Research Article

The Efficiency of Volatility Financial Model with Additive Outlier: A Monte Carlo Simulation

Intan Martina Md Ghani[1]*, and Hanafi A Rahim[2]

1 Affiliation [2]; hanafi@umt.edu.my
* Correspondence: intanmartina23@gmail.com

Received: 1st January 2019; Accepted: 10th January 2019; Published: 28th February 2019

Abstract: Observation that lies outside the overall pattern of its distribution is called outlier. The presence of outliers in time series data will effects on the modelling and also forecasting. Among the various types of outliers that effects the behavioral of finance series is additive outliers. This situation occurred because of recording errors, measurement errors or external factor. Therefore, the intention of this research is to investigate the effectiveness of volatility financial model with the presence of additive outliers. The appropriate approach in this paper is Autoregressive Moving Average-Generalized Autoregressive Conditional Heteroscedasticity (ARMA-GARCH) model. In this paper, data was simulated using ARMA (1, 0)-GARCH (1, 2) model via Monte Carlo method. There are three different sample size used in simulation study which are 500, 1000 and 1400. The comparison of effectiveness ARMA-GARCH model are based on MAE, MSE, RMSE, AIC, SIC and HQIC. The results of the numerical simulation indicate that when sample size increase, the effectiveness of ARMA-GARCH model diminished in the presence of additive outliers.

Keywords: Outliers; Financial, Behavioral finance; Additive outliers; Effectiveness; Simulation

About the Authors

Intan Martina Md Ghani is a postgraduate student in the Universiti Malaysia Terengganu. Highest academic degree: MSc in Mathematical Sciences, Universiti Malaysia Terengganu (2014). Research interests: modelling, GARCH, outlier, robust. Hanafi A Rahim is a Senior Lecturer in the School of Informatics and Applied Mathematics, Universiti Malaysia Terengganu. Highest academic degree: PhD in Applied Statistics in Universiti Teknologi Mara (2012). Research interests: applied statistics, GARCH, robust, time series analysis.

Public Interest Statement

Outlier is a very critical part in economy and business field. Its existence can give significant impact on volatility modelling and forecasting of financial series. Therefore, the sophisticated financial model that used among statisticians and economists is Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model. The finding of this research will help to governmental, investors, stock market traders and researchers to get an efficient volatility financial model to analyze financial series that contain outlier.

1. Introduction

The financial volatility model has been investigated by many researchers using financial time series data. Generally the financial time series consist of daily, weekly, monthly or yearly data. The series can be analyze and modelled by using Autoregressive (AR) model, Moving Average (MA) model, Autoregressive Moving Average (ARMA) model, Autoregressive Conditional Heteroscedasticity (ARCH) model, Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model and many other models. However, returns time series especially in economic, business and banking influenced by stylized facts.

There are two types of stylized facts that give significant impact on modelling which are volatility clustering and heavy-tails distribution. In statistics term, volatility clustering means unequal variance along the series. While the heavy-tails distribution occurred when the returns have excess kurtosis. This may cause by the existence of outliers. Over the past four decades the problem of outliers in the time series has begun identified by Fox (1972). Among the various types of outliers that effects to the behavioral of finance series is additive outliers (AO).

Previous researches have reported that the existing of outliers can give negative impacts such as bias to the GARCH parameters estimation (Sakata & White, 1998; Melo Mendes, 2000; Charles, 2008), on identification and estimation of the GARCH-type models (Carnero et al., 2007, 2012), and also on forecasting (Franses & Ghijsels, 1999; Carnero et al., 2007; Charles, 2008). Therefore, in an attempt to attain efficiency of the volatility financial model, most scholars applied ARMA (m,n)/GARCH(p,q) model.

Several studies have selected ARMA(m,n)-GARCH(p,q) model in modelling and forecasting such as in machine health condition (Pham & Yang, 2010) and stock exchange (Huq et al., 2013). While Behmiri and Manera (2015) used ARMA(p,q)-GARCH(2,2) model to estimate the persistence of volatility among metals. In another study, Liu and Shi (2013) and Sun et al. (2015) hybrid ARMA model with GARCH(-M) model in their research.

