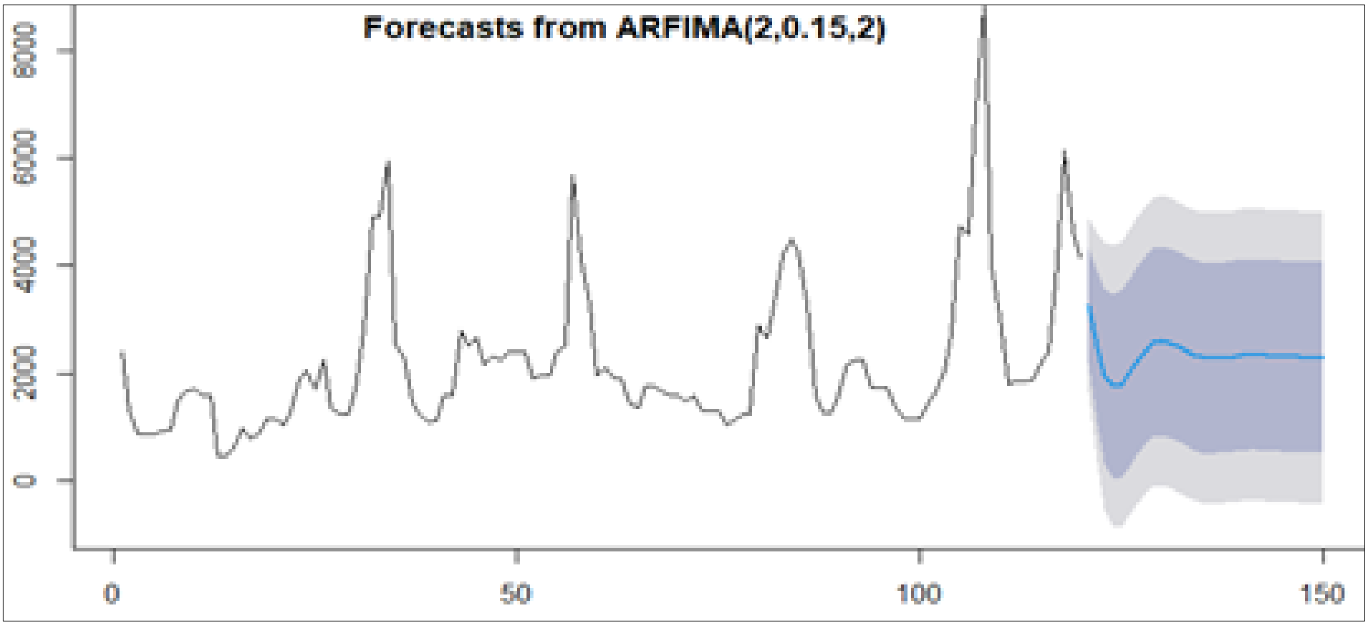

Fig 4: Graphical representation of Forecast of monthly mustard prices using ARFIMA (2, 0.15, 2) of Onion

