Indian Journal of Agricultural Sciences **92** (7): 825–30, July 2022/Article
https://doi.org/10.56093/ijas.v92i7.103633

Forecasting price of Indian mustard (*Brassica juncea*) using long memory time series model incorporating exogenous variable

RANJIT KUMAR PAUL[1]*, DIPANKAR MITRA[2], HIMADRI SHEKHAR ROY[1], A K PAUL[1] and M D YEASIN[1]

ICAR-Indian Agricultural Statistics Research Institute, New Delhi 110 012, India

Received: 17 August 2021; Accepted: 06 April 2022

ABSTRACT

The objective of present study was to investigate the efficiency of Autoregressive fractionally integrated moving average model with exogenous input (ARFIMAX) in forecasting price of Indian mustard [*Brassica juncea* (L.) Czern. & Coss]. The daily modal price and arrival data of mustard for two major markets of India, viz. Bharatpur and Agra were collected during 2008–2018 from AGMARKNET and used for the present investigation. It was observed that each of the price series under consideration is stationary but autocorrelation function of both the series decay in a hyperbolic pattern. This indicates possible presence of long memory in the price data. Moreover, the significant result of correlation between price and arrival indicate that arrival data could be used as exogenous variable to model and forecast the price for both markets. Accordingly, Autoregressive fractionally integrated moving average (ARFIMA) and ARFIMAX models were applied to obtain the forecasts. The forecast evaluation was carried out with the help of Relative mean absolute percentage error (RMAPE) and Root mean square error (RMSE). The residuals of the fitted models were used for diagnosis checking as well as to investigate the adequacy of developed model. To this end, a comparative study has also been made between the fitted ARFIMAX model and ARFIMA model for both in-sample and out-of-sample data to identify the best fitted model in order to forecast future prices. The model has demonstrated a good performance in terms of explained variability and predicting power.

Keywords: ARFIMA, ARFIMAX, Correlation, Long memory, Mustard, Stationarity

Price forecasting has a very important implication in current competitive and volatile market. Over last few decades, Box-Jenkins ARIMA model (Box *et al.* 2007) is efficiently used to obtain the forecasts. Even inclusion of exogenous information in the model to increase the forecast accuracy has been tried in literature. Paul *et al*. (2013) applied ARIMA with exogenous variable (ARIMAX) model for forecasting wheat yield in Kanpur district of Uttar Pradesh, India. Paul *et al*. (2014) have investigated the ARIMAX-GARCH (Generalized autoregressive conditional heteroscedastic) model for accommodating the heteroscedasticity in the wheat yield data. The evidence of presence of long memory has often been found in many economic and financial time series including agriculture. Long memory models can be thought of as complementary to the very well-known and widely applied ARMA models. To model time-series data showing long range persistency, ARFIMA model is commonly used. Doornik and Ooms (2007) have applied ARFIMA model for forecasting monthly core consumer price inflation in the US and quarterly overall consumer price inflation in the UK. Baillie and Morana (2012) developed Adaptive ARFIMA models for their applications in modeling Inflation. Paul (2014) and Paul *et al*. (2014) have applied ARFIMA model for forecasting agricultural commodity prices. The ARFIMAX model (Bierens 1987) is a generalization of the ARFIMA model, which is capable of incorporating an exogenous input variable (X).

For the purpose of price forecasting, ARFIMA model has been widely used in past. Mitra *et al*. (2018) investigated long memory model and Structural Break. Mitra and Paul (2021) applied long memory time series model for forecasting Price of Rice in India. India is the third largest rapeseed-mustard producer in the world after China and Canada with 12% of world's total production. This crop accounts for nearly one-third of the oil produced in India, making it the country's key edible oilseed crop. Therefore, forecasting the price of this crop is very important. In this study, ARFIMAX model has been applied to forecast the prices of mustard in two markets of India namely Bharatpur and Agra by considering market arrival as an exogenous variable.

MATERIALS AND METHODS

Long memory model: Long memory models are

[1]ICAR-Indian Agricultural Statistics Research Institute, New Delhi; [2]The Ministry of Statistics and Programme Implementation, Government of India, New Delhi. *Corresponding author email: ranjitstat@gmail.com

statistical models that describe strong correlation or dependence across time series data. This kind of phenomenon is often referred to as "long memory" or "long-range dependence". It refers to persisting correlation between distant observations in a time series. For scalar time series observed at equal intervals of time that are covariance stationary, so that the mean, variance, and autocovariances (ACV) (between observations separated by a lag j) do not vary over time, it typically implies that the ACV decay so slowly, as j increases, as not to be absolutely summable. However, it can also refer to certain non-stationary time series, including ones with an autoregressive unit root, that exhibit even stronger correlation at long lags (Robinson 2003).