In contrast, the study by Franses and Ghijsels (1999) indicated that when AO was corrected, the forecast of stock market volatility improved. After six years Charles and Darné (2005) extended this work to innovative outliers. Both studies was selected GARCH model in forecasting volatility and examine outlier's effect. The analysis of AO and other types of outliers were carried out by Urooj and Asghar (2017) which preferred AR(1) model. Although there were many researches about outliers, few of them focus on AO. So it is necessary to do deep research on the effectiveness of volatility financial model in the presence of AO via simulation.

The organization of this paper is organized as follows. In Section 2 the ARMA (m,n) model, GARCH(p,q) model and additive outlier (AO) are briefly described. The simulation study in order to evaluate the efficiency without AO and with AO performed in Section 3. The result and discussion of ARMA (1, 0)-GARCH (1, 2) model based on simulation study reported in Section 4. Finally, the conclusion are summarized in Section 5.

2. Methodology

2.1. Methods

In this section, the time series models involves two models which are Autoregressive Moving Average (ARMA) model and Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model.

2.1.1. ARMA Model

In 1976, Box and Jenkins proposed ARIMA (m,D,n) models where m is the number of autocorrelation terms, D is the number of differencing elements and n is the number of moving average terms. The letter "I" in ARIMA used to differentiate when the series are not stationary. However when the time series is stationary, we can model it using three classes of time series process: autoregressive (AR), moving-average (MA) and mixed autoregressive and moving-average (ARMA).

An autoregressive model of order m, denoted as AR (m), can be expressed as

$$\gamma_t = \mu + \phi_1 \gamma_{t-1} + \phi_2 \gamma_{t-2} + \ldots + \phi_m \gamma_{t-m} + u_t \tag{1}$$

The moving average of order n which denoted as MA (n) can be expressed as

$$\gamma_t = \mu + u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2} + \ldots + \theta_n u_{t-n} \tag{2}$$

where u_t $(t=1,2,3,\ldots)$ is a white noise disturbance term with $E(u_t)=0$ and $\mathrm{var}(u_t)=\sigma^2$.

The combination of AR (m) model and MA (n) model formed of ARMA (m,n) model which expressed as

$$\gamma_t = \mu + \phi_1\gamma_{t-1} + \phi_2\gamma_{t-2} + \ldots + \phi_m\gamma_{t-m} + \theta_1 u_{t-1} + \theta_2 u_{t-2} + \ldots + \theta_n u_{t-n} + u_t \quad (3)$$

or in sigma notation

$$y_t = C + \sum_{i=1}^{m}\phi_i y_{t-i} + \sum_{j=1}^{n}\theta_j \varepsilon_{t-j} \quad (4)$$

where y_t is the daily rubber SMR20 prices, C is a constant term, ϕ_i are the parameter of the autoregressive component of order m, θ_j are the parameters of the moving average component of order n, and ε_t is the error term at time t. The order m and n are non-negative integers.

2.1.2. GARCH Model

The time varying heteroscedasticity model that popular among researchers is GARCH model. After four years an extension from ARCH model was developed by Bollerslev (1986) namely GARCH model. The GARCH model is more parsimonious (use fewer parameters) than ARCH model (Poon and Granger, 2003). There are two part that consist in GARCH model which are mean equation, y_t; see Equation (5) and variance equation σ_t^2; see Equation (7). The general form for GARCH (p,q) model can be written as follows:

$$y_t = C + \varepsilon_t \quad (5)$$

$$\varepsilon_t = z_t \sigma_t \quad (6)$$

$$\sigma_t^2 = \eta + \sum_{i=1}^{p}\beta_i \sigma_{t-i}^2 + \sum_{j=1}^{q}\alpha_j \varepsilon_{t-j}^2 \quad (7)$$

where y_t is an observed data series, C is a constant value, ε_t is the residual, z_t is the standardized residual with independently and identically distributed with mean equal to zero and variance equal to one and σ_t is the square root of the conditional variance with non-negative process, η is the long-run volatility with condition $\eta > 0$, $\beta_i \geq 0; i=1,\ldots,p$ and $\alpha_j \geq 0; j=1,\ldots,q$.

