2024, 21(2): 0511-0523
https://doi.org/10.21123/bsj.2023.7489
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

Building a Sustainable GARCH Model to Forecast Rubber Price: Modified Huber Weighting Function Approach

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

Faculty of Ocean Engineering Technology and Informatics, Universiti Malaysia Terengganu, 21030 Kuala Nerus, Terengganu Darul Iman, Malaysia
*Corresponding Author.

Received 02/06/2022, Revised 07/05/2023, Accepted 09/05/2023, Published Online First 20/07/2023, Published 01/02/2024

Abstract

The unstable and uncertain nature of natural rubber prices makes them highly volatile and prone to outliers, which can have a significant impact on both modeling and forecasting. To tackle this issue, the author recommends a hybrid model that combines the autoregressive (AR) and Generalized Autoregressive Conditional Heteroscedasticity (GARCH) models. The model utilizes the Huber weighting function to ensure the forecast value of rubber prices remains sustainable even in the presence of outliers. The study aims to develop a sustainable model and forecast daily prices for a 12-day period by analyzing 2683 daily price data from Standard Malaysian Rubber Grade 20 (SMR 20) in Malaysia. The analysis incorporates two dispersion measurements (IQR/3 and Sn) and three levels of IO contamination 0%, 10%, and 20%. The results indicate that using the Huber weighting function with the IQR/3 measurement to build the AR(1)-GARCH(2,1) model leads to better sustainability. These findings have the potential to enhance the GARCH model by modifying the weighting function of the M-estimator.

Keywords: Autoregressive, Dispersion, Forecasting, GARCH, Huber.

Introduction

Forecasting volatility is crucial in the agricultural sector, mainly rubber. This may assist producers, traders and consumers in predicting future rubber prices. In recent years, there has been an increasing interest in the forecasting of natural rubber studies such as forecasting performance[1-3] and forecasting price[4-7]. In Malaysia's context, the Standard Malaysian Rubber Grade 20 (SMR 20) has chosen by researchers as empirical data in their research. However, the Malaysian natural rubber price fluctuates as a result of the world economy's decline[8]. These fluctuations can influence the volatility model as well as efficiency forecasting with the current volatility clustering.

Most researchers have used a time-varying volatility model such as the Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model to obtain reliable forecasts[1,9]. Although the GARCH model is very general, there are serious challenges, especially when there are outliers. Previous studies have found that outliers can have detrimental effects on parameter estimate[10-12], identification and estimation[13,14] and forecasting[13,15]. Therefore, robust methods are more preferred by researchers to reduce the influence of outliers. The well-known approach in robust methods is M-estimator.

The M-estimator has been one of the most interesting research subjects due to its down-weighting in minimizing residual function. The selection of the weighting function in the M-estimator will make model parameters less biased and forecast better performance during existing outliers[16]. Huber is the monotone weighting function that is most widely used in many areas[17-20]. Mathematically, the weighting function is dependent on standardized residuals with median absolute deviation (MAD) as a measure of dispersion. In contrast, [21] stated that MAD has two main flaws: low Gaussian performance (37%) and dependence on symmetric distributions. As a result, some robust dispersion, such as interquartile range, Q_n and S_n, have been proposed for heavy-tailed distributions.

A careful study of the literature reveals that the weighting function received little attention. A clear understanding of dispersion measurement in weighting functions is crucial to obtaining sustainable time series models, especially the GARCH model. The primary goal of this study was to build a sustainable GARCH model with applied two dispersion measurements (IQR/3 and S_n) in the Huber weighting function during the existence of an innovative outlier (IO). Besides, to forecast the 12-day rubber price in 2021.

The following section is structured as follows. The $\text{AR}(m)$ model, $\text{GARCH}(g,h)$ model, IO, central tendency and dispersion measurements, Huber weighting function in M-estimator and evaluation performance describes briefly in the section methodology. The simulation result discusses in the section on simulation study. The result of rubber price with forecasting performs in section empirical results. The summary of this paper includes in the section conclusion.

Materials and Methods

The conditional mean of the stationary time series model is the autoregressive (AR) model. The AR model of order m, $\text{AR}(m)$ can be expressed as

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_m Y_{t-m} + \varepsilon_t \qquad 1$$

where β_0 is a constant parameter with conditions $\beta_0 > 0, \beta_1, \beta_2, \ldots, \beta_m \geq 0$ are the constraints with non-negative integer and ε_t is the white noise, $\varepsilon_t \sim \text{WN}(0, \sigma^2)$, where WN is the white noise that's independent and identically distributed with a mean of zero.

Eq.1 is applicable only when the variance for the time series is constant. When the variance in time series data is non-constant, the generalized autoregressive conditional heteroscedasticity (GARCH) model established by [22] is appropriate. Suppose that $\varepsilon_t = z_t \sigma_t$, where $\{z_t\}$ is a $z_t \sim \text{N}(0,1)$. The symmetric $\text{GARCH}(g,h)$ model can be expressed as

$$\sigma_t^2 = \varphi_0 + \varphi_1 \sigma_{t-1}^2 + \cdots + \varphi_g \sigma_{t-g}^2 + \eta_1 \varepsilon_{t-1}^2 + \cdots + \eta_h \varepsilon_{t-h}^2 \qquad 2$$

where φ_0 is the constant parameter with conditions $\varphi_0 > 0, \varphi_1, \ldots, \varphi_g \geq 0$ and $\eta_1, \ldots, \eta_h \geq 0$. The $\text{GARCH}(g,h)$ model in Eq.2 specifies that the today's conditional variance depends on the first g past conditional variance and h past squared innovations. Problems arise when there are outliers in the data. There are several types of outliers such as innovative outliers.

Innovative Outlier:

An innovative outlier (IO), also known as internal change, is a data point that has an impact on subsequent observations[23]. The impact of IO is more complex than the impact of other forms of outliers[24,25]. For a stationary time series, IO will create a transient effect, while for a non-stationary time series, IO will produce a permanent level transition[26].

The dynamic pattern of the effect of IO outliers is represented as

$$\xi_{\text{IO}}(\text{M}) = \frac{1-\varphi_g \text{M}}{1-(\varphi_g+\eta_h)\text{M}} \qquad 3$$

As mentioned by [27], the existence of IO in the GARCH model becomes

$$e_t^2 = \omega_{\text{IO}} \xi_{\text{IO}}(\text{M}) I_t(\text{T}) + \varepsilon_t^2 \qquad 4$$

and can be considered as a regression model for ε_t^2, such as

$$e_t^2 = \omega_{\text{IO}} x_t + \varepsilon_t^2 \qquad 5$$

2024, 21(2): 0511-0523
https://doi.org/10.21123/bsj.2023.7489
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

where
e_t^2 is a series of observation ε_t^2,
ω_{IO} is the size of the IO's effect, which is $\omega_{\mathrm{IO}}(\mathrm{T}) = \frac{\sum_{t=\mathrm{T}}^{n} e_t^2 x_t}{\sum_{t=\mathrm{T}}^{n} x_t^2}$ with x_t represented as $x_t = \begin{cases} 0 & t \neq 1 \\ 1 & t = \mathrm{T}' \end{cases}$
$\xi_{\mathrm{IO}}(\mathrm{M})$ is the dynamic pattern of IO impact,
$I_t(\mathrm{T})$ is the indicator function that describes IO's effect as $I_t(\mathrm{T}) = \begin{cases} 1 & t = \mathrm{T} \\ 0 & \text{otherwise} \end{cases}$ with T is the location where IO emerged.

