J. Appl. Math. & Informatics Vol. **39**(2021), No. 1 - 2, pp. 239 - 249
https://doi.org/10.14317/jami.2021.239

PREDICTION OF U.S. GOLD FUTURES PRICES USING WAVELET ANALYSIS; A STUDY ON DEEP LEARNING MODELS[†]

D.H. LEE, D.H. KIM, J.H. YOON*

Abstract. This study attempts to predict the price of gold futures, a real financial product, using ARIMA and LSTM. The wavelet analysis was applied to the data to predict the price of gold futures through LSTM and ARIMA. As results, it is confirmed that the prediction performance of the existing model of predict was improved. the case of predict of price of gold futures, we confirmed that the use of a deep learning model that is not affected by the non-stationary series data is suitable and the possibility of improving the accuracy of prediction through wavelet analysis.

AMS Mathematics Subject Classification : 62P20, 91B84.
Key words and phrases : Gold future price, LSTM, ARIMA, Wavelet analysis.

1. Introduction

Gold is widely regarded as a safe asset; it is part of several important transactions in the market, both as a real product and a financial derivative. Historically, the increase in the value of gold has been inversely proportional to the value of the dollar. Moreover, gold is the preferred option for investors looking for a means of wealth accumulation and an alternative to the uncertainty of money. Researchers have studied the impact of US gold prices on Korean gold-related corporate stocks at 13% (see [16]). It should be noted that predicting the price of gold futures can enable us to comprehend the volatility of gold prices (see [14]). Moreover, predicting the price of gold futures is an important topic in the field of financial engineering.

Received August 22, 2020. Revised November 22, 2020. Accepted November 29, 2020.
*Corresponding author.
†The research of J.-H. Yoon was supported by the National Research Foundation(NRF) of Korea grant funded by the South Korean government (NRF-2019R1A2C108931012).

In this study, the price of gold futures was analyzed using Auto-Regressive Integrated Moving Average (ARIMA) and the Long Short-Term Memory (LSTM) model, which are widely used in big data analytics. To begin with, ARIMA is a type of time series analysis; it is a generalization of auto-regressive moving average (ARMA) that describes current time series values using past observations and errors. In his seminal work, Abdullah (see [1]) predicted the selling price of gold bullion in Malaysia from 2002–2007. The study demonstrated the effectiveness of ARIMA(2,1,2) for price prediction. In another study, Saeed et al. (see [13]) also predicted time series data for wheat production in Pakistan using ARIMA(2,2,1). Guha et al. (see [6]) used ARIMA to predict gold prices in India from November 2003 to January 2014. However, ARIMA has limitations when taking into consideration the random or sudden political and economic changes included in time series data.

The LSTM model addresses the vanishing gradient problem of recurrent neural networks (RNN). In a study by Kristjanpoller et al. (see [10]), the volatility of gold prices was predicted using GARCH applied with an ANN; the study confirmed that the prediction performance improved over GARCH. In another study, Jiang et al. (see [9]) predicted the Shanghai Composite Index and the Dow Jones Index using RNN and LSTM. The mean square error (MSE) was used to compare the RNN and LSTM predictions. The results suggested that LSTM exhibited smaller error than RNN. In another study, Tsai (see [15]) predicted Particulate Matter 2.5 (PM2.5) content using LSTM and evaluated the prediction performance using Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). The results suggest that the proposed approach can effectively forecast the value of PM2.5.

While there are several studies that analyze gold future prices using ARIMA and LSTM, very few studies have applied wavelet analysis correction for the prediction analysis of gold futures prices. Wavelet transforms were initially proposed as an alternative to the Fourier transform, which does not consider both the time and frequency domains concurrently. Fonseca Lemus et al. (see [5]) applied ARIMA with a wavelet transform to obtain improved forecasting results from Colombian financial market data. Murguia et al. (see [11]) analyzed chaotic time series data using wavelet transform by applying a statistical approach to experimental CTS. Antonini (see [2]) proposed a novel image compression method using wavelet transform and obtained good results from progressive transmissions and a very low bit rate compression. Similarly, wavelet transforms have found similar applications in a variety of fields. For example, Zang et al. (see [17]) developed a new approach to detect disease outbreaks based on wavelet transforms and found that it addressed the problem of negative singularities and long-term trends, which are issues that commonly occur in time series data.