From the general form of GARCH (p,q) model, the GARCH(1,2) model can defined as

$$\sigma_t^2 = \eta + \beta_1\sigma_{t-1}^2 + \alpha_1\varepsilon_{t-1}^2 + \alpha_2\varepsilon_{t-2}^2 \quad (8)$$

If $\beta_i + \alpha_j < 1$, then GARCH (p,q) model is covariance stationary. The volatility is called persistence whenever the value of $\sum_{i=1}^{p}\beta_i + \sum_{j=1}^{q}\alpha_j$ is close to one. The unconditional variance of the error terms

$$\mathrm{var}(\varepsilon_t) = \frac{\eta}{1-\beta-\alpha} \tag{9}$$

2.1.3. Additive Outlier

Additive outlier (AO) is a type of outlier that effect to data especially in financial series. The AO was identified by Fox (1972) in AR model. This outlier occurred because of recording errors, measurement errors or external factor. AO also defines as an external or exogenous change (Urooj & Asghar, 2017).

From Equation (7), GARCH(p,q) model can be written as an ARMA(m,n) model for ε_t^2 (Bollerslev, 1986) as follows:

$$\varepsilon_t^2 = \eta + \sum_{i=1}^{r} (\alpha_i + \beta_i)\varepsilon_{t-i}^2 + v_t - \sum_{j=1}^{s} \beta_j v_{t-j} \tag{10}$$

with $r = \max\{p,q\}$ and $v_t = \varepsilon_t^2 - \sigma_t^2; t = 1, 2, \ldots, n$ where ε_t^2 known as outlier free time series, while v_t known as outlier-free residuals.
The Equation (10) can be written as

$$\begin{aligned} \varepsilon_t^2 &= \frac{\eta}{1-\alpha(L)-\beta(L)} + \frac{1-\beta(L)}{1-\alpha(L)-\beta(L)} v_t \\ &= \frac{\eta}{1-\alpha(L)-\beta(L)} + \pi^{-1}(L) v_t \end{aligned} \tag{11}$$

with $\alpha(L) = \sum_{i=1}^{q} \alpha_i L^i, \beta(L) = \sum_{j=1}^{p} \beta_j L^j$ and $\pi(L) = \frac{1-\alpha(L)-\beta(L)}{1-\beta(L)}$.
According to Chen and Liu (1993), when AO presence in GARCH model becomes

$$e_t^2 = \varepsilon_t^2 + \omega_{AO}\xi_{AO}(L)\mathrm{I}_t(T) \tag{12}$$

with
e_t^2 is true series ε_t^2,
ω_{AO} is the magnitude effect of AO,
$\xi_{AO}(L)$ is the dynamic pattern of AO effect,
$\mathrm{I}_t(T)$ is the indicator function which can explain the effect of AO as

$$\mathrm{I}_t(T) = \begin{cases} 1 & t = T \\ 0 & \text{otherwise} \end{cases}$$

where T is the location of AO occurring.

3. Simulation Study

To achieve the objective in this research, we conduct a Monte Carlo simulation. The simulation of time series was written and generated using statistical package R version 3.5.1 that developed by R Core Team (2018). During this process, the GARCH modelled using *tseries* package (Trapletti & Hornik, 2018) and *fGarch* package (Wuertz et al., 2017) which consist of *garchSpec*, *garchSim* and *garchFit* in R software. There are two situations involves in this simulation: contaminated with 0%

AO (also known as without AO) and contaminated with 10% AO (also known as with AO). The sample size used are 500, 1000 and 1400. The general algorithm conducted as follows:

