Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF
AGRICULTURAL STATISTICS 74(2) 2020 99–106

Incorporation of Exogenous Variable in Long Memory Model: An ARFIMAX-GARCH Framework

Krishna Pada Sarkar, K.N. Singh, Achal Lama and Bishal Gurung

ICAR-Indian Agricultural Statistics Research Institute, New Delhi

Received 18 December 2019; Revised 13 January 2020; Accepted 13 January 2020

SUMMARY

In the present study exogenous variable is incorporated in the long memory model to give better forecast of time series. Autoregressive Fractionally Integrated Moving Average- Generalized Autoregressive Conditional Heteroscedastic (ARFIMA-GARCH) and Autoregressive Fractionally Integrated Moving Average with exogenous variable- Generalized Autoregressive Conditional Heteroscedastic (ARFIMAX-GARCH) models are studied for describing the volatile data. Brief description of the models are given along with parameter estimation procedure. As an illustration daily minimum market price of onion along with daily market arrival of Lasalgaon market of Maharashtra, India is taken. Comparative study of the fitted model is carried out by calculating the Root Mean Square Error (RMSE) and Relative Mean Absolute Percentage Error (RMAPE) from the validation set. The superior performance of the ARFIMAX-GARCH model than ARFIMA-GARCH model is demonstrated for data under study.

Keywords: Long memory, ARFIMA, ARFIMAX, ARFIMAX-GARCH, Forecasting.

1. INTRODUCTION

Commonly used ARIMA model assumes that the distant time series observation are independent of each other or nearly so. In terms of autocorrelation that means autocorrelation function decay rapidly in exponential rate toward zero (Box *et al.* 2007). But in some practical cases it has been seen that the observations in distant past are correlated. The correlation may be small but it may significantly affect the forecasting accuracy. To know this type of long range dependency ACF of observations is plotted over various lags. The later type of time series is said to possess long memory property. Autocorrelation function of most of the stationary (ARMA) time series process decays exponentially and in such cases autocorrelation can be approximately expressed as $|\rho_k| \approx |m|^k$, where $|m| < 1$ whereas for long memory process it decays at much slower hyperbolic rate and in this case autocorrelation coefficient can be expressed as $\rho_k \approx Hk^{2d-1}$, as k tends to infinity. Here ρ_k is autocorrelation at lag k, d and H are the long memory parameter and any constant respectively. For modeling of long memory time series Autoregressive Fractionally Integrated Moving Average (ARFIMA) model is used (Paul, 2014).

Sometimes besides the study or original time series, date on some auxiliary or exogenous variables are available or can easily be made available which may be significantly correlated with the original time series. These exogenous variables have significant influence on study variable and should be incorporated in the existing time series model for improving the model performance and forecasting accuracy (Paul *et al.* 2014). In this paper exogenous variable is included in the existing ARFIMA model resulting in the formulation of Autoregressive Fractionally Integrated Moving Average Model with Exogenous variable (ARFIMAX) model.

Another important consideration during forecasting of a time series is that the heteroscedasticity present in the data set. Linear models doesn't have the capacity to define the varying pattern in the conditional variances existing in the data. To cope with this situation, Engle (1982) have proposed the Autoregressive Conditional Heteroscedastic (ARCH) model by taking care of

Corresponding author: Achal Lama
E-mail address: achal.lama@icar.gov.in

significant autocorrelations present in the squared residual series. But the ARCH models provide reasonable forecasts only with a large number of parameters. To overcome the estimation of large number of parameter, a more generalized version of this model, Generalized ARCH (GARCH) models have been developed (Bollerslev, 1986). This model has been used extensively for modelling and forecasting of volatile agricultural data sets (Paul, *et al.* 2014, Lama, *et al.* 2015)

In this paper ARFIMA-GARCH and ARFIMAX-GARCH model are studied along with their parameter estimation technique. These models are applied on real data set by taking onion minimum market price. Comparative study of the models is also studied using RMSE and RMAPE criterion.