In this study, we predicted the price of US gold futures using ARIMA and LSTM; we confirm that gold futures price do not show a time pattern through a comparison of the prediction performance of the two models. In addition, we apply wavelet analysis to the gold futures data to improve the overall prediction accuracy of the model.

The remainder of the manuscript is organized as follows. Chapter 2 introduces models and briefly defines the time series prediction model ARIMA, deep learning model LSTM and wavelet. In Chapter 3, the model introduced in Chapter 2 is used to derive forecasts of gold futures prices and to graphically represent the forecasts and actual gold futures prices. In addition, the performance of predictive models is compared using the evaluation indicators. Chapter 4 draws conclusions based on the results of Chapter 3.

2. The Review of Models

In this section, we introduce the ARIMA model frequently used in previous studies and the Deep Learning model, which is known to have excellent predictive performance. Specifically, we review LSTM models, which improved the shortcomings of RNN. And, in the process of analyzing data, the data pre-processing is important as much as add a influential explanatory variable. Therefore, wavelet analysis is used in the process of pre-processing the data of the gold futures price used.

2.1. Auto-Regressive Integrated Moving Average (ARIMA). ARIMA (p, d, q) is a type of time series data analysis, which is assumes that the economy is moving according to past knowledge or experience. The ARIMA(p, d, q) model is a generalization of ARMA that describes the current time series value using past observation values and errors. It is an analytical technique used to predict the next indicator on a quarterly, semi-annual, or annual basis, or to review the indicator on a weekly or monthly basis and to monitor the trend for abnormalities. While the ARMA is applicable only to the stationary series, ARIMA(p, d, q) can be applied even when the analysis target exhibits the characteristics of weak non-stationary series.

Definition 2.1. The general format of the ARIMA(p, d, q) is as follows.

$$y_t = \theta_0 + \varphi_1 y_{t-1} + \cdots + \varphi_p y_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \cdots - \theta_q \epsilon_{t-q},\ t = 1,\ 2,\ 3\ ,\cdots \tag{1}$$

where y_t is a real value, ϵ_t is white noise, and φ_i $(i = 1, 2, 3, \ldots, p)$ and θ_j $(j = 1, 2, 3, \ldots, q)$ are the model parameters. It should be noted that p stands for the order of the auto regressive process, d is the order of the data stationary, and q is the order of the moving average process.

2.2. Long Short-Term Memory (LSTM). LSTM was proposed by Hochreither and Schmidhuder (see [7]) to address the vanishing gradient problem at a RNN. LSTM comprises the cell state and hidden state. Cell state is the role of maintaining memory. And, hidden state is the role of sending information to the next step. Figure 1 shows the visual representation of the LSTM structure.

Definition 2.2. The LSTM is defined by:

$$
\begin{aligned}
f_t &= \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f) \\
i_t &= \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i) \\
o_t &= \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o) \\
g_t &= tanh(W_{xg}x_t + W_{hg}h_{t-1} + b_g) \\
C_t &= f_t \circ C_{t-1} + i_t \circ g_t \\
h_t &= o_t \circ tanh(c_t)
\end{aligned}
$$