1. The ARMA(1,0)-GARCH(1,2) model specified using *garchSpec* function with set the true value of parameters: mu= 0.043, ar= -0.312, omega= 0.011, alpha1= 0.224, alpha2= -0.136 and beta= 0.913.
2. The GARCH process simulated 500 observations with mean=0 and standard deviation=1 using *garchSim*.
3. The parameters of the ARMA(1,0)-GARCH(1,2) model fitted using *garchFit* function in normal error distribution.
4. The efficiency of ARMA(1,0)-GARCH(1,2) model in 0% AO was evaluated.
5. About 10% from sample size contaminated as AO. The locations and magnitudes of AO are identified.
6. After contaminated data, the parameters of the ARMA(1,0)-GARCH(1,2) model fitted in normal error distribution.
7. The efficiency of ARMA(1,0)-GARCH(1,2) model in 10% AO was evaluated.
8. Steps (1) to (6) then repeated for different sample size, n=1000 and 1400.

3.1. Model Selection

When comparing among different sample size for different situations of ARMA(1,0)-GARCH(1,2) model, then we select an appropriate model based on Akaike Information Criterion (AIC) (Akaike, 1974), Schwarz's Information Criterion (SIC) (Schwarz, 1978) and Hannan-Quinn Information Criterion (HQIC) (Hannan & Quinn, 1979).

The AIC, SIC and HQIC can be computed as

$$\mathrm{AIC} = -2\ln(L) + 2k \tag{13}$$

$$\mathrm{SIC} = -2\ln(L) + \ln(N)k \tag{14}$$

$$\mathrm{HQIC} = -2\ln(L) + 2\ln(\ln(N))k \tag{15}$$

where L is the value of the likelihood function evaluated at the parameter estimates, N is the number of observations, and k is the number of estimated parameters. The minimum value of AIC, SIC and HQIC was selected as the better model when comparing among models.

3.2. Model Evaluations

The performance of ARMA(1,0)-GARCH(1,2) model are evaluated using three measures which are Mean Absolute Error (MAE), Mean Square Error (MSE) and Root Mean Square Error (RMSE).

$$\mathrm{MAE} = \frac{1}{T}\sum_{t=T_1}^{T}\left(\left|\sigma_t^2 - \hat{\sigma}_t^2\right|\right) \tag{16}$$

$$\mathrm{MSE} = \frac{1}{T}\sum_{t=T_1}^{T}\left(\sigma_t^2 - \hat{\sigma}_t^2\right)^2 \tag{17}$$

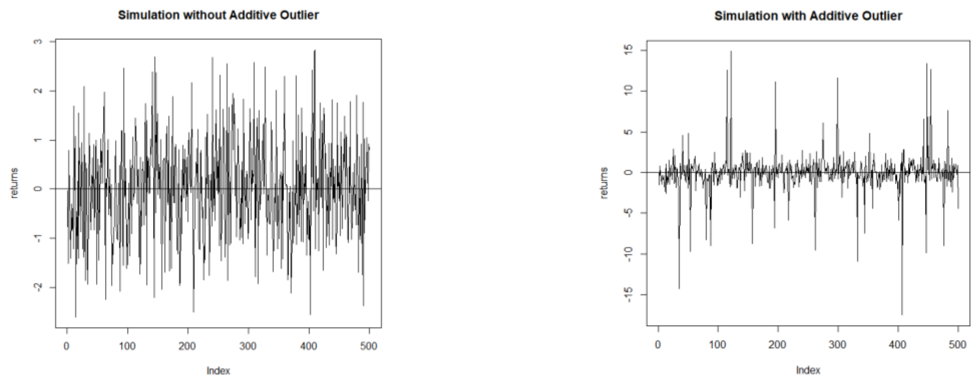
$$\mathrm{RMSE} = \sqrt{\frac{1}{T}\sum_{t=T_1}^{T}\left(\sigma_t^2 - \hat{\sigma}_t^2\right)^2} \tag{18}$$

where T is the number of total observations and T_1 is the first observation in out-of-sample. The σ_t^2 and $\hat{\sigma}_t^2$ is the actual and predicted conditional variance at time t, respectively. When comparing among different sample size for different situations of ARMA-GARCH models, the smallest value of MAE, MSE and RMSE are chosen as the best accurate model.

