Proceedings of Science and Mathematics
Faculty of Science, Universiti Teknologi Malaysia
Vol. 9, 2022, page 185-194

Forecasting The Crude Oil Price in Malaysia Using Geometric Brownian Motion

Muhamad Firdaus Lahmuddin*, Ani Shabri
Department of Mathematical Sciences, Faculty of Science
Universiti Teknologi Malaysia, 81310 Johor Bahru, Malaysia
*Corresponding author: muhamad.firdaus@graduate.utm.my

Abstract
Crude oil is a petroleum product made composed of naturally occurring hydrocarbon concentrations and other organic components. Crude oil is the most common raw material used in oil refineries, and its cost accounts for most of the overall operating expenses. In recent years, significant price volatility in crude oil has had a substantial influence on the economic balances including both oil exporting and importing countries. Forecasting is thus one of an accurate method of anticipating future pricing. One technique to mitigate the effects is to use Geometric Brownian Motion to forecast future crude oil prices (GBM). Geometric Brownian motion (GBM) is a stochastic continuous-time process, and it was chosen because of its potential for forecasting short-term price. This research used crude oil price data that consists of monthly price of crude oil (April 2002 – March 2022), obtained from World Bank. Drift and volatility parameters were estimated in this study. The objectives of research are to propose the best Geometric Brownian Motion model for predicting crude oil price data in Malaysia, determine the pattern of the projected data using the best model, and assess the accuracy of the model Geometric Brownian Motion in forecasting. The result shows that the log return of crude oil prices is normally distributed, and Geometric Brownian Motion models were suitable to forecast crude oil prices. The MAPE value for GBM is 7.97% and the MAPE value for Linear Regression is 14.61%. As a result, this show that GBM is more accurate in prediction than linear regression.
Keywords: Crude Oil; Geometric Brownian Motion; Linear Regression; MAPE.

1. Introduction

Crude oil is a fossil fuel that is processed into products like gasoline, diesel, and a range of other petrochemicals Crude oil is the most common raw material used in oil refineries, and its cost accounts for most of the overall operating expenses. Because crude oil prices change, it's hard to predict how much raw materials will cost. Brent and West Texas Intermediate are two categories that impact crude oil prices (WTI). WTI is a crude oil produced in Texas, whereas Brent is a standard value oil from the northern seas (Europe) (USA). Large price volatility in crude oil have had a considerable influence on the economic balances of both oil exporting and importing countries in recent years. As a result, forecasting is the most accurate means of predicting future pricing. Forecasting has been used to make judgments based on expected data for a long time. It might help them avoid stumbling blocks and prepare meticulously for business growth.

There are still several research issues that have yet to be resolved. One of the research gaps that effects crude oil production is the quick reduction in crude oil prices in Malaysia and other crude oil businesses in other countries, which causes the country's income to drop rapidly and, as a result, some crude oil companies incur huge losses. As a result, a forecasting model may be quite beneficial in this sector, and the approach of preferred for forecasting crude oil prices is Geometric Brownian Motion (GBM). To aid investors in anticipating crude oil price for a short length of time, a mathematical model as Geometric Brownian Motion (GBM) is required. GBM is a very accurate model, as evidenced by the MAPE value, and that it may be used to forecast future crude oil prices so that crude oil industries can make some planning to face the volatility of crude oil price in future.

2. Literature Review

2.1. Crude Oil in Malaysia

Energy security is defined as a country's or region's continuous and affordable access to energy sources. Earlier to the 1950s, most oil exploration in Malaysia took place inshore, until early evidence of oil supplies was discovered offshore in Peninsular Malaysia and the Seria oilfield off the coast of Brunei. Almost all offshore oil research has taken place since then. Malaysia's energy production is mostly reliant on four resources: oil, gas, coal, and hydro. Crude oil is a type of fossil fuel that is refined into useable products such as gasoline, diesel, and a variety of other petrochemicals. Malaysia's primary energy sources include petroleum and crude oil products, along with natural gas.