To overcome the effect of an innovative outlier, researchers, for example, Huber, have suggested using different types of measures instead of regular mean and variance as the basis for modelling the time series. Such measures were next to be discussed.

Measure of Central Tendency and Dispersion:

There are two measures were considered in the Huber weight function: central tendency and dispersion. The median was selected as a central tendency due to robustness against outliers. The interquartile range (IQR) is a dispersion measurement expressed as the distance between the 75% percentile ($\mathrm{Q}_{0.75}$) and the 25% percentile ($\mathrm{Q}_{0.25}$) of the data[28]. The IQR can be defined as

$$\mathrm{IQR} = \mathrm{Q}_{0.75} - \mathrm{Q}_{0.25} \qquad 6$$

This measure has a breakdown point of 25% [21,29]. The outcome of the simulation by [30] suggested that median and median absolute deviation (MAD) or IQR could be reasonable alternatives to mean and variance. Nevertheless, this paper used the IQR/3 which was suggested by Ghani and Rahim[31].

The measures of $\mathrm{S_n}$ were proposed by [21] as an alternative to MAD. This measure was convenient in the heavy-tailed and skewed distributions. The explanation of $\mathrm{S_n}$ also includes in the [32] research. The $\mathrm{S_n}$ can be defined as

$$\mathrm{S_n} = 1.1926\mathrm{med}_i\{\mathrm{med}_j|X_i - X_j|\} \qquad 7$$

where (med_j) is the inner median with $\lfloor(n+2)+1\rfloor$-th order statistic and (med_i) is the outer median with $\lfloor(n/1)+2\rfloor$-th order statistic. Huber in [33] has suggested a weight function to overcome the effect of heteroscedasticity in modelling time series data.

Huber Weight Function:

The Huber M-estimation[33] is a common robust estimation approach. The weight function, $w(d)$ in the M-estimator was used to reduce the effect of heteroscedasticity on the standard error of approximate coefficients. Huber is the well-known weight function in M-estimator. The weight function of Huber is defined as

$$w(d) = \begin{cases} 1 & ,\text{for } |d| \leq 1.345 \\ \frac{1.345}{|d|} & ,\text{for } |d| > 1.345 \end{cases} \qquad 8$$

with 1.345 as the default scaling constant for Huber, which produces 95% asymptotic efficiency for the normal distribution, ε_t and d term is the standardized residuals.

Generally, the standardized residual is formulated as

$$d = \frac{\varepsilon_i}{s} = \frac{x-\mu}{\sigma} \qquad 9$$

where in the conventional approach, the μ and σ in Eq.4 represent mean and median absolute deviation (MAD), respectively. In this paper, the researchers have suggested a modification to the Huber weight function.

Modified Huber Weight Function:

Even if the conventional approach is accurate, it can cause problems with location and dispersion measurements. Therefore, the modification of mean and MAD in Eq.9 is made as

$$\bar{d} = \frac{x - \mathrm{median}}{\mathrm{IQR}/3} \qquad 10$$

and

$$\bar{\bar{d}} = \frac{x - \mathrm{median}}{\mathrm{S_n}} \qquad 11$$

with median is the central tendency measurement, while IQR/3 and $\mathrm{S_n}$ are two dispersion of measurements. For the building of a sustainable GARCH model, the performance dispersion measurement was more considered.

Performance Evaluation:

The efficiency of various AR(m)-GARCH(g,h) model specifications were compared using Akaike's Information Criteria (AIC)[34,35]. While the mean absolute error (MAE), mean square error (MSE)[36], and root mean square error (RMSE) was used to evaluate the effectiveness of the IQR/3 and $\mathrm{S_n}$. The AIC, MAE, MSE and RMSE measures are calculated as follows:

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

2024, 21(2): 0511-0523
https://doi.org/10.21123/bsj.2023.7489
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

$$\mathrm{MAE} = \frac{1}{\mathrm{T}}\sum_{t=\mathrm{T}_1}^{\mathrm{T}}(|\sigma_t^2 - \hat{\sigma}_t^2|)$$
$$\mathrm{MSE} = \frac{1}{\mathrm{T}}\sum_{t=\mathrm{T}_1}^{\mathrm{T}}(\sigma_t^2 - \hat{\sigma}_t^2)^2$$
$$\mathrm{RMSE} = \sqrt{\frac{1}{\mathrm{T}}\sum_{t=\mathrm{T}_1}^{\mathrm{T}}(\sigma_t^2 - \hat{\sigma}_t^2)^2}$$

with L is the value of the likelihood function evaluated at the parameter estimates, k is the number of parameters to be estimated, T is the number of total observations, T_1 is the initial observation, σ_t^2 is the actual conditional variance at time t and $\hat{\sigma}_t^2$ is the predicted conditional variance at time t. The lower the AIC, MAE, MSE and RMSE values, the more accurate the dispersion measurement.

Simulation Study:

In this section, evidence on the dispersion measurement of the Huber weighting function in the different percentages of IO contamination is provided. Three types of IO percentage contamination will be examined: 0%, 10% and 20%. Seven sets of simulations of the AR(1)-GARCH(2,1) model with different time series lengths, T for T=200, 500 and 1000 are generated. The seven sets of simulations are summarized in Table 1:

Table 1. Grouping of different contamination IO and dispersion measurements

Model	Contamination IO	Dispersion measurements
Model 1	0%	Without weight
Model 2	10%	Without weight
Model 3	10%	IQR/3
Model 4	10%	S_n
Model 5	20%	Without weight
Model 6	20%	IQR/3
Model 7	20%	S_n

During this part, the AR(1)-GARCH(2,1) model used the *t series* package[37] and the *fGarch* package[38] in the R software version 3.6.3, which was developed by [39]. The general procedure of modifying Huber weight during IO contamination was conducted as follows:

1) The AR(1)-GARCH(2,1) model specified using *garchSpec* function with stipulated the true value of parameters: $\beta_0 = 0.0001622, \beta_1 = 0.2225, \varphi_0 = 0.0000033, \varphi_1 = 0.3725, \varphi_2 = 0.4810$ and $\eta_1 = 0.1388$.
2) In the beginning, the GARCH process simulated 200 observations with a mean is 0 and a standard deviation is 1 using *garchSim*.
3) About 10% of the sample size was contaminated as IO. The locations and magnitudes of IO are identified. The magnitude for each contaminated point was calculated by using normal distribution where the mean is 0 and the standard deviation is 16.
4) The modification of the Huber weight function with using IQR/3 as dispersion measurement was calculated to be 10% contamination of IO. The new data was determined based on the modified weighting the Huber function that was given to the 10% contamination of IO.
5) The modification of Huber weight function with using S_n as dispersion measurement was calculated to be 10% contamination of IO. The new data was determined based on the modified weighting Huber function that was given to the 10% contamination of IO.
6) Steps 3 to 5 are then repeated with increased contamination of IO to 20%.
7) The parameters of the AR(1)-GARCH(2,1) model for three situations fitted using *garchFit* function in normal error distribution.
8) The performance of the AR(1)-GARCH(2,1) model for three situations was evaluated.
9) Steps 1 to 8 then are repeated for different time series lengths, T=500 and T=1000. All-time series lengths were carried out for 1000 trials.

Table 2 presents the performance of the AR(1)-GARCH(2,1) model with percentage change via simulation analysis for T=200, 500 and 1000. The MAE, MSE, and RMSE values in model 2 increased to 1.1077, 4.3959, and 2.0966, respectively, when the data was contaminated with 10% IO. As IO contamination reached 20%, the three measures in model 5 increased as well ($\mathrm{MAE}_{20\%\ \mathrm{IO}}$ = 1.1518, $\mathrm{MSE}_{20\%\ \mathrm{IO}}$ = 4.6318, $\mathrm{RMSE}_{20\%\ \mathrm{IO}}$ = 2.1522). This result showed that all three dispersion measurements were higher than 0% IO (Model 1) during contamination with 10% IO and 20% IO.

2024, 21(2): 0511-0523
https://doi.org/10.21123/bsj.2023.7489
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

As Table 2 shows, the dispersion statistics of IQR/3 and S_n affect the value of MAE, MSE and RMSE. The minimum MAE value during contaminated 10% IO was recorded by model 3 at 0.4888, which dropped by 55.87 percent compared to model 2. Model 4 reveals that the MAE has likewise decreased by 0.7614 (-31.26%). For the MSE, model 3 recorded a minimum value of 0.2835 as compared to model 4 in the 10% of IO contamination. Model 3 was the highest percentage reduction for the RMSE at 74.61 percent contrasting with model 4 (57.19%).

In the 20% contamination of IO, model 6 indicates a minimum value for all three measures. For MAE and RMSE, model 6 reported a minimum value of 0.4956 and 0.5385, respectively, which declined by 56.97 percent and 74.98 percent compared with model 5. Otherwise, model 7 declined by 31.84 percent and 57.02 percent for MAE and RMSE, respectively. Model 6 and model 7 during contamination with 20% IO decreased the MSE to 0.29 and 0.8556, respectively; dropping by 93.74 percent and 81.53 percent. Although S_n is recognizable as a robust dispersion measurement and simplicity facilitates computation[21], however, our findings revealed that the IQR/3 showed more efficiency than S_n during contamination with 10% and 20% IO.

Table 2. Simulation of evaluation performance with a percentage change for model 1 to model 7

Time series length, T	Model	MAE	% MAE	MSE	% MSE	RMSE	% RMSE
200	Model 1	0.7865	-	0.9626	-	0.9811	-
	Model 2	1.1077	40.84	4.3959	356.67	2.0966	113.70
	Model 3	0.4888	-55.87	0.2835	-93.55	0.5324	-74.61
	Model 4	0.7614	-31.26	0.8056	-81.67	0.8976	-57.19
	Model 5	1.1518	46.45	4.6318	381.18	2.1522	119.37
	Model 6	0.4956	-56.97	0.2900	-93.74	0.5385	-74.98
	Model 7	0.7851	-31.84	0.8556	-81.53	0.9250	-57.02
500	Model 1	0.7679	-	0.9191	-	0.9587	-
	Model 2	1.1150	45.20	6.6408	622.53	2.5770	168.80
	Model 3	0.5024	-54.94	0.3005	-95.47	0.5482	-78.73
	Model 4	0.7709	-30.86	0.8258	-87.56	0.9087	-64.74
	Model 5	1.3630	77.50	10.4150	1033.17	3.2272	236.62
	Model 6	0.5408	-60.32	0.3482	-96.66	0.5901	-81.71
	Model 7	0.8642	-36.60	1.0481	-89.94	1.0238	-68.28
1000	Model 1	0.7957	-	0.9953	-	0.9977	-
	Model 2	1.1748	47.64	6.6377	566.87	2.5764	158.24
	Model 3	0.4909	-58.21	0.2818	-95.76	0.5308	-79.40
	Model 4	0.7897	-32.78	0.8732	-86.84	0.9345	-63.73
	Model 5	1.5672	96.96	16.5197	1559.69	4.0644	307.39
	Model 6	0.5426	-65.38	0.3496	-97.88	0.5913	-85.45
	Model 7	0.8852	-43.52	1.1128	-93.26	1.0549	-74.05

Note: % MAE, % MSE and % RMSE represent percentage changes of MAE, MSE and RMSE, respectively.

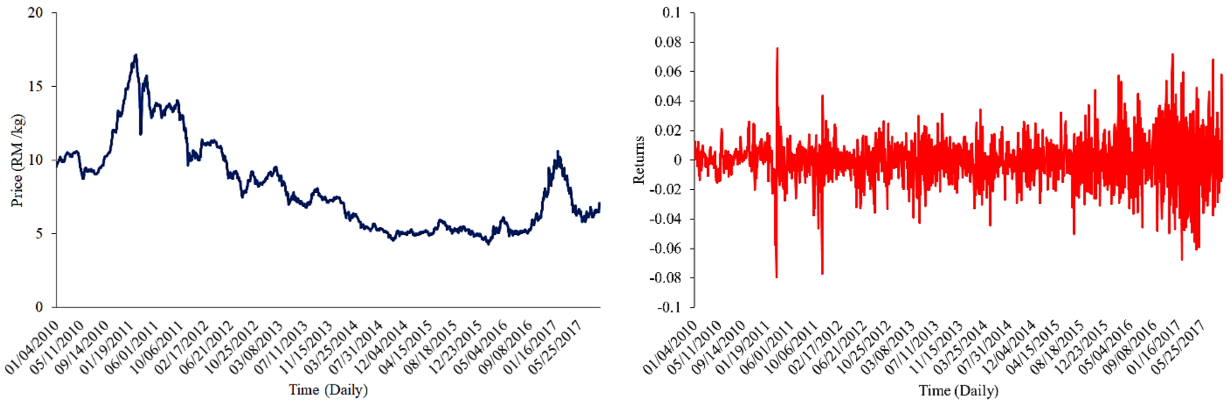
Results and Discussion

Empirical (Real Data) Results:

In this section, the daily price data for Standard Malaysian Rubber Grade 20 (SMR 20) are examined. These secondary data were obtained from the official website of the Malaysian Board of Rubber, which spans the period from 4th January 2010 to 30th December 2020. The empirical was executed using the *t series* package[36] and *fGarch* package[37] in R program version 3.6.3[38].