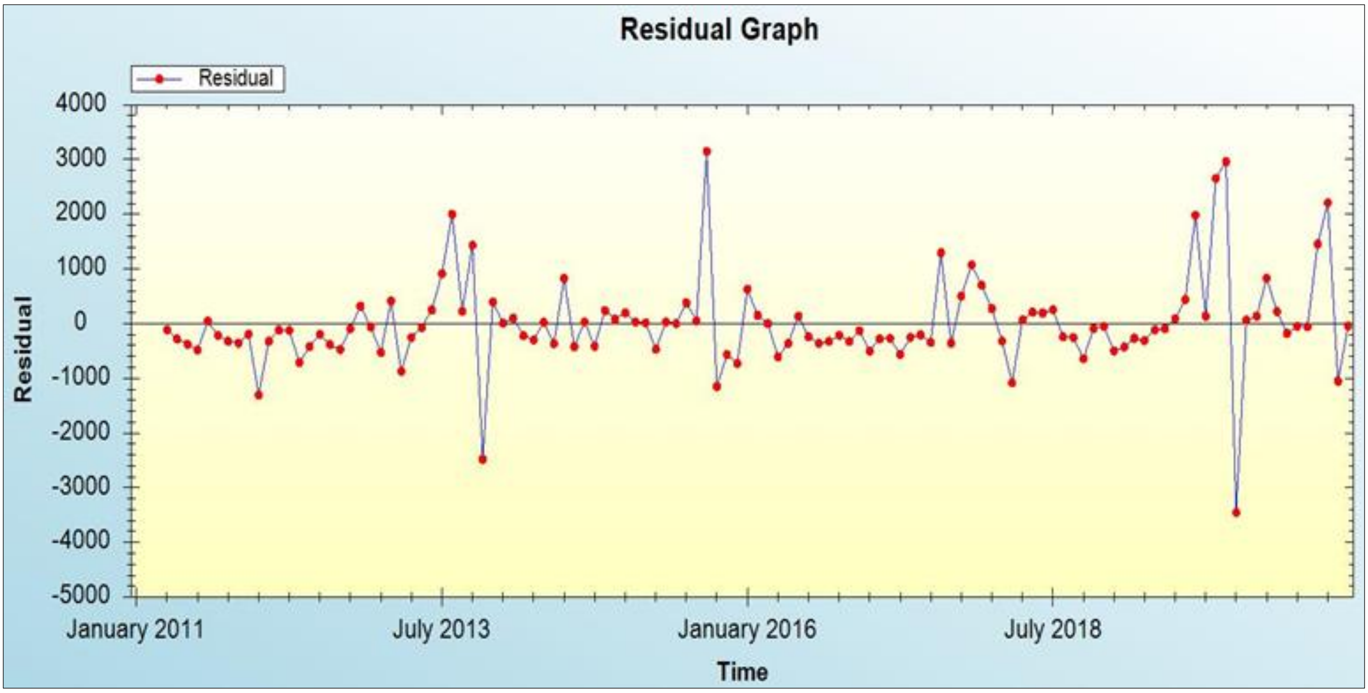

Fig 5: Residual Graphs of Onion

As we discussed here hybrid methodology contains both linear and non-linear part in the model. The linear part of the model estimated by ARFIMA model and for non-linear part residuals are required to check for their linearity assumption (Zhang 2003). Plot obtained in Fig 6.3.5. Indicates that the residuals of ARFIMA model are nonlinear and BDS test was confirmed that same as $p< 0.001$, hence we can apply ANN and SVM for non-linear part. We have tried different neural networks with different time delays with different hidden nodes on residuals and the results are given in table 6.6.

Table 6: MAE for Neural Network models for price of onion

Model parameters	MAE for Training	MAE for Testing
1:2s:1l	422.326	682.977
1:4s:1l	422.532	683.948
1:6s:1l	424.169	679.948
1:8s:1l	425.803	685.395
1:10s:1l	425.249	687.915
2:2s:1l	416.502	694.815
2:4s:1l	414.847	696.513
2:6s:1l	415.029	705.172
2:8s:1l	415.728	700.846
2:10s:1l	413.814	713.957
3:2s:1l	413.814	687.079
3:4s:1l	416.463	693.348
3:6s:1l	417.460	694.065
3:8s:1l	418.084	699.099
3:10s:1l	417.216	700.031
4:2s:1l	421.667	712.911
4:4s:1l	411.968	713.906
4:6s:1l	416.202	722.075
4:8s:1l	417.214	715.604
4:10s:1l	418.049	724.093
5:2s:1l	414.681	729.268
5:4s:1l	414.592	736.072
5:6s:1l	411.181	734.991
5:8s:1l	420.766	735.852
5:10s:1l	416.562	733.163
6:2s:1l	425.016	756.339
6:4s:1l	413.674	761.804
6:6s:1l	417.525	756.993
6:8s:1l	413.686	760.932
6:10s:1l	418.327	767.831

From the above table, the model 3:2s:1l was found to be the best one on the basis of minimum values of MAE for training=413.814 and testing= 687.079. From this selected model we have got the estimated values of residuals and fitted values of price of onion obtained by ARFIMA (2,15,2) then forecast value of price was obtained through hybrid approach i.e., ARFIMA (2,15,2)-ANN. The goodness of fit measure MAE for hybrid ARFIMA-ANN was found to be 497.673 as compare to 506.667 ARFIMA (2, 15, 2). Residuals obtained by using ARFIMA (2, 15, 2) were applied on the non-linear approach support vector machine using radial basis function as kernel. Forecast values of price obtained through ARFIMA (2.15, 2) were corrected by using the residuals through SVM and estimated the value MAE for hybrid ARFIMA-SVM. MAE for hybrid ARFIMA-SVM was found to be 460.383 as compare to 506.667 of ARFIMA (2, 15, 2) and 497.673 of hybrid ARIMA-ANN. Hence the performance of hybrid model found to be better than ARFIMA (2, 15, 2) alone.

For the purpose of forecast value of price through hybrid approach, we have got forecast of residuals through the best neural model (03:2s:1l) for January to December in the year 2021. Based on the forecasted value of residuals we found the forecast value of price through hybrid approaches and presented in Table 6.3.7 along with forecast values by ARFIMA (2, 15, 2).

