Universiti Teknologi MARA
Cawangan Kedah
Kampus Sungai Petani

2021 iCMS
International Conference on Computing, Mathematics and Statistics

e-PROCEEDINGS

of The 5th International Conference on Computing, Mathematics and Statistics (iCMS2021)

4-5 August 2021

Driving Research Towards Excellence

e-Proceedings of the 5th International Conference on Computing, Mathematics and Statistics (iCMS 2021)

Driving Research Towards Excellence

Editor-in-Chief: Norin Rahayu Shamsuddin

Editorial team:
Dr. Afida Ahamad
Dr. Norliana Mohd Najib
Dr. Nor Athirah Mohd Zin
Dr. Siti Nur Alwani Salleh
Kartini Kasim
Dr. Ida Normaya Mohd Nasir
Kamarul Ariffin Mansor

e-ISBN: 978-967-2948-12-4
DOI

Library of Congress Control Number:

Publication by
Department of Mathematical Sciences
Faculty of Computer & Mathematical Sciences
UiTM Kedah

TABLE OF CONTENT

PART 1: MATHEMATICS

PART 2: STATISTICS

PART 3: COMPUTER SCIENCE & INFORMATION TECHNOLOGY

PART 4: OTHERS

A COMPARATIVE STUDY OF MOVING AVERAGE AND ARIMA MODEL IN FORECASTING GOLD PRICE

Arif Luqman Bin Khairil Annuar [1], Hang See Pheng [2], Siti Rohani Binti Mohd Nor [3] and Thoo Ai Chin[4]
[1,2,3] Department of Mathematical Sciences, Faculty of Science
Universiti Teknologi Malaysia, [4] Azman Hashim International Business School,
Universiti Teknologi Malaysia, Johor Bahru
([1] arifluqman@graduate.utm.my, [2] sphang@utm.my, [3] sitirohani@utm.my, [4] acthoo@utm.my)

Technical analysis is becoming an important reference to the traders in financial markets such as stock markets, foreign exchange market, and gold market. Moving average is one of the most vastly used statistical model in technical analysis to project and forecast the trend of data in financial markets. This study aims to compare the performance of the simple moving average (SMA), moving average convergence divergence (MACD), and autoregressive integrated moving average (ARIMA) in forecasting gold price with the effect of COVID-19 pandemic. A dataset is collected from World Gold Council in 2020 and 2021. The comparison is done by different validations: accuracy rate, mean absolute percentage error (MAPE) and root mean square error (RMSE). The results indicating the accuracy rate of MACD is higher than SMA. SMA is found providing smaller forecasting error 1.264% compared to compare to ARIMA with 1.809%. The result favours to SMA as compared to ARIMA due to the dataset obtained during COVID-19 pandemic phase which observes the instability of economy.
***Keywords*:** Simple moving average, moving average convergence divergence, ARIMA, gold price forecasting

1. Introduction

Gold is a precious metal which is accepted as an alternative currency and commonly traded in the commodity market due to its sustainability and high liquidity value, (Makridou et al., 2013). Aside from its exploitation in industrial fields like jewelry industry, gold can be used as a hedge against inflation. In addition, the gold prices increased 6%, as reported in 2008 and early 2009, while other metal prices dropped and sent the global economy to recession, (Topal et al., 2010). There are many existing statistical techniques in predictive analytics to estimating the gold price in the global market, which include multiple linear regression, analysis of variance (ANOVA), and exponential smoothing. Since early 20th century, moving average is one of the most vastly used statistical tools to project and forecast the trend in a set of data up until today. Furthermore, many researchers from various backgrounds use the moving average method and ARIMA to analyze the trend of the gold data due to their characteristic to smooth out short-term fluctuations in trends and cycles in the market.

