Modelling and Forecasting Crude Oil Prices during COVID-19 Pandemic

Ernie Hendrawaty[1]*, Rialdi Azhar[2], Fajrin Satria Dwi Kesumah[1], Sari Indah Oktanti Sembiring[2], Mega Metalia[2]

[1]Department of Management, Faculty of Economics and Business, Universitas Lampung, Lampung, Indonesia, [2]Department of Accounting, Faculty of Economics and Business, Universitas Lampung, Lampung, Indonesia. *Email: ernie.hendrawaty@feb.unila.ac.id

ABSTRACT

Currently, the world suffers from the COVID-19 pandemic, which affects almost every aspect of daily life, giving rise to recession and affecting the world prices of crude oil. The study aims to model the high uncertainty of volatility as well as to forecast the daily prices of crude oil during the pandemic. One econometric model applied in this study is the Generalised Autoregressive Conditional Heteroscedasticity (GARCH) that allows more accurate and appropriate statistical analyses. Particularly, this study also discusses solving economic issues on the condition of any disturbances in the stability of daily crude oil prices. The findings suggest that the AR(1)-GARCH(1,1) model is a well-fitted model to predict relatively small errors. This model can act as a foundation for determining strategies in the future while facing such uncertain circumstances.

Keywords: Forecasting, COVID-19 Pandemic, Crude Oil Prices, Pandemic, Generalised Autoregressive Conditional Heteroscedasticity Model
JEL Classifications: C5, C53, C58, Q4, Q471

INTRODUCTION

Today's global economy is facing the worst circumstances as COVID-19 continues to spread. This pandemic has been affecting economic conditions such as trading, global supply chains and pressured asset pricing, and it forces multinational businesses to make difficult decisions due to limited information (Ayittey et al., 2020). Ivanov (2020) stated that a certain issue on risk of supply chain was marked by any disturbances and ripple effects that may have a high uncertainty. Furthermore, the pandemic affects people's consumption levels, which are decreasing, forcing markets to be more wary in their budgets. Therefore, while production is still ongoing, there is lesser demand, causing a high uncertainty in prices, such as that in the oil industry.

The COVID-19 pandemic has had the worst impact on crude oil, which plummeted and reached its first negative price. Some analysts argued that the decreasing price might be because some investors are worried about low demands. Besides, the economic mobility of some regions in lockdown will probably be paralysed, indicating a decreasing consumption level. Moreover, the effect of a macro economy on oil price volatility is a crucial factor for importers as well as oil-exporting countries (Drachal, 2016). Therefore, it becomes a necessity to forecast uncertain daily prices of crude oil with error level reduction. In economic statistics, forecasting a financial time-series data with high accuracy is one way to make better decisions.

Forecasting daily crude oil prices (COPs) is important and challenging because it might have a consequence on increasing and decreasing most economic and non-economic factors (Safari and Davallou, 2018). Abdulmajeed et al. (2020) stated that the applied mathematics model, artificial intelligence, big data and the forecasting method are potential tools predict the oil prices. Statistically, the Generalised Autoregressive Conditional

Heteroscedasticity (GARCH) forecasting model shows a good ability in forecasting the time-series dataset (Engle, 1982; Tse and Tsui, 2002). Ahmed et al. (2018) showed in their empirical study that GARCH model can be a fitted alternative to show the volatility behaviours.

2. METHODOLOGY AND STATISTICAL MODELLING

The observation data analysed in this study are the worldwide COPs. In this study, we refer to the COPs from late 2019 to May 2020. Bollerslev (1986) introduced GARCH(p,q) to model the behaviour of volatility that can be equipped to have a good measurement for forecasting model. Some procedures are applied to satisfy the fittest model of GARCH(p,q) to predict its short future prices.