Fig. 1 (a) exhibits a clear trend of 1878 daily observations price of the rubber SMR 20 (in RM per kilogram) in Malaysia from 4th January 2010 to 11th September 2017. When daily prices are transformed to log returns, the plot of daily rubber SMR 20 returns clearly shows volatility clustering, as seen in Fig. 1 (b).

2024, 21(2): 0511-0523
https://doi.org/10.21123/bsj.2023.7489
P-ISSN: 2078-8665 - E-ISSN: 2411-7986



(a) **(b)**

Figure 1. Plot of daily (a) price and (b) returns on the SMR 20 for T=1878.

Table 3 illustrates the descriptive statistics of the daily returns for SMR 20. In daily returns, the range is between -0.07929 and 0.07568. The expected returns showed that -0.000171 per day. In daily returns, excess kurtosis occurs, which is 3.2659 greater than the usual value of 3. This may clarify that the data includes heavier tails and distributes them as leptokurtic.

Table 3. Descriptive statistics of the daily price and returns of SMR 20

Statistics	Observation	Minimum	Maximum	Mean	Skewness	Kurtosis
Price	1878	4.265	17.17	8.1821	0.8231	-0.1320
Returns	1877	-0.07929	0.07568	-0.000171	-0.07025	3.2659

The first step in time series data is to test the unit root. The R output is based on Phillips-Perron (PP) test[40], Augmented Dickey-Fuller (ADF) test[41] and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test[42] are shown in Table 4. As Table 4 shows, rejecting the null hypothesis of a unit root for the PP test and ADF test at the 1% level of significance. This indicates that the series is stationary where there is no unit root. Meanwhile, the *p*-value for KPSS test failed to reject the null hypothesis of a unit root at 10% level of significance. This indicates that the series is stationary and there is no unit root. As all three tests show that the series is stationary and has no unit root, the following step is carried out.

Table 4. Test of unit root

Test	Value	Truncation lag parameter	*p*-value	Decision
Phillips-Perron	-1498.4	8	<0.01	Reject null hypothesis
Augmented Dickey-Fuller	-10.816	12	0.01	Reject null hypothesis
Kwiatkowski-Phillips-Schmidt Shin	0.14167	8	0.1	Fail to reject null hypothesis

In this paper, the AR(1) model is used as the conditional mean. Table 5 provides the result of ARCH effect using Engle's Lagrange Multiplier (LM) test[43]. The *p*-value for the LM test indicates that the null hypothesis—that there was no ARCH effect—was rejected at the 5% level of significance. From Table 5, the result showed the presence of ARCH effect in daily returns SMR 20. This can be explained that heteroscedastic appearing in residuals.

Table 5. Test for ARCH effect

λ^2	df	*p*-value
317.05	12	$< 2.2 \times 10^{-16}$

Therefore, the AR(1) is required to be combined with the conditional variance, i.e. the GARCH model. Four models were produced from different conditional variance specifications in GARCH(g, h) models, where g and h order were either 1 or 2. The four different specification models were compared using AIC criteria to determine the

https://doi.org/10.21123/bsj.2023.7489
P-ISSN: 2078-8665 - E-ISSN: 2411-7986

best one. Table 6 shows a comparison of four models. The GARCH(2,1) model exhibits the lowest AIC value. The best in-sample part, according to Table 6, was the GARCH(2,1) model.

Table 6. Specifications of GARCH(g, h) model

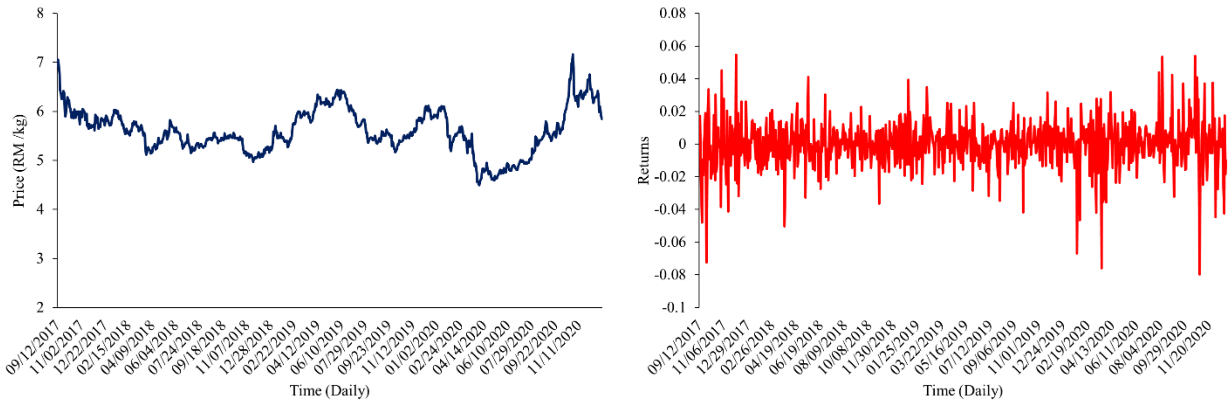
Model	GARCH (1,1)	GARCH (1,2)	GARCH (2,1)	GARCH (2,2)
AIC value	-5.9605	-5.9593	-5.9647	-5.9636

To obtain the best fit between AR(1) and the four specifications GARCH(g, h) model, the AIC criteria in Table 7 were shown. The result in Table 7, showed that the AR(1)-GARCH(2,1) model reported the smallest value of AIC compared to the three types of specifications AR(1)-GARCH(g, h) models.

Table 7. Selection criteria of AR(1)-GARCH(g, h) model for T=1878

Model	AR(1)-GARCH (1,1)	AR(1)-GARCH (1,2)	AR(1)-GARCH (2,1)	AR(1)-GARCH (2,2)
AIC value	-5.9774	-5.9756	-5.9797	-5.9786

For the out-of-sample part, the pattern of the 805 daily SMR 20 price can be seen in Fig. 2 (a). Because the LM test implied an ARCH effect in daily returns SMR 20, volatility clustering was evident in the returns presented in Fig. 2 (b).