This study focuses on the application of three types of moving average models; simple moving average (SMA), moving average convergence divergence (MACD), and auto regressive integrated moving average (ARIMA) to forecast the gold prices during COVID-19 pandemic. The data for the gold price can be observed through any gold trading centres. However, according to World Gold Council (2020), most of global gold trading volumes are traded in the three most important gold trading centres which are London OTC market, US futures market, and Shanghai Gold Exchange (SGE). Despite the gold market is constantly volatile, the scope of this study covers on the data in which the gold price moves in its highest volatility phase because high volatility in gold price will indicate the fluctuation in stock market. In addition to the forecasting purpose, this study will also attempt to compare the performance and test the reliability of both models in the forecasting the gold price. The fitness of both models can be observed to conclude whether the models are reliable to be used as forecasting model. Furthermore, all three models used in this study do not take economic factors such as inflation and currency price movements, into accounts. Thus, if one or both models fail to be tested as a fit model, we can conclude that a model solely constructed only by using a single dependent variable, in this case: gold price, cannot be assumed as reliable without including economic factors.

2. Literature Review

Given the prospect of profitability which can be made in gold market, many researchers, especially economists, have worked on developing forecasting models and strategies to gain advantages in predicting the future price of that highly in-demand commodity. These forecasting models were developed using various mathematical approaches including linear regression and moving averages. However, researchers have taken in consideration of the simplicity of those mathematical approaches to be used in the forecasting model, hence, a few modifications and improvisations were made.

One of the famous forecasting methods involving gold price is auto-regressive integrated moving average (ARIMA) model which is widely proposed and discussed by many researchers. Some researchers also suggested that ARIMA model to be the most accurate forecasting model and better than any other forecasting models. Khaemasunun (2007) proposed two forecasting models: Multiple-Regressive model and ARIMA model to forecast gold price in Thailand. The study concluded that the ARIMA model is the best model to forecast the gold price in Thailand while the multiple-regressive model contributed many factors which affected the gold price. However, one of the limitations of implementing ARIMA model into the dataset is dealing with the outliers because the model can significantly underestimate extreme values below or above the mean. In a more detailed research, Hassani et al. (2015) compared the ARIMA model with several forecasting models including exponential smoothing (ETS), trend and seasonal components (TBATS), a fractionalized ARIMA model (ARFIMA), vector autoregression (VAR), Bayesian autoregression (BAR), and Bayesian VAR model (BVAR). Among the forecasting models proposed in the study, the ETS model presented the best forecasting results over 24 months in terms of root mean squared errors.

The other common forecasting method is generalized autoregressive conditional heteroskedasticity (GARCH) models which is mainly applied to forecast the volatility of the price. Kristjanpoller et al. (Kristjanpoller, Fadic, & Minutolo, 2014) improvised by using Artificial Neural Network GARCH model (ANN-GARCH), concluding that the forecasted outcomes are better, in term of the accuracy, than GARCH model. The ANN-GARCH model able to determine the factors that variables involved in the results which increases the accuracy of forecasts compared to GARCH model.

3. Methodology

3.1 Data collection

In this study, the dataset for gold price was collected from the World Gold Council website (World Gold Council, 2021). The selected dataset of the daily close price of gold in the financial market from 1st January 2020 to 26th February 2021. These time series are chosen based on the observations of high volatility of the gold price in 2020 due to the result of economic uncertainty caused by the COVID-19 pandemic outbreak. This study aims to investigate the gold price movement with effects of the global COVID-19 pandemic. A significant observation in this pandemic period is the all-time highest price for gold is recorded on 7th August 2020, hitting at US$2,067.15. With the effects of the COVID-19 pandemic, most of major currencies, especially US dollar, deflated due to shutdown of businesses and economies. As mentioned earlier by Makridou et al. (2013), gold investment increased tremendously among investors when global financial markets crash, and the global economy is in recession due to less trustworthiness in other financial investments.