4. Results and Discussions

The results begin with the plot of returns for ARMA (1,0)-GARCH(1,2) model which simulates using *garchSim* function. The plot of returns without AO for sample size 500 is shown in Figure 1(a). When contaminated with 10% AO, we can see that there are large negative values especially on observation 407 which is -17.483.

Figure 1. Plot of returns for sample size, n=500

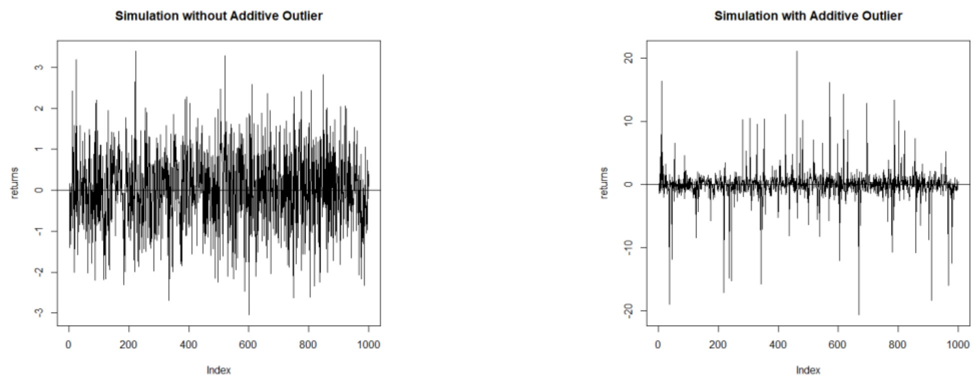


(a) (b)

(**a**) Simulation without additive outlier; (**b**) Simulation with additive outlier

Figure 2(a) and Figure 2(b) illustrates the plot of simulation without AO and with AO for sample size 1000, respectively. From the Figure 2(b), it is apparent that on observation 668 there are large negative values compared to Figure 2(a) which is -20.6930.

Figure 2. Plot of returns for sample size, n=1000

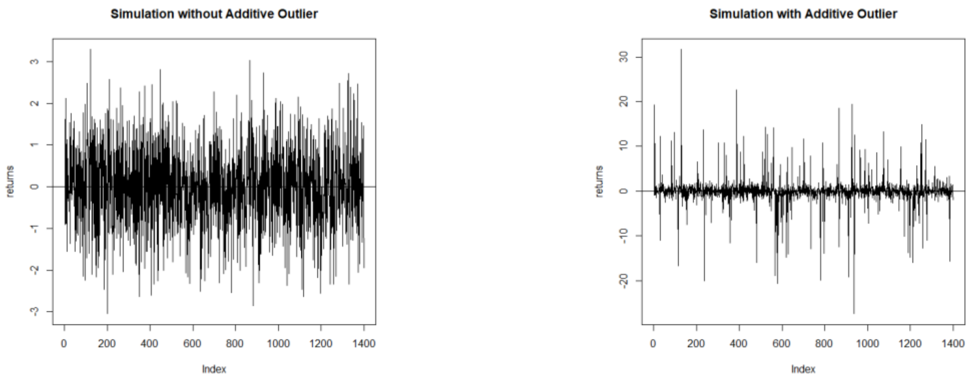


(a) **(b)**

(a) Simulation without additive outlier; (b) Simulation with additive outlier

The plot of returns without AO and with AO for 1400 observations depicted in Figure 3(a) and Figure 3(b), respectively. It appears from Figure 3(b) that, there are large negative values of returns especially on observation 937 which is -27.4280.

