MJFAS **MALAYSIAN JOURNAL OF**
FUNDAMENTAL AND APPLIED SCIENCES
PRINT ISSN: 2289-5981 | ONLINE ISSN: 2289-599X

RESEARCH ARTICLE

Forecasting model for crude oil price with structural break

Arifah Bahar [a, b], Norshela Mohd Noh [a, b,*], Zaitul Marlizawati Zainuddin [a, b]

[a] *UTM Centre for Industrial and Applied Mathematics, Univeristi Teknologi Malaysia, 81310 Johor Bharu Malaysia*
[b] *Jabatan Sains Matematik, Fakulti Sains Univeristi Teknologi Malaysia, 81310 Johor Bharu Malaysia*

* *Corresponding author: shelanoh@gmail.com, arifah@utm.my*

Abstract

Nowadays, in unstable economic environment, oil refining company is facing fluctuating crude oil price that causes unstable profit margin. Fluctuating crude oil price leads to difficulty in forecasting raw material procurement. Inaccurate forecast leads to inefficient decision making in optimizing refining company profit margin. In order to overcome an inaccurate in forecasting raw material procurement, an appropriate study of forecasting model is needed. Thus the objective of this study is to model fluctuating crude oil price based on geometric Brownian motion and mean reverting Ornstein-Uhlenbeck process and also to forecast fluctuating crude oil price with structural break. In modeling crude oil price, the information on whether the structural break exists is very crucial due to the long memory property might be camouflaged by the existence of the structural break. In this study, we employed long memory test to West Texas Intermediate (WTI) daily data from 2nd January 1986 to 31st August 2016 using log periodogram regression of Geweke and Porter-Hudak (1983). Bai and Perron test was applied to find break date. The result indicates that crude oil price is characterized by structural breaks. With the assumption that future price is affected by today's price, we modeled and forecasted crude oil price using geometric Brownian motion and mean reverting Ornstein-Uhlenbeck process for 14 days, 30 days and 6 months. Results showed that forecasting crude oil price is highly accurate for short term with geometric Brownian motion compared to mean reverting Ornstein-Uhlenbeck process.

Keywords: Crude oil price, long memory, structural break, stochastic differential equation, forecasting crude oil price

INTRODUCTION

Nowadays, in unstable economic environment, many industries are facing fluctuating procurement of raw material prices which causes unstable profit margin. As we know, crude oil is the main raw material used by oil refining companies where most of the total operation cost is from the cost of crude oil. Fluctuating crude oil price leads to difficulty in forecasting raw material procurement price. Inaccurate forecast leads to inefficient decision making in optimizing refining company profit margin. For example, the well known refining companies, Exxon Mobil, BP and Shell were facing declining net income from 2006 to 2009. One of the reasons is due to combination of high crude oil price [1]. Therefore, the changes of crude oil price will affect the operation of refining companies. In order to overcome an inaccurate forecasting raw material procurement prices, an appropriate study of forecasting model is needed. Thus the objective of this study is to model fluctuating crude oil price based on geometric Brownian motion and mean reverting Ornstein-Uhlenbeck process and also to forecast fluctuating crude oil price with structural break. The reasons we are choosing geometric Brownian motion is because the pattern of crude oil price grow exponentialy, volatile and follow the random walk. While the crude oil price volatility exhibits mean reverting behavior that suit with Ornstein-Uhlenbeck process.

In order to study crude oil price behavior, we need to think the dependence between points in the time series data. It considered having long memory or long range dependence property if the

statistical dependence decays more slowly than exponential decay. According to Ding and Granger [2], a time series have a long memory property if auto correlation function declining slowly as lag $\rho(h)$ behave as

$$\rho(h) \approx h^{2d-1}, h \rightarrow \infty \ and \ \sum_{h=0}^{\infty}|\rho(h)| = \infty \qquad (1)$$

which mean that as the lag tends to infinity their sum of lag diverges. However, the appearance of strong persistence in this time series might also be due to structural breaks. In modeling crude oil price, the information on whether the structural break exists is very crucial. This is due to the long memory property might be camouflage by the existence of the structural break. That is, if the structural break exist and not taken into the model, the process might show the property of long memory which is not the truly long memory. Thus in this study, we need to investigate the crude oil prices data are true long memory or short memory with structural breaks in order to model and forecast accurately crude oil price using geometric Brownian motion and mean reverting Ornstein-Uhlenbeck process.

