Journal of Oilseed Brassica, 8 (2) : 181-186, July 2017

Price forecasting of mustard using ARIMA and EGARCH models

Priyanka Solanki[1], Rajesh Sharma[2] and Madhu Sharma[3*]

[1]*Institute of Agri Business Management, SKRAU, Bikaner 334006, Rajasthan, India*
[2]*College of Agriculture, SKRAU, Bikaner 334006, Rajasthan, India*
[3] *Directorate of Planning, Monitoring and Evaluation, SKRAU, Bikaner 334006, Rajasthan, India*
**Corresponding author: madhu.sharma.ae@gmail.com*
(Received: 21 April 2017; Revised: 14 May 2017; Accepted: 14 June 2017)

Abstract

Autoregressive Integrated Moving Average (ARIMA) and Exponential GARCH (EGARCH) model was studied along with their estimation procedures for modelling and forecasting of mustard price. For forecasting mustard price ARIMA (0,1,1) model is used which gives reasonable and acceptable forecasts but the study has revealed that the AR(1)-EGARCH(1,1) model outperformed the price forecasting models for mustard prices primarily due to its ability to capture asymmetric volatility pattern.

Key words: *ARIMA, EGARCH, mustard, forecasting, price*

Introduction

Mustard, Brassica species, is an annual, cool season crop that is native to the temperate regions of Europe and one of the first domesticated crops. Mustard/ Rapeseed cultivation is done widely throughout the world (Oplinger *et al*, 1991).

Mustard crop accounts for nearly one third of the oil produced in India, making it the country's key edible oilseed crop. Due to the gap between domestic availability and actual consumption of edible oils, India has to resort to import of edible oils. It is the major source of income especially even to the marginal and small farmers in rainfed areas. Since these crops are cultivated mainly in the rainfed and resource scarce regions of the country, their contribution to livelihood security of the small and marginal farmers in these regions is also very important. By increasing the domestic production substantial import substitution can be achieved. So, the crop has the importance for farmers as well as for the nation. Accurate forecasting about the prices will help the farmer to plan the area under the crop and the traders to plan their decisions.

Prices of the agricultural commodities are important both economically and politically in almost all countries. Agricultural commodity prices strongly influence not only the farmers' income but also consumers, agri business industry and policy makers as they are quite volatile in nature. India has a long history of policies aimed at smoothing out the price volatility for the consumers and income volatility for the farmers. But now there is need to understand the complexity of commodity price dynamics that is more urgent against the backdrop of current tendencies to remove the traditional schemes to sustain in the globalized markets. To capture these unforeseen variations in the prices of the agricultural commodities accurate forecasting models are extremely important for efficient planning and monitoring. Over the period, there have been continuous refinements in price forecasting models so that more and more accurate price forecasting can be done for the benefit of farmers and other organisations. The study on finding a best suitable method out of existing advance models of price forecasting is a useful exercise for planners, agriculture departments and other stake holders working for price forecasting.

Thus, the present study was an attempt to identify the best suited model for the price forecasting of mustard in the Tonk district of Rajasthan.

Materials and Methods

The secondary data of monthly wholesale mustard prices for Tonk *mandi* were collected from the AGMARKNET site. The data of the mustard prices for the period from January 2006 to February 2016 was utilized for the analysis purpose.

Auto Regressive Integrated Moving Average (ARIMA) model and Generalized Auto Regressive Conditional Heteroscadasticity (GARCH) model were used to identify the best fitted model for the mustard crop. Exponential GARCH model was also used as it is advance form of GARCH model. The details of these models are as follows:

Auto Regressive Integrated Moving Average (ARIMA) model

Box and Jenkins introduced this procedure in the year 1970. The set of models introduced by them are popularly known as ARIMA models. This technique is used to forecast future values of a series based on completely its own past values. ARIMA models are most popular models for forecasting a time series which can be made to be 'stationary' by differencing if necessary required. A random variable that is a time series is stationary if its statistical properties are all constant over time.

ARIMA methodology attempts to describe the movements in a stationary time series as a function of 'Autoregressive and Moving Average' parameter. These are referred to as AR parameter (Autoregressive) and MA parameters (Moving Averages).

Autoregressive (AR) Model

An AR model with only single parameter may be written as

$$Y_{(t)}=A_{(1)}*Y_{(t-1)}+E_{(t)}$$

Where $Y_{(t)}$= time series under investigation; $A_{(1)}$= the autoregressive parameter of order 1; $Y_{(t-1)}$=the time series lagged 1 period; $E_{(t)}$=the error term of the model.

