A hybrid approach for forecasting mustard price having long-memory property

Rajeev Ranjan Kumar[1], Girish Kumar Jha[2,*], Ronit Jaiswal[1] and Kapil Choudhary[1]

[1]ICAR-Indian Agricultural Statistics Research Institute, New Delhi 110 012, India
[2]ICAR-Indian Agricultural Research Institute, New Delhi 110 012, India

For the modelling of time series data having long memory properties, we generally use the autoregressive fractionally integrated moving average (ARFIMA) model. This model performs well compared to the autoregressive integrated moving average (ARIMA) model. However, it cannot capture the nonlinear property of the data. In order to achieve the desired and accurate forecasts, hybridizing the existing forecasting models is an important technique. The hybrid time-series model combines the strength of individual models. Accordingly, this study proposes a hybrid model based on ARFIMA and extreme learning machine (ELM) for agricultural time-series data with long memory properties. For evaluation of the proposed model, the daily mustard price (₹ /q) of Agra and Bharatpur markets from 1 January 2016 to 31 January 2020 was used. Empirical results show that the forecasting performance of the proposed hybrid model based on ARFIMA and ELM is better than the existing models.

Keywords: Hybrid model, long memory, mustard, price forecasting, time-series data.

THE oilseed sector plays a vital role in the Indian agricultural economy[1,2]. Apart from cereals, oilseeds are one of the most important crops grown in the country[1]. They account for 13% of the gross cultivated area, 3% of the gross national product and 10% of all agricultural commodities in terms of value. In India, mainly nine oilseed crops are grown, of which seven are edible, viz. soybean, groundnut, sunflower, sesame, rapeseed-mustard, safflower and niger, and two are non-edible, viz. castor and linseed. Among the seven edible oilseeds grown in the country, mustard contributes approximately 25% of the total production. It is primarily grown in Rajasthan, Uttar Pradesh, Gujarat, Haryana, Madhya Pradesh and West Bengal[1].

Given the importance of oilseed crops with regard to food security and their significant adverse economic and social effects, it is vital to understand the dynamics of their prices. The prices of oilseed crops are determined by various factors, including adverse weather conditions, natural disasters, shifts in demand and supply (e.g. due to agricultural policy changes), etc. Such variables cannot be quantified using the same norm and have different effects on various oilseed crops in wholesale markets, which further makes forecasting their prices extremely challenging. High persistence over relatively long periods and dependency structure through time are crucial factors in modelling various agricultural price series. These data series are distinguished by their unique non-periodic cyclical patterns. They are behavioural in the sense that current values are influenced not just by recent values, but also by those from earlier periods of time[3]. There has been growing interest in capturing such a feature referred to as long memory. In this study, we model mustard price having as long memory property. For this, a hybrid model has been proposed based on the autoregressive fractionally integrated moving average (ARFIMA) model[4] and extreme learning machine (ELM)[5] model.

For nonstationary time-series data, the autocorrelation function (ACF) gradually decreases, whereas it declines rapidly in the stationary data series[3]. However, even after making them stationary using appropriate differencing, some real-world time series data do not exhibit these properties. In this type of time-series data, there is evidence of a connection between distant data points, and as the number of lags increases, dependence between apart observations decreases gradually. Such a time-series process is known as a fractionally integrated process or long memory process[3]. The first attempt to parameterize the long memory character represented as memory parameter d was by Granger and Joyeux[4].

For the testing of long memory parameters, two major methods of estimation are used in the literature, i.e. parametric and semi-parametric. The parametric method for long memory estimation is computationally expensive and prone to misspecification. On the other hand, a semi-parametric estimate uses d as the main parameter to prevent problems with other parameter descriptions. In addition to parametric and semiparametric methods, heuristic approaches such as the R/S statistic, displaying ACF plots, variance plots, and nonparametric methods such as wavelet methodology are used to test long memory in the time-series data. In this study, we used a semi-parametric Geweke and Porter-Hudak (GPH) test for long memory in the data series[6].