Long memory models can be thought of as complementary to the very well-known and widely applied stationary and invertible autoregressive and moving average (ARMA) models, whose ACV are not only summable but decay exponentially fast as a function of lag j (Beran 1995). Such models are often referred to as "short memory" models, because there is negligible correlation across distant time intervals. These models are often combined with the most basic long memory ones, however, because together they offer the ability to describe both short and long memory features in many time series.

Let y_t (t = 1, 2, ..., n) be an equally spaced, real valued and covariance stationary time-series process so that the mean

$$\mu = E(y_t) \tag{1}$$

and lag–k ACV (or variance when k = 0

$$\lambda(k) = Cov(y_t, y_{t+k}) \tag{2}$$

do not depend on t.

Further consider that the autocorrelation function of the time-series with a time lag of k is given as

$$\rho_k = Cov(y_t, y_{t+1})/var(y_t) \tag{3}$$

The series y_t (t = 0, 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. GPH (Geweke and Porter-Hudak) test is the most commonly used test for testing long memory process.

ARFIMA Model: The Autoregressive fractionally integrated moving-average (ARFIMA) model is used for modeling time-series in presence of long memory. Fractional integration is a generalization of integer integration. In integer integration a time-series is usually presumed to be integrated of order zero or one but in case of fractional integration process the parameter can take any fractional value from 0 to 1. For example, an autoregressive moving-average process integrated of order d [denoted ARFIMA (p, d, 1)] can be represented as

$$\phi(B)\,(1-B)^d y_t = \theta(B) e_t \tag{4}$$

where e_t is an independently and identically distributed (*i.i.d.*) ran dom variable having zero mean and constant variance; B denotes the lag operator; ϕ(B) and θ(B) denote finite AR and MA polynomials in the lag operator of order p and q respectively having roots outside the unit circle and for any value of d we have

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

with binomial coefficients

$$\binom{d}{j} = \frac{d!}{j!(d-j)!} = \frac{\Phi(d+1)}{\Phi(j+1)\Phi(d-j+1)}$$

where Φ (.) represents the gamma function.

For $d \in$ (0,0.5), the ARFIMA process is said to exhibit long memory or long range positive dependence.

For $d \in$ (–0.5,0), the process exhibits intermediate memory or long range negative dependence.

The process is short memory for d = 0 corresponding to a standard ARMA process.

ARFIMAX Model: The fractionally integrated autoregressive moving average with exogenous variable (ARFIMAX) models was used for forecasting in the context of realized volatility analysis (Degianankis 2008).

The general for of ARFIMAX (p, d, q) model is written as:

$$\phi(B)\,(1-B)^d (y_t - x'_t \beta) = \theta(B) e_t \tag{5}$$

where ϕ (B), B, θ (B), e_t are defined earlier; x_t is a vector of explanatory variables and β is a vector of unknown parameters. The parameters of this models are estimated by minimizing the Schwarz's Bayesian criterion (SBC) criterion. SBC of an ARFIMAX (p, d, q) is computed as $\{y_t,\ t = 1, 2, ..., T\}$, can be represented as:

$$\text{SBC} = -2T^{-1} L_T\left(\{y_t\}; \hat{\psi}^{(T)}\right) + \breve{\psi}\, T^{-1} log(T) \tag{6}$$

where $\psi = (d, \varphi_1, \varphi_2, ..., \varphi_p, \varphi_1, ..., \theta_1, \beta)$; L_T(.), maximized value of the log-likelihood function; $\hat{\psi}^{(T)}$, maximum likelihood estimator of ψ based on a sample of size T and $\breve{\psi}$ denotes the dimension of ψ.

RESULTS AND DISCUSSION

For the present study daily arrival (q) and average modal price (₹/q) data of mustard for Bharatpur and Agra markets were collected from AGMARKNET (https://agmarknet.gov.in/). The bifurcation of the price and arrival data for these two markets as training and test data set is given as:

Market	Training set	Test set
Bharatpur (1761 observations)	23rd December, 2008 to 26th December, 2017 (1581 observations)	27th December, 2017 to 8th December, 2018 (180 observations)
Agra (2864 observations)	7th January, 2008 to 4th April, 2018 (2664 observations)	4th April, 2018 to 11th December, 2018 (200 observations)

A perusal of (Table 1) indicates that average price remains high in Bharatpur as compared to Agra market. The maximum price of mustard reached to more than ₹4900/q in Bharatpur market and ₹4500/q in Agra market. The variability in price as measured in terms of coefficient of variations (CV) reveals that Agra market has higher price variability than Bharatpur market; but in arrival, Bharatpur market has higher variability than Agra market. Fig 1 (first plot) depicts daily arrival (in tonnes) and average modal price (₹/q) of mustard in Bharatpur market of India for the period 23 December 2008 to 8 December 2018.