2. MODELS DESCRIPTION

2.1 ARFIMA model

The ARFIMA model of order (p, d, q), denoted by ARFIMA (p, d, q) for the stationary process $y_t; (t = 1, 2 \ldots, n)$ with mean μ and variance σ^2 can be represented as (Granger and Joyeux, 1980)

$$\varphi(L)(1-L)^d y_t = \theta(L)e_t \quad (1)$$

Here e_t is an i.i.d random variable having zero mean and constant variance σ^2. L is the lag operator, d is the fractional difference operator known as long memory parameter which can take any fractional value between -0.5 to 0.5 (Hosking, 1981), $\varphi(L)$ and $\theta(L)$ are the finite Autoregressive (AR) and Moving Average (MA) polynomials of order p and q respectively and represented by

$$\varphi(L) = 1 - \varphi_1 L - \cdots - \varphi_1 L^p = 1 - \sum_{i=1}^{p} \varphi_i L^i \quad (2)$$

$$\theta(L) = 1 - \theta_1 L - \cdots - \theta_q L^q = \left(1 + \sum_{i=1}^{q} \theta_i L^i\right) \quad (3)$$

$(1-L)^d$ can be expanded using binomial series expansion in the following way

$$(1-L)^d = 1 - dL + \frac{L^2 d(d-1)}{2!} - \cdots = \sum_{i=0}^{\infty} \binom{d}{i} (-1)^i L^i \quad (4)$$

where $\binom{d}{i}$ is represented as $\binom{d}{i} = \frac{d!}{i!(d-i)!}$

The model has total $p + q + 3$ parameters $\mu, \sigma^2, d, \boldsymbol{\varphi} = (\varphi_1, \varphi_2, \ldots, \varphi_p)'$ and $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_q)'$. The parameters are restricted in $\mathcal{R}^{p+q+3}$ dimensional space.

2.2 ARFIMAX model

ARFIMAX model with k exogenous variables is given below. Let $y_t; (t = 1, 2, \ldots, n)$ is a stationary process with mean μ and variance σ^2 and $x_j, j = 1,2, \ldots, k$ be the k exogenous variable. Then by following Degiannakis (Degiannakis, 2008) the ARFIMAX with k exogenous variable can be written as

$$\varphi(L)(1-L)^d (y_t - \mu - \boldsymbol{x}_t' \boldsymbol{\beta}) = \theta(L)e_t \quad (5)$$

Where $\boldsymbol{\beta}$ is the vector of k coefficients corresponding to k exogenous variables and rest notations are same as of ARFIMA model. The model has total of $p + q + k + 3$ parameters $\mu, \sigma^2, d, \boldsymbol{\varphi} = (\varphi_1, \varphi_2, \ldots, \varphi_p)'$, $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_q)'$ and $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_k)'$. For the present study daily market prices of commodities are taken as study or original time series and daily market arrival of the commodity a the exogenous variable. Market arrival generally affects the market price of commodities in opposite direction.

2.3 Parameter estimates of ARFIMA and ARFIMAX model

The parameters of the model are estimated using maximum likelihood estimation (MLE) technique whereby likelihood function based on observed sample observations is maximized. The exact likelihood function based on n observations $y_n = (y_1, y_2, \ldots, y_n)'$ for the ARFIMA model is given by

$$f(\boldsymbol{y}_n|\boldsymbol{\eta}) = (2\pi\sigma^2)^{-n/2}|V_n|^{-1/2} \exp\left[\frac{-(\boldsymbol{y}_n - \mu\boldsymbol{1}_n)' V_n^{-1} (\boldsymbol{y}_n - \mu\boldsymbol{1}_n)}{2\sigma^2}\right] \quad (6)$$

where $\boldsymbol{\eta} = (\boldsymbol{\varphi}, \boldsymbol{\theta}, d, \mu, \sigma^2)$ is the vector of dimension $(p + q + 3)$ and $\sigma^2 \boldsymbol{V}_n$ is the variance covariance matrix of $\boldsymbol{Y}_n$. Similarly the likelihood function of ARFIMAX model based on the n observations on study variable y, $\boldsymbol{y}_n = (y_1, y_2, \ldots, y_n)'$ and on each k exogenous variable $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k$ is given by:

$$L(\boldsymbol{y}_n|\boldsymbol{\eta}^*) = (2\pi\sigma^2)^{-n/2}|\boldsymbol{V}_n|^{-1/2}\exp\left[\frac{-(\boldsymbol{y}_n - \mu\boldsymbol{1}_n - \boldsymbol{X\beta})'\boldsymbol{V}_n^{-1}(\boldsymbol{y}_n - \mu\boldsymbol{1}_n - \boldsymbol{X\beta})}{2\sigma^2}\right] \quad (7)$$

Where $\boldsymbol{\eta}^* = (\boldsymbol{\varphi}, \boldsymbol{\theta}, \mu, d, \boldsymbol{\beta}, \sigma^2)$ be a vector of $(p+q+k+3)$ parameters, **X** is $n \times k$ matrix of covariates, $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_{k-1})'$ and $\sigma^2\boldsymbol{V}_n$ is the variance covariance matrix of the sample observations.

