Original Research Article

MODELING AND FORECASTING OF ALL INDIA MONTHLY AVERAGE WHOLESALE PRICE VOLATILITY OF ONION: AN APPLICATION OF GARCH, EGARCH AND TGARCH MODELS

ABSTRACT

The study utilized log returns of all India monthly average wholesale prices (Rs/Q) of onion over period Jan-2010 to Dec-2021 and employed the autoregressive moving-average (ARMA), generalized autoregressive conditional heteroscedastic (GARCH), exponential GARCH (EGARCH) and threshold GARCH (TGARCH) techniques with different error distribution such as normal and student-t. The Autoregressive Conditional Heteroscedastic (ARCH) effect has been identified using the Lagrange multiplier test. The Ljung-Box test has been used for testing the autocorrelation exists in a time series. Root mean square error (RMSE), mean absolute percentage error (MAPE), and R-square are used for comparison of the aforementioned models. The fitted models' residuals have been examined for diagnostic purpose. The study has revealed that the ARMA (2,1) model is the better fitted model for mean equation while in the variance equation, GARCH (1,1) and EGARCH (1,1) model with student-t distributions are suitable for modelling the symmetric and asymmetric patterns of log returns on the basis of smaller value of AIC (Akaike information criterion) and BIC (Bayesian information criterion).

Keywords: ARIMA, EGARCH , GARCH, and TGARCH.

INTRODUCTION

The ability to predict future volatility is a crucial tool in financial economics, particularly in risk management and asset allocation, as it can assist investors to reduce losses. The volatility has four often seen properties, it may be estimated using the log return as the underlying series. (1) volatility clusters means that the variance of the series changes over different periods. (2) Since volatility changes continuously, volatility leaps are uncommon. (3) Volatility does not diverge to infinity; it fluctuates within a predetermined range. (4) Asymmetric effects, or significant price change, appear to have a higher impact on volatility than comparable price increase, is also known as the leverage effect, is frequently seen in financial time series. The earlier theoretical models on volatility made the assumption of homoscedastic regression models, which do not accurately capture the characteristics of volatility. Engle (1982) introduced the ARCH model to more accurately capture the characteristics of volatility. The ARCH model had drawback of requiring numerous parameters to adequately capture the dynamics of conditional variance. As a result, Bollerslev (1986) introduced the GARCH model, a generalized extension to the ARCH model that allowed for a more flexible lag structure and potentially required fewer model parameters. Leptokurtosis and volatility clustering are two often seen traits that can be capture by both the ARCH and GARCH. Due to their symmetry, these models have the drawback of not accounting for the leverage effect, which states that a negative shock in asset return will have a bigger impact on volatility of series than an equally significant positive shock, has led to the proposal of numerous asymmetric extensions to GARCH. The exponential GARCH (EGARCH) introduced by Nelson (1991) and GJR-GARCH introduced by Glosten, Jaganathan and Runkle (1993) are example of asymmetric extensions. Burark et al. (2012) used coriander wholesale pricing data on the Kota market in Rajasthan from April 2000 to May 2011 to investigate the performance of the ARIMA and exponential smoothing model. Ali (2013) investigated the efficacy of various asymmetric models, including the EGARCH model, IGARCH model, TGARCH model, GJR-GARCH, NGARCH model AVGARCH model and APARCH model to analyse daily data of faecal indicator bacteria densities near Huntington Beach in Ohio, the United States.

MATERIALS AND METHODS

Generalized Autoregressive Conditional Heteroscedastic (GARCH) model

The GARCH models to capture the significant properties like heteroscedasticity. Take into account a commodity price time series with a sample of N observations. Suppose r_t stand for the continually compounded returns, sometimes known as the log change between prices at time t and t-1. The return at time t can be described using the GARCH model as

$r_t = u_t + \sigma_t \varepsilon_t$ where $\varepsilon_t \sim N(0,1)$

Where u_t is conditional mean, σ_t^2 is conditional variance and ε_t is standardized residuals at time t. conditional variance is modeled as

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i u_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2$$

$\alpha_0 > 0, \alpha_i \geq 0, i = 1 \ldots q; \beta_j \geq 0, j = 1 \ldots p; \sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j < 1$ for ensuring as stationary condition. Enocksson and Skoog (2012) noted some drawbacks in GARCH model. The most important is that GARCH model is unable to account for asymmetric behavior.