3.2 Simple moving average (SMA)

Simple moving average is a method which calculates the average of selected range of data by the number of periods in that range. It commonly used to observe the trend of a data set. Hence, it is very useful to forecast long-term trends. It is represented by an equation of

$$SMA = (A_1 + A_2 + \cdots + A_n) \,/\, n \tag{1}$$

where A_n stands for the average data in period while n is the number of periods in the data sets.

3.3 Moving average convergence divergence (MACD)

Moving Average Convergence Divergence (MACD) is another example of improvisation of simple moving average method and widely benefited by financial traders and investors in trading market (Jian& Junseok, 2018). It consists of three exponential moving averages (EMA) which extends to a period of 9, 12, and 26 days. These three EMA are translated into a momentum oscillator by taking away the shorter moving average from the longer moving average. This can be seen from the equation below:

MACD = EMA 1 (12-day periods) – EMA 2 (16-day periods) (2)

while EMA 3 (9-day periods) are being used as a signal to indicate the buy and sell call in trading. Figure 3 displays the framework of forecasting the gold price using MACD which begin with calculating the exponential moving average of 9, 12, and 26 days before constructing the MACD histogram.

3.4 Autoregressive integrated moving average (ARIMA)

Auto regressive integrated moving average model abbreviated as ARIMA (p, d, q) model is an improvised moving average method. It combines the auto regressive (AR) model and the moving average (MA) model, where d is the difference term while p, q are the delay parameters. Auto regressive integrated moving average (ARIMA) model converted non-stationary sequence in the financial time series into stationary time series through d-order difference before established as a differential auto regressive moving average model that is ARIMA (p, d, q) model. Using ARIMA (p, d, q) model to model the time series data enforces a combination of AR, MA, and ARMA models with different orders to enable the model delivers information for time series, thus increase the accuracy of forecasting. The modified stationary sequence, Y_t^* can be expressed as:

$$Y_t^* = (1 - B)^d Y_t \tag{3}$$

In ARIMA model, the future values of a variable are assumed to be a linear function of few pasts, current observations and random errors, (Zhang, 2003). The structure of the ARIMA model can be shown as:

$$y_t = \theta_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} \tag{4}$$

where y_t and ε_t are the actual value and random error at time period t, ϕ_i and θ_j are model parameters. Meanwhile, p and q are integers are often referred to as orders of the model. Random errors, ε_t, are assumed to be independently and identically distributed with a mean of zero and a constant variance of σ^2.

3.4.1 Autocorrelation function (ACF)

Implementing ACF into ARIMA model provides a better improvement to measure the amount of linear dependence between observations in a time series that are separated by a lag, Zakaria et al. (2012). Box et al. (2015) also proposed the usage of autocorrelation function and partial autocorrelation function in a dataset as the basic tools to identify the order of the ARIMA model. Autocorrelation function (ACF) formula is given as below:

$$\boldsymbol{r_k} = \frac{\sum_{t=1}^{n}(y_t - \bar{y})^2 (y_{t+k} - \bar{y})}{\sum_{t=1}^{n}(y_t - \bar{y})^2} \qquad (5)$$

while t_{rk} statistic is $t_{rk} = \frac{r_k}{S_{rk}}$ where $S_{rk} = \sqrt{\frac{1+2\sum_j^k r_j^2}{n}}$.

3.4.2 Partial autocorrelation function (PACF)

PACF is used to measure the degree of association between Y_t and Y_{t-k}, when the effects of other time lags $(1, 2, 3, \dots, k-1)$ are removed. The sample PACF is given by:

$$\boldsymbol{r_{kk}} = \frac{r_k - \sum_{j=1}^{k-1}(r_{k-1,j})(r_{k-j})}{1 - \sum_{j=1}^{k-1}(r_{k-1,j})(r_j)} \qquad (6)$$

while t_{rkk} statistic is $t_{rkk} = \frac{r_{kk}}{S_{rkk}}$ where $S_{rkk} = \sqrt{\frac{1}{n}}$.