2.2. Forecasting Methods for Crude Oil

2.2.1. Novel Hybrid Method

User search data (USD), which represents investor interest, has been extensively studied and found to be linked to crude oil price changes over a range of frequency ranges [1]. A unique hybrid technique for crude oil price forecasting was used which to combine bivariate empirical mode decomposition (BEMD) with user search data and machine learning. Initially, BEMD is used to deconstruct crude oil price data and USD into a finite number of components at the same time. Furthermore, a random vector functional link (RVFL) network is used to model and forecast each component, and the corresponding final findings are derived using an ensemble model. Hence, the suggested technique is experimentally tested using the Brent crude oil spot price. The robustness of the forecasting predictions is assessed using a variety of evaluation criteria. In terms of prediction accuracy, the suggested strategy statistically surpasses typical forecasting machine learning approaches and equivalent counterparts (using a USD or EMD-based method).

2.2.1. Support Vector Machine

A novel support vector machine-based technique for crude oil price forecasting (SVM) was investigated in [2]. Data sampling, sample pre-processing, training, and learning, and out-of-sample forecasting are all steps in the development of a support vector machine model for time series forecasting. To assess SVM's predicting abilities, we compare its results to those of ARIMA and BPNN. The findings of the experiment suggest that SVM beats the other two approaches and is a viable contender for crude oil price forecasting.

Support Vector Machines, according to [3] are supervised learners based on statistical learning theory. Their approach used major economic factors that influence crude oil prices as inputs, and the price of crude oil as output. Findings from the West Texas Intermediate (WTI) dataset, which spans 24 years, was used in their technique. The data then transferred to a Slantlet algorithm, which identifies several features to be used as variables in a hybrid system that includes Auto Regressive Moving Averages (ARMA) and SVMs. The test data were quite encouraging, demonstrating that support vector machines may be used to forecast crude oil prices with a high degree of accuracy.

2.2.2. Compumetric Method

Computers are used in compumetric forecasting approaches to determine the underlying model that generates the prediction. Humans, rather than algorithms, often create or specify forecasting models. To see whether the models they supply create credible predictions, compumetric approaches are used. The outcomes of two compumetric techniques, genetic programming, and artificial neural networks, are studied and compared to a random walk prediction, according to [4]. The findings show that genetic programming outperforms random walk predictions, whereas neural network predictions perform poorly.

2.3. Geometric Brownian Motion

Geometric Brownian motion (GBM) is a continuous stochastic process. To characterize commodity price movements, the GBM model should be utilized. GBM is used to model variations in crude oil prices in the procedure. GBM is a random walk geometry process, in which the random walk is the exponential shape of the random walk geometry asymmetry. GBM is significant in mathematically modelling financial processes. It is the continuous model's derivative from the discrete model and may be used to forecast how stock values will change in the future in the limited time frame.

2.3. Applications Geometric Brownian Motion

2.3.1. Stock Price Forecasting

One of the commonly used application Geometric Brownian Motion applications in Malaysia is stock price prediction. [5] stated that to aid investors in anticipating share prices for a short length of investing time, a mathematical model as simple as Geometric Brownian Motion (GBM) is required. Their findings suggest that GBM is a very accurate model, as evidenced by the MAPE value, and that it can be used to forecast future share values in Bursa Malaysia estimated for two weeks.

As study "Forecasting Nestle Stock Price by using Brownian Motion Model during Pandemic Covid-19", [6] found that all GBM simulations' Mean Absolute Percentage Error (MAPE) values are less than 10%, indicating that MAPE estimates are very accurate and stated that it is preferable to increase the number of stock data because this will improve the model's performance.

3. Methodology

3.1. Description of Data

In this study, the data that will be used is the Crude Oil Monthly Price of Malaysia through the Index Mundi website source from World Bank. The range of the data are from April 2002 until March 2022.

3.2. Crude Oil Log Return

The formula of crude oil return is defined as following:

$$\boldsymbol{R_t = ln(\frac{P_t}{P_{t-1}})} \tag{1}$$

where:

R_t : Crude oil log return at time t
P_t : Actual crude oil price at time t
P_{t-1} : Actual crude oil price at time t-1

3.3. Normality Test

The Lilliefors test is used to test normally or not normal distributed data.