2.1. Stationary Satisfaction
To satisfy the requirements of the GARCH(p,q) model, the first condition is to have a stationary dataset. Statistically, one measurement is by checking the data plot; if the fluctuation of the dataset is not stable around zero, it is considered as non-stationary (Gunarto et al., 2020). Dickey and Fuller (1979) introduced the Augmented Dickey-Fuller (ADF) test to check stationary data as mathematically present as follows.

$$DF_\tau = \frac{\gamma_i}{\widehat{Se}_{\gamma_1}} \qquad (1)$$

The hypothesis is defined as.
H_0: $DF\tau > 2.57$ = non-stationary
H_0: $DF\tau < 2.57$ = stationary

In addition, Tsay (2005) tested stationary dataset by computing the autocorrelation function (ACF) and the partial autocorrelation (PACF), where a non-stationary dataset can be identified by their decay movement for any given lags. Since most of financial data series are not stationary in both the mean and variance, transformation into a stationary dataset should be done by applying the method of difference (Ambya et al., 2020).

2.2. Differencing
Granger and Joyeux (1980) introduced the method of differencing to transform a non-stationary time-series dataset into stationary to stabilise its mean and volatility. The mathematic equation is as follows.

$$a(B) = (1-B)^d \qquad (2)$$

where B is defined as backward operator; d is the number of differencing; and a(B) is the integrating filter of order d. Once a stationary dataset has been met, the stable movement in mean and volatility model of GARCH can be applied after the confirmation that the model introduced in this study has been free from the ARCH effect (Tsay, 2014).

2.3. ARCH Effect Test
It is worth-noting that in modelling time-series for financial data, the probability of having a heteroscedasticity is quite high (Engle,

1982), making the estimation parameters of the forecasting model less accurate. The presence of the ARCH effect is examined by computing the Lagrange Multiplier (LM) test (Lee and King, 1993), and the order of ARCH can be determined by applying the Wong and Li (1995) test. If the probability value of the LM test is significant (<0.001) at any given orders, then the heteroscedasticity involved in the model requires a long memory to process its larger order (Ahmad et al., 2016). Since the estimation of the variance changes in the ARCH(q) model (P = 0) has a short memory process, the GARCH model (P > 0) is then applied as the squared residuals in past data estimated the variances Tsay (2005).

2.4. The Mean and Variance Model of AR(p)-GARCH(p,q)
The mean model of AR(p) is defined to have lag order of p, and the order conditional variance and its squared residuals are presented as order p and q, respectively. Equations 3 and 4 mathematically present the purposed models.

$$COPCOV_t = \varphi + \sum_{i=1}^{p}\phi_i COPCOV_{t-i} + \varepsilon_t \qquad (3)$$

$$\sigma_t^2 = \alpha + \sum_{i=1}^{q}\beta_i \hat{\delta}^2{}_{t-i} + \sum_{j=1}^{p}\delta_j COPCOV_{t-j} \qquad (5)$$

If the mean square error (MSE) and root mean square error (RMSE) are relatively small in association with the statistical description model, the models are assumed to have a well-fitted measurement to forecast (Azhar et al., 2020).

3. RESULTS AND DISCUSSION

3.1. Data Description and Stationary Diagnostics
The data observation is taken from the COPs data during the Covid-19 pandemic, with a study period from December 2019 to May 2020. A total of 152 sample data are observed to examine the impact on COPs, which are becoming very volatile, and to forecast both the mean and volatility for the next 10 days.

The analysis is started by plotting the time-series data to visually understand the behaviour, i.e. stationary or non-stationary. Figure 1 shows that in the first 50 observations, the COPs was relatively stable around $60. As the pandemic spread around the globe, it decreased gradually until approximately the 80th data. Afterwards, the COVID-19 pandemic caused COPs to plummet that it only reached around $15 and was down to a negative price on the 101st day of observation, which was the first time ever in the history of COPs. However, for the sake of statistical analysis in this study, the 101st data is assumed to be similar to the previous one. Furthermore, the data climbed due to a positive reaction in the market, which is shown by the rocket increase in a few days ahead. Nevertheless, the COPs could not maintain the positivity as the COVID-19 pandemic dropped it into the lowest price, nearly reaching zero in the 125th data. In the meantime, the market returned to positive as the world economy has been rebuilt,

described by the gradual upward movement of COPs to the last date of study, reaching approximately $40 per barrel.