This study will primarily focus on the daily crude oil prices of West Texas Intermediate (WTI) collected online from U.S. Energy Information Administration (EIA) from 2nd January 1986 to 31st August 2016 which consists of 7736 observations. First, we checked for long memory and structural break for the whole data. Then we modeled and forecasted crude oil price with geometric Brownian motion and Ornstein-Uhlenbeck process models for 14 days, 30 days

and 6 months using the data after structural break date were found. By conducting this study, both academic and industry will gain benefits. In academic, we can figure out which stochastic differential equation is more accurate in forecasting crude oil price and industry gains a better understanding of the crude oil future price to mitigate risk in future procurement raw material in optimizing oil refining industry profit margin.

LITERATURE REVIEW

In modeling energy prices, Barros et al., [3] showed that presence of structural breaks is important to consider because the result suggest long memory property if the breaks are not allowed. In testing long memory to crude oil spot price, Aloui and Mabrouk [4] found that WTI from January 1986 to July 2007 did not display long memory in daily return while volatility exhibit strong evidence of long memory. Cunado et al., [5] found no evidence of long memory in the crude oil future price return from 1983 to 2008, but they found strong evidence of long memory in the volatility process using nonparametric, semiparametric and parametric approach. While Mostafei et al., [6] running the Perron 1997 unit root test with structural break and conclude that Iranian light and heavy crude oil from1980 to 2009 have one break date. Jibrin et al., [7] modelled WTI and Brent crude oil price using ARFIMA models and identified long memory characteristic and discovered three structural breaks in both time series. However they did not confirm that subseries of the time series also discovered was long memory. Then, Yusof et al., [8] concluded that though the time series exhibit long memory property, we have to make sure the subseries (after detecting structural breaks) also display long memory property. Most of these studies only focus on either structural break or long memory except for [7]. There is less study on crude oil price with long memory, structural breaks, modeling and forecasting using stochastic differential equations. Thus, in this study, we will investigate whether WTI crude oil price has the property of long memory with and existence of structural breaks. We know that crude oil price behavior is similar to stock price behavior, [9], [10] and [11] indicates that geometric Brownian motion method is more effective to forecast stock prices. However they did not compare with mean reverting Ornstein-Uhlenbeck process. Therefore we will model and forecast crude oil price with structural breaks using geometric Brownian motion and mean reverting Ornstein-Uhlenbeck process.

METHODOLOGY

Before we model and forecast WTI crude oil price, we need to check whether the daily crude oil price data from 2nd January 1986 to 31st August 2016 in Fig. 1 is true long memory or short memory with structural break. This time series in Fig. 1 indicates stable price, followed by gradual increase and volatile.
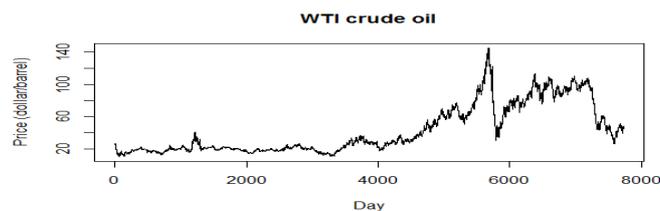

Fig. 1 WTI crude oil price from 2nd Jan 1986 to 31st August 2016.

Then we check for autocorrelation function in Fig. 1. It is known that short memory with structural breaks exhibit long memory properties. Slowly declining autocorrelation function in Fig. 2 is perhaps because of the presence of long memory as mentioned in Eq. (1).
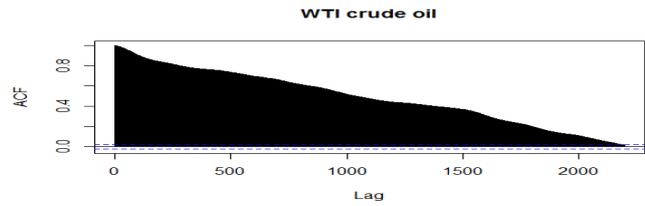

Fig. 2 Autocorrelation function for WTI crude oil price.

The series may now stationary by taking the first difference as in Fig. 3. The autocorrelation function for the first difference series in Fig. 4 shows exponential decay which means that the crude oil price may consider has short memory property.
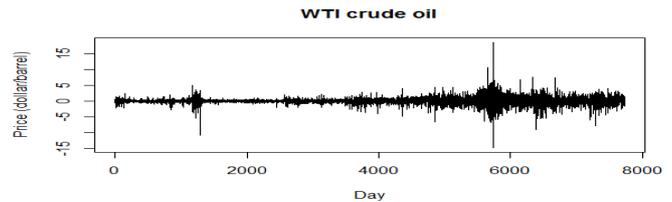

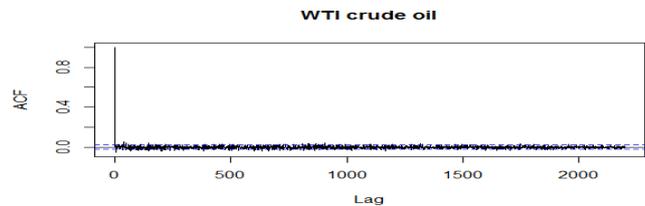**Fig. 3** Time plot of the first difference series of WTI crude oil price.