ARFIMA model is used to model series with long memory properties. Integer integration is an extension of fractional integration. A time series should often be integrated with order 0 or 1. However, in the fractional integration technique, the parameter can take any fractional value from 0 to 1. An autoregressive integrated moving average (ARIMA) process of order d, for example (ARIMA(p, d, q)) may be expressed as

$$\phi(B)X_t = (1 - B)^{-d}\theta(B)\varepsilon_t,$$

where ε_t is an independent and identically distributed (*i.i.d.*) random variable with zero mean and constant variance, B the lag operator, and $\phi(B)$ and $\theta(B)$ denote finite autoregressive

*For correspondence. (e-mail: girish.stat@gmail.com)

and moving average polynomials in the lag operator of order p and q respectively.

When parameter $d = 0$, the process is considered stationary; when $0 < d < 0.5$, the process is said to have a long memory. The sum of absolute values of its autocorrelation approaches a constant in the range $-0.5 < d < 0$. In this situation, the process has a negative reliance on distant observations and is hence referred to be 'anti-persistent' or having 'intermediate memory'. It can be demonstrated that the ACF of long memory processes decreases hyperbolically rather than exponentially, as would be expected for stable autoregressive moving average (ARMA) (p, q) models. The value of d determines the decay rate.

Huang *et al.*[5] proposed ELM, a cutting-edge innovative machine learning technique for single-layer feedforward neural networks (SLFNs). The ELM model has been widely used in various fields to solve estimation problems, and it is now gaining popularity in financial time series. The ELM model is simple to use, and no parameters other than the predefined network architecture need to be tuned[7]. It circumvents many of the challenges that gradient-based learning algorithms face, such as the difficulty of learning epochs, learning rate and local minima. ELM is quicker than other learning algorithms, such as artificial neural networks (ANNs) and support vector machine (SVM). In large and complicated applications, which are difficult to achieve with typical neural network models, the ELM technique completes the majority of the training in seconds or minutes. The hidden layer is randomly initialized by ELM to employ nonlinear mapping functions to translate the input data into a feature space. Nonlinear mapping functions in ELM can be any nonlinear piecewise continuous function.

To model the long memory property in the time-series data, the ARFIMA model has been widely used in the literature. It shows good performance compared to the ARIMA model. However, it is unable to capture the nonlinear property of the data. For this we generally used nonlinear models like ANNs, SVM, ELM, etc. ELM shows good generalization power compared to ANNs and avoids the over-fitting problem[8]. So, in this work, we employed the ELM model to detect nonlinear patterns in the data. The ARFIMA and ELM models have succeeded in their respective linear and nonlinear domains. However, none of these is a universal paradigm that can be applied to all situations, because it is extremely difficult to fully understand data features in a real-world scenario. As a result, for practical purposes, a hybrid technique with both linear and nonlinear modelling skills may be a useful option. It is considered that the time-series data X_t may be divided into linear and nonlinear components[9–12].

$$X_t = L_t + N_t,$$

where X_t is the time-series data under consideration, L_t the linear autoregressive component and N_t is the nonlinear component. Accordingly, to fit the linear component of the data series, we used the ARFIMA model and obtained the residuals. To check the existence of nonlinear components in the residuals, we used the Brock–Dechert–Scheinkman (BDS) test. If there is evidence of nonlinearity, then the residuals were modelled using ELM. Let e_t be the residual at time t using the ARFIMA model. Then,

$$e_t = X_t - \hat{L}_t,$$

where $\hat{L}_t$ is the forecast of the ARFIMA model at time t. With k input nodes, the ELM model for the residuals will be

$$e_t = f(e_{t-1}, e_{t-2}, ..., e_{t-k}) + \varepsilon_t,$$

where f is a nonlinear function and ε_t is a random error. Let the forecast from the nonlinear component at time t be $\hat{N}_t$. Then the combined forecast will be

$$\hat{X}_t = \hat{L}_t + \hat{N}_t.$$

As a result, the suggested hybrid technique would use the strengths of the ARFIMA and ELM models.