Exponential Generalized Autoregressive Conditional Heteroscedastic (EGARCH) model

Nelson (1991) introduced the Exponential GARCH model to capture asymmetric shock to conditional variance, in which natural logarithm of conditional variance is allowed to fluctuate over time as a function of lagged error terms rather than lagged squared error. The Exponential GARCH (p,q) model expressed as:

$$log\sigma_t^2 = c + \sum_{i=1}^{p} g(Z_{t-1}) + \sum_{j=1}^{q} \beta_j log\sigma_{t-j}^2$$

$$g(Z_{t-1}) = \gamma_1 Z_{t-1} + \alpha_1(|Z_{t-1}| - E(|Z_{t-1}|))$$

Define $Z_{t-1} = \frac{\varepsilon_{t-1}}{\sigma_{t-1}}$ and the conditional variance's logarithm is defined as:

$$\log(\sigma_t^2) = c + \sum_{j=1}^{q} \beta_j \log \sigma_{t-1}^2 + \sum_{i=1}^{p} \alpha_i \left(\left| \frac{\varepsilon_{t-1}}{\sigma_{t-1}} \right| - E\left(\left| \frac{\varepsilon_{t-1}}{\sigma_{t-1}} \right| \right) \right) + \sum_{i=1}^{p} \gamma_1 \frac{\varepsilon_{t-1}}{\sigma_{t-1}}$$

Where α_i show the symmetric effect, β_j evaluate the persistence in conditional volatility shock and reflects the asymmetric performance. EGARCH (1,1) can be expressed as:

$$\log \sigma_t^2 = \omega + \beta_1 \log \sigma_{t-1}^2 + \alpha_1 \left| \frac{\varepsilon_{t-1}}{\sigma_{t-1}} \right| + \gamma \frac{\varepsilon_{t-1}}{\sigma_{t-1}}$$

If $\gamma < 0$, then leverage effect present, and bad news (negative shocks) cause more volatility than good news (positive shocks) of the equal size and $\gamma > 0$, it suggests that for same modulus, good news produce more volatility than bad news. The volatility shock is asymmetric when $\gamma \neq 0$.If on the other hand $\gamma = 0$, then the model is symmetric.

The exponential structure of the EGARCH model ensure that conditional variance is always positive even if parameter value are negative, hence there is no need for parameter restrictions to impose non-negativity.

TGARCH model

Using dummy variables, the TGARCH model alters the original GARCH specification. This model's primary goal is to account for asymmetries in terms of bad and good news by incorporating a multiplicative dummy variable into a variance equation to determine whether there is statistically significant variation from bad news. The TGARCH model introduced by Glosten, Jagannathan and Runkle (1993), is also known as GJR-GARCH model. GJR-GARCH model is written by

$$\sigma_t^2 = c + \sum_{j=1}^{p} \beta_j \sigma_{t-1}^2 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{k=1}^{r} \gamma_k \varepsilon_{t-k}^2 I_{t-k} (\varepsilon_{t-k} < 0)$$

The asymmetry in TGARCH model is captured by the indicator term's sing, as Patrick, Stewart and Chris (2006) highlight in their work. It is an indicator term defined as:

$$I_t = \begin{cases} 1, if\ \varepsilon_t < 0, & badnews \\ 0, if\ \varepsilon_t \geq 0, & goodnews \end{cases}$$

GJR-GARCH (1,1) model is written by:

$$\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 + \gamma \varepsilon_{t-1}^2 I_{t-1}$$

Where I_t =1 if ε_t is negative and otherwise equal to zero. According to TGARCH model, volatility rise in response to negative news ($\varepsilon_{t-1} < 0$) and fall in response to positive news ($\varepsilon_{t-1} \geq 0$). The impact of positive news is denoted by α_1 but impact of negative news is represented by $\alpha_1 + \gamma$. If $\gamma > 0$ parameter of leverage effect then statistically significant leverage effect is present. The shock is asymmetric if $\gamma \neq 0$ and shock is symmetric if $\gamma = 0$. The persistence of volatility shocks determined by $\alpha_1 + \beta_1 + \gamma/2$.

Error distribution

Two distributions namely, 1) normal error distribution and 2) student-t error distribution are introduced as

Normal distribution

The probability density function of Z_t is given as follows,

$$f(Z_t) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left\{-\frac{1}{2}\left(\frac{Z_t-\mu}{\sigma}\right)^2\right\}$$

where μ is mean and σ is standard deviation.

Student t-distribution

The probability density function of Z_t is given as follows,

$$f(Z_t) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\Gamma\left(\frac{v}{2}\right)\sqrt{(v-2)\pi}}\left(1+\frac{z_t^2}{v-2}\right)^{-\frac{1}{2}(v+1)}$$

Where v is the number of degrees of freedom, $2 < \nu \leq \infty$, and is Γ gamma function. When $\nu \rightarrow \infty$, the student-t distribution nearly equals to the normal distribution. The lower the v, the fatter the tails.