(a) **(b)**

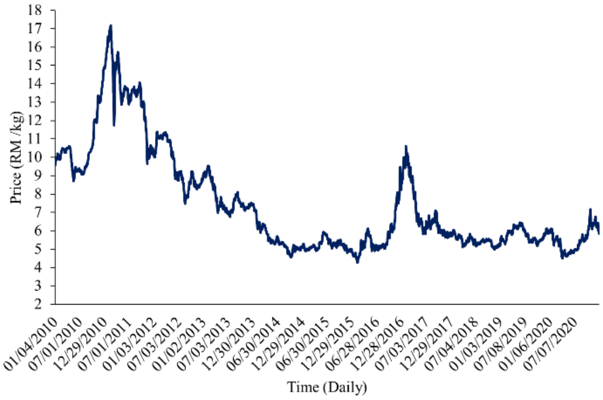
Figure 2. Plot of daily (a) price and (b) returns on the SMR 20 for T=805.

The performance of forecasting for various specifications AR(1)-GARCH(g, h) model is provided in Table 8. Based on the AIC criteria, the AR(1)-GARCH(2,1) model showed the lowest AIC value. The best model out of the three was determined to be AR(1)-GARCH(2,1).

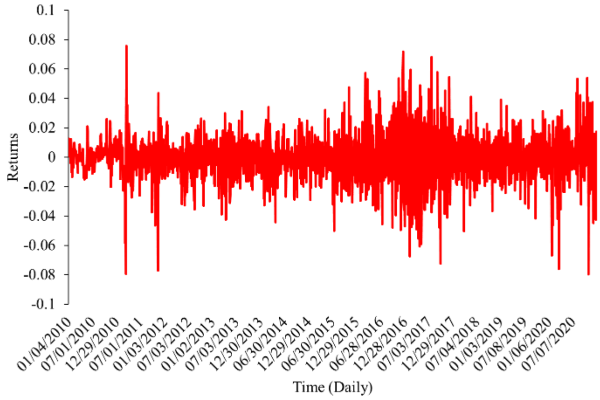
Table 8. Selection criteria of AR(1)-GARCH(g, h) model for T=805

Model	AR(1)-GARCH(1,1)	AR(1)-GARCH(1,2)	AR(1)-GARCH(2,1)	AR(1)-GARCH(2,2)
AIC value	-5.6761	-5.6745	-5.6764	-5.6763

The overall pattern of the 2683 daily SMR 20 price and returns from 4th January 2010 to 30th December 2020 illustrates in Fig. 3 (a) and Fig. 3 (b), respectively.



(a)



(b)

Figure 3. Plot of daily (a) price and (b) returns on the SMR 20.

The estimated conditional mean and conditional variance equation from daily SMR 20 is AR(1)-GARCH(2,1) was obtained. The five models expressed as

$$Y_t(\text{Model 1}) = \underset{(0.7381)}{0.0001622} + \underset{(9.8807^{****})}{0.2225}Y_{t-1} \quad (12)$$

$$\sigma_t^2(\text{Model 1}) = \underset{(3.5143^{***})}{0.0000033} + \underset{(3.419^{****})}{0.3725}\sigma_{t-1}^2 + \underset{(4.722^{****})}{0.4810}\sigma_{t-2}^2 + \underset{(8.2301^{****})}{0.1388}\varepsilon_{t-1}^2 \quad (13)$$

$$Y_t(\text{Model 3}) = \underset{(2.1906^{**})}{0.0003} + \underset{(0.6731)}{0.0136}Y_{t-1} \quad (14)$$

$$\sigma_t^2(\text{Model 3}) = \underset{(1.2876)}{0.0000021} + \underset{(1.3622)}{0.4591}\sigma_{t-1}^2 + \underset{(1.1703)}{0.3860}\sigma_{t-2}^2 + \underset{(2.7845^{***})}{0.1042}\varepsilon_{t-1}^2 \quad (15)$$

$$Y_t(\text{Model 4}) = \underset{(0.7223)}{0.0002} - \underset{(-1.8941^{*})}{0.0447}Y_{t-1} \quad (16)$$

$$\sigma_t^2(\text{Model 4}) = \underset{(3.1718^{***})}{0.0000216} + \underset{(2.8587^{***})}{0.3667}\sigma_{t-1}^2 + \underset{(1.8221^{*})}{0.2467}\sigma_{t-2}^2 + \underset{(7.2588^{****})}{0.2403}\varepsilon_{t-1}^2 \quad (17)$$

$$Y_t(\text{Model 6}) = \underset{(1.1609)}{0.0002} - \underset{(-8.0925^{****})}{0.1609}Y_{t-1} \quad (18)$$

$$\sigma_t^2(\text{Model 6}) = \underset{(1.3438)}{0.0000033} + \underset{(1.1543)}{0.3308}\sigma_{t-1}^2 + \underset{(1.8564^{*})}{0.5222}\sigma_{t-2}^2 + \underset{(2.8029^{***})}{0.0894}\varepsilon_{t-1}^2 \quad (19)$$

$$Y_t(\text{Model 7}) = \underset{(0.1624)}{0.00004} - \underset{(-13.3632^{****})}{0.3209}Y_{t-1} \quad (20)$$

$$\sigma_t^2(\text{Model 7}) = \underset{(4.1942^{****})}{0.000062} + \underset{(1.8406^{*})}{0.1889}\sigma_{t-1}^2 + \underset{(2.4245^{**})}{0.2715}\sigma_{t-2}^2 + \underset{(7.0246^{****})}{0.2438}\varepsilon_{t-1}^2 \quad (21)$$

The bracket values under the coefficient for all equations show the *t*-statistics, which are statistically significant at the level of 0.1% (****), 1% (***) and 5% (**) and 10% (*).

To ensure that IQR/3 is more efficient than S_n, the seven models were compared based on performance evaluation. Table 9 shows the performance of the AR(1)-GARCH(2,1) model based on the percentage contamination of IO and dispersion statistics in the Huber weight function. The MAE, MSE, and RMSE values increased to 0.304604, 6.342343, and 2.518401, respectively, when the daily returns of SMR 20 were contaminated with 10% IO. As IO contamination reached 20%, the three measures increased as well ($MAE_{model\ 5}$ = 0.634847, $MSE_{model\ 5}$ = 10.88619, $RMSE_{model\ 5}$ = 3.299422). This result showed that all three dispersion measurements in model 2 and model 5 were higher than in model 1.