The descriptive data shows that the dataset is non-stationary because of the unstable movement of the mean and variance. Hence, it indicates that it is necessary to test the stationary statistically by applying ADF test as presented on Table 1.

From Table 1, the P-value of the zero mean of lag 3 is not significant (>0.05), which suggests that the result agrees with the previous data description, proving that the mean and variance of the data series is not stationary. Furthermore, Figure 2 confirms what has been proven statistically from Table 1, which shows that all three graphs do not satisfy the dataset as stationary. Figure 2a shows that the distribution data are not normal as they exceed the interval curve, and Figure 2b depicts the autocorrelation moves too smooth. These first two graphs from Figure 2 affirm that the mean and variance of the dataset is non-stationary. Figure 3c meets the stationary dataset, as the mean of PACF is around zero after lag 1.

3.2. Conversion of a Stationary Dataset

The dataset of the world COPs has been confirmed as non-stationary. The next step in time-series modelling is to convert it

Figure 1: Plotting of the world COPs data series
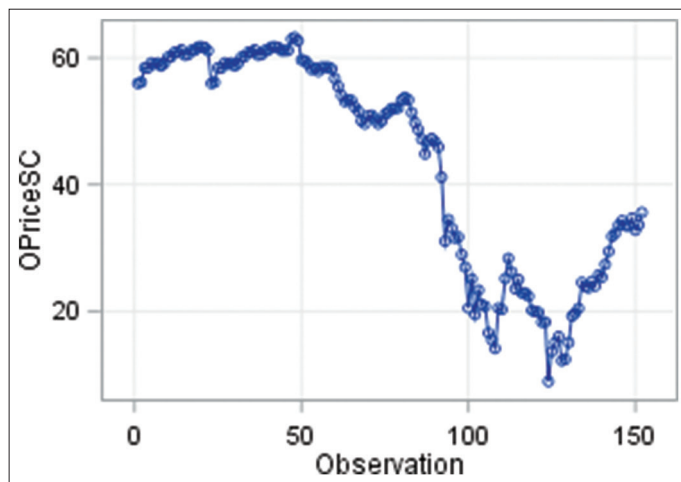
to stationary by applying the differencing method with one or more lag(s). This study then conducts differencing 1 to test whether it does turn to a stationary dataset. As shown in Figure 3, all graphs satisfy the stationary condition. Once differencing 1 is conducted, the plotting of dataset shown on Figure 3a volatiles around zero, wherein there are some data that reaches two standard deviations. It is clear as the COVID-19 pandemic makes the fast movement of COPs but still maintains the track of zero circle. Figure 3b and 3c show a rapid down trend after lag 1 in the area of zero and the dataset is normally distributed and fitted the curve shape shown on Figure 4.

The ADF unit-root test is then examined to prove the stationary dataset statistically. Table 2 shows statistical significance ($P < 0.05$) in the zero mean, which suggests that after conducting differencing 1, the dataset is already statistically stationary. As a result, the stationary dataset allows us to further attempt the stages in modelling forecasting COPs.

3.3. Mean and Variance Model

Azhar et al. (2020) stated that heteroscedasticity is frequent in the financial time-series data. Hence, although the ARIMA(p,d,q) model can fit the data for forecasting, the ARCH effect is required to examine whether the model involves heteroscedasticity. We skipped the process modelling ARIMA (p,d,q) in this study on purpose, but we went further tested the availability of an unstable homoscedasticity using the ARCH-LM test as follows.

From Table 3, the ARIMA(p,d,q) model involves the ARCH effect as the test of portmanteau Q and Lagrange Model (LM) identifies to have p-value of less than their respective P-value. This diagnosis concludes to reject the null hypothesis, as the dataset involving to model ARIMA(p,d,q) has the effect of ARCH. Therefore, it is then necessary to model mean and variance more accurately, of which is required to be generalised. Furthermore, GARCH(p,q) model is applied to generalise the conditional effect of heteroscedasticity to have a good fit measurement for modelling variance and its forecasting, while the AR(p) model for the mean model is carried out.