Model Selection

Model choice, we compare various ARMA-GARCH model specifications and then choose the most appropriate model using Akaike information criteria (AIC) and Bayesian information criteria (BIC). Calculating the AIC and BIC is as follows:

AIC = −2ln (residual sum of squares) + 2k

BIC = −2ln (residual sum of squares) + ln (N) k

Where N is the number of observations, and k is the number of estimated parameters.

Model testing

The model volatility forecasting performance is examination using four statistical measures: Relative deviation percentage (RD%), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE) and R-square, are defined by the formula:

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=T_1}^{T}(O_i - E_i)^2} \qquad MAPE = \frac{100}{N}\sum_{i=1}^{n}\frac{O_i-E_i}{O_i}$$

$$RD\% = \frac{O_i-E_i}{O_i} \times 100 \qquad R^2 = 1 - \frac{\sum(O_i-E_i)^2}{\sum(O_i-\bar{O})^2}$$

Where, O_i, $\bar{O}$ and E_i are the observed, mean and predicted values and N is the number of observations for which estimation has been done.

RESULTS AND DISCUSSION

Time series data on all India monthly average wholesale prices of onion over period Jan-2010 to Dec-2021 (total number of observations 144), collected from agriculture market (Source:https://agmarknet.gov.in/).The first 132 observations (Jan-2010 to Dec-2020) are used for model building and parameter estimation, while the last 12 values (Jan-2021 to Dec-2021) are used for post-validity checking. The log returns (r_t) are calculated as the continuously compounded returns which are the first differences of log prices of consider time series.

$$r_t = log\left(\frac{P_t}{P_{t-1}}\right)$$

where P_t and P_{t-1} are prices at the current month and previous month respectively.

Table 1: Log returns descriptive statistics for all India monthly average wholesale prices of onion.

N	Mean	Median	Maximum	Minimum	SD	Skewness	Kurtosis
144	0.61	0.37	62.85	-74.22	21.95	-0.17	1.09

From this table, the skewness is -0.17 (negative skewed), is not zero, indicate that log return is asymmetric, and kurtosis is 1.09, is smaller than three which means that the distribution of log returns is flat (platykurtic) relative to the normal. Plots of log returns and square returns shown in fig. 1, indicate that presence of autocorrelation in log returns and square return

Fig. 1: Plots of log returns and absolute returns for all India monthly average wholesale price of onion.