Fig. 4 Autocorrelation function of the first difference series.

Long memory and structural break test

The log periodogram regression (GPH) of Geweke and Porter-Hudak (1983)

Apart from time series have a short memory property if auto correlation function declining exponentially as in Fig. 4, we need to run for long memory formal test to find the fractional difference parameter using Geweke and Porter-Hudak (1983). The GPH estimator is based on the regression equation using the periodogram function as an estimate of the spectral density. Under the null hypothesis of no long memory, the t-statistic

$$t_{d=0} = \hat{d} \left(\frac{\pi^2}{6 \sum_{j=1}^{g(n)} (U_j - \bar{U})^2} \right)^{-\left(\frac{1}{2}\right)} \qquad (2)$$

Where $U_j = \ln\left[4 sin^2\left(\frac{\omega_j}{2}\right)\right]$ and $\bar{U}$ is the sample of $U_j, j = 1, \ldots, g(n)$. We found that fractional difference parameter d for WTI crude oil price return is 0.000045 which is close to 0. It indicates that there is no correlation between various points in the crude oil price data. In other word we may say that the behavior of crude oil price follows random walk.

Bai and Perron test

The main assumption used to estimate a model is that the structure of the conditional mean does not change. In particular, parameters are constant. However, in realistic assumption, the parameter does change as in oil price data in Fig. 1. Bai and Perron [12] test was applied to find the structural break in crude oil price data. We found the following structural breaks at 11th Nov 1999, 18th Feb 2005 and 13th Oct 2009 which indicates that this time series has three discontinuities in the data generating process. These three breaks may be attributed to event in Table 1 [7]. Neglect and misspecification of structural break will lead to poor forecast of future crude oil price.
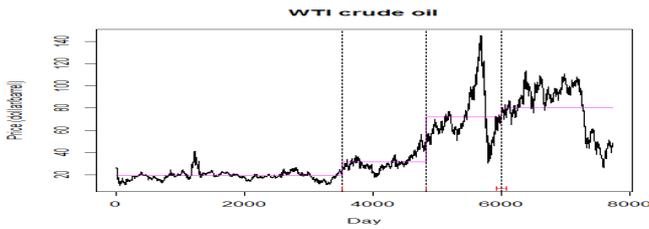
Fig. 5 WTI daily crude oil price with structural breaks from 2nd Jan 1986 to 31st Aug.

Table 1 WTI Structural break dates.

No	WTI Structural break point	Date	Event
1	3517	11th Nov 1999	Asian financial crisis
2	4833	18th Feb 2005	OPEC limit production capacity
3	6001	13th Oct 2009	Global financial collapse

On the whole, we can conclude that the WTI crude oil price time series is characterized by structural breaks and not by long memory. From these result we will model and forecast WTI crude oil price using geometric Brownian motion and mean reverting Ornstein-Uhlenbeck process.

Modeling and forecasting crude oil price

We choose crude oil price data after structural break of year 2009 to model and forecast future crude oil price. The data is from 5th Feb 2010 to 31st Aug 2016 consists of 1657 observations as in Fig. 6 are used for the training set in modeling crude oil price using geometric Brownian motion and Ornstein-Uhlenbeck process models. While 1st Sept 2016 - 21st Sept 2016 (14 days), 1st Sept 2016 - 13th Oct 2016 (30 days) and 1st Sept 2016 - 19th May 2017 (6 months) are used as a testing set. In this study, $X(t)$ refer to the crude oil price and analyzed data correspond to the log transformation data.
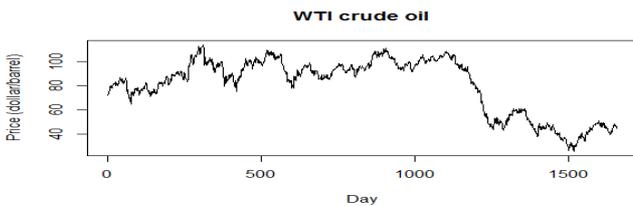

Fig. 6 WTI crude oil price from 5th Feb 2010 to 31st Aug 2016.