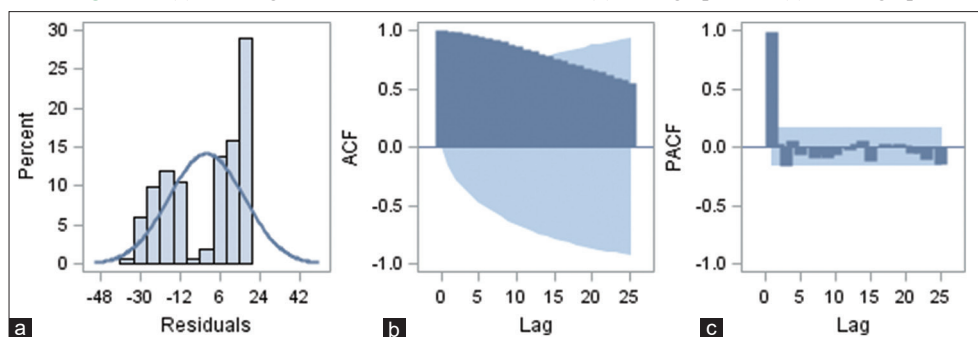Figure 2: (a) Fit Diagnostics of Normal Distribution; (b) ACF graph; and (c) PACF graph

Table 1: ADF unit-root examination

Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
Zero mean	3	−0.6569	0.5359	−1.1348	0.2326		
Single mean	3	−1.5613	0.8274	−0.9652	0.7647	0.8179	0.8623
Trend	3	−4.0049	0.8838	−1.1335	0.9190	0.7609	0.9899

Figure 3: (a) Plotting dataset, (b) ACF graph, (c) PACF graph and (d) Normal distribution graph after differencing 1
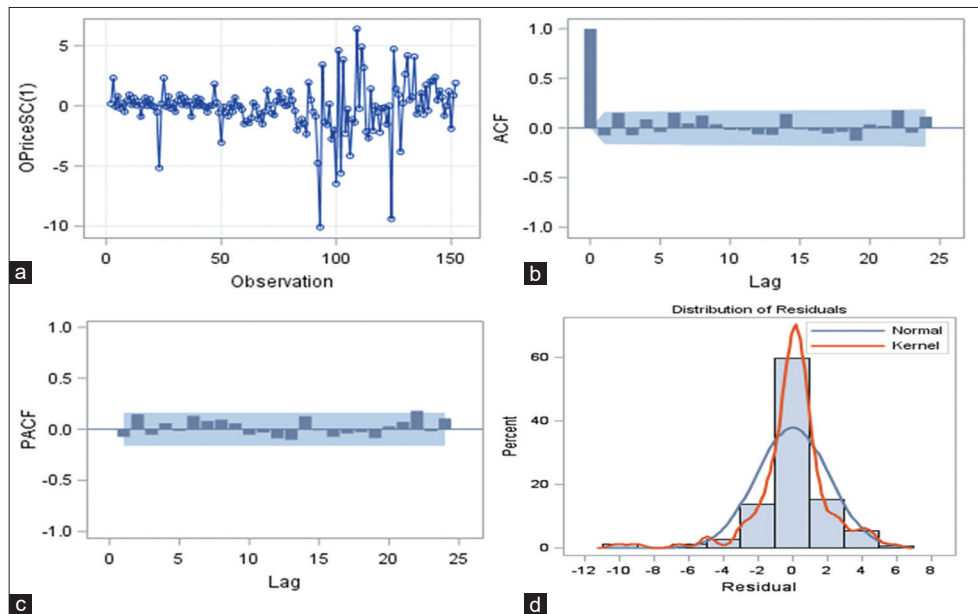