4. Discussion and Results Analysis

This section presents the experimental results of simple moving average (SMA) analysis, moving average convergence divergence (MACD) forecasting analysis, and autoregressive integrated moving average (ARIMA) forecasting analysis for the gold price data.

4.1 Simple moving average (SMA) forecasting

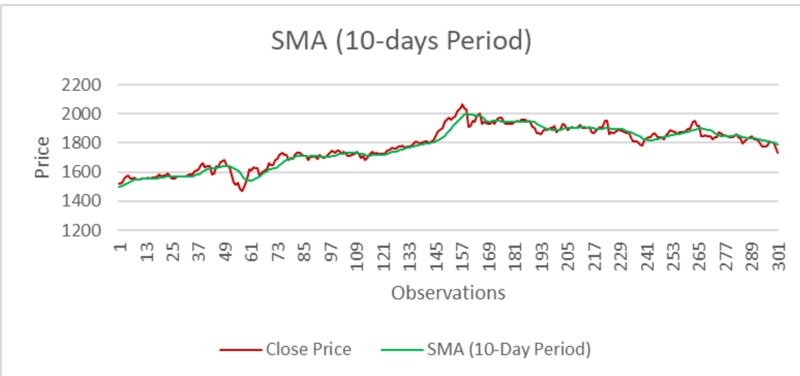


Figure 1: The SMA (10-day period) and actual gold daily close price

Figure 1 visualizes the simple moving average in 10-days period for the daily gold price from 1st January 2020 to 26th February 2021 in 301 observations. It can be observed that SMA forecast the 10-days period tracks closely the daily closing price of gold while it lags in following the trend of the daily closing price. However, there are some points the actual gold daily close price breaks above and below the SMA (10-day period) which indicates the change in the trend of the gold price, (Metghalchi et al., 2016).

The accuracy rate is determined by the comparison of the SMA buy-sell signal and the actual price trend on a 5-day buy-and-hold strategy. The SMA buy-sell signal is indicated by the intercepts of the actual price and the SMA (10-day period). The buy signal is triggered when the actual price crosses above the SMA (10-day period) while the sell signal is noticed when the actual price crosses below the SMA (10-day period). In all 301 observations, there are 31 observations where the actual price intercepts with the SMA (10-day period).

The accuracy rate of applying SMA to seek buy and sell signal against the actual price trend is recorded at 0.4516 or 45.16% which is significantly low to adopt this method for forecasting the gold price.

4.2 Moving average convergence divergence (MACD) forecasting

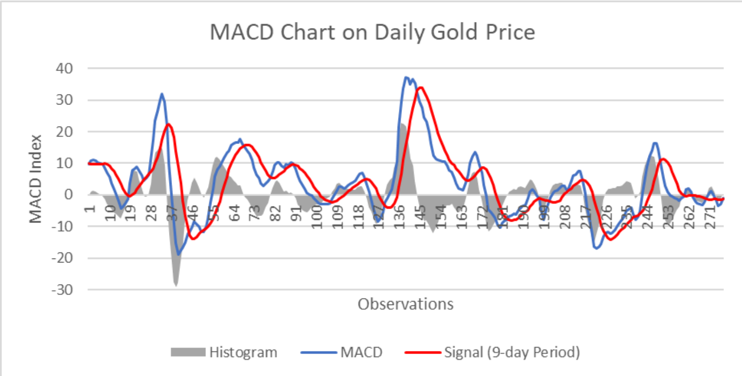


Figure 2: MACD and signal (9-day period) from 1st January 2020 until 26th February 2021

The accuracy rate is calculated by the comparison of the MACD buy-sell signal and the actual price trend. As mentioned previously, the MACD buy-sell signal is indicated by the intercepts between the MACD line and the signal line, SMA (9-day period). In all 301 observations, there are 30 interception points which represents the buy and sell signal respectively. Meanwhile, the actual price trend is indicated by the difference between the price on the test date and the price on the 5th day after the test date. From the results obtained, it is found that there are 18 out of 30 predicted trends well matching with the actual trends. Therefore, the accuracy rate of applying MACD to seek the buy and sell signal against the actual price trend is recorded at 0.600 or 60.00% which is relatively low for a 5 day buy-and-hold strategy.