For the present study, daily mustard prices (₹/q) of Agra and Bharatpur markets from 1 January 2016 to 31 January 2020 were collected from the Agricultural Marketing Information Network (AGMARKNET) (https://agmarknet.gov.in/) website. Figure 1 *a* and *b* shows the time plot of the mustard price series for the Agra and Bharatpur markets respectively. The descriptive statistics of mustard price shows that the average mustard price of Agra and Bharatpur markets is 3696 and 3735 ₹/q respectively. There is not much difference in the average price for both the markets, which could be due to the fact that both are production markets of mustard. The values of standard deviation, minimum, maximum, coefficient of variation, skewness and kurtosis were slightly

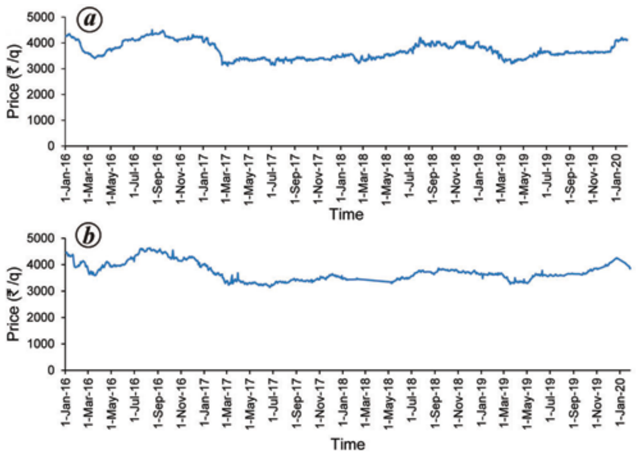

Figure 1. Time plot of daily mustard price in (***a***) Agra market and (***b***) Bharatpur market (1 January 2016 to 31 January 2020).

higher in the Bharatpur market than in the Agra market. The Jarque–Bera test result showed that both series were not normally distributed. In this study, we applied a logarithmic transformation to stabilize the variance of the data.

Autocorrelation functions of both series are highly persistent over long lag and decay slowly towards zero (hyperbolic rate), which indicates the possible presence of long memory properties in the data.

For selecting an appropriate technique for modelling and forecasting the data, it is vital to check whether the time-series data under consideration are linear. If there is solid evidence of nonlinearity in the dynamics of the data-generating process, then in addition to linear models, nonlinear models should also be used for forecasting the data. To test the linearity of the series, we used the BDS test. The test result indicates that both price series are nonlinear. In other words, if a linear model is applied to the mustard price series of both markets, then some hidden structure is left unaccounted for the residuals of the fitted model. Thus, the nonlinear model can be more suitable for forecasting mustard prices in both markets.

To test the long memory, we employed the semiparametric GPH test to the both price series[6]. The GPH estimate of fractional integration parameter (d) was 0.14 and 0.16 for the Agra and Bharatpur markets respectively, and was significant at 5% level of significance.

Long memory and structural breaks are often confused. Several studies have shown that structural breaks may provide a false long memory component in data series[13]. Therefore, we also analysed the possibility of spurious long memory in the data series. We used the Qu test for long memory against spurious long memory[14]. The test statistics were 0.80 and 0.84 for the Agra and Bharatpur markets respectively, and both were insignificant at a 5% level of significance. The results of the Qu test confirm the presence of true long memory in both series.

The suggested technique for modelling and predicting long memory processes is inspired by the fact that agricultural pricing data frequently contains both linear and nonlinear patterns, and no single model can capture all patterns in the data. ARFIMA is a simple method for modelling long memory properties in the data. It is a parametric approach to modelling long memory dynamics[4]. However, this model can only capture the linear component of the data series. To capture the nonlinear properties present in the data series, we used an artificial intelligence technique, viz. ELM. The proposed model uses the strength of the ARFIMA model as well as ELM.