H_0 : Sample data is normally distributed
H1 : Sample data is not normally distributed
Statistical test:

$$D_{max} = Max \mid F_t - F_s \mid \tag{2}$$

where:

D_{max} : minimum deviation

F_t : the hypothesized distributed function normally distributed
F_s : cumulative distribution function of the sample data

H0 is approved if D max < Dα, n (value α = 0.05) is true, indicating that the sample data is regularly distributed. When using Microsoft Excel software to assess normality, if the P value > 0.05, H0 is accepted, and the sample data is normally distributed.

3.4. Volatility

Volatility is the rate of movement of oil prices raw. The formula of the volatility value is as follows:

$$\hat{\sigma} = \frac{s_r}{\Delta t} \tag{3}$$

The equation for $\bar{R}$ and S_r as follows:

$$\boldsymbol{\bar{R}} = \frac{\sum_{t=1}^{n}(\boldsymbol{R_t})}{\boldsymbol{n}} \tag{4}$$

$$\boldsymbol{S_r} = \sqrt{\frac{\sum_{t=1}^{n}(\boldsymbol{R_t} - \boldsymbol{\bar{R}})^2}{\boldsymbol{n}}} \tag{5}$$

where:

$\boldsymbol{\hat{\sigma}}$	: value estimation volatility
$\boldsymbol{S_r}$	: standard deviation crude oil
$\Delta \boldsymbol{t}$	: time interval in calculating the log return value
$\boldsymbol{\bar{R}}$	: average log return
$\boldsymbol{R_t}$	: log return at t

3.5. Drift

Drift is the expectation of the rate of movement of oil prices raw. The formula of drift is as follows:

$$\boldsymbol{\hat{\mu}} = \frac{\boldsymbol{\bar{R}}}{\Delta \boldsymbol{t}} + \frac{\boldsymbol{\hat{\sigma}}^2}{\boldsymbol{2}} \tag{6}$$

where:

$\boldsymbol{\hat{\mu}}$	: value drift estimation
$\boldsymbol{\bar{R}}$	: average log return
$\Delta \boldsymbol{t}$	: time interval in calculating the log return value
$\boldsymbol{\hat{\sigma}}$	: volatility value

3.5. Geometric Brownian Motion

Geometric Brownian motion (GBM) is an important method, and it is one of stochastic process continuous time forecast. The GBM model should be used to describe commodity price changes. The process crude oil price changes are modelled using GBM. GBM is a random walk geometry process, in which the random walk is the exponential shape of the random walk geometry asymmetry. The geometric model equation Brownian motion can be used to calculate each crude oil price forecast at time t as follows

$$\boldsymbol{F_t} = \boldsymbol{F_{t-1}} \boldsymbol{e}^{(\hat{\mu} - \frac{1}{2}\hat{\sigma}^2)dt + \hat{\sigma}\varepsilon\sqrt{dt}} \tag{7}$$

where:

F_t : Crude oil price forecast at t
F_{t-1} : Crude oil price forecast at t-1
μ : drift value
σ : volatility value

3.7. Mean Absolute Percentage Error (MAPE)

The Mean Absolute Percentage Error (MAPE) is the average absolute percentage of prediction mistakes. When measuring forecasting accuracy, MAPE is a crucial component to consider. MAPE will display the magnitude of forecasting mistakes in comparison to actual values. If the MAPE value is calculated using a method, the smaller the forecasting, the better the forecasting method. Equation of MAPE as follows:

$$\boldsymbol{MAPE} = \sum_{\boldsymbol{t=1}}^{\boldsymbol{N}} \frac{\frac{|\boldsymbol{P_t} - \boldsymbol{F_t}|}{\boldsymbol{P_t}}}{\boldsymbol{N}} \times \mathbf{100} \tag{8}$$

where:

$\boldsymbol{P_t}$: actual crude oil price at time t
$\boldsymbol{N}$: total data crude oil price
$\boldsymbol{F_t}$: forecast actual crude oil price at time t

4. Results and discussion

4.1. Data Collection

At this stage, crude oil price in Malaysia was collected from the Index Mundi website sources by the World Bank. The monthly crude oil price data is collected from April 2002 until March 2022, totalling 20 years, and is presented in Malaysian Ringgit.