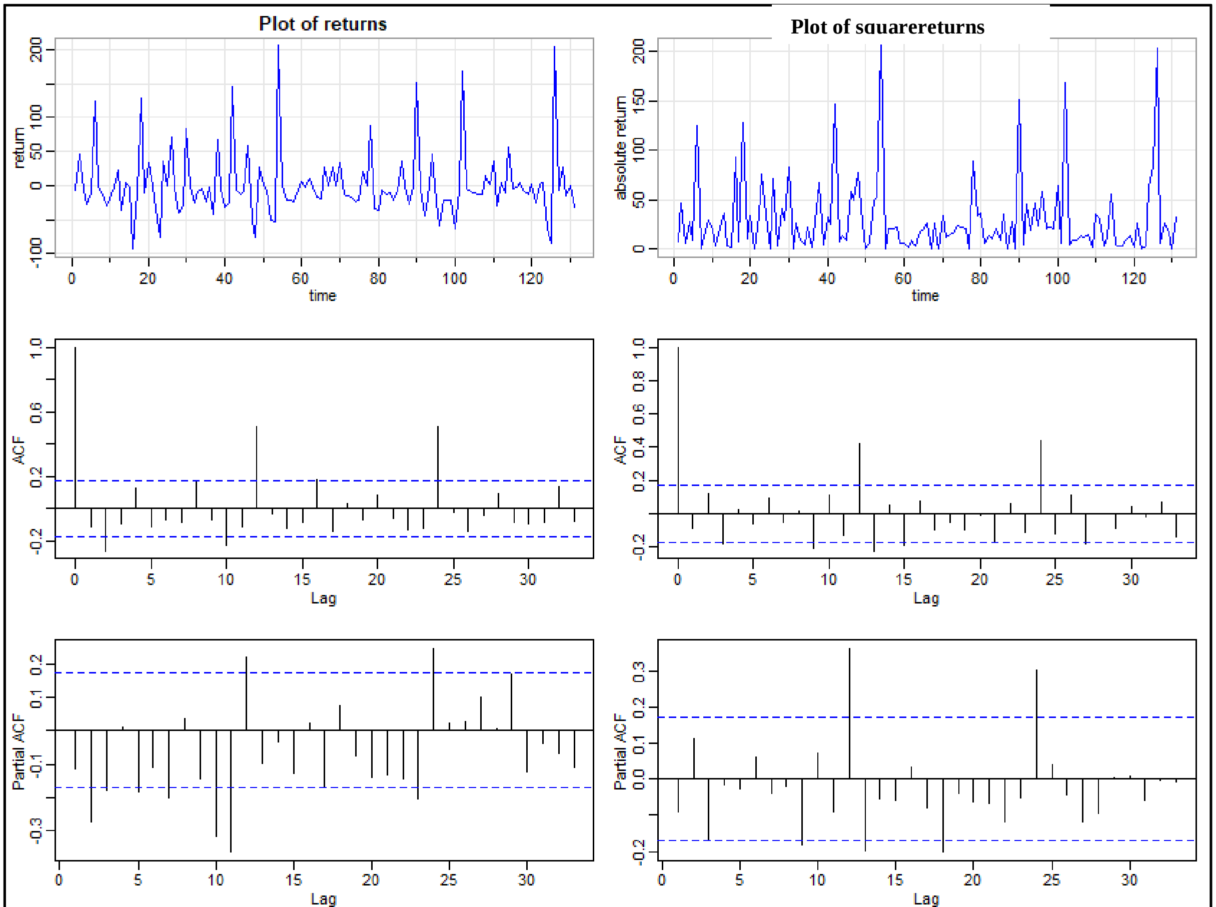


Table 2: Lagrange multiplier test for log returns of all India monthly average wholesale prices of onion.

	Null Hypothesis	Test Statistics	P-value
LM	Presence ARCH effect	1.75	0.002
KPSS	Time series is stationary	0.03	0.1

Lagrange multiplier and Kwiatkowski–Phillips–Schmidt–Shin (KPSS) tests are used to check the presence of ARCH effect and stationarity around a deterministic trend in log returns time series. Log returns time series has stationary and presence of ARCH effect are shown in above Table2. Based on the assumption of 5% significance level, all of the p-values are less than 0.05, indicate that not reject null hypothesis otherwise rejected.

ARMA-GARCH, ARMA-EGARCH and ARMA-TGARCH models are used to estimate and forecast the log returns of all India monthly average wholesale prices of onion under the different error distributions i.e. normal and student-t distributions and then compare the forecasting performance measures such as RD (%), RMSE, MAPE and R-square and choose the appropriate volatility forecast model.

Selection of models for mean equation and variance equation

Selection of suitable mean equation ARMA (p, q) and variance equation GARCH (p, q), EGARCH (p, q) and TGARCH (p, q) models for log returns. The tentative order of p and q can be determined by looking at the partial and autocorrelation of log returns.

Compare ARMA (2,1)-GARCH (1, 1), ARMA (2,1)-EGARCH (1, 1) and ARMA (2,1)-TGARCH (1, 1) models with both the distributions of error term. After comparing AIC and BIC values for different models, the more suitable ARMA (2,1)-GARCH (1, 1) and ARMA (2,1)-EGARCH (1, 1) model with Student-t distribution will be picked up with the smallest value of AIC and BIC shown in Table 3. The estimated parameters of ARMA (2,1)-GARCH (1, 1) and ARMA (2,1)-EGARCH (1, 1) model with Student-t distribution shown in table 4 & table 5 respectively.

Table 3: Model Selection Criteria.

Model	Error Distribution	AIC	BIC
ARMA (2,1)-GARCH (1,1)	ND	4.74	4.82
ARMA (2,1)-GARCH (1,1)	**STD**	**4.69**	**4.78**
ARMA (2,1)-EGARCH (1,1)	ND	4.70	4.79
ARMA (2,1)-EGARCH (1,1)	**STD**	**4.68**	**4.75**
ARMA (2,1)- TGARCH (1,1)	ND	4.73	4.83
ARMA (2,1)-TGARCH (1,1)	STD	4.72	4.83

Table 4: ARMA-GARCH model's Parameter Estimates with Student-t distribution.

Mean Equation	Coefficient	S.E	t-statistics	P-value
Φ_0	0.16	0.17	0.97	0.03
AR(1) Φ_1	1.14	0.14	7.85	0.00
AR(2)Φ_2	-0.34	0.06	-5.47	0.00
MA(1)θ_1	-0.75	0.14	-5.15	0.00
Variance Equation				
Ω	0.0435	0.358	1.215	0.02
α_1	0.027	0.024	1.091	0.002
β_1	0.9078	0.066	13.63	0.000

V	8.919	3.520	2.533	0.011
$\alpha_1 + \beta_1$	0.934			

$$Y_t = 0.169 + 1.144Y_{t-1} - 0.347Y_{t-2} + \varepsilon_t + 0.751\varepsilon_{t-1}$$