2.4 Testing for ARCH Effects (ARCH-LM Test)

After estimating the parameters of ARFIMA and ARFIMAX model, residuals are further checked for presence of ARCH effect. Autoregressive Conditional Heteroscedastic- Lagrange Multiplier(ARCH-LM) test is employed for this purpose (Engle, 1982). Let ε_t the residual series obtained from the ARFIMA or ARFIMAX model. The Lagrange Multiplier (LM) test for squared series $\{\varepsilon_t^2\}$ may be used to check for conditional heteroscedasticity. The test is equivalent to usual F-statistic for testing H_0: $a_i = 0, i = 1, 2, \ldots, q$ in the linear regression

$$\varepsilon_t^2 = a_0 + a_1\varepsilon_{t-1}^2 + \cdots + a_q\varepsilon_{t-q}^2 + e_t, \quad t = q+1 \ldots T \quad (8)$$

where e_t denotes the error term, q is the prespecified positive integer, and T is the sample size. Let $SSR_0 = \sum_{t=q+1}^{T}(\varepsilon_t^2 - \bar{\omega})^2$, where $\bar{\omega} = \sum_{t=q+1}^{T}\varepsilon_t^2/T$ is sample mean of $\{\varepsilon_t^2\}$, and $SSR_1 = \sum_{t=q+1}^{T}\hat{e}_t^2$, where $\hat{e}_t^2$ is the least square residual of the above regression model. Then, under H_0 the ARCH-LM test statistic is

$$F = \frac{(SSR_0 - SSR_1)/q}{SSR_1/(T-q-1)} \quad (9)$$

This test statistic is asymptotically distributed as chi-square distribution with q degrees of freedom. The H_0 is rejected if $F > \chi_q^2(\alpha)$, where $\chi_q^2(\alpha)$ is the upper $100(1-\alpha)\%$ point of the chi square distribution.

2.5 GARCH model

The process $\{\varepsilon_t\}$ is said to have ARCH (q) if the conditional distribution of $\{\varepsilon_t\}$ provided with the available information up to $t-1$, (ψ_{t-1}) is denoted as

$$\varepsilon_t|\psi_{t-1} \sim N(0, h_t) \text{ and } \varepsilon_t = h_t^{1/2}\zeta_t \quad (10)$$

where ζ_t is independently and identically distributed with zero mean and constant variance. The conditional variance is represented by,

$$h_t = \alpha_0 + \sum_{i=1}^{q}\alpha_i\,\varepsilon_{t-i}^2,\ \alpha_0 > 0, \alpha_i \geq 0 \sum_{i=1}^{q}\alpha_i < 1 \quad (11)$$

Where $\alpha_0 > 0, \alpha_i \geq 0$ for all of i and $\sum_{i=1}^{q}\alpha_i < 1$ are required to be satisfied to ensure non-negativity and finite unconditional variance of stationary $\{\varepsilon_t\}$ series. In Generalized ARCH (GARCH) model, the conditional variance is also a linear function of its own lags. The GARCH (p, q) process has the following form provided the conditions, $\alpha_0 > 0, \alpha_i \geq 0$ and $\beta_j \geq 0$ for all i and j (Bollerslev, 1986).

$$h_t = \alpha_0 + \sum_{i=1}^{q}\alpha_i\,\varepsilon_{t-i}^2 + \sum_{j=1}^{p}\beta_j h_{t-j} \quad (12)$$

The GARCH (q, p) process is said to be weakly stationary iff

$$\sum_{i=1}^{q}\alpha_i + \sum_{j=1}^{p}\beta_j < 1 \quad (13)$$

The conditional variance h_t in GARCH model (Engle, 1982) has the property that unconditional autocorrelation function of ε_t^2 if exists, decay slowly toward zero. Hence, a GARCH (p, q) is more parsimonious model of conditional variance as compared to higher order ARCH (q).

2.6 Estimation of Parameters of GARCH model

Similar to ARFIMA and ARFIMAX models, method of maximum likelihood estimation is used for estimating the parameters of the ARCH or GARCH model. The log likelihood function of a sample of T observations, apart from constant, is

$$L_T(\boldsymbol{\theta}) = T^{-1}\sum_{t=1}^{T}(\log h_t + \varepsilon_t^2 h_t^{-1}) \quad (14)$$

Where $h_t = a_0 + \sum_{i=1}^{q}\alpha_i y_{t-1}^2 + \sum_{j=1}^{p}\beta_j h_{t-j}$ and $\boldsymbol{\theta}' = (\alpha_0, \alpha_1, \alpha_2, \ldots, \alpha_q, \beta_1, \beta_2, \ldots, \beta_p)$ is the vector of parameters of the GARCH model.