Geometric Brownian motion (GBM)

$$dX(t) = \mu X(t)dt + \sigma X(t)dW(t) \qquad (3)$$

where X(t) is the crude oil price at time t with drift μ, volatility σ and W(t) is random value at time t. Using Ito's formulation, it has solution

$$X_t = X_0 \, e^{\left(\left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma W_t\right)} \qquad (4)$$

Mean reverting Ornstein-Uhlenbeck process (OU)

$$dX(t) = \lambda(\mu - X(t))dt + \sigma dW(t) \qquad (5)$$

where λ is mean reverting speed, μ is long term mean. As crude oil price X(t) increase or decrease some mean value, the drift term will become negative or positive. This will make dX(t) more likely be negative and positive so X(t) will decrease and increase. The

difference between GBM and OU is in OU; the diffusion term does not depend on X(t) so the X(t) may become negative. OU has solution

$$X_t = X_0 \, e^{-\lambda t} + \mu\left(1 - e^{-\lambda t}\right) + \sigma\sqrt{\frac{1 - e^{-2\lambda t}}{2\lambda}} N_{0,1} \qquad (6)$$

In modeling crude oil price, the value cannot be negative so we solve the problem by transformation of the initial stochastic differential equation using the Ito's formula to a process [13]. Eq. (4) and (6) are the crude oil price model that is able to predict crude oil price at specific time t.

A maximum likelihood estimation (MLE) method is considered in this study due to provide consistent approach in parameter estimation by maximizing the density function of geometric Brownian function and mean reverting Ornstein-Uhlenbeck process. Root mean square error (RMSE) is used to find the accuracy of modeling method. The smaller value of RMSE, the better the model. RMSE formula is

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n}(X_t - \hat{X}_t)^2}{n}} \qquad (7)$$

To measure the accuracy of forecasting method, we use mean absolute percentage error (MAPE) which widely used in statistics. MAPE formula is

$$MAPE = \frac{\sum\left|\frac{X_t - \hat{X}_t}{X_t}\right|}{n} \qquad (8)$$

RESULT

We estimate the parameters by maximizing the density function and we obtained the result as in Table 2. Using these parameters, we model the fluctuating WTI crude oil price based on geometric Brownian motion and mean reverting Ornstein-Uhlenbeck process.

Table 2 Parameter estimations using MLE.

Model	μ	σ	
GBM	-0.00005	0.02166	
Model	**λ**	**μ**	**σ**
OU	0.001	69.189	1.475

From Table 3 we found that root mean square error for geometric Brownian motion is smaller than mean reverting Ornstein-Uhlenbeck process which indicates that geometric Brownian motion is better model for forecasting fluctuating WTI crude oil price.

Table 3 Root mean square error for modelling with GBM and OU process.

RMSE(GBM)	RMSE(OU)
15.983	26.636

Fig. 7, 8 and 9 show forecasting WTI crude oil price using geometric Brownian motion for 14 days, 30 days and 6 months.
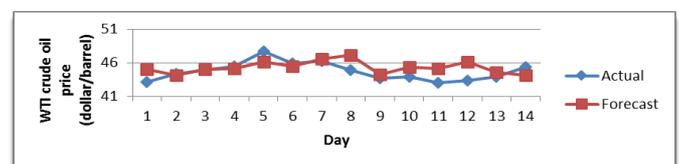
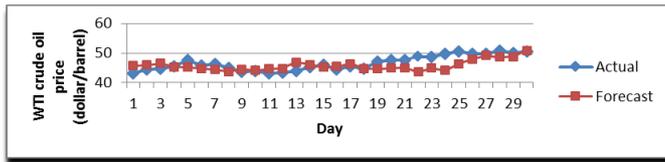

Fig. 7 Forecasting WTI crude oil for 14 days using GBM.

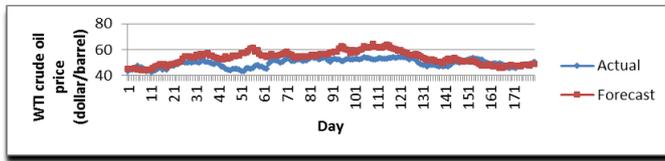**Fig. 8** Forecasting WTI crude oil for 30 days using GBM.

Fig. 9 Forecasting WTI crude oil for 6 months using GBM.

Due to the values of mean absolute percentage error in Table 4 is less than 10 meaning that geometric Brownian motion is a highly accurate model to forecast future WTI crude oil price for a short term.

Table 4 Mean absolute percentage error for forecasting using GBM.

No	Duration	MAPE(GBM)
1	14 days	2.544
2	30 days	3.944
3	180 days	9.451

We also did a comparison with the forecast crude oil price using all the data without considering the structural break in Table 5 and we found that forecasting crude oil price with structural break gives the better result.