ARFIMA (p, d, q) is defined by the orders of the autoregressive (p) and moving average (q) components of the model, as well as the non-integer order of differencing (d). The parameters are calculated using the greatest likelihood function, which minimizes the total measure of errors. To find the appropriate ARFIMA model, Akaike's information criterion (AIC) is used. ARFIMA (1, 0.11, 1) and ARFIMA (1, 0.13, 1) are the most adequate models for the Agra and Bharatpur markets respectively, based on the AIC values and the concept of model parsimony.

In the next step, residuals were extracted from the best-fitted ARFIMA model. The BDS test was used to determine the presence of a nonlinear pattern in the residuals. The results of the BDS test indicate that the residuals of the chosen ARFIMA model exhibit nonlinear patterns. Therefore, the obtained residuals were modelled through a nonlinear ELM model. To find the best ELM architecture of the residuals, we tested all possible combinations of 1–7 input nodes and 2–10 hidden nodes, and each combination was trained 50 times. The overall average mean square error (MSE) of each ELM model was computed on the testing set. The model which had the lowest average MSE was chosen as optimal. The number of optimal input and hidden nodes were found to be five and seven respectively, for the Agra market. For the Bharatpur market, three input nodes and five hidden nodes were found to be optimal. Finally, the forecasted value of the ARFIMA and ELM models was summed to get the forecast of the hybrid model.

In the process of fitting pure ELM to the original dataset, we found 6–7–1 and 4–6–1 as the optimal architecture for the Agra and Bharatpur markets respectively. Using the root mean square error (RMSE) and mean absolute error (MAE), Table 1 reports the in-sample and out-of-sample forecast evaluation results. For the Agra market, the in-sample RMSE and MAE values of the proposed hybrid model are 0.29 and 0.21 respectively, while the out-of-sample RMSE and MAE values are 25.56 and 19.56 respectively; these values are lower compared to the individual ARIMA and ELM models. Similar results have also been reported for the Bharatpur market. It demonstrates that, in terms of prediction accuracy, the proposed hybrid model outperforms all competing models.

Table 1. In-sample and out-of-sample forecasting performance of ARFIMA, ELM and the proposed Hybrid model

	ARFIMA	ELM	Hybrid model
In-sample			
Agra			
RMSE	1.02	0.42	0.29
MAE	0.93	0.44	0.21
Bharatpur			
RMSE	1.18	0.49	0.39
MAE	0.84	0.34	0.31
Out-of-sample			
Agra			
RMSE	89.67	31.32	25.56
MAE	73.62	21.87	19.39
Bharatpur			
RMSE	96.65	45.81	38.28
MAE	90.43	41.89	34.79

Note: All RMSE and MAE values should be multiplied by 10^{-3}.

Conflict of interest: The authors declare that they have no competing interests.

1. Jha, G. K. *et al.*, *Edible Oilseeds Supply and Demand Scenario in India: Implications for Policy*, Indian Agricultural Research Institute, New Delhi, 2012, pp. 1–6.
2. Renjini, V. R. and Jha, G. K., Oilseeds sector in India: a trade policy perspective. *Indian J. Agric. Sci.*, 2019, **89**, 73–81.
3. Beran, J., *Statistics for Long-Memory Processes*, Chapman and Hall Publishing Inc., New York, USA, 1995, pp. 21–31.
4. Granger, C. W. J. and Joyeux, R., An introduction to long-memory time series models and fractional differencing. *J. Time Ser. Anal.*, 1980, **1**, 15–29.
5. Huang, G. B., Zhu, Q. Y. and Siew, C. K., Extreme learning machine: theory and applications. *Neurocomputing*, 2006, **70**, 489–501.
6. Geweke, J. and Porter-Hudak, S., The estimation and application of long memory time series models. *J. Time Ser. Anal.*, 1983, **4**, 221–238.
7. Huang, G. B., Zhou, H., Ding, X. and Zhang, R., Extreme learning machine for regression and multiclass classification. *IEEE Trans. Syst. Man, Cybern., Part B*, 2012, **42**, 513–529.
8. Wang, J., Lu, S., Wang, S. H. and Zhang, Y. D., A review on extreme learning machine. *Multimed. Tools Appl.*, 2021, 1–50.
9. Chaâbane, N., A hybrid ARFIMA and neural network model for electricity price prediction. *Int. J. Electr. Power Energy Syst.*, 2014, **55**, 187–194.
10. Jha, G. K. and Sinha, K., Agricultural price forecasting using neural network model: an innovative information delivery system. *Agric. Econ. Res. Rev.*, 2013, **26**, 229–239.
11. Jha, G. K. and Sinha, K., Time-delay neural networks for time series prediction: an application to the monthly wholesale price of oilseeds in India. *Neural Comput. Appl.*, 2014, **24**, 563–571.
12. Zhang, H., Nguyen, H., Vu, D. A., Bui, X. N. and Pradhan, B., Forecasting monthly copper price: a comparative study of various machine learning-based methods. *Resour. Policy*, 2021, **73**, 102189.
13. Choi, K. and Zivot, E., Long memory and structural changes in the forward discount: An empirical investigation. *J. Int. Money Financ.*, 2007, **26**, 342–363.
14. Qu, Z., A test against spurious long memory. *J. Bus. Econ. Stat.*, 2011, **29**, 423–438.