Table 1 Summary statistics of price and arrival in studied markets

Statistic	Agra		Bharatpur	
	Arrival (t)	Price ₹ (q)	Arrival (t)	Price ₹ (q)
Mean	772.75	3224.84	150.18	3482.20
Median	820.00	3240.00	115.70	3443.00
Maximum	1800.00	4500.00	1106.00	4908.00
Minimum	9.90	1800.00	0.80	2199.00
SD	349.80	643.52	127.79	580.36
Skewness	0.11	0.02	2.37	-0.05
Kurtosis	2.48	1.95	11.31	2.89
CV (%)	45.27	19.95	85.10	16.67
Observations	2864	2864	1761	1761

Considering the overall movement of the time series data there is a little average increment in the price data while for the arrival dataset the overall pattern is nearly decreasing.

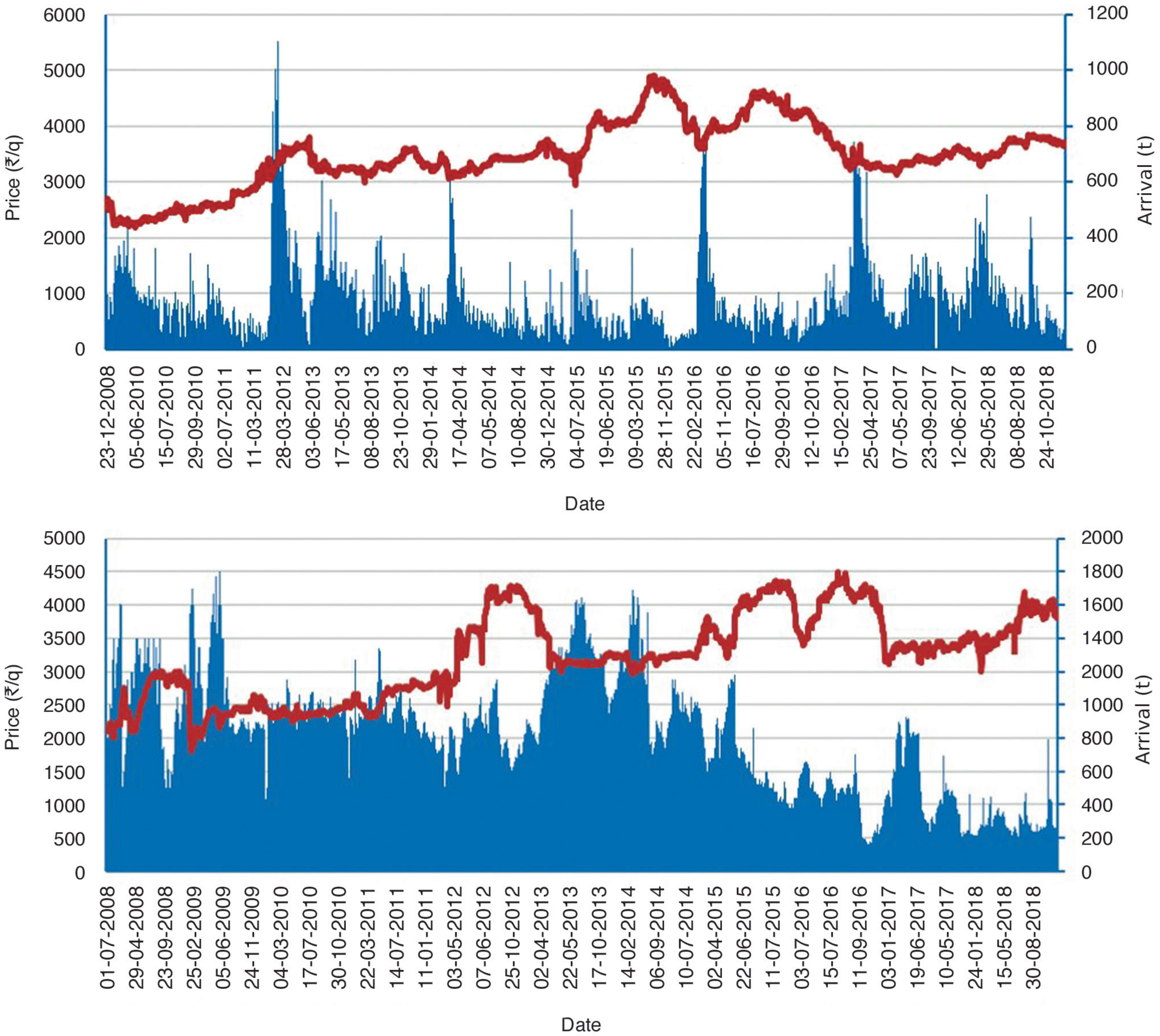

Fig 1 Pattern of arrival and price in Bharatpur (first plot) and Agra market (second plot).

In parallel fashion, Fig 1 (second plot) represents daily arrival (in tonnes) and average modal price (₹/q) of mustard in Agra market of India for the period 7 January, 2008 to 11 December, 2018. This figure shows that the general tendency of price data is increasing while the arrival is decreasing over the years depicting negative correlation with the arrival data.

Test for correlation: The estimated values of correlation coefficients between price and arrival data for both markets are found out to be -0.185 and -0.307 respectively in Bharatpur and Agra. The price and arrival of mustards for both markets are negatively correlated which support the conventional economic theory of price and supply. Further these correlations are found to be significant at least 1% level of significance implying that the arrival data could be used to model and forecast the price using their dependency.

Test for stationarity: The stationarity of the series under consideration are ensured by applying Augmented Dickey Fuller (ADF) test, Phillip-Perron (PP) unit root test and Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test. In ADF and PP test, the null hypothesis is: The series under consideration is nonstationary whereas in KPSS test, the null hypothesis is: The series under consideration is stationary. All of these tests confirm that both the arrival and price data for Bharatpur as well as Agra markets have no unit root i.e. they represent stationary processes.