4.2. Calculation of Crude Oil Return

After acquiring the data, the first stage is completed. The price of crude oil is used to assess the log return on price in crude oil. Using equation (1), the log returns of crude oil from April 2002 until March 2022 were calculated and shown as below:

$$R_2 = ln\left(\frac{P_2}{P_1}\right) = ln(\frac{97.55}{96.63}) = 0.00947581505790314$$

$$R_3 = ln\left(\frac{P_3}{P_2}\right) = ln(\frac{93.06}{97.55}) = -0.0471206206526177$$

$$\vdots$$

$$R_{240} = ln\left(\frac{P_{240}}{P_{239}}\right) = ln(\frac{472.19}{391.73}) = 0.186808619994563$$

Table 4.1 shows a piece of result of crude oil log return from R_2 until R_{30}, but the total return is until R_{240}. After obtaining the crude oil log return value, the next step is to plot the log return. This graph is helpful in determining the pattern of crude oil log return. The following plot depicts crude oil log return graph.

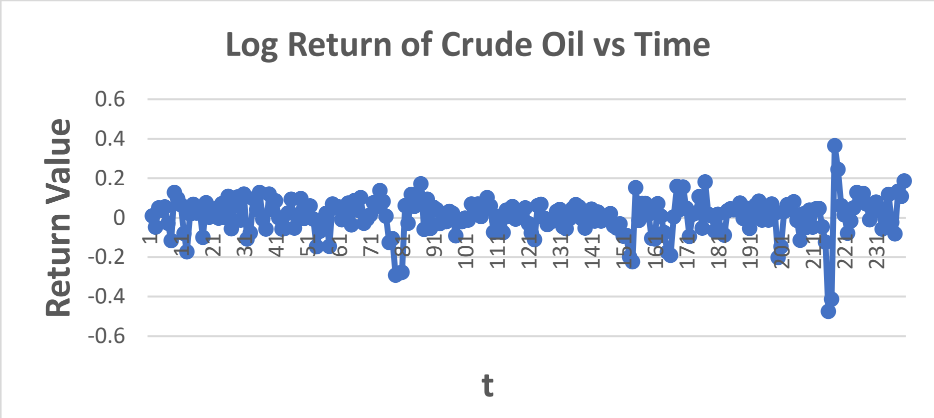

Figure 4.1: Log Return of Crude Oil vs Time

Figure 4.1 shows that the log return of the crude oil price versus time. The plot patterns are up and down which are some values are negative and some of them are positive.

4.3. Normality Test

At this stage, a test of normality is performed from April 2002 to March 2022, which has been obtained from the previous stage. A normality test is performed to determine whether the log return of crude oil is normally distributed or not. Here is a step to test the normality of crude oil log returns.

$$\begin{aligned} D_{max} &= \text{maks}|F_t - F_s| \\ &= 0.09689 \end{aligned}$$

$$\begin{aligned} D_{\alpha,n} &= D_{0.05,239} \\ &= \frac{0.886}{\sqrt{239}} \\ &= 0.05731 \end{aligned}$$

The value of $D_{max} > D_{0.05,239}$, then we can conclude that the log return of crude oil data is normally distributed. Histogram of crude oil log return is present in the Figure 4.2:

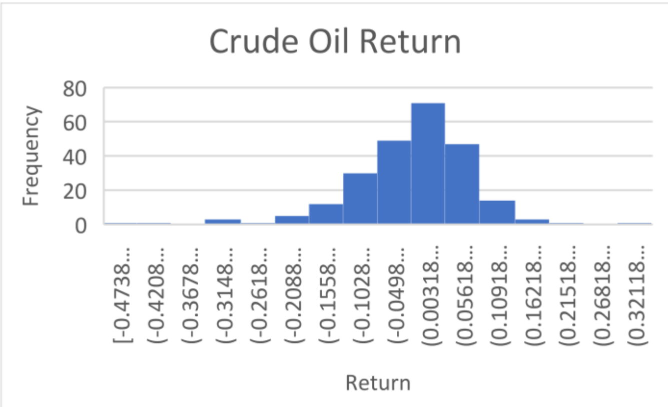

Figure 4.2: Histogram of Crude Oil Return

4.4. Parameter Estimation

The estimate of volatility parameters and drift from a geometric model of Brownian motion is done at this step. The values of the drift and volatility parameters are constant, and this is the value that will be used to forecast crude oil prices in the future. Calculating the average log return and standard deviation of log return is the first step in determining the value of volatility. By using the results of the crude oil log return computation in the preceding calculation, the calculation of average log return, R¯using equation (4) are presented below:

$$\overline{\mathrm{R}} = \frac{\sum_{t=2}^{240}(\mathrm{R_t})}{239}$$

$$= \frac{\mathrm{R_2} + \mathrm{R_3} + \mathrm{R_4} + \cdots + \mathrm{R_{239}} + \mathrm{R_{240}}}{239}$$

$$= \frac{0.009476 - 0.047121 + 0.050191 + \cdots + 0.108221 + 0.186809}{239}$$

$$= 0.006638$$

After the calculation of log return has been done, the standard deviation of log return needs to be calculated. By using equation (5), the calculation of standard deviation of log return have been done and presented as below:

$$\mathrm{S_r^2} = \sum_{t=2}^{n} \frac{(\mathrm{R_t} - \overline{\mathrm{R}})^2}{\mathrm{n} - 1}$$

$$= \frac{((0.009476 - 0.006638\,)^2 + (-0.047121 - 0.006638\,)^2 \cdots + (0.186809 - 0.006638\,)^2}{238}$$

= 0.008853

Therefore,
$\mathrm{S_r}$ = 0.094092

The calculation parameter $\hat{\sigma}$ is the following step after getting the standard deviation value of the crude oil log return. The result was derived using equation (3) and the value shown as below:

$$\hat{\sigma} = \frac{0.094092}{1}$$

$$= 0.094092$$

Then, the value of $\hat{\mu}$ was calculated using equation (3.6) as shown below:

$$\hat{\mu} = \frac{0.006638}{1} + \frac{0.094092^2}{2} = 0.011065$$

4.5. Crude Oil Forecasting

The price of crude oil is forecasted to be at this level. The forecasting method has been conducted by using the Geometric Brownian Motion formula. Geometric Brownian Motion forecasting is divided by two classify as training part and testing part. The geometric model equation Brownian motion can be used to calculate each crude oil price forecast at time t and the equation is equation (7).

4.5.1. Training Part

$$F_t = F_{t-1} e^{(\hat{\mu} - \frac{1}{2}\hat{\sigma}^2)dt + \hat{\sigma}\varepsilon\sqrt{dt}}$$
$$F_1 = F_0 e^{(0.011065 - \frac{1}{2}(0.094092)^2)dt + 0.094092\varepsilon\sqrt{dt}}$$

$F_1 = 110.9871$

$\vdots$

$F_{227} = F_{226}e^{(0.011065-\frac{1}{2}(0.094092)^2)dt+0.094092\varepsilon\sqrt{dt}}$

$F_{227} = 218.9226$

4.5.2. Testing Part

$F_t = F_{t-1}e^{(\hat{\mu}-\frac{1}{2}\hat{\sigma}^2)dt+\hat{\sigma}\varepsilon\sqrt{dt}}$

$F_{228} = F_{227}e^{(0.011065-\frac{1}{2}(0.094092)^2)dt+0.094092\varepsilon\sqrt{dt}}$

$F_{228} = 220.2537$

$\vdots$

$F_{239} = F_{238}e^{(0.011065-\frac{1}{2}(0.094092)^2)dt+0.094092\varepsilon\sqrt{dt}}$

$F_{239} = 220.1721$

The simulation results based on the testing part are shown in the table 4.6 below:

Simulation	Crude Oil price
154.77	259.61
163.31	273.99
171.22	296.88
167.05	307.68
153.41	290.70
155.32	303.58
168.13	341.56
182.90	333.62
170.64	307.47
147.55	351.55
131.63	391.73
130.70	472.19