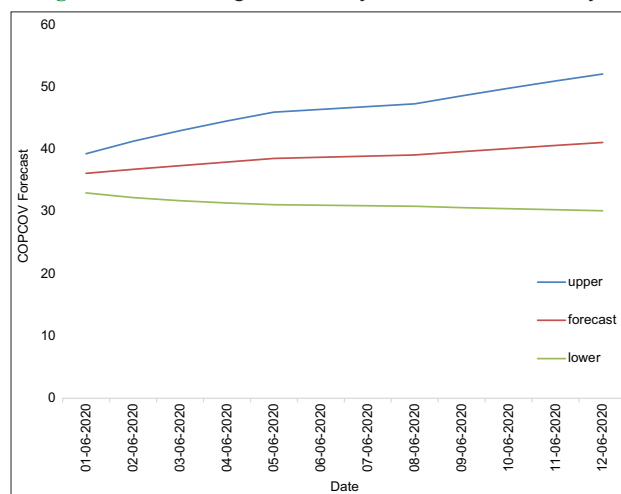**Figure 4:** Forecasting data of daily world COPs for 10 days

Table 2: Augmented unit-root (ADF) Test (d=1)

Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
Zero mean	3	−106.042	0.0001	−5.47	<0.0001		
Single mean	3	−110.888	0.0001	−5.53	<0.0001	15.27	0.0010
Trend	3	−111.595	0.0001	−5.53	<0.0001	15.33	0.0010

Table 3: ARCH-LM disturbances for daily world COPs based on OLS residuals

Order	Q	Pr > Q	LM	Pr > LM
1	101.8220	<0.0001	96.7166	<0.0001
2	179.9966	<0.0001	98.3812	<0.0001
3	226.4789	<0.0001	99.9454	<0.0001
4	260.5028	<0.0001	100.6403	<0.0001
5	278.6486	<0.0001	101.1346	<0.0001
6	288.6359	<0.0001	101.3129	<0.0001
7	290.7221	<0.0001	101.9782	<0.0001
8	290.9554	<0.0001	101.9849	<0.0001
9	291.8320	<0.0001	102.8089	<0.0001
10	295.0178	<0.0001	102.9625	<0.0001
11	300.6530	<0.0001	103.0377	<0.0001
12	307.3686	<0.0001	103.0938	<0.0001

Table 4 shows that the AR(1)-GARCH(1,1) model fits to the stationary dataset, which is indicated by a P < 0.005 for each parameter estimation. The model can be equated as follows.

AR(1) for modelling the mean: $COPCOV_t = 55.8089 - 0.9680\ COPCOV_{t-1} + e_t$

GARCH(1,1) for modelling the variance:
$$\sigma_t^2 = 0.3199 + 0.3105\varepsilon_{t-1}^2 + 0.6676\sigma_{t-1}^2$$

The AR(1)-GARCH(1) model is statistically a fit measurement as shown in Table 5. It is worth-noting that the MSE value is relatively small at 4.91458, which means that the mean error is also considerably small. The RMSE is then calculated as 2.21688,

which is significantly small relative to its unconditional variance. This implies that the variance of both the model and dataset is significantly close to each other. The r-square is also identified to have a considerably significant value of 98.27%. In fact, it can be summed up that the model gives an accurate prediction as a forecasting model.

The persistency of the model can be analysed by summing both parameter estimates of ARCH and GARCH. If the coefficients are close to 1, the conditional variance of the model is distributed constantly. This implies that the model provides a better prediction.

3.4. Forecasting the Daily World COPs

Since 1 June 2020, with the ongoing COVID-19 pandemic, most countries have come to decide to reopen the economy. The role of the AR(1)-GARCH(1) model is to predict the next 10 days of the world COPs, which is from June 1st to 12th 2020. As shown in Figure 4, there is a gradual upward trend in the daily COPs forecast, but the increasing trend is still not approaching the initial price before the pandemic was coming. However, as regulators attempt to stabilise the economy, the increasing world price of oil is expected to continue as predicted from the model.