The GARCH (1,1) model is represented by conditional variance equation as:

$$\sigma_t^2 = 0.435 + 0.027\varepsilon_{t-1}^2 + 0.9078\sigma_{t-1}^2$$

We note that all ARMA (2,1)-GARCH (1,1) model parameters are significant. ARCH term $\alpha_1 = 0.027$ and GARCH term $\beta_1 = 0.907$. The fact that the GARCH term has a large value indicate that its impact on conditional variance shocks takes a very long time (long memory process). The stationary condition $\alpha_1 + \beta_1 < 1$ is met, as shown by the sum of ARCH and GARCH term $\alpha_1 + \beta_1 = 0.9348$.

Table 5: ARMA-EGARCH model's Parameter Estimates with Student-t distribution.

	Coefficient	**S.E**	**t-statistics**	**P-value**
Mean Equation				
Φ_0	06.218	0.111	1.956	0.05
AR(1) Φ_1	1.214	0.094	12.820	0.00
AR(2)Φ_2	-0.338	0.054	-6.176	0.00
MA(1)θ_1	-0.852	0.073	-11.529	0.00
Variance Equation				
Ω	0.155	0.041	3.739	0.001
α_1	0.031	0.024	2.752	0.005
β_1	0.922	0.007	133.95	0.000
γ	-0.111	0.040	-2.727	0.001
V	16.525	13.854	1.192	0.000
$\alpha_1 + \beta_1$	0.953			

The EGARCH (1,1) model is represented by conditional variance equation as:

$$log\sigma_t^2 = 0.155 + 0.922log\sigma_{t-1}^2 + 0.031\left|\frac{\varepsilon_{t-1}}{\sigma_{t-1}}\right| - 0.110\frac{\varepsilon_{t-1}}{\sigma_{t-1}}$$

We note that all ARMA (2,1)-EGARCH (1,1) model parameters are significant, and the shock persistence parameter is very close to unity i.e. (β_1=0.9223), showing that volatility shock has long memory and the variance has long memory.The leverage effect parameter (= -0.110) is also statistically significant and negative in the EGARCH (1,1) model, indicating that previous negative shocks have a stronger impact on future volatility than positive shocks of comparable magnitude.

Diagnostic checking for selected models

Diagnostic checking for selected suitable ARMA (2,1)-GARCH (1, 1) and ARMA (2,1)-EGARCH (1, 1) models with Student-t distribution for log returns series.

The standardized residuals are used to determine whether the model is appropriately specified. The ACF and PACF plots (Fig. 2 and 3) of the standardized residuals derived using ARMA (1,2)-GARCH (1,1) and ARMA (1,2)-EGARCH (1,1) revealed that autocorrelations are not substantially different from zero.