Table 9. Evaluation performance for model 1 to model 7 of the SMR 20 price

Model	MAE	MSE	RMSE
Model 1	0.010456	0.000219	0.014783
Model 2	0.304604	6.342343	2.518401
Model 3	0.005702	0.000039	0.006263
Model 4	0.009675	0.000138	0.011732
Model 5	0.634847	10.88619	3.299422
Model 6	0.006703	0.000055	0.007401
Model 7	0.011707	0.000199	0.014120

From the data in Table 10, it is apparent that the four models affect the value of MAE, MSE and RMSE. The minimum MAE value in model 3 presents 0.005702, dropped by 98.13 percent compared with model 2 in Table 9. While model 4 also dropped by 0.009675 (-96.82%). For the MSE, model 3 recorded a minimum value of 0.000039 as compared to model 4 (0.000138) in the 10% of IO contamination. Model 3 was the highest percentage reduction for the RMSE at 99.75 percent contrasting with model 4.

In the 20% contamination of IO, model 6 indicates a minimum value for all three measures. For MAE and RMSE, model 6 reported a minimum value of 0.006703 and 0.007401, respectively, which declined by 98.94 percent and 99.78 percent compared with model 5 in Table 9. Otherwise, model 7 declined by 98.16 percent and 99.57 percent for MAE and RMSE, respectively. The MSE of model 6 and model 7 declined by 100 percent with 0.000055 and 0.000199, respectively. Even though S_n is well-known for its reliable dispersion measurement and ease of calculation[21] our findings demonstrated that the IQR/3 is more efficient than S_n.

Table 10. Evaluation performance for model 1 to model 7 of the SMR 20 price

Model	MAE	%	MSE	%	RMSE	%
Model 3	0.005702	-98.13	0.000039	-100.00	0.006263	-99.75
Model 4	0.009675	-96.82	0.000138	-100.00	0.011732	-99.53
Model 6	0.006703	-98.94	0.000055	-100.00	0.007401	-99.78
Model 7	0.011707	-98.16	0.000199	-100.00	0.014120	-99.57

Forecasting:

Table 11 presents the 12-days ahead forecast results for the price of SMR 20 with a percentage change for Model 1, Model 3, Model 4, Model 6 and Model 7 based on the *forecast* package[44]. The forecast price of SMR 20 on 4th January 2021 for Model 1 is increased by 0.03265 per cent to RM6.127 per kilogram as compared to RM 6.125 per kilogram. However, model 3 and model 4 increased the forecast price to RM6.1253 per kilogram and RM6.1262 per kilogram, respectively, rising by 0.0049 percent and 0.01959 per cent. As the IO contamination increases to 20%, model 6 and model 7 yield the forecast price of RM6.1254 per kilogram and RM6.1263 per kilogram, an increase of 0.00653 percent and 0.02122 percent, respectively against RM 6.125 per kilogram for the same time.

It appears from Table 11 that the difference between actual and forecast prices (in model 1) of SMR 20 was 0.002, which ranged from 0.03115 percent to 0.03265 percent from 5th to 19th January 2021. When contaminated with 10% IO, model 3 and model 4 resulted in a forecast price increase by ranged 0.00467 percent to 0.0049 percent and 0.01716 percent to 0.01959 percent, respectively. As the level of IO contamination raises to 20%, the difference between models 6 to actual price was 0.0004. In the meantime, the forecast price of SMR 20 is increased by a range from 0.02122 percent to 0.02406 percent for model 7. According to the results in Table 11, the dispersion statistics of IQR/3 showed more efficiency than S_n during contamination with 10% and 20% IO.

2024, 21(2): 0511-0523
https://doi.org/10.21123/bsj.2023.7489
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

Table 11. Results of daily actual and forecast price with a percentage change for selected models

Date	Actual price	Forecast price				
		Model 1	**Model 3**	**Model 4**	**Model 6**	**Model 7**
4/1/2021	6.1250	6.1270 (0.03265)	6.1253 (0.00490)	6.1262 (0.01959)	6.1254 (0.00653)	6.1263 (0.02122)
5/1/2021	6.2350	6.2370 (0.03208)	6.2353 (0.00481)	6.2362 (0.01925)	6.2354 (0.00642)	6.2365 (0.02406)
6/1/2021	6.1950	6.1970 (0.03228)	6.1953 (0.00484)	6.1962 (0.01937)	6.1954 (0.00646)	6.1964 (0.02260)
7/1/2021	6.4200	6.4220 (0.03115)	6.4203 (0.00467)	6.4212 (0.01869)	6.4204 (0.00623)	6.4215 (0.02336)
8/1/2021	6.3700	6.3720 (0.03140)	6.3703 (0.00471)	6.3712 (0.01884)	6.3704 (0.00628)	6.3714 (0.02198)
11/1/2021	6.2100	6.2120 (0.03221)	6.2103 (0.00483)	6.2112 (0.01932)	6.2104 (0.00644)	6.2114 (0.02254)
12/1/2021	6.2000	6.2020 (0.03226)	6.2003 (0.00484)	6.2011 (0.01774)	6.2004 (0.00645)	6.2014 (0.02258)
13/1/2021	6.2650	6.2670 (0.03192)	6.2653 (0.00479)	6.2661 (0.01756)	6.2654 (0.00638)	6.2664 (0.02235)
14/1/2021	6.2200	6.2220 (0.03215)	6.2203 (0.00482)	6.2211 (0.01768)	6.2204 (0.00643)	6.2214 (0.02251)
15/1/2021	6.3900	6.3920 (0.03130)	6.3903 (0.00469)	6.3912 (0.01878)	6.3904 (0.00626)	6.3914 (0.02191)
18/1/2021	6.4100	6.4120 (0.03120)	6.4103 (0.00468)	6.4111 (0.01716)	6.4104 (0.00624)	6.4114 (0.02184)
19/1/2021	6.3900	6.3920 (0.03130)	6.3903 (0.00469)	6.3911 (0.01721)	6.3904 (0.00626)	6.3914 (0.02191)

Note: The actual and forecast price represent in RM per kilogram. The value in the bracket represents percentage change.

Conclusion

The two dispersion measurements (IQR/3 and S_n) which were applied in the Huber weighting function for AR(1)-GARCH(2,1) model were reported in this paper. The following conclusions can be made:

a) The AR(1)-GARCH(2,1) model using IQR/3 as a dispersion measurement in the Huber weighting function was found to be the sustainable GARCH model to forecast SMR 20 price.
b) The forecast price of SMR 20 is more sustained when the IQR/3 measure is applied to Huber weight function during contamination 10% and 20% IO.

Therefore, the two measurements identified in the Huber weight function assist in our understanding of the role of suitable dispersion measurement in obtaining a sustainable model. Although M-estimator has some weighting, this work focuses on Huber weights in the M-estimator. Considerably further work would have to be undertaken to test the order type weighting function which can make the time series modelling and forecasting sustained with contaminate other types of outliers.