The study is carried out after the 10-day prediction price. Consequently, the prediction model can be compared with its real data to check its accuracy. Table 6 shows a comparison between the predicted prices from the established model and the real price of world COPs from 1 June to 12 June 2020.

From Table 6, the predicted prices are well fitted with the respective real prices for the first eight days. However, for the last two days, the gaps are wider. This is because the AR(1)-GARCH(1,1) model has a wider confidence interval for a longer time. Therefore, it is only suitable in forecasting data for a short period.

4. DISCUSSION

Forecasting models on oil prices has been widely applied as a benchmark to determine what strategies to take in the future. Crude oil as a global commodity becomes a very crucial element in the world economy. The increasing and decreasing daily COPs could be a determinant of the price levels of other commodities. In previous study on the macroeconomic changes in oil price, Hamilton (1983) found that a shock on oil price is a factor that contributes to the recession in the United States of America. Furthermore, oil prices that experience upward trends might increase companies' production costs, reduce the profit and affect the stock prices (Apergis and Miller, 2009).

As we encounter many economic issues caused by the fluctuation of oil prices, forecasting data could be used as an alternative to prepare for the worst. As the COVID-19 pandemic has caused the decrease in oil price, it also affects to decrease the consumption level in communities dramatically. On the other hand, it is almost impossible to stop the production of crude oil as the high cost of its exploration. Therefore, an alternative solution is to save production on the available oil tanks at the time of oil price decreasing.

The dramatic decrease on world COPs gives a lesson on how important the speed of the oil supply chain is to reach selling tanks. Government, as buyers, should prepare the budget to use at any time when a decrease in oil prices occurs, so the buy action is highly likely to be a benefit. Oil refinery companies should also provide spare tanks to save production even if there is zero demand. Time scheduling on oil production based on forecasting the economic condition could also be a wise strategy in planning. The decreasing on oil price could be a sign of a declined demand; therefore, during this time, it is proper to slow down the production.

5. CONCLUSION

The COVID-19 pandemic has caused a dramatic effect on most sectors, including the economic sector. The global COPs are the most affected, dropping it into the lowest price in history, reaching a negative price during the outbreak. The ultimate aim of this study was to construct the fitted model of mean and variance in order to forecast the daily and future COPs once lockdowns are lifted globally by most countries, which is the first 10 days of June 2020.

Initially, the dataset time series was non-stationary; therefore, it is necessary to make it stationary by conducting the method of differencing. Once it is stationary, before the mean and variance model prediction, heteroscedasticity is checked to identify its

Table 4: Parameter estimates for AR(1)-GARCH(1,1) Model

Variable	DF	Estimate	Standard error	t-value	Approx Pr > |t|
Intercept	1	55.8089	5.4480	10.24	<0.0001
AR1	1	−0.9680	0.009045	−107.02	<0.0001
ARCH0	1	0.3199	0.1007	3.17	0.0015
ARCH1	1	0.3105	0.0776	4.00	<0.0001
GARCH1	1	0.6676	0.0763	8.75	<0.0001

Table 5: GARCH estimates

SSE	747.015418	Observations	152
MSE	4.91458	Uncond var.	14.5842314
Log likelihood	−305.58044	Total R-square	0.9827
SBC	636.280281	AIC	621.160878
MAE	1.39231386	AICC	621.571837
MAPE	5.61445866	HQC	627.302905
		Normality test	452.1857
		Pr > Chi-square	<0.0001

Table 6: Comparison between predicted and real prices of crude oil

Forecasted prices	Daily prices
36.2186	35.49
36.8464	36.88
37.4541	37.33
38.0424	37.42
38.6117	39.49
39.1629	38.17
39.6963	38.98
40.2127	39.54
40.7125	36.43
41.1963	36.24

involvement in the model. If the model has such an issue and requires a long memory process of a larger order of p, then the GARCH model is applied. Therefore, the model suggested in this study is AR(1)-GARCH(1,1) which has a relatively small MSE and RMSE at 4.91458 and 2.21688, respectively. The R-square of 99.87 also means the model provides a good ability in forecasting the dataset.