Table 4.6: Simulation Result of Crude Oil

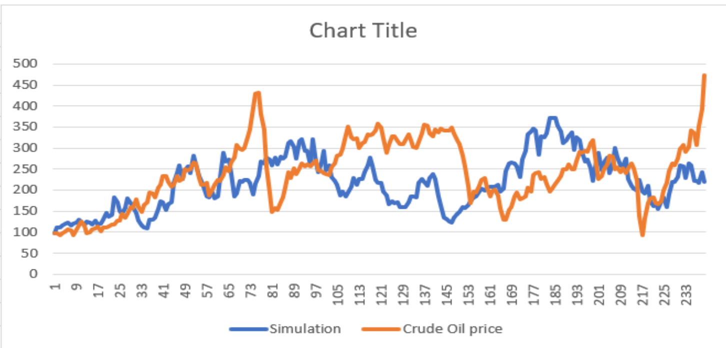

Figure 4.3: Forecasted Crude Oil Price

4.5. Model Validation

After receiving the outcomes of crude oil price predictions, the Mean Absolute Percentage Error (MAPE) value prediction is counted at a later stage using equation (8) to determine the Brownian motion geometric model's level of accuracy in forecasting crude oil prices. MAPE test also divided by two parts as training part and testing part which is forecasting part.

The value for training MAPE is 54.48% and average value for forecasting MAPE is 7.97%. This show that the forecasting or testing MAPE is better than training MAPE. As a result, we may deduce that many period data needed to get more accurate prediction with less MAPE value.

Conclusion

In this research, we learn that the Geometric Brownian Motion (GBM) model is capable of forecasting future outcomes. The MAPE values and the prediction graph display have led to this conclusion. The graph reveals that the forecasting plot pattern is quite similar with the real crude oil. The MAPE value of the forecasting portion in GBM is satisfying since it is 7.97 percent, which is extremely accurate according to the table of accuracy. As a result, GBM forecasting is a good choice to perform a forecast since it has a lower error rate which is less than 10%. Therefore, GBM is strongly recommended to the crude oil decision maker and trading investors to learn more about it so that they may could do a better forecast in the future and take precautionary measures if something unfavourable occurs and help them in planning to avoid losses.

Acknowledgement

I would like to express my special thanks of appreciation to my supervisor, PM Dr Ani Shabri for supervising and guiding me during my research on this study. His assistance of giving comments, advice and suggestions during research has contributed to success of this report. I also would like to thank to my beloved family and my fellow friends for the continuous support and encouragement.

References

[1] Zhang, C., Jiang, F., Wang, S., & Shang, W. (2020, December). A Novel Hybrid Approach with A Decomposition Method and The RVFL Model for Crude Oil Price Prediction. In *2020 IEEE International Conference on Big Data (Big Data)* (pp. 4446-4449). IEEE

[2] Xie, W., Yu, L., Xu, S., & Wang, S. (2006, May). A new method for crude oil price forecasting based on support vector machines. In *International conference on computational science* (pp. 444-451). Springer, Berlin, Heidelberg.

[3] Khashman, A., & Nwulu, N. I. (2011, January). Intelligent prediction of crude oil price using Support Vector Machines. In *2011 IEEE 9th International Symposium on Applied Machine Intelligence and Informatics (SAMI)* (pp. 165-169). IEEE.

[4] Kaboudan, M. A. (2001, May). Compumetric forecasting of crude oil prices. In *Proceedings of the 2001 congress on evolutionary computation (IEEE Cat. No. 01TH8546)* (Vol. 1, pp. 283-287). IEEE.

[5] Abidin, S. Z., Old, C., & Martin, T. (2010). Effects of New Zealand general elections on stock market returns..

[6] Hamzah, S. R., Halul, H., Jeng, A., & Sha'ari, U. A. S. (2021). Forecasting Nestle Stock Price by using Brownian Motion Model during Pandemic Covid-19. *Malaysian Journal of Science Health & Technology*, *7*(2), 58-64..

[7] Zhang, J. L., Zhang, Y. J., & Zhang, L. (2015). A novel hybrid method for crude oil price forecasting. *Energy Economics*, *49*, 649-659.