Fig. 2: plots of standardized residuals of ARMA-GARCH model

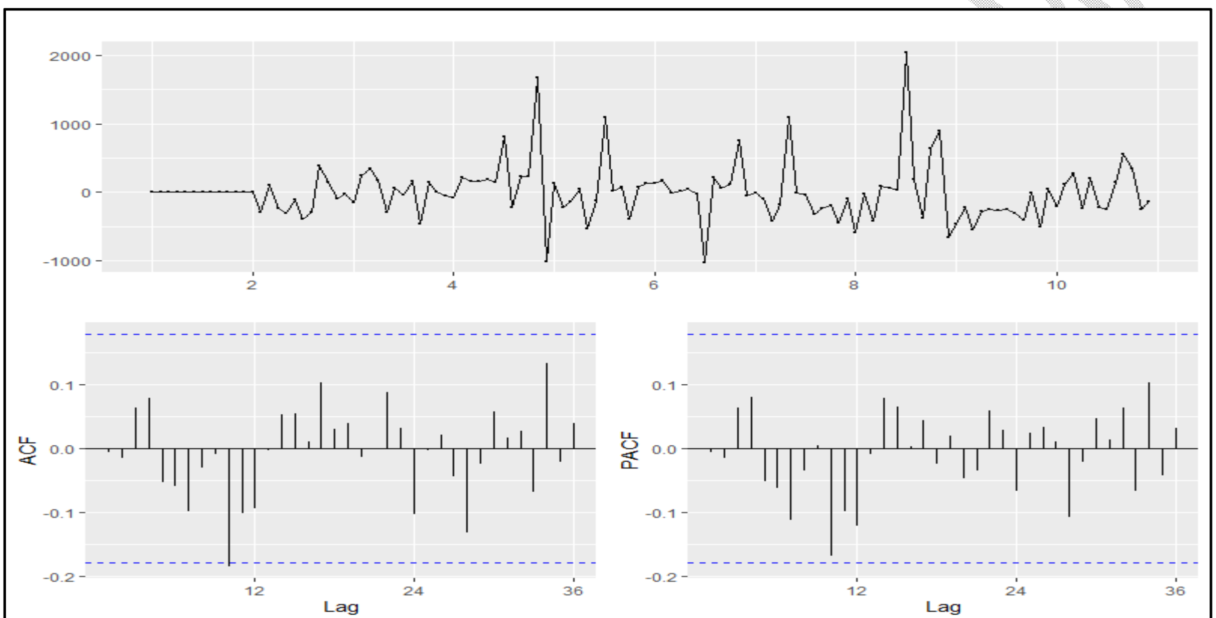
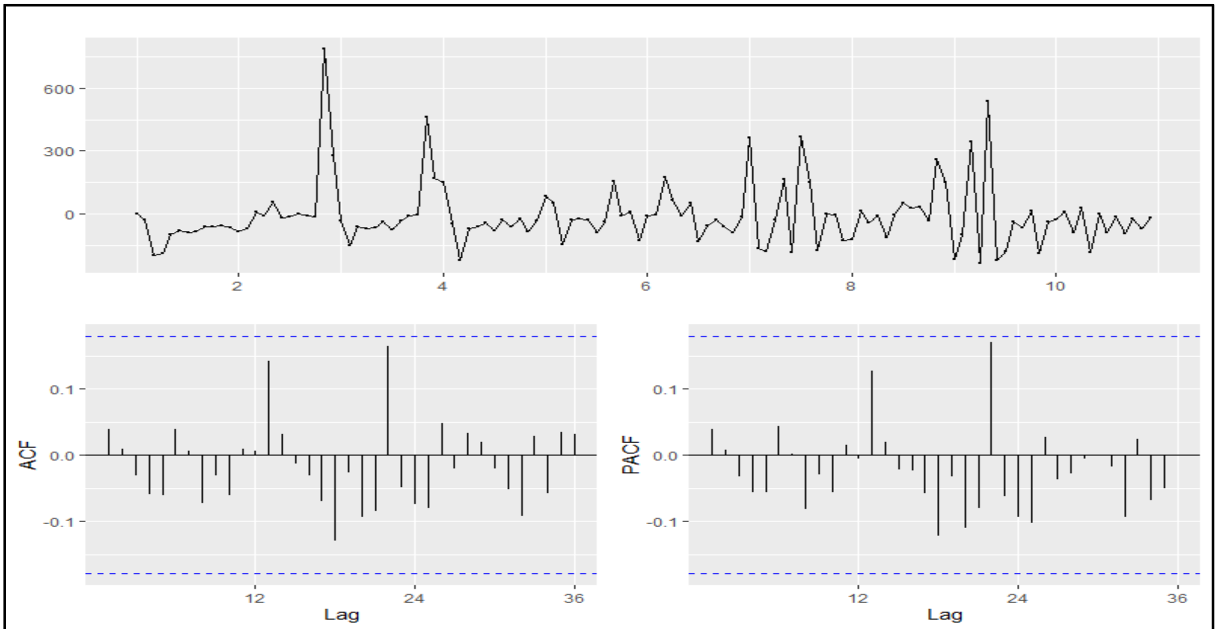


Fig. 3: plots of standardized residuals of ARMA-EGARCH model



To confirm the appropriateness of the model, the statistical tests, Ljung-Box test and ARCH-LM test are used for checking the autocorrelation and ARCH effect exists in standardized residuals of selected suitable models. The p-value of both the tests are found to be greater than 0.05 (at 5% level of significance) for all the lags as shown in Table 6, accepting the null hypothesis of no autocorrelation and no ARCH effect in ARMA (1,2)-GARCH (1,1) and ARMA (1,2)-EGARCH (1,1) models residuals.

Table 6: Diagnostic checking of ARMA-GARCH and ARMA-EGARCH models with Student-t distribution.