ACF plot: After testing the statistical associationship between the arrival and price data the next step was to see the statistical dependency between the price data which was achieved through observing plots of Autocorrelation function (ACF). The ACF plots of the price data for Bharatpur and Agra markets indicated slow hyperbolic decay of autocorrelation towards zero indicating possible presence of long memory. Since the autocorrelation functions are significant at distant lags (even after 400 lags), there is a clear indication of long memory in the price data.

Fitting ARFIMAX and ARFIMA model: ARFIMAX model is fitted to the prices data sets for both markets in order to capture the long memory pattern and the influence of arrival data. The best model is selected based on minimum Akaike Information Criterion (AIC) and Bayesian Information Criteria (BIC) value. Accordingly, ARFIMA model is also fitted to the price dataset and the best model is chosen on the basis of minimum AIC and BIC value. The parameter estimates of ARFIMAX and ARFIMA models for both Bharatpur and Agra markets along with their standard errors (in brackets) are given in Table 2. A perusal data indicates that the long memory parameter is significant at 5% level of significance in both ARFIMAX and ARFIMA models for two markets. Moreover, The AR and MA parameters are also significant at 0.1% level of significance in ARFIMAX and ARFIMA models. The explanatory variable (X) is significant in ARFIMAX model for Bharatpur market but, in case of Agra market it is not significant implying that ARFIMAX model is not the good fitted model in case of Agra market data.

Forecast evaluation and comparison: One-step ahead forecast of price for the last 180 observations (Bharatpur market) and for last 200 observations (Agra market) was calculated using the best selected ARFIMAX and ARFIMA model. The accuracy of the fitted models was measured in terms of Relative Mean Absolute Percentage Error (RMAPE) and Relative Mean Squared Error (RMSE) criteria as:.

$$\text{RMAPE} = \frac{1}{h}\sum_{(t=1)}^{h} |y_t - \hat{y}_t| \, y_t \times 100 \tag{7}$$