This simply means that any given value $Y_{(t)}$ can be explained by some function of the previous value, $Y_{(t-1)}$ plus some unexplainable random error $E_{(t)}$.

Moving Average Models

A second type of box-Jenkins model is called a moving average model. Although these models look very similar to the AR model, the concept behind them is quite different. Moving average parameter relate what happens in period t only to the random error that occurred in past time period, i.e. $E_{(t-1)}$ $E_{(t-2)}$ etc. rather than to $Y_{(t-1)}$,$Y_{(t-2)}$,$Y_{(t-3)}$ as in the autoregressive approaches.

A moving average model with one MA term may be written as follows:

$$Y_{(t)}=-B_{(1)*}E_{(t-1)+}E_{(t)}$$

The term $B_{(1)}$ is called an MA of order 1. The negative sign in front of the parameter is used for convention only. The above model simply says that any given value of $Y_{(t)}$is directly related only to the random error in the previous period, $E_{(t-1),}$and to the current error terms, $E_{(t)}$.

Mixed Models (ARIMA)

ARIMA methodology also allows model to be built that incorporate both autoregressive and moving average parameters together. These models are often referred to as "mixed model" although this makes for a more completed forecasting tool, the structure may indeed simulate the series better and produce a more accurate forecasting. Pure model imply that the structure consists only of AR or MA parameter- not both. The model developed by this approach are usually called ARIMA model because they use a combination of autoregressive (AR) integration i.e., referring to the reverse process of differencing to produce the forecasting, and moving average (MA) operation. An ARIMA model is usually stated as ARIMA (p,d,q) this represent the order of the autoregressive component (p), the number of differencing operators (d), and the highest order of the moving average terms (q).

Generalized Autoregressive Conditional Heteroscedastic (GARCH) model

Autoregressive conditional heteroscedastic (ARCH) model, was introduced by Engle in 1982. ARCH models are quite useful in analyzing the time series data which exhibit volatility or clustering and are characterized by varying variance. This model allows the conditional variance to change over time as a function of squared past errors leaving the unconditional variance constant. The presence of ARCH type effects in financial and macro-economic time series is well established fact. The combination of ARCH specification for conditional variance and the Autoregressive (AR) specification for conditional mean has many appealing features, including a better specification of the forecast error variance.

The ARCH (q) model for series (ε_t) is defined by specifying the conditional distribution of $_t$ given information available up to time t.

The process (ε_t) is ARCH (q), if the conditional distribution of (ε_t) given available information $_{t-1}$ is

$$(\varepsilon_t)\ \psi_{t-1} \sim N(O, h_t) \text{ and } h_t = a_0 + \sum_{i=0}^{q} a\ \varepsilon^2_{t-1}$$

Where, $a_o > 0$, $a_i > 0$ for all I and $\sum_{i=0}^{q} a < 1$

Generalized ARCH (GARCH) Model

In order to overcome the limitation of the ARCH model, Bollerslev (1986) and Taylor (1986) independently proposed the Generalized ARCH (GARCH) model in which conditional variance is also a linear function of its own lags. This model is also a weighted average of past squared residuals but it has declining weights that never go completely to zero. It gives parsimonious models that are easy to estimate and even in its simplest form, has proven surprisingly successful in predicting conditional variances. A general GARCH model has the following functional form:

$$\varepsilon_t = \epsilon_t h_t^{1/2} h_t$$

$$h_t = a^0 + \sum_{i=0}^{q} a_i\ \varepsilon^2_{t-1} + \sum_{j=1}^{p} b_j h_{t-j}$$

Exponential GARCH (EGARCH) Model

The exponential GARCH or EGARCH model was first developed by Nelson (1991), and the logarithm of conditional variance for this model is given by:

This specification makes the effect exponential instead of quadratic and therefore, the estimates of the conditional variance are guaranteed to be non-negative. The EGARCH model allows for the testing of asymmetries.

Forecasting Accuracy Measure

To compare the accuracy of models Mean Absolute Percentage Error (MAPE) is used. MAPE measures the absolute error as a percentage of actual value rather than per period. It usually results in elimination of the problem for interpreting the measure of accuracy relative to the magnitude of the actual and forecast values, as MAD does.

$$\text{MAPE} = \frac{(\text{Sum} |X_t - F_t|)}{\text{Sum}(X_t)}$$

Where, X_t is the actual value; F_t is the forecasted value

Results and Discussion

ARIMA and GARCH models are used for price forecasting and based on the MAPE value best model is selected.