ARMA - GARCH				
	Ljung-box squared residual		ARCH LM Test	
	Statistics	P-value	Statistics	P-value
Lag [3]	0.14	0.70	0.02	0.87
Lag [5]	0.34	0.97	0.21	0.96
Lag [7]	0.51	0.99	0.26	0.99
ARMA-EGARCH				
Lag [3]	0.10	0.74	0.02	0.89
Lag [5]	0.25	0.98	0.14	0.97
Lag [7]	0.35	0.99	0.20	0.99

Forecasting performance

The four statistical measures: Relative deviation percentage (RD%), RMSE (Root mean squared error) MAPE (mean absolute percentage error) and R-square are used to compare the volatility forecasting performance of selected ARMA (1,2)-GARCH (1,1) and ARMA (1,2)-EGARCH (1,1) models. Relative deviation percentage (RD%) of these selected models is shown in table 7. The corresponding plot of observed and predicted values from selected models is given in Fig. 5. ARMA (2,1)- EGARCH (1,1) model is selected as appropriate volatility forecasting model on base of smaller values RMSE (148.01) and MAPE (5.28) of ARMA (2,1)-EGARCH (1,1) as compare to RMSE (196.94) and MAPE (6.99) of ARMA (2,1)-GARCH (1,1) model while R-square (0.85) of ARMA (2,1)-EGARCH (1,1) is higher than R-square (0.82) of ARMA (2,1)-GARCH (1,1) model are shown in fig. 4.

Table 7: Observed and Predicted of onion prices (Rs/Quintal) for India monthly average wholesale prices of onion for the year 2021 by ARMA-GARCH and ARMA-EGARCH models

Month	Observed	ARMA – GARCH		ARMA- EGARCH	
		Predicted	RD (%)	Predicted	RD (%)
Jan-21	3155.65	3502.23	-10.98	3375.53	-6.97
Feb-21	3663.68	3463.24	5.47	3463.45	5.47
March-21	2708.56	2800.25	-3.39	2526.15	6.73
April-21	1806.54	2003.23	-10.89	1947.89	-7.82
May-21	1812.64	1900.23	-4.83	1787.29	1.40
June-21	2120.73	1923.21	9.31	2283.27	-7.66
July-21	2337.94	2122.31	9.22	2222.38	4.94
August-21	2335.5	2224.15	4.77	2498.85	-6.99
Sept-21	2255.46	2154.26	4.49	2324.25	-3.05
Oct-21	3036.44	3212.78	-5.81	3185.88	-4.92
Nov-21	3224.9	3355.84	-4.06	3154.14	2.19
Dec-21	2851.67	2545.12	10.75	3000.12	-5.21

Fig. 4: Plots of RMSE, MAPE and R-square of ARMA-GARCH and ARMA-EGARCH models

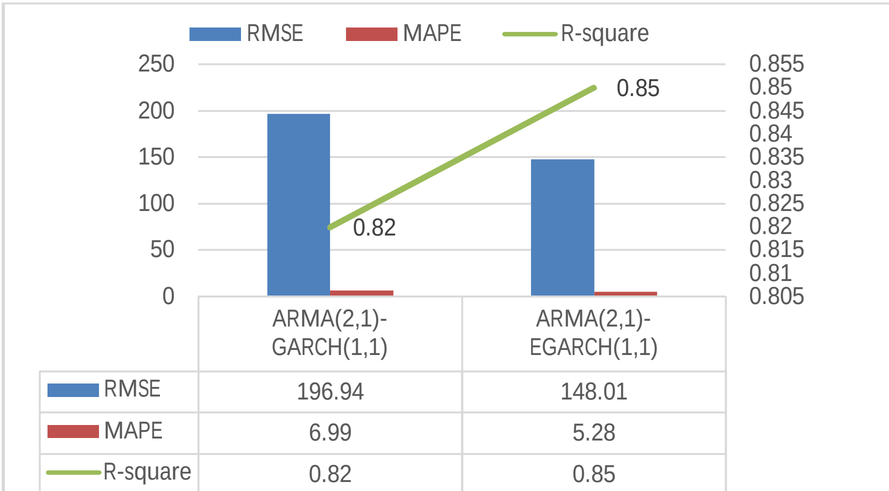
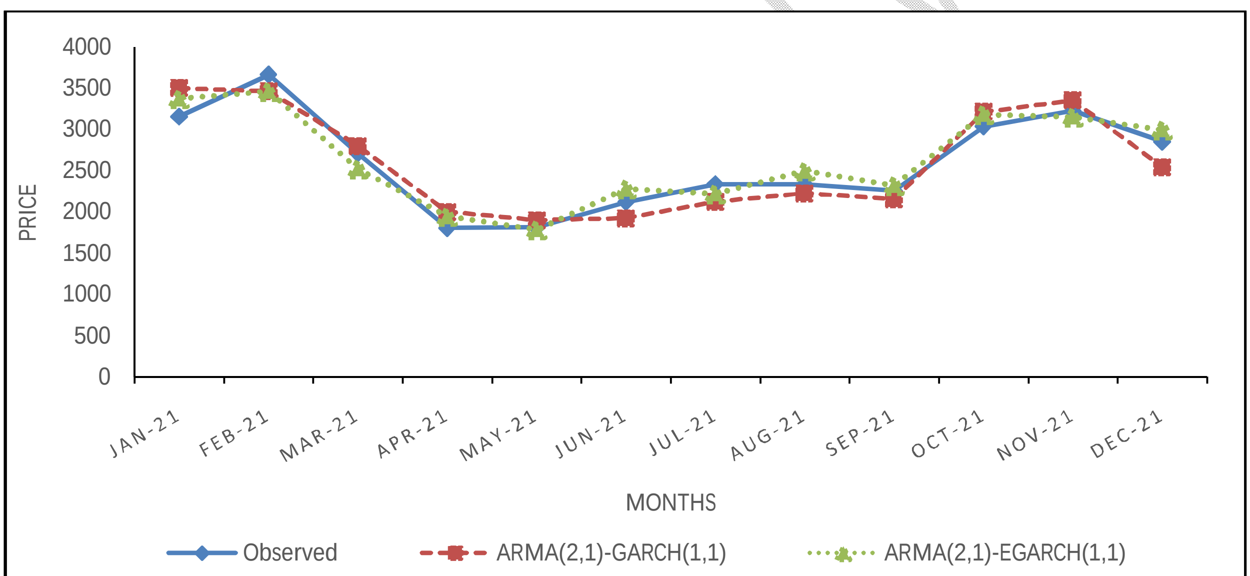