3. AN ILLUSTRATION

For the present study the daily minimum market price (rupees per quintal) data of onion of Lasalgaon market of Maharashtra, India for the period of Jan, 2012 to Oct, 2019 has taken from National Horticultural Research and Development Foundation (http://nhrdf.org/en-us/) along with daily market arrival (quintal). The data set consists of 1773 data points from where last 35 observations are used for model validation purpose. Time plots for the series under study are plotted and shown in Fig. 1 and Fig. 2. The descriptive statistics of the data series is also given in Table 1. For checking the stationarity ADF (Dickey and Fuller, 1979) and PP (Phillips and Perron, 1988) test has been employed on both original as well as return series and the results (Table 2) indicate that return series are stationary. The correlation between minimum price and arrival data are calculated and it is found to be -0.3534 (p value <0.001) and that for return series 0.0777 (p value 0.001) indicating that both the correlations are statistically significant.

The long range dependency of the return series is visualized by plotting ACF (Fig. 3)and PACF (Fig. 4) of the data series. Long memory parameter is tested by GPH (Geweke and Porter-Hudak, 1983) and Sperio (Reisen, 1994) test for both the original series and return series (Table 3).

After confirming the presence of long memory in the return series ARFIMA and ARFIMAX model with market arrival as exogenous variable are fitted using maximum likelihood estimation (MLE) of parameter approach. Residuals obtained from the fitted models are then tested for ARCH effects using usual Ljung Box test (Ljung and Box, 1978) on squared residuals and ARCH-LM test (Table 4). After knowing the presence of ARCH effects in residuals of the both the model, GARCH model is fitted by MLE. The parameter estimates along with their standard errors and significant p values of ARFIMA-GARCH and ARFIMAX-GARCH models are given in Table 5 and 6 respectively. Model comparison criterions AIC, BIC and log likelihood values are also given in table 5 and 6. The residuals obtained from the ARFIMA-GARCH and ARFIMAX-GARCH are checked for possible presence of autocorrelation by Ljung Box test, ACF plot (Fig. 5) and PACF plot (Fig. 6) The obtained results indicate the independent and identicalness of the residual. In sample forecasting for validation set is carried out for both the model and RMSE and RMAPE values are obtained and given in Table 7.

4. CONCLUDING REMARKS

In the present study long memory model with exogenous variable is investigated with the help of onion daily price data of Lasalgaon market. One exogenous variable namely daily market arrival is incorporated after checking the significant correlation with the price series. From the ARCH-LM test it has found that the price series has volatility. ARFIMA-GARCH and ARFIMAX-GARCH is model is fitted accordingly to the data set under consideration. Residual analysis of the fitted models indicates that residuals are uncorrelated. The lower AIC and BIC value and increased log likelihood value of the ARFIMAX-GARCH model as compared to the ARFIMA-GARCH model are obtained. These values establish the superiority of the ARFIMAX-GARCH model in terms of modelling the price series in hand. Further, the lower RMSE and RMAPE values in ARFIMAX-GARCH model than the competing one indicates its enhanced forecasting efficiency. The reduction in RMSE and RMAPE values in ARFIMAX-GARCH model is not substantial as compared to the ARFIMA-GARCH model, this can be attributed mainly to low correlation between the market arrival and return price series. But, the data series having high correlation with exogenous variable is expected to give better results. This study has highlighted the importance of adding exogenous variable in the model for improving the modelling and forecasting efficiency.

REFERENCES

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

Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics*, **31**, 307-327.

Box, G.E.P., Jenkins, G.M. and Reinsel G.C. (2007). *Time-Series Analysis: Forecasting and Control,* 4th edition. Willey Publication.

Brockwell, P.J. and Davis, R.A. (1991). *Time Series: Theory and Methods*, 2nd Edition. Springer, New York.

Cheong, C.W., Isa, Z., and Nor, A.H.S.M. (2007). Modelling financial observable-volatility using long memory models. *Applied Financial Economics Letters*, **3**(**3**), 201-208.

Contreras-Reyes, J.E., and Palma, W. (2013). Statistical analysis of autoregressive fractionally integrated moving average models in R.*Computational Statistics*, **28**(**5**), 2309-2331.

Degiannakis, S. (2008). ARFIMAX and ARFIMAX-TARCH realized volatility modeling. *J. App. Statist.*, **35**(**10**), 1169-1180.

Dickey, D. and Fuller, W. (1979). Distribution of the estimators for autoregressive time series with a unit root. *J. Amer. Statist. Assoc.*, **74**, 427-431.

Engle, R.F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom in1ation. *Econometrica*, **50**, 987-1007.