4.3 Autoregressive integrated moving average (ARIMA) forecasting

In the implementation of autoregressive integrated moving average (ARIMA) method, three main steps to be considered – model identification, parameter identification, and diagnostic checking. Firstly, in the model identification step, the dataset stationarity needs to be tested.

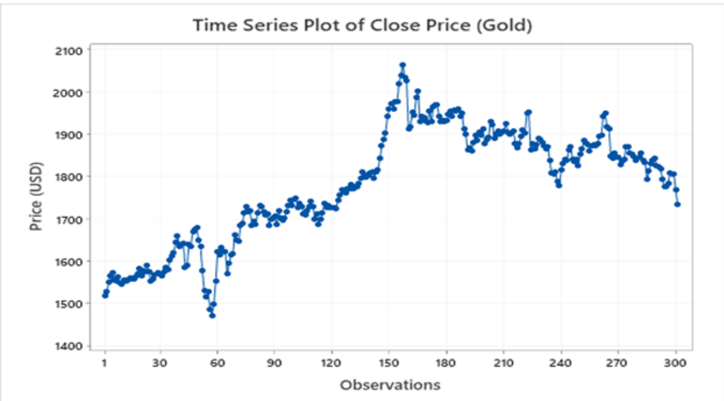
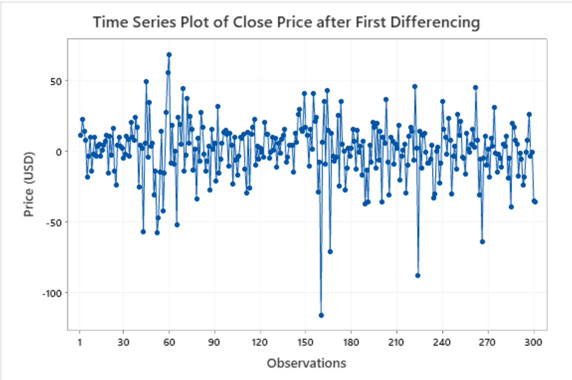


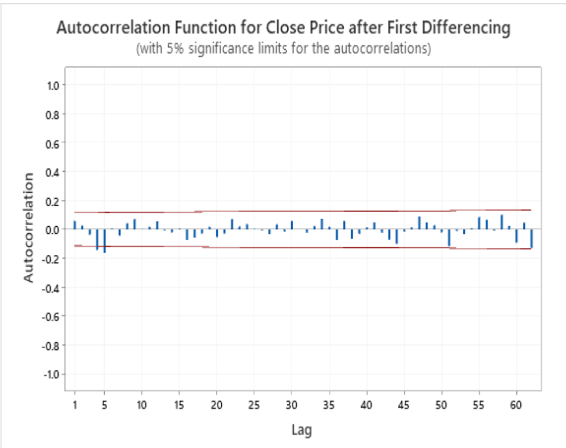
Figure 3: Time series plot for daily close price of gold from January 2020 until February 2021

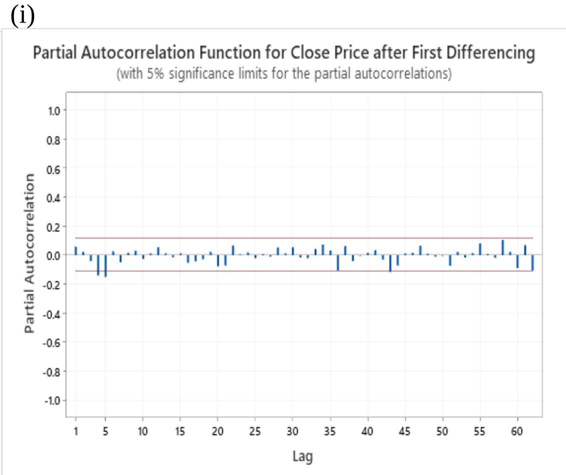
Based on Figure 3, there is an outlier dated on 31st March 2020 which is contributed by the early phase of COVID-19 outbreak. The sudden drop in the gold price observes the presence of outliers in the dataset which affected the performance of ARIMA. One of the limitations of ARIMA is the difficulty in forecasting when the outliers present in the dataset because the outliers lie outside of the trend captured by the model, (Grogan, 2020).