Figure 3. Plot of returns for sample size, n=1400



(a) **(b)**

(a) Simulation without additive outlier; (b) Simulation with additive outlier

The descriptive statistics of the simulation without AO are presented in Table 1. Data from this table provides the value of kurtosis in situation without AO are between the normal value, $-3 \leq x \leq 3$. This shows that the heavy tail does not exist in the simulation data for sample size 500, 1000 and 1400. However in situation with AO, the kurtosis value for sample size 500, 1000 and 1400 are 15.594292, 19.835252 and 23.1385, respectively. Therefore there is excess kurtosis in simulation which are larger than the normal value of 3. This can explain that when data is 10% contaminated, there exist heavier tails and distributed as leptokurtic.

Table 1. Descriptive Statistics for simulation without AO and with AO.

	n=500	**n=1000**	**n=1400**
	Without AO		
Mean	0.031453	0.017099	-0.001947
Variance	1.065112	0.989746	0.978074
Standard deviation	1.032043	0.994860	0.988976
Kurtosis	-0.110890	-0.076448	-0.072850
Skewness	0.080195	-0.007068	-0.011625
	With AO		
Mean	-0.033656	-0.051505	0.016678
Variance	6.411266	7.906186	10.983996
Standard deviation	2.532048	2.811794	3.314211
Kurtosis	15.594292	19.835252	23.138500
Skewness	-0.252945	-0.780921	0.133900

Source: Author's calculation using R software.

As illustrated in Table 2, the different sample size for both situations (without AO and with AO) was compared based on AIC, SIC and HQIC. In situation without AO, the value of AIC and SIC shows decrease of 3.43% and 3.42%, respectively from sample size 500 to 1400. However, for HQIC criteria there was an increase of 10.75% from sample size 500 to 1400 in situation with AO.

From the Table 2, it is apparent that when the sample size increase, the AIC, SIC and HQIC value in ARMA(1,0)-GARCH(1,2) model without AO is smaller than in ARMA(1,0)-GARCH(1,2) model with AO.

Table 2. Comparison Sample Size of Selection Criteria.

Criteria	**Sample size (n)**	**Without AO**	**With AO**
AIC	500	2.9228250	4.7116980

	1000	2.8364160	4.9163730
	1400	2.8226750	5.2316980
SIC	500	2.9225420	4.7114140
	1000	2.8363440	4.9163020
	1400	2.8226380	5.2316620
HQIC	500	2.9426710	4.7315440
	1000	2.8476070	4.9275650
	1400	2.8310770	5.2401000

Source: Author's calculation using R software.

Table 3 provides the result of comparison of different sample size and model evaluation for different situation (without AO and with AO). For MAE criteria, there was a decrease of 3.14% from sample size 500 to 1400 in situation without AO. While in situation with AO, the value of MSE and RMSE shows an increase of 72.39% and 31.3%, respectively from sample size 500 to 1400.

From Table 3, it is obvious that the value of MAE, MSE and RMSE in ARMA(1,0)-GARCH(1,2) model with AO is larger than in ARMA(1,0)-GARCH(1,2) model without AO.

Table 3. Comparison Sample Size of Model Evaluation.

Criteria	Sample size (n)	Without AO	With AO
MAE	500	0.8148005	1.3348890
	1000	0.7968617	1.3820510
	1400	0.7891958	1.5145180
MSE	500	1.0628440	6.3589060
	1000	0.9871511	7.8971140
	1400	0.9767424	10.9622200
RMSE	500	1.0309430	2.5216870
	1000	0.9935548	2.8101800
	1400	0.9883028	3.3109250

Source: Author's calculation using R software.

5. Conclusions

In this paper, the aim was to assess the effectiveness of ARMA(1,0)-GARCH(1,2) model with the presence of AO via simulation. The most obvious finding emerged from this paper is that whenever sample size increase, the efficiency of ARMA(1,0)-GARCH(1,2) model diminished in the presence of 10% AO. These findings enhance our understanding of the effects of contamination by outliers especially AO towards model estimation and model evaluation in forecasting. Further research might explore the other types of outliers that effects on the behavioral finance series such as innovative outliers, level shift outliers and temporary change outliers based on different specification of ARMA(m,n)-GARCH(p,q) model.