6. ACKNOWLEDGMENTS

REFERENCES

Abdulmajeed, K., Adeleke, M., Popoola, L. (2020), Online forecasting of COVID-19 Cases in Nigeria using limited data. Data in Brief, 30, 105683.

Ahmad, N., Ahmed, R.R., Vveinhardt, J., Streimikiene, D. (2016), Empirical analysis of stock-returns and volatility: Evidence from Asian stock markets. Technological and Economic Development of Economy, 22(6), 808-829.

Ahmed, R.R., Vveinhardt, J., Streimikiene, D., Ghauri, S.P., Ashraf, M. (2018), Stock returns, volatility, and mean reversion in emerging and developed financial markets. Technological and Economic Development of Economy, 24(3), 1149-1177.

Ambya, A., Gunarto, T., Hendrawaty, E., Kesumah, F.S.D., Wisnu, F.K. (2020), Future natural gas forecasting model and its policy implication. International Journal of Energy Economics and Policy, 10(5), 64-70.

Apergis, N., Miller, S. (2009), Do structural oil-market shocks affect stock prices? Energy Economy, 31(4), 569-575.

Ayittey, F.K., Ayittey, M.K., Chiwero, N.B., Kamasah, J.S., Dzuvor, C. (2020), Economic Impacts of Wuhan 2019-nCov on China and the world. Journal of Medical Virology, 92, 473-475.

Azhar, R., Kesumah, F.S.D., Ambya, A., Wisnu, F.K., Russel, E. (2020), Application of short-term forecasting models for energy entity stock price (Study on Indika Energy Tbk, JII). International Journal of Energy Economics and Policy, 10(1), 294-301.

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

Dickey, D.A., Fuller, W.A. (1979), Distribution of the estimators for autoregressive time series with a unit root. Journal of American Statistical Association, 74, 427-431.

Drachal, K. (2016), Forecasting spot oil price in a dynamic model averaging framework-have the determinants changed over time? Energy Economics, 60, 35-46.

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

Granger, C.W.J., Joyeux, R. (1980), An introduction to long-memory time series models and fractional differencing. Journal of Time Series Analysis, 1(1), 15-29.

Gunarto, T., Azhar R., Tresiana, N., Supriyanto, S., Ahadiat, A. (2020), Accurate estimated model of volatility crude oil prices. International Journal of Energy Economics and Policy, 10(5), 228-233.

Hamilton, J.D. (1983), Oil and the macro-economy since World War II. Journal of Political Economy, 91(2), 228-248.

Ivanov, D. (2020), Viable Supply Chain Model: Integrating Agility, Resilience and Sustainability Perspectives-Lessons from and Thinking Beyond the COVID-19 Pandemic, Annals of Operations Research.

Lee, J.H.H., King, M.L. (1993), A locally most mean powerful based score for ARCH and GARCH regression disturbances. Journal of Business and Economics Statistics, 11, 17-27.

Safari, A., Davallou, M. (2018), Oil price forecasting using a hybrid model. Energy, 148, 49-58.

Tsay, R.S. (2005), Analysis of Financial Time Series. Hoboken, New Jersey: John Wiley and Sons, Inc.

Tsay, R.S. (2014), In: Balding, D.J., Cressie, N.A., Fitzmaurice, G.M., Goldstein, H., Johnstone, I.M., Molenberghs, G., Scott, D.W., Smith, A.F., Tsay, R.S., Weisberg, S., editors. Multivariate Time Series Analysis with R and Financial applications. Hoboken: John Wiley & Sons, Inc.

Tse, Y.K., Tsui, A.K.C. (2002), A multivariate generalized autoregressive conditional heteroscedasticity model with time-varying correlations. Journal of Business and Economic Statistics, 20(3), 51-62.

Wong, H., Li, W.K. (1995), Portmanteau test for conditional heteroscedasticity, using ranks of squared residuals. Journal of Applied Statistics, 22, 121-134.