Figure 4 shows the model identification and parameter identification where Figure 4 (i) is the time series plot for first difference of daily close price of gold determines the d value for the ARIMA model (p, d, q) which is 1, Figure 4 (ii) and Figure 4 (iii) are the autocorrelation function and partial autocorrelation function for close price of gold after first differencing which determine the value of p and q in the parameter of ARIMA model (p, d, q).



(i)





(ii) (iii)

Figure 4: The model identification and parameter identification in ARIMA

Table 1: Ljung Box Chi-Square Statistics for ARIMA (4, 1, 4) model.

Lag	12	24	36	48
Chi-Square	2.67	8.44	16.09	28.90
DF	3	15	27	39
P-Value	0.445	0.905	0.951	0.882

In diagnostic checking, the adequacy of ARIMA (4, 1, 4) model should be tested. To test the adequacy of ARIMA (4, 1, 4) model, an adequacy test is required which is Ljung-Box test. As shown above in Table 1, Ljung-Box Chi-Square table for ARIMA (4, 1, 4) model is obtained. For a model to be determined as adequate, the p-values for all lags consist of 12, 24, 36, and 48 should be greater than 0.05 which in this case, the criteria are met.

5. Comparison of performances between moving average methods and ARIMA

The comparison between all three methods will be divided into two parts. The first part will be the comparison between simple moving average (SMA) and moving average convergence divergence (MACD) while the second part will be between simple moving average (SMA) and autoregressive integrated moving average (ARIMA).

5.1 Comparison of performances between SMA and MACD

Table 2: Comparison of accuracy rate between SMA and MACD.

	No. of buy/sell signals	Accuracy rate (%)
Simple Moving Average (SMA)	31	45.16
Moving Average Convergence Divergence (MACD)	30	60.00

Table 2 illustrates the comparison between simple moving average (SMA) and moving average convergence divergence (MACD). These methods are compared by their accuracy rates of the buy and sell signals generated by both moving averages. The accuracy rates comparison method is one of the commonly used method, to compare the moving average convergence divergence (MACD) method to other methods because MACD acts as only an indicator in the financial market rather than as a conventional forecasting tool like SMA and ARIMA, hence, mean absolute percentage error (MAPE) and root mean square error (RMSE) cannot be obtained for MACD. The accuracy rates refer to the percentage of the accurate prediction signals and the actual trend of the daily gold price.

As shown in Table 2, the accuracy rate obtained is 60% for moving average convergence divergence (MACD) which is greater than 45.16% of accuracy rate obtained for simple moving average (SMA). In other research papers involving the accuracy rate of MACD, the results obtained usually score around 60% to 80%. Thus, it can be concluded that MACD is better than SMA in modelling and forecasting the daily gold price because the buy and sell strategy can benefit more from the signals generated by MACD than SMA.

5.2 Comparison of performances between SMA and ARIMA

Table 3: Comparison of MAPE and RMSE between SMA and MACD.