Geweke, J. and Porter-Hudak, S. (1983). The estimation and application of long- memory time series models. *Journal of Time Series Analysis,* **4**, 221-238.

Granger, C.W.J. and Joyeux, R. (1980). An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis,* **4**, 221-238.

Helms, B.P., Kaen, F.R. and Rosenman, R.E. (1984). Memory in commodity futures contracts. *The Journal of Futures Markets*, **10**, 559-567.

Hosking, J.R.M. (1981). Fractional differencing. *Biometrika*, **68**, 165-176.

Jin, H.J. and Frechette, D. (2004). Fractional integration in agricultural future price volatilities. *American Journal of Agricultural Economics*, **86**, 432-443.

Lama, A., Jha, G.K., Paul, R.K. and Gurung, B. (2015) Modelling and forecasting of price volatility: An application of GARCH and EGARCH models. *Agricultural Economics Research Review*, **28(1)**, 73-82.

Ljung, G.M. and Box, G.E.P. (1978). On a measure of a lack of fit in time series models. *Biometrika*, **65**(**2**), 297-303.

Palma, W. (2007). *Long-memory time series: theory and methods* (Vol. 662). John Wiley and Sons.

Paul, R.K. (2014). Forecasting wholesale price of pigeon pea using long memory time-series models. *Agricultural Economics Research Review*, **27**(**2**), 167-176.

Paul, R.K., Gurung, B. and Paul, A.K. (2014). Modelling and forecasting of retail price of arhar dal in Karnal, Haryana. *Ind. J. Agril. Sci.*, **85**(**1**), 69-72.

Paul, R.K., & Himadri, G. (2014). Development of out-of-sample forecasts formulae for ARIMAX-GARCH model and their application. *J. Ind. Soc. Agril. Statist.*, **68**(**1**), 85-92.

Phillips, P.C.B. and Perron, P. (1988). Testing for unit roots in time series regression. *Biometrika*, **75**, 335-346.

Reisen, V.A. (1994) Estimation of the fractional difference parameter in the ARFIMA (p, d, q) model using the smoothed periodogram. *Journal of Time Series Analysis*, **15**(**1**), 335-350.

Robinson, P.M. (1994). Semi-parametric analysis of long-memory time series. *The Annals of Statistics*, **22**, 515-539.

Yang, K., Chen, L., and Tian, F. (2015). Realized volatility forecast of stock index under structural breaks. *Journal of Forecasting*, **34**(**1**), 57-82.