Table 7: Experimental Results of forecast of Price of Onion

Lead Period	Actual values of Price	Forecast Price by ARFIMA (2,0.15,2)	Forecast Price by Hybrid Approach using ANN	% Deviation	Lead Period	Actual values of Price	Forecast Price by ARFIMA (2,0.15,2)	Forecast Price by Hybrid Approach using ANN	% Deviation
1	2400	2264.929	---	---	61	2100	2659.061	2704.821	36.953
2	1300	2326.394	---	---	62	1900	1449.527	1470.102	29.995
3	850	1324.705	---	---	63	1900	1717.810	1747.157	8.044
4	850	935.354	964.187	13.433	64	1450	1870.373	1899.195	0.042
5	850	1104.960	1128.597	32.776	65	1350	2019.148	2054.711	41.704
6	925	1274.113	1301.424	40.694	66	1750	1683.638	1714.109	26.971
7	950	1403.523	1432.682	50.809	67	1750	1582.311	1611.745	7.900
8	1500	1423.831	1451.364	3.242	68	1675	1963.245	1991.127	13.778
9	1675	1863.285	1892.033	12.957	69	1600	2005.069	2033.852	21.423
10	1700	1993.227	2021.194	18.894	70	1600	1893.669	1924.539	20.283
11	1600	1925.938	1956.416	22.276	71	1475	1785.153	1813.585	13.349
12	1600	1761.584	1790.116	11.882	72	1575	1773.473	1801.627	22.144
13	450	1725.523	1753.947	289.76	73	1300	1676.638	1704.779	8.239
14	450	746.964	771.199	71.377	74	1300	1767.254	1796.630	38.2023
15	650	739.729	766.892	17.983	75	1300	1549.935	1577.552	21.350
16	975	1075.632	1097.881	12.603	76	1025	1538.222	1567.303	20.561
17	775	1458.854	1487.867	91.983	77	1125	1565.332	1592.477	55.363
18	900	1288.941	1317.080	46.342	78	1250	1345.256	1372.680	22.015
19	1150	1317.491	1345.701	17.017	79	1250	1427.494	1455.695	16.455
20	1150	1509.684	1535.445	33.517	80	2900	1568.704	1595.657	27.652
21	1050	1494.578	1521.907	44.944	81	2650	1574.442	1602.987	44.724
22	1300	1365.111	1393.221	7.171	82	3300	2967.688	3002.199	13.291
23	1900	1553.346	1581.511	6.763	83	4200	2765.276	2795.599	15.284
24	2050	2080.909	2110.684	2.960	84	4500	3086.592	3126.916	25.549
25	1700	2197.384	2228.765	31.103	85	4250	3763.393	3797.551	15.609
26	2250	1801.161	1832.802	18.542	86	3300	3937.348	3975.781	6.452
27	1350	2190.557	2222.586	64.636	87	1600	3573.368	3613.092	9.487
28	1225	1450.832	1476.691	20.546	88	1275	2645.034	2680.242	67.515
29	1250	1286.755	1318.768	5.501	89	1275	1177.567	1206.628	5.362
30	1675	1401.151	1426.669	14.825	90	1600	1038.093	1066.428	16.358
31	2800	1856.107	1887.015	32.607	91	2175	1379.649	1405.443	12.159
32	4900	2864.019	2898.863	40.839	92	2250	1889.939	1923.017	11.585
33	4900	4633.167	4674.096	4.610	93	2250	2456.226	2490.325	10.681
34	5950	4475.930	4516.126	24.098	94	1725	2469.629	2502.013	11.201
35	2550	4981.814	5029.085	97.219	95	1725	2329.679	2361.674	36.908
36	2300	1878.126	1904.657	17.188	96	1725	1785.667	1813.160	5.111
37	1475	1429.281	1467.444	0.512	97	1350	1745.5	1774.313	2.858
38	1250	1132.538	1148.597	8.112	98	1150	1818.433	1845.835	36.728
39	1090	1276.917	1311.272	20.300	99	1150	1549.806	1579.057	37.309
40	1125	1386.133	1417.281	25.981	100	1150	1387.755	1416.789	23.199
41	1575	1522.607	1553.458	1.367	101	1400	1431.628	1458.512	26.827
42	1600	1937.211	1967.203	22.951	102	1650	1482.424	1509.860	7.847
43	2800	1945.349	1973.943	29.502	103	2050	1715.215	1744.358	5.718
44	2500	2886.853	2921.054	16.842	104	2700	1930.625	1959.876	4.396
45	2650	2589.110	2618.337	1.194	105	4750	2234.271	2265.457	16.094
46	2150	2525.420	2561.491	19.139	106	4600	2735.561	2768.361	41.718
47	2300	2033.522	2061.189	10.383	107	6900	4422.746	4462.241	2.994
48	2250	2135.466	2167.584	3.663	108	8850	4204.389	4241.787	38.524
49	2400	2170.867	2200.748	8.302	109	4000	5840.852	5891.279	33.431
50	2400	2334.301	2368.229	1.324	110	3050	7403.506	7448.589	86.214
51	2400	2351.844	2384.470	0.647	111	1800	2948.265	2987.423	2.051
52	1900	2333.801	2366.809	24.568	112	1850	1619.922	1664.425	7.531
53	1950	1889.612	1919.884	1.544	113	1850	1015.187	1023.765	44.661