	MAPE	RMSE
Simple Moving Average (SMA)	0.01264	27.6403
Autoregressive Integrated Moving Average (ARIMA)	0.018093	35.57732

Table 3 displays the comparison of mean absolute percentage error (MAPE) and root mean square error (RMSE) between simple moving average (SMA) and autoregressive integrated moving average (ARIMA). The values of both MAPE and RMSE for simple moving average (SMA) are 1.2640 % and 27.64023 respectively which are relatively lower as compared to autoregressive integrated moving average (ARIMA). The result favors to SMA as compared to ARIMA due to the outliers in the dataset which contribute to the underperforming of ARIMA in forecasting the gold price. One of the limitations of ARIMA is the difficulty in forecasting when the outliers present in the dataset because the outliers lie outside of the trend captured by the model, (Grogan, 2020).

6. Conclusion

The simple moving average method performs better than autoregressive integrated moving average due to the better values of MAPE and RMSE while the comparison between SMA and MACD sees that the accuracy rate for MACD is better than SMA. Although ARIMA is an advanced method as compared to SMA, the forecasting result for ARIMA underperformed due to certain limitations. The dataset chose ranged in the COVID-19 phase which observes the global economy including gold market experiences recession and economy instability. Due to the uncertainty of the economy, the gold price dropped to $1570 US dollar, particularly on 31st March 2020, as reported by World Gold Council. The sudden drop in the gold price observes the presence of outliers in the dataset which affected the performance of ARIMA.

Acknowledgment

The authors wish to thank the Malaysian Ministry of Education (MOE), Universiti Teknologi Malaysia (UTM) and Research Management Centre (RMC) for financial sponsorship to this work through UTM Tier 2 with grant funding number R.J130000.2654.17J37.

References

"World Gold Council (2020, November). Major global trading hubs. https://www.gold.org/what-we-do/gold-market-structure/global-gold-market#:~:text=The%20three%20most%20important%20gold,OTC%20and%20exchange%2Dtraded)

"World Gold Council (2021, January). https://www.gold.org/goldhub/data/gold-prices.

Box, G. E., Jenkins, G. M., Reinsel, G. C., and Ljung G.M. (2015). Time Series Analysis: Forecasting and Control. *John Wiley & Sons.*

Makridou, G., Atsalakis, G. S., Zopunidis, C., and Andrisopoulus, K. (2013). Gold Price Forecasting with a Neuro-Fuzzy-Based Inference System. *Int. J. Financial Engineering and Risk Management,* 1(1):35-34

Jian, W., & Junseok, K. (2018). Predicting Stock Price Trend Using MACD Optimized by Historical Volatility. Mathematical Problems in Engineering, 1(1), 1-12

Zhang, G. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing,* 50(1): 59-175.

Hassani, H., Silva, E. S., Gupta, R., and Segnon, M. K. (2015). Forecasting of the price of gold. *Applied Economics,* 47(39):4141-4152.

Mui, H. W., and Chu, C. W. (1993). Forecasting the spot price of gold: Combined forecast approaches versus a composite forecast approach. *Journal of Applied Statistics,* 20(1): 13-23.

Grogan, M. (2020, September 5). Limitation of ARIMA: Dealing with Outliers – Towards Data Science. https://towardsdatascience.com/limitations-of-arima-dealing-with-outliers-30cc0c6ddf33

Metghalchi, M., Chen, C-P., and Hajilee, M. (2016). Moving average trading rules for NASDAQ composite index. *Applied Finance Letters,* 5(2): 45-57.

Khaemasunun, P. (2007). Forecasting Thai Gold Price, College of Innovation, Thammasat University, Thailand.

Shafiee, S., and Topal, E. (2010). An Overview of Global Gold Market and Gold Price Forecasting. *Resources Policy,* 35(1): 178-189.

Zakaria, S., Al-Ansari, N., Knutsson, S., and Al-Badrany, T. (2012). ARIMA Models for weekly rainfall in the semi-arid Sinjar District at Iraq. *Journal of Earth Sciences and Geotechnical Engineering,* 2(3).

2021 ICMS

INTERNATIONAL CONFERENCE ON COMPUTING, MATHEMATICS AND STATISTICS

e ISBN 978-967-2948-12-4

9 7 8 9 6 7 2 9 4 8 1 2 4