Table 2 Parameter estimates of ARFIMAX and ARFIMA models

Parameter	ARFIMAX(1,d,1)		ARFIMA (1,d,1)	
	Bharatpur	Agra	Bharatpur	Agra
Intercept	3592.860 (423.674)***	4028.729 (462.238)***	3461.330 (376.298) ***	3180.590 (289.563)***
d	0.125 (0.062)*	0.124 (0.062)*	0.114 (0.053)*	0.135 (0.068)*
X	-0.098 (0.035)**	-0.248 (0.153)		
AR(1)	0.996(0.002)***	0.997 (0.089)***	0.996 (0.002)***	0.770 (0.043)***
MA (1)	0.736 (0.051)***	0.695 (0.078)***	0.723 (0.053)***	0.521 (0.031)***
AIC	14688.400	24170.600	14694.000	23861.200
BIC	14720.600	24206.000	14720.800	23742.920
Forecast performance of two models				
	RMAPE		*RMSE*	
	In sample	*Out-of-sample*	*In sample*	*Out-of-sample*
ARFIMA Model				
Bharatpur	0.014	0.048	103.562	233.985
Agra	0.012	0.083	93.068	364.276
ARFIMAX Model				
Bharatpur	0.014	0.043	103.813	227.075
Agra	0.012	0.080	93.526	357.223

* Significant at 5%; ** significant at 1%; *** significant at 0.1%; the value in the parenthesis is the SE of coefficient.

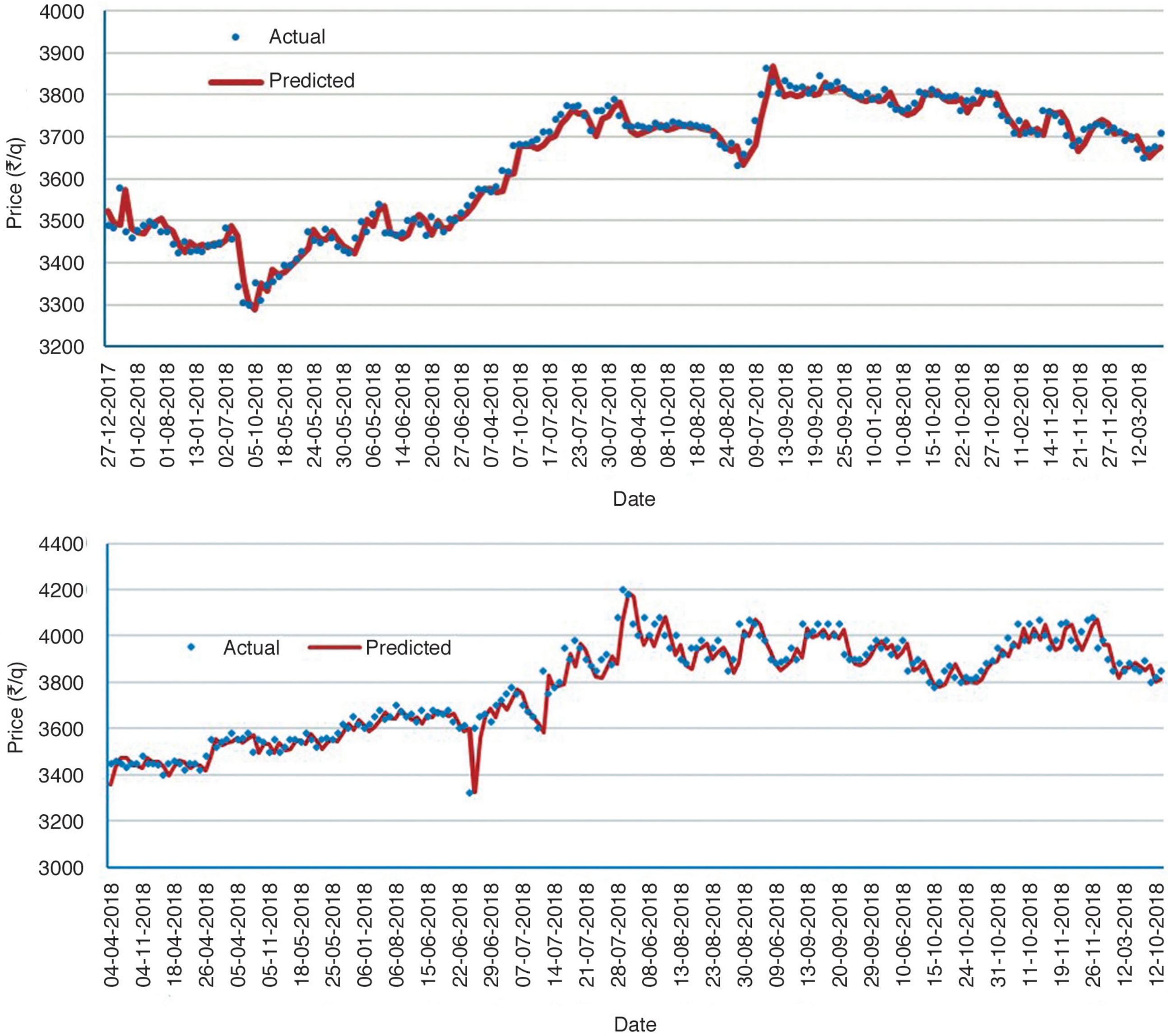

Fig 2 Actual vs predicted plot in Bharatpur (first plot) and Agra (second plot) Market.