ACKNOWLEDGEMENT. We thank the PG School, ICAR-Indian Agricultural Research Institute, New Delhi, and ICAR-Indian Agricultural Statistics Research Institute, New Delhi, for assistance while carrying out this study.

Received 1 August 2022; accepted 16 December 2022

doi: 10.18520/cs/v124/i5/632-635

Agricultural weeder with nail assembly for weed control, soil moisture conservation, soil aeration and increasing crop productivity

Asesh Kumar Ghorai*

B107–108, Pushpakalay, Jagannathpur, Kolkata 700 126, India

Agricultural weeder with nail assembly, popularly known as CRIJAF Nail Weeder, controls germinating and young weeds. It performs best at field capacity (FC) and has low draft (8–12 kg at FC) requirement. Its operation improves soil hydrothermal regimes and aeration (oxygen diffusion rate, 303 μg^{-2} O_2 m^{-2} s^{-1}). It has 5–6 detachable nails, each at 3 cm distance, and has option to attach one scrapper or one tine. Introducing a boat in place of its front wheels and addition of two conical rotors in the mainframe makes it suitable to control weeds in transplanted rice. It requires 12–18 man-days/ha for operation, controls 85–90% weeds, produced 33–40 q/ha jute fibre, 4.5–5 t/ha of upland and transplanted rice, 3.0–4.5 t/ha of wheat and 15 q/ha of mustard. More than 55,000 units have been distributed by the Department of Agriculture, Government of West Bengal.

Keywords: CRIJAF nail weeder, manual weeder, soil air, soil moisture, soil temperature, weed control.

In field and horticultural crops, usually 30–40% of the total cost of cultivation is consumed by the manual weeding process alone. Thereby it minimizes the net income from crop husbandry. Recently, new invasive weeds are creating newer concerns and their management is challenging. The environmental concerns about the use of herbicides in agriculture are well known. Mechanical control of weeds is a viable alternative in the long run. Currently, workforce availability is low during peak hours in the agricultural sector. Hence we have developed an agricultural weeder with nail assembly for simultaneous weeding, thinning, line arrangement and soil mulching in broadcast crops. The fine nails of the weeder scratch the upper surface of the soil and conserve soil moisture (5–15%). This saves the crops from long drought spells and increases water productivity under limited irrigation. It keeps the soil cooler by 1–5°C and increases soil aeration[1]. Using additional components like a scrapper helps weed out established weeds and tine helps in line-making after final soil preparation. Operating the weeder after seed sowing and fertilizer application is helpful in mixing seeds and fertilizers with the soil for proper germination and improving nutrient use efficiency. In jute, it saves up to 100–135 man days/ha depending on weed densities. Reducing workforce requirements in manual

*e-mail: aseshghorai@gmail.com