Funding: This research received no external funding.

Acknowledgments: We are grateful to the Universiti Malaysia Terengganu for their support.

6. References

Akaike, H. (1974). A New Look at the Statistical Model Identification. In Selected Papers of Hirotugu Akaike (Ed.), *Springer Series in Statistics (Perspectives in Statistics)* (pp. 215–222). New York, NY: Springer.

Behmiri, N. B., & Manera, M. (2015). The role of outliers and oil price shocks on volatility of metal prices. *Resources Policy*, *46*, 139-150.

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

Carnero, M. A., Peña, D., & Ruiz, E. (2007). Effects of outliers on the identification and estimation of GARCH models. *Journal of Time Series Analysis*, *28*(4), 471–497.

Carnero, M. A., Peña, D., & Ruiz, E. (2012). Estimating GARCH volatility in the presence of outliers. *Economics Letters*, *114*(1), 86–90.

Charles, A. (2008). Forecasting volatility with outliers in GARCH models. *Journal of Forecasting*, *27*, 551–565.

Charles, A., & Darné, O. (2005). Outliers and GARCH models in financial data. *Economics Letters*, *86*, 347–352.

Chen, C., & Liu, L.-M. (1993). Joint Estimation of Model Parameters and Outlier Effects in Time Series. *Journal of the American Statistical Association*, *88*(421), 284–297.

Fox, A. J. (1972). Outlier in Time Series. *Journal of the Royal Statistical Society. Series B (Methodological)*, *34*(3), 350–363. Retrieved from https://www.jstor.org/stable/2985071

Franses, P. H., & Ghijsels, H. (1999). Additive outliers, GARCH and forecasting volatility. *International Journal of Forecasting*, *15*(1), 1–9.

Hannan, E. J., & Quinn, B. G. (1979). The Determination of the Order of an Autoregression. *Royal Statistical Society. Series B (Methodological)*, *41*(2), 190–195.

Huq, M. M., Rahman, M. M., Rahman, M. S., Shahin, A., & Ali, A. (2013). Analysis of Volatility and Forecasting General Index of Dhaka Stock Exchange. *American Journal of Economics*, *3*(5), 229–242.

Liu, H., & Shi, J. (2013). Applying ARMA-GARCH approaches to forecasting short-term electricity prices. *Energy Economics*, *37*, 152–166.

Melo Mendes, B. V. D. (2000). Assessing the bias of maximum likelihood estimates of contaminated garch models. *Journal of Statistical Computation and Simulation*, *67*(4), 359–376.

Pham, H. T., & Yang, B. S. (2010). Estimation and forecasting of machine health condition using ARMA/GARCH model. *Mechanical Systems and Signal Processing*, *24*(2), 546–558.

Poon, S. H., & Granger, C. W. (2003). Forecasting Volatility in Financial Markets: A Review. *Journal of Economic Literature*, *41*(2), 478–539.

R Core Team. (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing. Retrieved from https://www.r-project.org/

Sakata, S., & White, H. (1998). High Breakdown Point Conditional Dispersion Estimation with

Application to S & P 500 Daily Returns Volatility. *Econometrica*, *66*(3), 529–567.

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, *6*(2), 461–464.

Sun, H., Yan, D., Zhao, N., & Zhou, J. (2015). Empirical investigation on modeling solar radiation series with ARMA-GARCH models. *Energy Conversion and Management*, *92*(March), 385–395.

Trapletti, A., & Hornik, K. (2018). tseries: Time Series Analysis and Computational Finance.

Urooj, A., & Asghar, Z. (2017). Analysis of the performance of test statistics for detection of outliers (additive, innovative, transient, and level shift) in AR (1) processes. *Communications in Statistics-Simulation and Computation*, *46*(2), 948-979.

Wuertz, D., Setz, T., Chalabi, Y., Boudt, C., Chausse, P., & Miklavoc, M. (2017). fGarch: Rmetrics - Autoregressive Conditional Heteroskedastic Modelling.