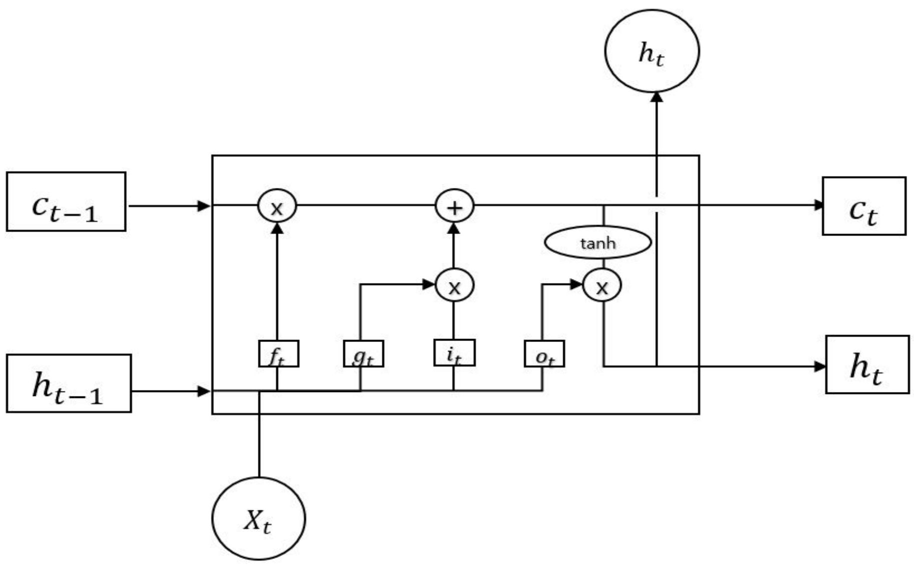


FIGURE 1. Structure of the LSTM(Olah, 2015)

The first step of LSTM is the forget gate layer that determines whether cell state is remembered. In this step, the f_t form is calculated, where in the f_t is applied to the weighted sum of the bias to determine whether the amount of information to forget. The second step is the input gate. In this step, the g_t and i_t form is calculated, where in the g_t and i_t is decide the amount of new information. The next step is the output gate. The output gate determines the value which utilizes the information passing in through the cell state. In this step, the o_t form is calculated, where in the o_t is decide the amount of information of output. The final step is the hidden state. In this step, the h_t form is calculated, After applying tanh activation to the cell state, it is output to the output gate. For more details, see [7].

2.3. Wavelet Analysis. First proposed by Morlet (see [3]), the wavelet has overcome the limitations posed by Fourier transforms, which do not consider both time and frequency domains at the same time. Wavelet analysis can decompose complex information in the signal into wavelets. In other words, wavelet analysis are used for the decomposition and reconstruction of the information contained in the signal. The most used empirically are orthogonal wavelets such as the Haar, Daublets, Symmlets and coiflets (Daubechies (1992). In this paper, we use Haar Wavelet analysis among the available multifarious wavelet analysis.

In wavelet analysis, we must note two functions; the ϕ mother function, also referred to as the scaling function, and the ψ function, which is referred to as the wavelet function. Using these two functions, the data is approximated in a block format.

The wavelet is defined by Ref. [4] and [5] as given below.

Definition 2.3. Wavelet is defined by:

$$\Psi_{(a,b)}(x) = \frac{1}{\sqrt{a}}\psi\left(\frac{x-b}{a}\right) \tag{2}$$

Wavelet transform is given by:

$$W(a,b) = \int_{-\infty}^{\infty} f(x)\Psi_{(a,b)}(x)dx \ where \ \Psi_{(a,b)}(x) = \frac{1}{\sqrt{a}}\psi\left(\frac{x-b}{a}\right) \tag{3}$$

where wavelet transform is the convolution of the wavelet. The wavelet function is designed to strike a balance between time domain and frequency domain; here, $a, b \in \{1, 2, \cdots\}$ is a dilatiation parameter and b is a translation parameter, and $f(x)$ are the signals in which period is T. $f(x) = \frac{1}{2}a_0 + \sum_{k=1}^{\infty}\left(a_k cos\frac{2\pi kt}{T} + b_k sin\frac{2\pi kt}{T}\right)$.