ARIMA Model

Testing the stationarity: The first step for applying the ARIMA model is to check whether the series is stationarity or not. By examining the visual inspection

Table 1: Stationarity Test for checking White Noise

Autocorrelation Check for White Noise									
To Lag	Chi-Square	DF	Pr>ChiSq	Autocorrelations					
6	536.23	6	<.0001	0.962	0.912	0.859	0.809	0.755	0.703
12	788.85	12	<.0001	0.653	0.605	0.563	0.531	0.508	0.485
18	910.01	18	<.0001	0.454	0.416	0.380	0.353	0.332	0.313
24	973.50	24	<.0001	0.296	0.278	0.263	0.255	0.250	0.248

of the autocorrelation function indicated that the mustard price series is non stationary, since the ACF decays very slowly. The result for the stationarity test is given in table 1. In this case, the white noise hypothesis is rejected based on the autocorrelation test. The p value for the test of the first twenty four autocorrelations is observed as <0.0001, which significantly rules out the assumption of stationarity of the series. Therefore, the data series is non-stationary in nature.

Table 2: Augmented Dickey-Fuller Unit Root Tests Values after single order differencing

Lags	Rho	Pr < Rho
1	-136.223	0.0001
5	-53.5592	<.0001

As the series is non-stationary, the next step is to transform it to a stationary series by first differencing. After first differencing, Augmented Dickey-Fuller procedure was used to test the null hypothesis that means data series is non-stationary in nature and the alternate hypothesis depicts the series is stationary in nature. As p-value < 0.05 that means the null hypothesis is rejected which conclude that data series is stationary. The result of Augmented Dickey Fuller Unit Root Test is shown in table 2.

Once the mustard price series has become stationary after first differencing then different models for AR and MA combination were estimated and the model with minimum AIC and SBC was selected. After comparing various ARIMA models, ARIMA (0,1,1) model was selected. The t value provides significance of the tests for the parameter estimates and indicates whether some terms in the model may be unnecessary. In this case, the value for the moving average is 3.43 which is highly significant as shown in the table 3. Then the forecasted price of the mustard using ARIMA (0,1,1) was estimated as given in the table. The Mean Absolute Percentage Error of the ARIMA (0,1,1) is 6.2.

GARCH Model

Testing the ARCH Effects: The Q statistics test was performed for analysing the changes in variance across time using lag windows, ranges from 1 through 12 as shown in table 4. Since the p-value for the test statistics are less than 0.0001 for all lag windows, it strongly indicates heteroscedasticity. The Lagrange Multiplier (LM) test results shown below in table 4 can help in determining the order of ARCH Model appropriate for the data. The tests are significant (p<.0001) through order 12, indicates that price series are volatile and need to be modelled using ARCH or GARCH models.

Table 4: Q and LM Tests' result for ARCH Disturbances

Tests for ARCH Disturbances Based on OLS Residuals				
Order	Q	Pr > Q	LM	Pr > LM
1	103.9830	<.0001	96.8511	<.0001
2	180.0565	<.0001	97.9534	<.0001
3	232.7259	<.0001	97.9602	<.0001
4	271.4314	<.0001	98.2609	<.0001
5	297.0415	<.0001	98.8151	<.0001
6	311.7927	<.0001	99.5941	<.0001
7	318.7187	<.0001	99.7814	<.0001
8	321.7404	<.0001	99.8258	<.0001
9	323.5150	<.0001	100.3399	<.0001
10	324.7409	<.0001	100.3735	<.0001
11	325.5968	<.0001	100.3740	<.0001
12	325.9457	<.0001	100.8275	<.0001

The basic ARCH (q) model is a short memory

Table 3: Estimate Statement Output of ARIMA (0,1,1)

Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	ApproxPr > |t|	Lag
MU	19.89894	17.91986	1.11	0.2690	0
MA1,1	-0.30097	0.08763	3.43	0.0008	1

Table 5: Parameter Estimates for AR (1) - GARCH (1,1)

Parameter Estimates					
Variable	DF	Estimate	StandardError	t Value	ApproxPr > |t|
Intercept	1	1468	3002	0.49	0.6247
AR1	1	-1.0057	0.0127	-79.05	<.0001
ARCH0	1	26043	4633	5.62	<.0001
ARCH1	1	0.1030	0.2013	0.51	0.6090
GARCH1	1	0	0		

process in which only recent q squared residuals are used to estimate the changing variance. The GARCH model (p>0) allows long range memory processes, which use all the past squared residuals to estimate the current variance. The LM test suggests the use of GARCH model would be appropriate instead of the ARCH model.