Fig. 5: Plot of observed and predicted India monthly average wholesale prices (Rs/Quintal) of onion for the year 2021 by ARMA-GARCH and ARMA-EGARCH models



Conclusion

This study utilized all India monthly average wholesale prices of onion over period Jan-2010 to Dec-2021 and used ARMA-GARCH, ARMA-Exponential GARCH and ARMA-Threshold GARCH models with two different error distribution namely, normal and student-t. ARMA(Autoregressive moving average) (2,1) model selected as better fitted model for mean equation of log on the basis of least value AIC and BIC, while in variance equation, symmetric and asymmetric log return behaviors may be described by simple GARCH (1,1) and Exponential GARCH (1,1) models with student-t error distribution respectively. ARMA (2,1)-EGARCH (1,1) model selected as appropriate volatility forecasting model for all India monthly average wholesale prices of onion on the basis of smaller values of statistical measures such as RD (%), RMSE and MAPE, and higher value of R-square.

REFERENCES

Bhardwaj S P, Paul, R K, Singh, D R and Singh K N. An empirical investigation of arima and garch models in agricultural price forecasting. *Economic Affairs.* 2014;**59**(3):415-428. DOI 10.5958/0976-4666.2014.00009.6.

Bollerslev, T. Generalised autoregressive conditional heteroskedasticity. *Journal of Econometrics.* 1986; **31**: 307-27. http://dx.doi.org/10.1016/0304-4076(86)90063-1.

Bryant, P G and Smith, M A. Practical Data Analysis: Case Studies in Business Statistics, Richard Irwin. Inc. Burr Ridge, Illinois.1995; https://doi.org/10.2307/2685436.

Burark, S S, Sharma, H and Meena, G L. Market integration and price volatility in domestic market of coriander in Rajasthan. *Indian Journal of Agricultural Marketing.* 2012; **27**: 121-131.

Chadwick, M and Bastan, M. News impact for Turkish food prices. *Central Bank Review,* 2017;**17**(2):55-76. https://doi.org/10.1016/j.cbrev.2017.05.001.

Engle, R. Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica.* 1982;**50**(4):987–1007. https://doi.org/10.2307/1912773.

Farid, S Kayani, G M Naeem, M A and Shahzad, S J H. Intraday volatility transmission among precious metals, energy and stocks during the COVID-19 pandemic. *Resources Policy.* 2021; **72**:0301-4207. https://doi.org/10.1016/j.resourpol.2021.102101.

Glosten, L R Jagannathan, R And Runkle, D. On the Relation Between the Expected Values and the Volatility of The Nominal Excess Return on Stocks. *Journal of Finance.* 1993; **48:**1779–1801. https://doi.org/10.1111/j.1540-6261.1993.tb05128.x.

Ibrahim, A O and Bruno, S S. Modeling and forecasting volatility in the global food commodity prices. *Agricultural Economics. 2011;***57**(3):132-139. https://doi.org/10.17221/28/2010-AGRICECON.

Nelson, D B. Conditional Heteroscedasticity in Asset Returns: A New Approach. *Econometrica*. 1991; **59**(2):347-370. https://doi.org/10.2307/2938260.

Paul, R K Prajneshu and Ghosh H. GARCH Nonlinear Time Series Analysis for modeling and forecasting of India's volatile spices Export Data. *Journal of the Indian Society of Agricultural Statistics. 2009;* **63**(2): 123-131.