$$\text{RMSE} = \sqrt{\frac{1}{h}\sum_{t=1}^{h}\left(y_t - \widehat{y_t}\right)^2} \qquad (8)$$

RMAPE and RMSE was calculated for both in sample and out-of-sample part considering the best fitted ARFIMAX and ARFIMA (Table 2). Data indicates that for in sample part both ARFIMA and ARFIMAX models perform similarly over different markets. While for out-of-sample part ARFIMAX model performs better than the fitted ARFIMA model in both Bharatpur and Agra market. The prediction performance of the fitted model has been displayed in Fig 2 for the test dataset for both the markets.

Application of long memory test reveals that both the price series exhibit long memory, accordingly, ARFIMA and ARFIMAX models are fitted to each of the price series considering arrival of mustard in that market as an explanatory variable and the best model was selected on the basis of minimum AIC and BIC value. Long memory series was characterized by their ability to remember distant past and their ability to make decisions on the basis of such memories. The study has revealed that the long memory models, e.g. ARFIMA and ARFIMAX could be used successfully for modelling as well as forecasting of mustard in different markets. The model has demonstrated a good performance in terms of explained variability and predicting power. In general, it is true that the price and arrival of a commodity in markets are negatively correlated. In the present analysis it was also observed that the price and arrival of mustard in both the studied markets are negatively correlated. Forecast accuracy and comparison among the fitted ARFIMAX and ARFIMA models have been carried out in terms of lower RMAPE and RMSE criteria. In case of Bharatpur market lower RMAPE and RMSE values for out-of-sample part indicate better performance of the fitted ARFIMAX models over ARFIMA model. The findings of the present study provides direct support for the potential use of accurate forecasts in decision making for the wholesalers, retailers, farmers as well as consumers. While for Agra market ARFIMA model performs better in out-of-sample part in terms of lower RMSE and RMAPE values than ARFIMAX model due to the fact that coefficient corresponding to explanatory variable included in the model was not statistically significant. Moreover, the adequacy of the model was checked in terms of examining the residuals. The mean of the residuals has been found out to be approximately zero. The distribution of residuals was found out to be Gaussian and there was no autocorrelation. Further study may be carried out for incorporation of more number of exogenous variables along with capturing nonlinearity present in price series.

REFERENCES

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

Baillie R T and Morana C. 2012. Adaptive ARFIMA models with applications to Inflation. *Economic Modelling* **29**(6): 2451–59.

Box G E P, Jenkins G M and Reinsel G C. 2007. *Time-Series Analysis: Forecasting and Control,* 3rd edn. Pearson Education, India.

Doornik J A and Ooms M. 2007. Inference and forecasting for ARFIMA models with an application to US and UK inflation. *Studies in Nonlinear Dynamics and Econometrics* **8**(2): 1–23.

Mitra D, Paul R K, Paul A K and Bhar L M. 2018. Forecasting time series allowing for long memory and structural break. *Journal of the Indian Society of Agricultural Statistics* **72**(1): 49–60

Mitra D and Paul R K. 2021. Forecasting of price of rice in India using long memory time series model. *National Academy of Science Letter* **44**: 289–93

Paul R K, Prajneshu and Ghosh H. 2013. Statistical modelling for forecasting of wheat yield based on weather variables. *Indian Journal of Agricultural Science* **83**(2): 180–83

Paul R K. 2014. Forecasting wholesale price of pigeon pea using long memory time-series models, *Agricultural Economics Research Review* **27**(2): 167–76.

Paul R K, Gurung B and Paul A K. 2014. Modelling and forecasting of retail price of arhar dal in Karnal, Haryana. *The Indian Journal of Agricultural Sciences* **85**(1): 69–72.

Paul R K, Ghosh H and Prajneshu. 2014. Development of out-of-sample forecast formulae for ARIMAX-GARCH model and their application. *Journal of the Indian Society of Agricultural Statistics* **68**(1): 85–92.

Robinson P M. 2003. *Time Series with Long Memory*. Oxford university press, Oxford.