[8] Kaboudan, M. A. (2003). Short-term compumetric forecast of crude oil prices. In *Modeling and Control of Economic Systems 2001* (pp. 365-370). Elsevier Science Ltd

[9] Rahim, K. A., & Liwan, A. (2012). Oil and gas trends and implications in Malaysia. *Energy Policy*, *50*, 262-271.

[10] Elsayed, E. A., & Liao, H. (2003). An accelerated degradation rate model for reliability prediction at field conditions. In *IIE Annual Conference. Proceedings* (p. 1). Institute of Industrial and Systems Engineers (IISE)

[11] Marathe, R. R., & Ryan, S. M. (2005). On the validity of the geometric Brownian motion assumption. *The Engineering Economist*, *50*(2), 159-192.

[12] Djauhari, M. A., Asrah, N. M., Li, L. S., Djakaria, I., Badruzzaman, S. T. A. I. K., & KM10, J. R. S. (2020). Forecasting model of electricity consumption in Malaysia: A geometric Brownian motion approach. *Solid State Technology*, *63*(3), 40-46.

[13] Yu, L., Wang, S., & Lai, K. K. (2008). Forecasting crude oil price with an EMD-based neural network ensemble learning paradigm. *Energy economics*, *30*(5), 2623-2635.

[14] Gharbi, R. B., Elsharkawy, A. M., & Karkoub, M. (1999). Universal neural-network-based model for estimating the PVT properties of crude oil systems. *Energy & fuels*, 13(2), 454-458.

[15] Omar, A., & Jaffar, M. M. (2011, September). Comparative analysis of Geometric Brownian motion model in forecasting FBMHS and FBMKLCI index in Bursa Malaysia. In *2011 IEEE Symposium on Business, Engineering and Industrial Applications (ISBEIA)* (pp. 157-161). IEEE.

[16] Agustini, W. F., Affianti, I. R., & Putri, E. R. (2018, March). Stock price prediction using geometric Brownian motion. In *Journal of physics: conference series* (Vol. 974, No. 1, p. 012047). IOP Publishing.

[17] Teoh, Y. K., Abu Hasan, S., & Hamdan, N. (2017). Forecasting the financial times stock exchange Bursa Malaysia Kuala Lumpur Composite Index using Geometric Brownian Motion/Teoh Yeong Kin, Suzanawati Abu Hasan and Nashni Hamdan. *Journal of Computing Research and Innovation*, *2*(1), 45-49.

[18] El-Sebakhy, E. A., Sheltami, T., Al-Bokhitan, S. Y., Shaaban, Y., Raharja, P. D., & Khaeruzzaman, Y. (2007, March). Support vector machines framework for predicting the PVT properties of crude oil systems. In *SPE Middle East oil and gas show and conference*. OnePetro.

[19] Hahn, W. J., DiLellio, J. A., & Dyer, J. S. (2014). What do market-calibrated stochastic processes indicate about the long-term price of crude oil?. *Energy Economics*, *44*, 212-221.

[20] ZAKIA, H. I. (2017). Prediksi harga komoditas minyak mentah menggunakan model geometric brownian motion (Doctoral dissertation, Institut Teknologi Sepuluh Nopember).

[21] Abidin, S. N. Z., & Jaffar, M. M. (2014). Forecasting share prices of small size companies in Bursa Malaysia using geometric Brownian motion. *Applied Mathematics & Information Sciences*, *8*(1), 107.

[22] Bahar, A., Noh, N. M., & Zainuddin, Z. M. (2017). Forecasting model for crude oil price with structural break. *Malaysian Journal of Fundamental and Applied Sciences*, *13*(4-1), 421-24.

[23] Nwafor, C. N., & Oyedele, A. A. (2017). Simulation and hedging oil price with geometric Brownian motion and single-step binomial price model. *European Journal of Business and Management*, *9*(9), 68-81.

[24] Postali, F. A., & Picchetti, P. (2006). Geometric Brownian motion and structural breaks in oil prices: a quantitative analysis. *Energy Economics*, *28*(4), 506-522.