Different ARCH and GARCH model were calculated and their parameter estimates are given in table 5. The t-values in the GARCH estimates of GARCH 1 was not available indicating that the data was not following GARCH.Similarly other combinations of AR and GARCH were also computed but the data series was not following this model. Then the next step was to test the EGARCH model.

EGARCH Model

The result of EGARCH model is shown table 6.

Table 6: Test Results for AR (1) – EGARCH (1,1)

Exponential GARCH Estimates			
MAE	113.904246	Observations	122
MAPE	4.06918443	Total R-Square	0.9408
		AIC	1613.1579
		Normality Test	601.4378
		Pr>ChiSq	<.0001

Table 7: Parameter Estimates for AR (1) - EGARCH (1,1)

Parameter Estimates					
Variable	DF	Estimate	StandardError	t Value	ApproxPr > |t|
Intercept	1	2751	241.1768	11.41	<.0001
AR1	1	-0.9669	0.0314	-30.77	<.0001
EARCH0	1	21.0156	0.6367	33.01	<.0001
EGARCH1	1	-0.9889	0.0507	-19.52	<.0001

Table 8: Actual and Forecasted price of mustard by AR (1) - EGARCH (1,1) model (‘/quintal)

Month	Actual Price	Forecasted Price using ARIMA (0,1,1)	Forecasted Price using AR (1) - EGARCH (1,1)
January 2016	4066.44	4545.03	4468.60
February 2016	3893.04	3942.30	4022.91
March 2016		3898.11	3855.25
April 2016		3918.01	3818.71
May 2016		3937.91	3783.38

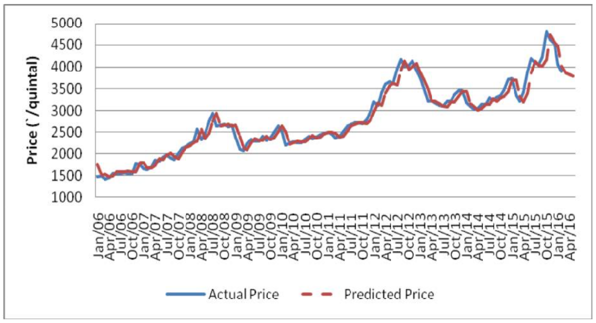

Figure 1: Graph showing actual and forecasted prices of mustard by AR (1)-EGARCH (1,1) model

The EGARCH estimates show that the Mean Absolute Error (MAE) is approximately 113 and Mean Absolute Percentage Error (MAPE) in the given price data series is approximately 4. The t-value in the EGARCH estimates, as shown in table 7 of EGARCH, is significant at <.0001 indicating that the EGARCH1 is the best suited model.

The forecast obtained from applying AR(1) - EGARCH (1,1) are given in the table for month of March to May 2016 in table 8. The graph shown in figure 1 depicts the actual versus the forecasted values. On the basis of comparison of MAPE of both the models, AR (1)-EGARCH (1,1) is best-fitted model.

Conclusions

The performance of ARIMA and EGARCH hasbeen studied using monthly wholesale price of the mustard. The EGARCH model has forecasted the volatility better than the ARIMA model. EGARCH was employed in addition to ARCH and GARCH models in order to capture asymmetry pattern of the data. The EGARCH model has outperformed the various models for the present data set as far as modelling and forecasting is concerned. Hence, the empirical results have supported the theory that EGARCH model can capture asymmetric volatility and therefore is more suitable model for price forecasting of mustard prices in Tonk district of Rajasthan.

References

AGMARKNET. Retrieved from http://agmarkweb.dacnet.nic.in/SA_Month_PriMar.aspx.

Bollerslev T. 1986. Generalized autoregressive conditional heteroscedasticity. *J Econom* **31**: 307- 327.

Box GEP and Jenkins G. 1970. Time Series Analysis. In: Forecasting and Control. San Francisco, CA, Holden-Day.

Engle RF. 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of UK inflation. *Econometrica* **50**: 987-1008.

Nelson DB. 1991. Conditional heteroskedasticity in asset returns: a new approach, *Econometrica* **59**: 347–370.

Oplinger ES, Oelke EA, Putnam DH, Kelling KA, Kaminsid AR, Teynor TM, Doll JD and Durgan BR. 1991. Mustard. In: Alternative Field Crops Manual. Retrieved from: https://hort.purdue.edu/newcrop/afcm/mustard.html

Taylor SJ. 1986. Forecasting. In: Modeling financial time series. New York, John Wiley & Son, pp 234-239.