Table 5 Mean absolute percentage error for forecasting using GBM with and without structural break.

No	Duration	MAPE(with structural break)	MAPE(without structural break)
1	14 days	2.544	6.064
2	30 days	3.944	9.150
3	180 days	9.451	17.695

CONCLUSION

WTI daily crude oil prices are studied in this paper for modeling and forecasting using geometric Brownian motion and mean reverting Ornstein-Uhlenbeck process. The first difference data was checked for auto correlation function and it showed rapidly declining autocorrelation function which indicates it may have short memory property. Using the log periodogram regression (GPH) of Geweke and Porter-Hudak (1983) we found that fractional difference parameter d for WTI crude oil price is close to 0. It indicates that the behavior of crude oil price follows random walk. Bai and Perron test was applied to find structural break and we found the following break points, 11th Nov 1999, 18th Feb 2005 and 13th Oct 2009 in WTI crude oil price. From these tests, we can conclude that the WTI crude oil price time series is characterized by structural breaks. These time series have three structural breaks meaning that the series has three discontinuities in the data generating process. We choose the data after structural break 2009 to model based on geometric Brownian motion and Ornstein-Uhlenbeck process. We found that geometric Brownian motion gives better model than Ornstein-Uhlenbeck process to forecast WTI crude oil price. By comparing mean absolute percentage error between data and forecasting value, we can conclude that geometric Brownian motion provides a highly accurate model to forecast future WTI crude oil price for a short term. We also found that forecasting crude oil price with structural break gives the better result than without considering structural break. The extension of this study is to consider other than MLE in parameter estimation method.

ACKNOWLEDGEMENT

REFERENCES

[1] A. Andrews, R. Pirog, and M. F. Sherlock, "The U . S . Oil Refining Industry : Background in Changing Markets and Fuel Policies," *Congr. Res. Serv.*, 1–39, 2010.

[2] C. W. J. Granger and Z. Ding, "Varieties of long memory models," *J. Econom.*, 73(1), 61–77, 1996.

[3] C. P. Barros, G. M. Caporale, and L. a. Gil-Alana, "Long Memory in German Energy Price Indices," *SSRN Electron. J.*, 1–25, 2012.

[4] C. Aloui and S. Mabrouk, "Value-at-risk estimations of energy commodities via long-memory, asymmetry and fat-tailed GARCH models," *Energy Policy*, 38(5), 2326–2339, 2010.

[5] J. Cunado, L. A. Gil-Alana, and F. P. de Gracia, "Persistence in some energy future markets," *J. Futur. Mark.*, 30(5), 409–507, 2010.

[6] H. Mostafei, A. A. R. Sani, and S. Askari, "A methodology for the choice of the best fitting continuous-time stochastic models of crude oil price," *Int. J. Energy Econ. Policy*, 2, 137–142, 2013.

[7] S. A. Jibrin, Y. Musa, U. A. Zubair, and A. S. Saidu, "ARFIMA Modelling and Investigation of Structural Break (s) in West Texas Intermediate and Brent Series," *CBN J. Appl. Stat.*, 6(2), 59–78, 2015.

[8] F. Yusof, I. L. Kane, and Z. Yusop, "Structural break or long memory: An empirical survey on daily rainfall data sets across Malaysia," *Hydrol. Earth Syst. Sci.*, 17(4), 1311–1318, 2013.

[9] Z. A. Siti Nazifah and M. J. Maheran, "A Review on geometric Brownian Motion in forecasting the share prices in Bursa Malaysia," *World Appl. Sci. J.*, 17, 87–93, 2012.

[10] S. N. Z. Abidin and M. M. Jaffar, "Forecasting share prices of small size companies in Bursa Malaysia using geometric Brownian motion," *Appl. Math. Inf. Sci.*, 8(1), 107–112, 2014.

[11] R. D. Estember and M. J. R. Maraña, "Forecasting of Stock Prices Using Brownian Motion – Monte Carlo Simulation," in *Proceedings of the 2016 International Conference on Industrial Engineering and Operations Management Kuala Lumpur, Malaysia, March 8-10, 2016*, 2016, pp. 704–713.

[12] B. Y. J. Bai and P. Perron, "Estimating and testing linear models with multiple structural changes," *Econometrica*, 66(1), 47–78, 1998.

[13] E. Platen and N. Bruti-Liberati, "Exact Simulation of Solution of SDEs," in *Numerical Solution of Stochastic Differential Equations with Jumps in Finance*, vol. 64, no. 2001, Springer Verlag Berlin Heidelberg, 2010, p. 61.