Definition 2.4. Haar scaling function is defined as follows:

$$\phi(x) := \begin{cases} 1, & x \in [0,1) \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

Definition 2.5. Haar wavelet mother function is defined as follows:

$$\phi(x) := \begin{cases} 1, & x \in [0, \frac{1}{2}) \\ -1, & x \in [\frac{1}{2}, 1) \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

Definition 2.6. Haar wavelet is defined as follows:

$$\psi(x) = \phi(2x) - \phi(2x-1). \tag{6}$$

3. Main Results

This section discusses the prediction results and analyzes the predictions of gold futures prices using LSTM and ARIMA with wavelet analysis. The historical data of Gold Futures were downloaded from the website(https://investing.com/)

for the period between January 2015 and November 2019. The data were divided into training data from January 2015 to May 2018, and test data from June 2018 to November 2019. Each figure is a figure of test data from June 2018 to November 2019, and forecasts represent forecast data. The horizontal axis indicates days, and the vertical axis indicates the price of gold futures per each troy ounce . In this study, we used open source wavelet analysis software (PyWavelets 1.0.2), ARIMA (statsmodels 0.9.0), LSTM model provided by Keras, and TensorFlow 1.4.0. The analysis was conducted using python in Jupiter lab 1.2.4 environment.

In the case of Wavelet, it was used in the data pre-processing process. At this time, prediction was performed on the generated value using Wavelet, and inverse transform was applied to the predicted result. In the case of ARIMA estimation results, Results of ARIMA estimation, ARIMA (1, 1, 1) is the most suitable model to be used for forecasting gold futures prices. In the case of LSTM, epoch was run 200, 500, 1000, 2000, 5000, 10000 times. As a result, the best prediction performance was confirmed when 200 epochs were performed. Also, in the case of batch size, it was set between 300 and 400. In general, in the case of batch size, the larger the value, the more stable learning can be.

Root Mean square Error(RMSE)	$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y_i}-y_i)^2}$
Mean absolute Error(MAE)	$\frac{1}{n}\sum_{i=1}^{n}\lvert\hat{y_i}-y_i\rvert$
Root Mean square percent Error(RMSPE)	$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\frac{\hat{y_i}-y_i}{y_i})^2} \times 100$
Mean Absolute Percent Error(MAPE)	$\frac{1}{n}\sum_{i=1}^{n}\lvert\frac{\hat{y_i}-y_i}{y_i}\rvert \times 100$

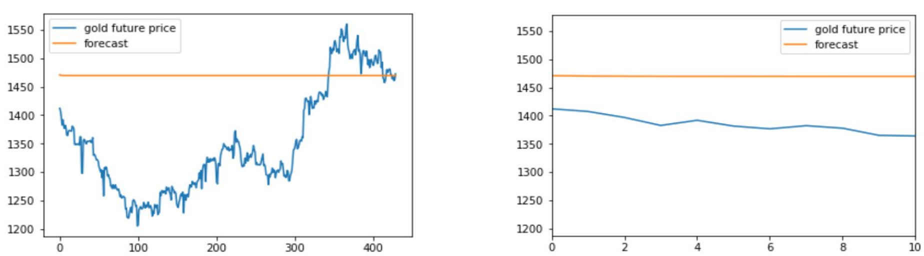
Table 1. Evaluation Index

	LSTM	LSTM by using Wavelet	ARIMA	ARIMA by using Wavelet
RMSE	34.56	30.85	150.13	107.94
MAE	20.89	20.83	132.49	97.98
RMSPE(%)	2.57	2.29	11.19	8.04
MAPE(%)	1.54	1.54	11.13	7.99

Table 2. Result of Evaluation Index

Table 1 lists the evaluation indices used for the evaluation of the prediction performance. The results in Table 1 indicate that the closer the value of the evaluation index is to 0, the better the predictive performance. RMSE is given the high weighted value of relatively errors compared to MAE . As regards MAPE, it indicates the ratio of MAE relative to the real value. Also, as regards RMSPE, it indicates the ratio of RMSE relative to the real value. Table 2 presents the comparison of the evaluation indexes obtained using the predicted values of each model. The results from Table 2 indicate the RMSE and MAE values of ARIMA and LSTM using wavelet; The results of RMSE and MAE or RMSPE and MAPE in Table 2 show that for the LSTM and ARIMA, applying wavelet to the prediction on each method reduces the varinace of the error. In addition, in the MAPE results of Table 2, the MAPE of ARIMA is 11.13%, while ARIMA, after applying Wavelet analysis, shows a 3.14% decline in MAPE from 11.13% to 7.99%. This can be attributed to the fact that wavelet analysis improves the performance when solving long-term trends.