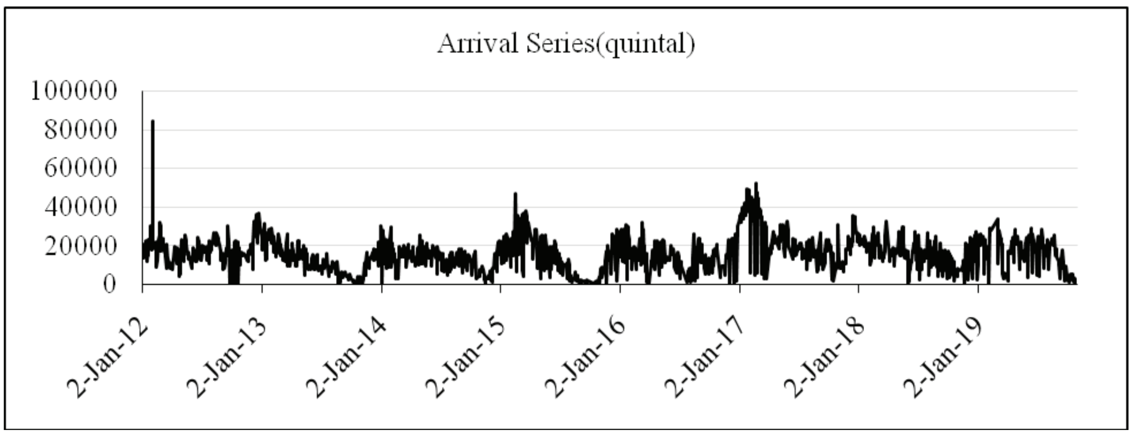

Fig. 1. Time plot for market arrival series

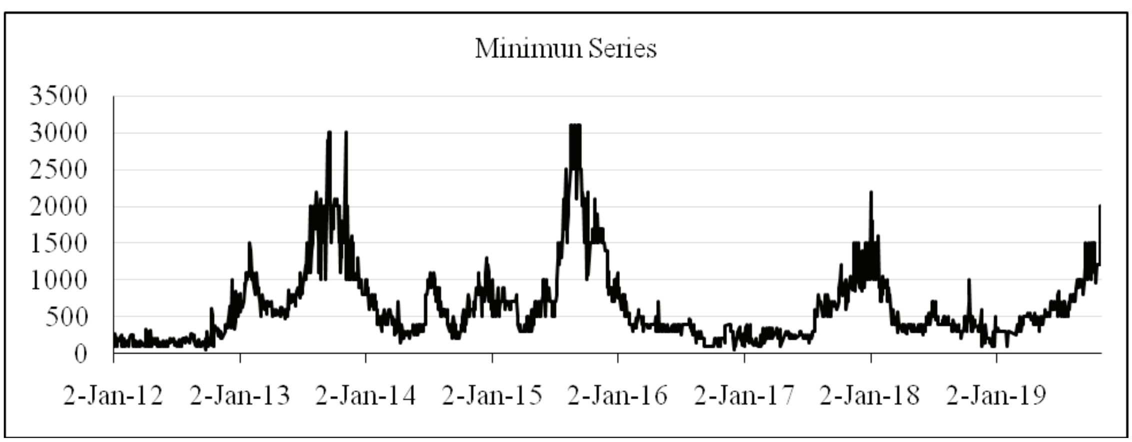

Fig. 2. Time plot for minimum price series

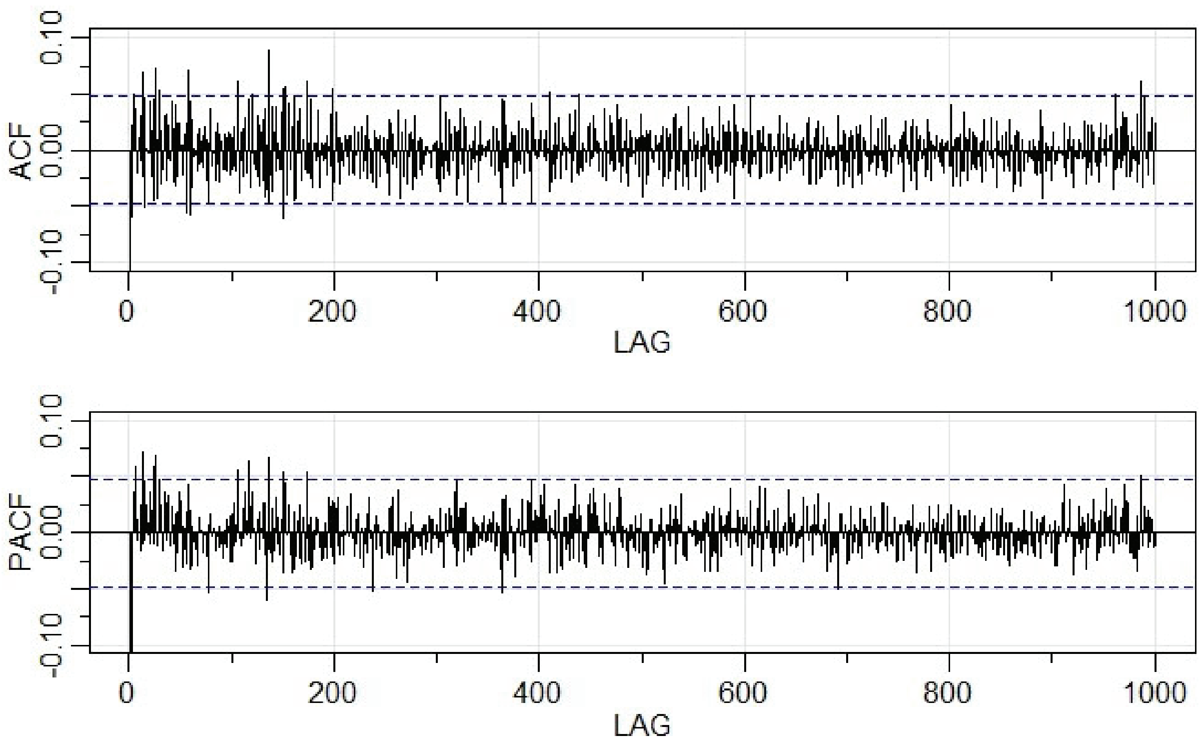

Fig. 3. ACF plot for return price series

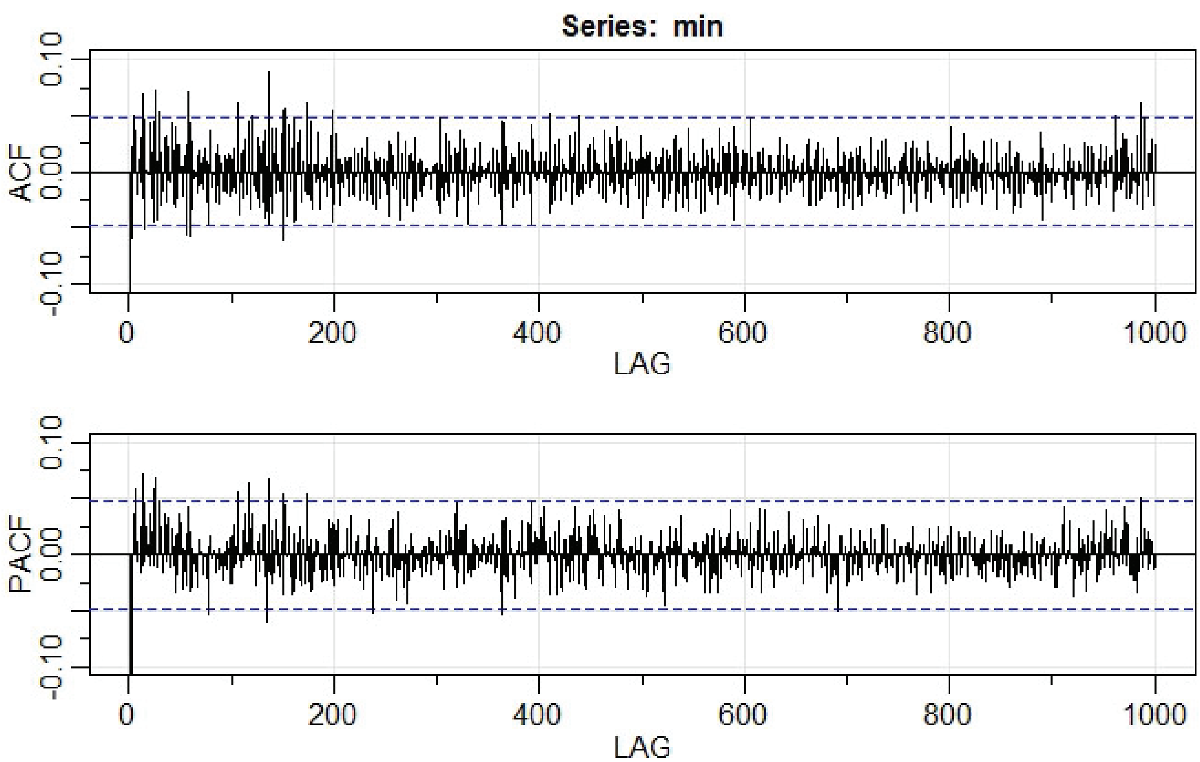

Fig. 4. PACF plot for return price series

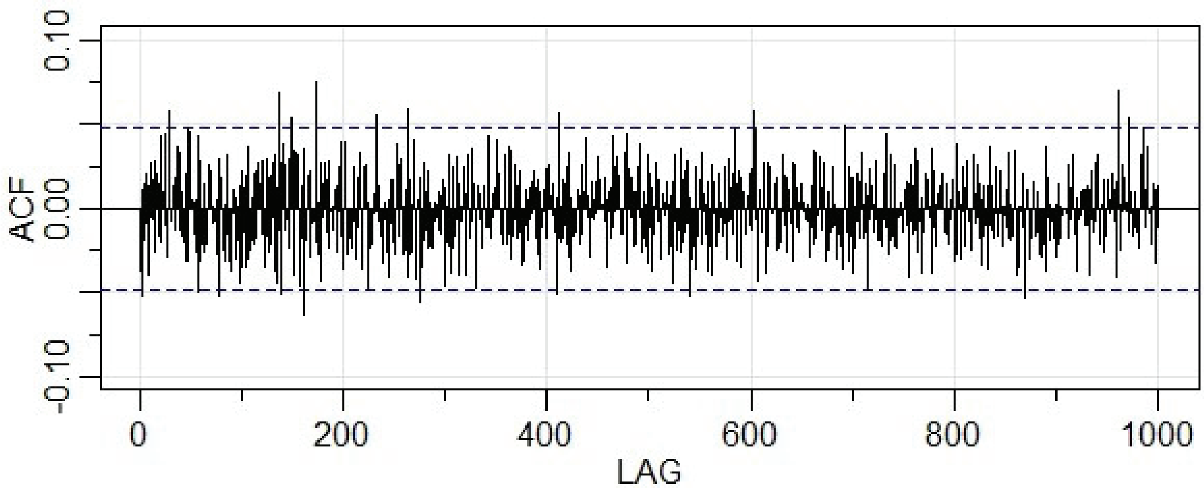

Fig. 5. ACF plot of ARFIMA-GARCH residuals

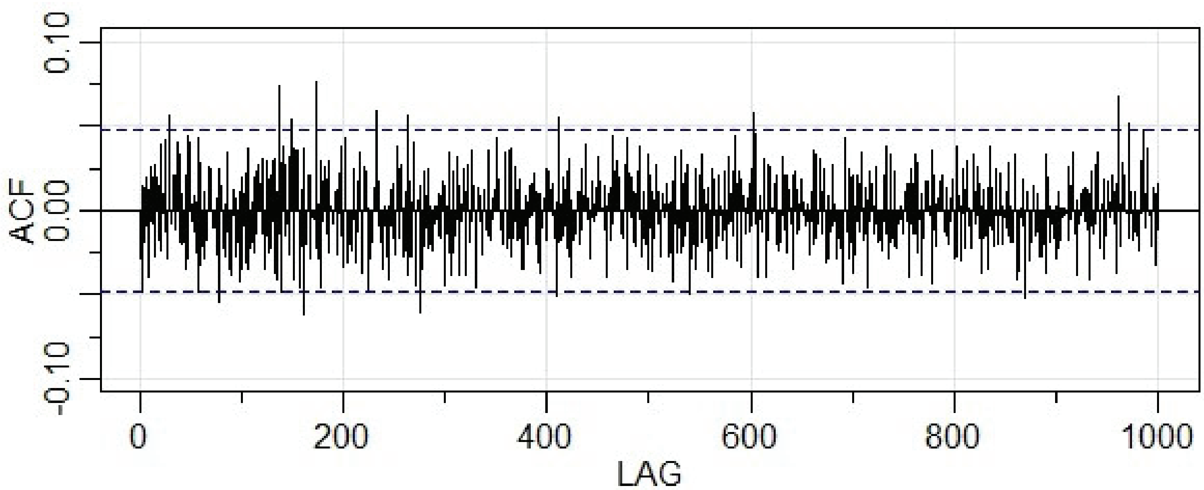

Fig. 6. ACF plot of ARFIMAX-GARCH residuals

Table 1. Descriptive Statistics of the data

Statistics	Arrival	Minimum price
Min	33	50
Max	84250	3100
Mean	15863.51	631.71
St. Deviation	8590.32	513.94
CV(%)	54.15	81.36
Skewness	0.74	1.93
Kurtosis	2.70	4.51

Table 2. Test for stationarity