54	1950	1918.372	1949.580	0.021	114	1875	1591.442	1626.585	12.076
55	2400	1988.349	2017.232	15.948	115	2175	2021.071	2055.607	9.632
56	2500	2413.846	2447.075	2.116	116	2400	2183.829	2219.950	2.066
57	5700	2517.089	2549.561	55.270	117	4050	2425.788	2458.461	2.435
58	4100	5204.393	5248.598	28.014	118	6150	2566.275	2596.632	35.885
59	3300	3833.218	3865.943	17.149	119	4650	3903.088	3939.517	35.943
60	1975	2264.929	964.186	13.434	120	4150	5660.411	5701.045	22.603

Table 8: MAE of different models for price of onion

Data	ARFIMA	ANN	SVM	ARFIMA-ANN	ARFIMA-SVM
Training	446.406	413.814	398.936	446.767	382.648
Testing	647.276	687.079	640.284	659.587	612.735

From the above table, the value of MAE under training set for different models ARFIMA (2,15,2), ANN (03:2s:1l), SVM, ARFIMA-ANN and ARFIMA-SVM are found to be 446.406, 413.814, 398.936, 446.767 & 382.648 respectively, whereas the value of MAE under testing set are found to be 647.276, 687.079, 640.284, 659.587 & 612.735 respectively. Based on these results, the model ARFIMA-SVM can be recommended for forecasting of price of agricultural commodity because of the minimum value of MAE both under training and testing set.

4. Conclusion

For linear Structure, long memory time series has been analysed by using ARFIMA model. In case long memory time series, ARFIMA model is not always adequate for linear and non-linear dataset. In this context, hybrid methods were evolved for both linear and non-linear part which can be an effective way to improve forecasting performance. Based on the results obtained from this work combines model gives better accuracy for forecasting of price of agricultural commodity. This approach can be further extended by using some other machine learning techniques for varying autoregressive and moving average orders so that practical validity of the model can be well established.

Acknowledgement

The authors would like to thank the Major and Co Guide of Dibrugarh University, Dibrugarh for their help in this work.

Funding: This research received no external funding.

References

1. Aladag CH, Egrioglu E, Kadilar C. Improvement in Forecasting Accuracy Using the Hybrid Model of ARFIMA and Feed Forward Neural Network, American Journal of Intelligent Systems. 2012;2(2):12-17.
2. Granger CWJ, Joyeux R. An Introduction to Long-Range Time Series Models and Fractional Differencing. Journal of Time Series, Anal. 1980;1:15-30.
3. Geweke J, Porter HS. The estimation and application of long memory time series models. Journal of time series analysis. 1983;4(4):221–238.
4. Hosking JRM. Fractional Differencing. Biometrika, Vol.68, No, pp.165-176 Kuensch, H.R. (1987). Statistical aspects of self-similar processes. In Proceedings of the first World Congress of the Bernoulli Society. 1981;1:67–74.
5. Magsood A, Aqil SM. Long memory forecasting an application to KSE share index, Pak. J Statist. 2014;30(3):333-344.
6. Ramalingam S. A unique book on long memory time series is reviewed. J Statist. Comput. Sim. 2010;80(4):475-475.
7. Reisen VA, Lopes S. Some simulations and applications of forecasting long-memory timeseries models. Journal of Statistical Planning and Inference. 1999;80(1):269-287.
8. Reisen V, Abraham B, Lopes S. Estimation of Parameters in ARFIMA Processes: A Simulation Study, Communication in statistics-Simulation and Compution. 2001;30(4):787-803.
9. Robinson PM, *et al*. Gaussian semi-parametric estimation of long-range dependence. The Annals of statistics. 1995;23(5):1630–1661.
10. Robinson PM. Log-periodogram regression of time-series with long-range dependence. Ann. Statist. 1995;23(3):1048-1072.
11. Sowell F. Maximum likelihood estimation of stationary univariate fractionally integrated time series models, Journal of Econometrics. 1992;53:165-188.
12. Sun R, Chen Y, Li, Q. The modeling and prediction of Great Salt Lake elevation time series based on ARFIMA, International Design Engineering technical conferences &computers and information in engineering, Las Vegas, Nevada, USA. 2007.
13. Wiri L, Godwin LT. Autoregressive Fractional Integrated Moving Average (ARFIMA(p,d,q)) Modelling of Nigeria Exchange Rate, Asian Journal of Pure and Applied Mathematics. 2022;4(1):28-35.
14. Zhang GP. Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing. 2003;50:159-175.