Figure 2 shows the prediction results of gold futures prices using ARIMA. While there are limitations in that ARIMA can only be applied for short-term detection of small changes; comprehensively identifying of changes in policy and economics still remains a challenge (see [7]). The solid orange line in Figure 2 (a) does not predict all future gold futures prices over time, which indicates that the test data (2018.06.-2019.11.). This result shows that it is not a sufficiently trained model. Figure 2 (b) is an enlarged representation of Figure 2 (a), which indicates that it is nearly impossible to predict the price movement of gold futures by using ARIMA. In other words, ARIMA is not suitable for analysis of period of non-stationary series.

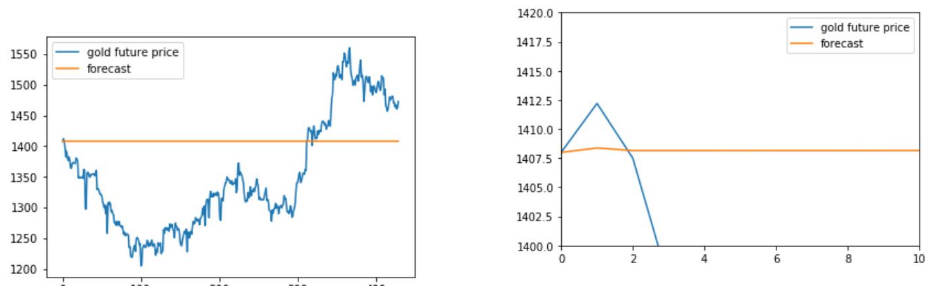


(A) ARIMA (B) ARIMA (Zoom)

Figure 2. Forecast of ARIMA

Figure 3 presents the ARIMA analysis results obtained after applying the wavelet analysis to gold futures price data. From Figure 3 (a), the performance does not seem to improve significantly; however, on closer examination, as shown in Figure 3 (b), the prediction performance improves in the short term. Thus, it can be inferred that the non-stationary series learning problem of ARIMA has been addressed a little improvement, through wavelet analysis. Nevertheless, it is inappropriate to use ARIMA for gold futures price predictions, as indicated by the orange solid line. because price of gold futures show that characteristic of non-stationary series

Next, let us consider a case where LSTM predicts the gold futures price data. As shown in Figure 4, the LSTM exhibits better predictions than ARIMA. In order words, the results of prediction using ARIMA and LSTM show that there is a minor difference. because, it was not influenced by non-stationary series. In



(A) ARIMA Model by using Wavelet

(B) ARIMA Model by using Wavelet (Zoom)

FIGURE 3. Forecast of ARIMA by using Wavelet

addition, we can observe a large error in ($x \approx 360$) near June 2019, which can be regarded as an indicator of both prices and factors such as changes in political and economic conditions.