Series	ADF test		PP test	
	d	p-value	Z	$p -$ value
Original Series				
Arrival	-4.807	<0.01	-810.78	<0.01
Minimum price	-2.66	0.298	-55.812	<0.01
Return Series				
Arrival	-10.336	<0.01	-1853.40	<0.01
Minimum Price	-11.246	<0.01	-2100.40	<0.01

Table 3. Long memory test

Test	GPH		Sperio	
Series	d	*Std. Error*	d	*Std. Error*
Minimum series	0.912*	0.136	0.907*	0.062
Minimum return series	0.354**	0.107	0.298**	0.048

* and ** denotes the significance at 5% and 1% level respectively

Table 4. Testing of ARCH effects

Test	Ljung Box test on squared residuals		ARCH-LM test	
Model	Chi square	p value	F value	p value
ARFIMA	13.607	0.0002	13.634	0.0002
ARFIMAX	16.254	<0.001	16.285	<0.001

Table 5. Parameter estimates of ARFIMA-GARCH model

Parameters	Estimate	Std. Error	p-Value
Mean equation			
Constant	0.0459	0.0147	0.002
d	0.2065	0.0668	0.002
MA 1	0.5893	0.0689	<0.001
Variance equation			
Constant	0.001	0.0003	<0.001
ARCH(1)	0.1108	0.0136	<0.001
GARCH(1)	0.8882	0.0118	<0.001
Log Likelihood		-173.371	
AIC		0.2042	
BIC		0.2168	

Table 6. Parameter estimates of ARFIMAX-GARCH model

Parameters	Estimate	Std. Error	p-Value
Mean equation			
Constant	0.0437	0.0147	0.002
d	0.2154	0.0767	0.005
MA 1	0.6143	0.0774	<0.001
Arrival	0.0021	0.0005	<0.001
Variance equation			
Constant	0.0015	0.0004	<0.001
ARCH(1)	0.1154	0.0146	<0.001
GARCH(1)	0.8823	0.0130	<0.001
Log Likelihood		-161.494	
AIC		0.1905	
BIC		0.2031	

Table 7. RMAPE and RMSE values of the fitted models

Criteria	RMAPE	RMSE
ARFIMA-GARCH	15.46	256.58
ARFIMAX-GARCH	15.42	256.42