Author’s Declaration

- Conflicts of Interest: None.
- We hereby confirm that all the Figures and Tables in the manuscript are ours. Furthermore, any Figures and images, that are not ours, have been included with the necessary permission for re-publication, which is attached to the manuscript.
- Ethical Clearance: The project was approved by the local ethical committee in Universiti Malaysia Terengganu, 21030 Kuala Nerus, Terengganu Darul Iman, Malaysia.

Author's Contribution Statement

I. M. M. G. and H. A. R. designed the study and drafted the result. I. M. M. G. wrote the manuscript, performed simulations and analyzed the data. H. A. R. read and approved the final manuscript.

References

1. Fong YC, Khin AA, Lim CS. Determinants of natural rubber price instability for four major producing countries. Pertanika J Soc Sci Humanit. 2020; 28(2): 1179–97. http://www.pertanika.upm.edu.my/pjssh/browse/regular-issue?article=JSSH-5107-2019
2. Zhu Q, Zhang F, Liu S, Wu Y, Wang L. A hybrid VMD–BiGRU model for rubber futures time series forecasting. Appl Soft Comput J. 2019; 84. https://dx.doi.org/10.1016/j.asoc.2019.105739
3. Mah PJW, Buhary FN, Abdullah NH, Saad SAM. A Comparative study of univariate time series modelling for natural rubber production in Malaysia. Malaysian J Comput. 2018; 3(2): 108–18. https://ir.uitm.edu.my/id/eprint/43415/
4. Mathew S, Murugesan R. Indian natural rubber price forecast–An Autoregressive Integrated Moving Average (ARIMA) approach. Indian J Agric Sci. 2020; 90(2): 418–22. https://dx.doi.org/10.56093/ijas.v90i2.103067
5. Jong LJ, Ismail S, Mustapha A, Abd Wahab MH, Idrus SZS. The combination of autoregressive integrated moving average (ARIMA) and support vector machines (SVM) for daily rubber price forecasting. IOP Conf Ser Mater Sci Eng. 2020; 917(1). https://dx.doi.org/10.1088/1757-899X/917/1/012044
6. Norizan NFHBM, Yusof ZBM. Forecasting natural rubber price in Malaysia by 2030. Malays J Soc Sci and Hum. 2021; 6(9): 382–390. https://dx.doi.org/10.47405/mjssh.v6i9.986
7. Muda NZ, I SNI. Comparison of GBM, GFBM and MJD models in Malaysian rubber prices forecasting. Malaysian J Fundam Appl Sci. 2023; 19(1):73-81. https://dx.doi.org/10.11113/mjfas.v19n1.2763
8. Ramli N. Rubber Price Volatility and Inequality in Economic Development. In: Issues and Challenges in the Malaysian Economy. Bingley: Emerald Publishing Limited; 2019. p. 79–93. https://doi.org/10.1108/978-1-83867-479-320191006
9. Ramli N, Md Noor AHS, Sarmidi T, Said FF, Azam AHM. Modelling the volatility of rubber prices in ASEAN-3. Int J Bus and Soc. 2019; 20(1): 1–18. http://www.ijbs.unimas.my/index.php/volume-11-20/volume-20-no-1-2019/547-modelling-the-volatility-of-rubber-prices-in-asean-3
10. Cai G, Wu Z, Peng L. Forecasting volatility with outliers in Realized GARCH models. J Forecast. 2021; 40(4): 667–85. https://dx.doi.org/10.1002/for.2736
11. Carnero MA, Pérez A. Outliers and misleading leverage effect in asymmetric GARCH-type models. Stud Nonlinear Dyn Econom. 2021; 1–47. https://dx.doi.org/10.1515/snde-2018-0073
12. Aronne A, Grossi L, Bressan AA. Identifying outliers in asset pricing data with a new weighted forward search estimator. Rev Contab e Financ. 2020; 31(84): 458–72. https://dx.doi.org/10.1590/1808-057x201909620
13. Hotta LK, Trucíos C. Inference in (M)GARCH Models in the presence of additive outliers: specification, estimation, and prediction. In: Lavor C,Gomes FAM, editors. Adv Math Appl. Springer International Publishing; 2018 : 179–202. https://doi.org/10.1007/978-3-319-94015-1_8
14. Dutta A. Impacts of oil volatility shocks on metal markets: A research note. Resour Policy. 2018; 55: 9–19. https://dx.doi.org/10.1016/j.resourpol.2017.09.003
15. Wang J, Yang W, Du P, Niu T. Outlier-robust hybrid electricity price forecasting model for electricity market management. J Clean Prod. 2020; 249: 119318. https://dx.doi.org/10.1016/j.jclepro.2019.119318
16. Grossi L, Nan F. Robust forecasting of electricity prices: simulations, models and the impact of renewable sources. Technol Forecast Soc. 2019; 141: 305–318. https://dx.doi.org/10.1016/j.techfore.2019.01.006
17. Bachmann C, Tegtmeier L, Gebhardt J, Steinborn M. The "sell in May" effect: an empirical investigation of globally listed private equity markets. Manage Financ. 2019; 45(6): 793–808. https://dx.doi.org/10.1108/MF-07-2018-0322
18. da Cunha AS, Peixoto FC, Prata DM. Robust data reconciliation in chemical reactors. Comput Chem Eng. 2021;145. https://dx.doi.org/10.1016/j.compchemeng.2020.107170
19. de Menezes DQF, Prata DM, Secchi AR, Pinto JC. A review on robust M-estimators for regression analysis. Comput Chem Eng. 2021; 147: 107254. https://dx.doi.org/10.1016/j.compchemeng.2021.107254
20. Polat E. The effects of different weight functions on partial robust M-regression performance: A simulation study. Commun Stat Simul Comput. 2020; 49(4): 1089–104. https://dx.doi.org/10.1080/03610918.2019.1586926
21. Rousseeuw PJ, Croux C. Alternatives to the median absolute deviation. J Am Stat Assoc. 1993; 88(424): 1273–1283. https://dx.doi.org/10.1080/01621459.1993.10476408
22. Bollerslev T. Generalized autoregressive conditional heteroskedasticity. J Econometrics. 1986; 31(3): 307–