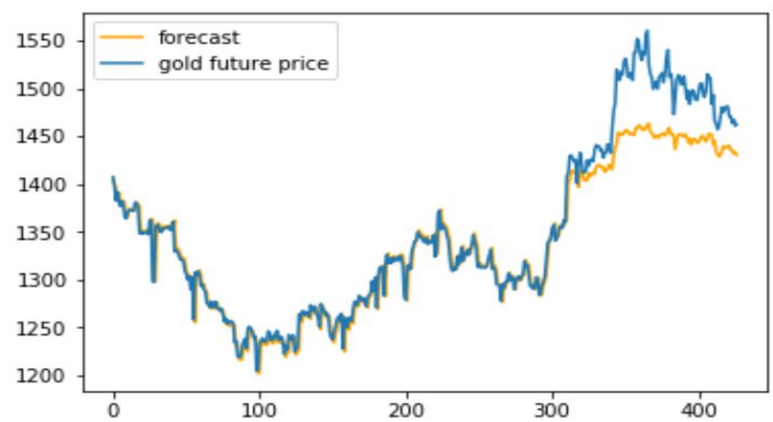


FIGURE 4. Forecast of LSTM

Figure 5 shows the graph of LSTM prediction results after the application of wavelet analysis. Although a significant difference cannot be observed from the results of Figure 4, from an RMSE and MAE standpoint (as indicated in Table 2), the LSTM predictions exhibit a certain degree of performance improvement. Similar to Figure 4, we can observe a large error in ($x \approx 360$) near June 2019, which indicates that it was influenced by economical and political.

4. Concluding Remarks

ARIMA, LSTM and wavelet analysis are used to predict the price of gold futures. This study proposes a method of predicting data used for analysis

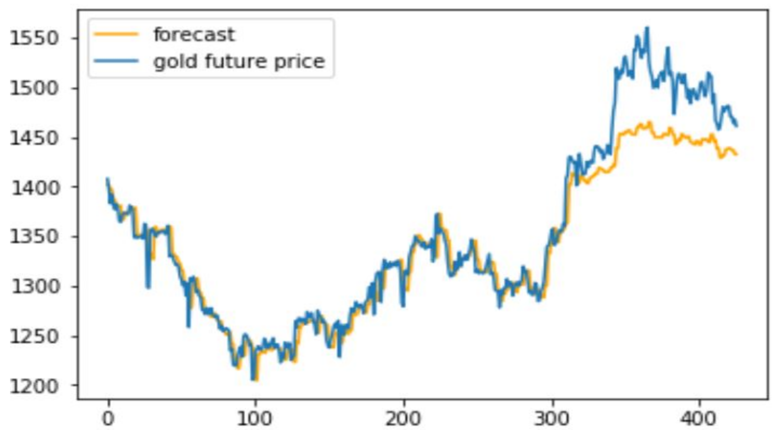


FIGURE 5. Forecast of LSTM by using Wavelet

through wavelet analysis through a kind of data pre-processing, rather than adding meaningful explanatory variables to improve the accuracy of prediction.

The influence of wavelet analysis is investigated as follows using evaluation indicators. The impact of data on gold futures prices is determined by comparing the performance of predictions through ARIMA and LSTM and the performance of predictions between ARIMA and LSTM when wavelet analysis is used.

As a result, RMSE, MAE, MAPE, and RMSPE can all get good results from using LSTM as a predictive model. In the case of ARIMA, previous studies had good results, but for the period of gold futures price considered in this study, they showed overfitting of the training data and failure to learn properly. In this study, we can see that ARIMA shows poor results. In other words, ARIMA shows that in the case of unstable time series data and prediction results may be inaccurate.

In the case of LSTM, it is possible to confirm that there is an overall error of 2 troy ounces through MAE, and about 1.5% of the total data can be checked through MAPE. In addition, when wavelet analysis was used as a pre-processing of the existing data, RMSE can confirm 3.7 troy ounce reduction and RMSPE can see 0.28% error reduction. In addition, from the results in Figures 4 and 5, you can see the increase in error from June 2019. In this regard, it can be thought that it was influenced by FOMC. FOMC interest rate changes can be viewed through the following website: (cf. https://kr.investing.com/economic-calendar/interest-rate-decision-168)

These results make it possible for LSTM to see the result of reduced error variance when making predictions using data with wavelet analysis.

Therefore, in order to predict the futures price of gold, we confirmed that the use of a deep learning model that is not affected by the characteristics of time series data is suitable and the possibility of improving the accuracy of prediction through wavelet analysis.