327. https://www.sciencedirect.com/science/article/abs/pii/0304407686900631, https://dx.doi.org/10.1016/0304-4076(86)90063-1
23. Chowdhury KP. Supervised machine learning and heuristic algorithms for outlier detection in irregular spatiotemporal datasets. J Environ Inform. 2019; 33(1): 1–16. http://www.jeionline.org/index.php?journal=mys&page=article&op=view&path%5B%5D=201700375 https://dx.doi.org/10.3808/jei.201700375
24. Chakrabarty M, Ray P. Changes of base-year and Indian GDP growth: an agnostic look. Indian Growth Dev Rev. 2021; 14(3): 281–301. https://dx.doi.org/10.1108/IGDR-08-2020-0124
25. Adedoyin FF, Bekun F V., Driha OM, Balsalobre-Lorente D. The effects of air transportation, energy, ICT and FDI on economic growth in the industry 4.0 era: Evidence from the United States. Technol Forecast Soc Change. 2020; 160: 120297. https://dx.doi.org/10.1016/j.techfore.2020.120297
26. Byers JW, Popova I, Simkins BJ. Robust estimation of conditional risk measures using machine learning algorithm for commodity futures prices in the presence of outliers. J Commod Mark. 2021; 24: 100174. https://dx.doi.org/10.1016/j.jcomm.2021.100174 .
27. Chen C, Liu, L-M. Joint estimation of model parameters and outlier effects in time series. J Am Stat Assoc. 1993; 88(421): 284–297. https://dx.doi.org/10.1080/01621459.1993.10594321
28. Vinutha HP, Poornima B, Sagar BM. Detection of outliers using interquartile range technique from intrusion dataset. In: Satapathy, S., Tavares, J., Bhateja, V., Mohanty J, editors. Information and Decision Sciences, Advances in Intelligent Systems and Computing. Singapore: Springer; 2018. p. 511–518. https://doi.org/10.1007/978-981-10-7563-6_53
29. Jones PR. A note on detecting statistical outliers in psychophysical data. Attention, Perception, Psychophys. 2019; 81: 1189–96. https://doi.org/10.3758/s13414-019-01726-3
30. Singh D, Singh B. Investigating the impact of data normalization on classification performance. Appl Soft Comput. 2020; 97: 105524. https://dx.doi.org/10.1016/j.asoc.2019.105524
31. Ghani IMM, Rahim HA. Weighting temporary change outlier by modified Huber function with Monte Carlo simulations. J Phys Conf Ser. 2020; 1529. https://iopscience.iop.org/article/10.1088/1742-6596/1529/5/052050
32. Sakata S, White H. High Breakdown Point Conditional Dispersion Estimation with Application to S & P 500 Daily Returns Volatility. Econometrica. 1998; 66(3): 529–67. https://dx.doi.org/10.2307/2998574
33. Huber PJ. Robust Estimation of a location parameter. Ann Math Stat. 1964; 35(1): 73–101. https://www.jstor.org/stable/2238020 https://dx.doi.org/10.1214/aoms/1177703732
34. Akaike H. A new look at the statistical model identification. IEEE Trans. Automat Control. 1974; 19(6): 716-723. Available from: https://dx.doi.org/10.1109/TAC.1974.1100705
35. Juma AA, AL-Mohana FAM. A modified approach by using prediction to build a best threshold in ARX model with practical application. Baghdad Sci J. 2019; 16(4): 1049–63. https://dx.doi.org/10.21123/bsj.2019.16.4(Suppl.).1049
36. Yousif AH, Ali OA. Proposing robust LAD-Atan penalty of regression model estimation for high dimensional data. Baghdad Sci J. 2020 ; 17(2): 550–5. https://dx.doi.org/10.21123/bsj.2020.17.2.0550
37. Trapletti A, Hornik K. tseries: Time Series Analysis and Computational Finance R package version 0.10-51. 2022 [cited 2021 Aug 28]; 1-54. https://cran.r-project.org/web/packages/tseries/ .
38. Wuertz D, Setz T, Chalabi Y, Boudt C, Chausse P, Miklavoc M. fGarch: Rmetrics - Autoregressive conditional heteroskedastic modelling [Internet]. 2020 [cited 2021 Aug 28]: 1-51. https://cran.r-project.org/package=fGarch/
39. R Core Team. R: A language and environment for statistical computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2021 [cited 2021 Aug 28]. https://www.r-project.org/ .
40. Phillips PCB, Perron P. Testing for a unit root in time series regression. Biometrika. 1988; 75(2): 335–346. https://dx.doi.org/10.2307/2336182
41. Dickey DA, Fuller WA. Distribution of the estimators for autoregressive time series with a unit root. J Am Stat Assoc. 1979; 74(366): 427–431. https://dx.doi.org/10.1080/01621459.1979.10482531
42. Kwiatkowski D, Phillips PCB, Schmidt P, Shin Y. Testing the null hypothesis of stationarity against the alternative of a unit root. How sure are we that economic time series have a unit root? J Econom. 1992; 54(1–3): 159–178. https://dx.doi.org/10.1016/0304-4076(92)90104-Y
43. Engle RF. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica. 1982; 50(4): 987–1007. https://dx.doi.org/10.2307/1912773
44. Hyndman R, Athanasopoulos G, Bergmeir C, Caceres G, Chhay L, O'Hara-Wild M, et al. forecast: Forecasting functions for time series and linear models. 2020; [cited 2021 Aug 28]. https://cran.r-project.org/web/packages/forecast/

2024, 21(2): 0511-0523
https://doi.org/10.21123/bsj.2023.7489
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

بناء نموذج GARCH مستدام للتنبؤ بسعر المطاط: نهج دالة الترجيح المعدلة لهوبر

إنتان مارتينا محمد غني، حنفي رحيم

كلية تكنولوجيا هندسة المحيطات والمعلوماتية ، جامعة ماليزيا تيرينجانو ، 21030 كوالا نيروس ، تيرينجانو دار الإيمان ، ماليزيا

الخلاصة

الطبيعة غير المستقرة وغير المؤكدة لأسعار المطاط الطبيعي تجعلها شديدة التقلب وعرضة للقيم المتطرفة ، والتي يمكن أن يكون لها تأثير كبير على كل من النمذجة والتنبؤ. لمعالجة هذه المشكلة ، يوصي البحث بنموذج هجين يجمع بين نموذج الانحدار الذاتي (AR) ونموذج التباين الشرطي الشرطي المعمم (GARCH). يستخدم النموذج وظيفة الترجيح Huber لضمان بقاء القيمة المتوقعة لأسعار المطاط مستدامة حتى في وجود القيم المتطرفة. تهدف الدراسة إلى تطوير نموذج مستدام والتنبؤ بالأسعار اليومية لمدة 12 يومًا من خلال تحليل 2683 من بيانات الأسعار اليومية من الدرجة 20 للمطاط الماليزي القياسي (SMR 20) في ماليزيا. يشتمل التحليل على قياسين للتشتت (IQR / 3 و Sn) وثلاثة مستويات من تلوث IO (0٪ ، 10٪ ، و 20٪). تشير النتائج إلى أن استخدام دالة الترجيح Huber مع قياس IQR / 3 لبناء نموذج GARCH (1) -AR (1)،1 يؤدي إلى استدامة أفضل. هذه النتائج لديها القدرة على تعزيز نموذج GARCH عن طريق تعديل وظيفة الترجيح لمقدر M.

الكلمات المفتاحية: الانحدار الذاتي ، التشتت ، التنبؤ ، GARCH، هوبر.