References

1. L. Abdullah, *ARIMA model for gold bullion coin selling prices forecasting*, International Journal of Advances in Applied Sciences **1** (2012), 153-158.
2. M. Antonini, M. Barlaud, P. Mathieu, I. Daubechies, *Image coding using wavelet transform*, IEEE Transactions on image processing **1** (1992), 205-220.
3. A. Bernardino, J. Santos-Victor, *A real-time gabor primal sketch for visual attention*, Iberian Conference on Pattern Recognition and Image Analysis 2005, DOI: 10.1007/1149242941.
4. L. Chun-Lin, *A tutorial of the wavelet transform*, NTUEE, Taiwan, 2010.
5. A. Fonseca Lemus, *Wavelet analysis on financial time series*, Doctoral dissertation, Universidad del Rosario, 2018.
6. B. Guha, G. Bandyopadhyay, *Gold price forecasting using ARIMA model*, Journal of Advanced Management Science **4** (2016), 117-121.
7. S. Hochreiter, J. Schmidhuber, *Long short-term memory*, Neural computation **9** (1997), 1735-1780.
8. H. Jang, B.K. Seo, *Monetary policy rate expectation and energy prices during the FOMC announcement period*, Finance Research Letters (2019), Article 101093.
9. Q. Jiang, C. Tang, C. Chen, X. Wang, Q. Huang, *Stock price forecast based on LSTM neural network*, International Conference on Management Science and Engineering Management 2018, 393-408.
10. W. Kristjanpoller, M.C. Minutolo, *Gold price volatility: A forecasting approach using the artificial neural network–garch model*, Expert Systems with Applications **42** (2015), 7245-7251.
11. J.S. Murguia, E. Campos-Cantón, *Wavelet analysis of chaotic time series*, Revista mexicana de física **52** (2006), 155–162.
12. P.K. Narayan, S. Narayan, X. Zheng, *Gold and oil futures markets: Are markets efficient?*, Journal Abbreviation **87** (2010), 3299-3303.
13. N.A.D.E.E.M. Saeed, A.S.I.F. Saeed, M.U.H.A.M.M.A.D. Zakria, T.M. Bajwa, *Forecasting of wheat production in Pakistan using ARIMA models*, International Journal of Agriculture and Biology **2** (2000), 352-353.
14. N. Sopipan, P. Sattayatham, B. Premanode, *Forecasting volatility of gold price using markov regime switching and trading strategy*, Journal of Mathematical Finance **2** (2012), 121-131.
15. Y.T. Tsai, Y.R. Zeng, Y.S. Chang, *Air pollution forecasting using RNN with LSTM*, In 2018 IEEE 16th Intl Conf on Dependable, Autonomic and Secure Computing, 16th Intl Conf on Pervasive Intelligence and Computing, 4th Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech) (2018), 1074-1079.
16. B.J. Yim, J.H. Heo, *An empirical study on the united states gold futures price change impact on the south korean gold-related on corporate value*, Academic Society of Global Business Administration **11** (2014), 85–102.
17. J. Zhang, F.C. Tsui, M.M. Wagner, W.R. Hogan, *Detection of outbreaks from time series data using wavelet transform*, AMIA Annual Symposium Proceedings 2003, 2003:748-52.

Donghui Lee is currently a master program in the Department of Mathematics at Pusan Nationa University. His research interest is research related to Financial Mathematics using Deep Learning and Machine Learning.

Department of Mathematics, Pusan National University, Busan 46241, Korea.
e-mail: eorn04@pusan.ac.kr

Donghyun Kim received the bachelor degree in Mathematics from University of Ulsan, Ulsan, South Korea, in 2017. He is currently a Ph.D. student at the Pusan National University, Busan, South Korea. His research interests include option pricing under stochastic volatility and applied partial differential equation (PDE).

Department of Mathematics, Pusan National University, Busan 46241, Korea.
e-mail: donghyunkim@pusan.ac.kr

Ji-Hun Yoon received Ph.D. from Yonsei University. He is currently a professor at Pusan Nationa University. His research interests focus on the Financial Mathematics and Stochastic Differential Equation.

Department of Mathematics, Pusan National University, Busan 46241, Korea.
e-mail: yssci99@pusan.ac.kr