A Study of Prediction of Airline Stock Price through Oil Price with Long Short-Term Memory Model

Jae Won Choi[1], Youngkeun Choi[2]*
Department of Computer Science, University of Texas at Dallas, Richardson, TX, USA[1]
Division of Business Administration, Sangmyung University, Seoul, Korea[2]

***Abstract*—This study aims to present a model that predicts the stock price of an airline by setting the economic and technical information of oil as features and taking advantage of the LSTM method. In this study, oil price data for about seven years from January 4, 2016, to April 14, 2023, were collected through FinanceDataReader. The collected data is a total of 1,833 days of AA stock price data. The price data consists of six categories: Date, Open, High, Low, Close, Volume, and Change (price is based on dollars). Data is stored every 24 hours, so it was judged to be most suitable for short-term price prediction (24 hours later) to be conducted in this study. In this paper, normalized closing price data was trained for 50 epochs. As a result of learning, the loss value converged close to 0. The MSE measured by the accuracy of the model shows a result of 0.00049. The significance of this study is as follows. First, it is meaningful in that it can present indicators such as more sophisticated predictions and risk management to airline companies. Oil price as our selected feature can compensate for the poor performance of a simple model and its limitations on overfitting.**

Keywords*—Stock price prediction; airline; oil; long short-term memory

I. Introduction

For a considerable period, scholars have been interested in forecasting future stock price movements [1]. While proponents of the efficient market hypothesis argue that accurate stock price prediction is unattainable, empirical evidence suggests that by employing appropriate variables and suitable models, it is feasible to predict future stock prices and detect patterns of stock price movements with a relatively high level of precision. LSTM (Long Short-Term Memory) is a powerful technique for predicting stock prices due for several compelling reasons [2]. Firstly, stock prices exhibit sequential patterns, and LSTMs are designed to capture long-term dependencies in time-series data. They can effectively model the sequential relationships in historical stock price data, enabling them to capture relevant trends, patterns, and seasonality for accurate future price forecasting. Secondly, LSTMs incorporate a memory cell that selectively retains or forgets information over time. This memory cell helps in preserving important historical information while filtering out irrelevant noise, making LSTMs robust in handling noisy and volatile stock price data where multiple factors may affect price movements. Thirdly, stock prices are influenced by a wide range of factors with non-linear relationships, and LSTMs are capable of modeling such non-linearity. With non-linear activation functions and multiple gates, LSTMs can capture complex patterns in the data, allowing them to capture intricate relationships in stock price data that linear models may miss. Fourthly, LSTMs can automatically learn relevant features from raw data without relying on manual feature engineering. This is particularly advantageous for stock price prediction as relevant features may change over time and may be difficult to identify manually. LSTMs can learn to represent important features from historical data, enhancing the accuracy of predictions. Lastly, LSTMs are adaptable and can be trained on different time scales, such as daily, weekly, or monthly data, depending on the prediction task requirements. They can also be trained on various stock markets or different stocks, making them flexible for different prediction scenarios. Finally, LSTMs can undergo training using extensive historical stock price data, which is often accessible for multiple stocks. This enables them to capture prolonged patterns and trends, rendering them suitable for predicting stock prices over extended timeframes. Consequently, the ability of LSTMs to handle sequential data, capture long-term dependencies, model non-linear relationships, learn relevant features, adapt to different time scales, and scale to large datasets makes them a formidable tool for stock price prediction.

Selecting crucial factors for predicting stock prices using LSTM is imperative to capture relevant information that can impact stock prices [3]. Economic indicators, financial statements, market sentiment, and news are significant factors that can significantly affect stock prices. Incorporating these relevant factors into the LSTM model enables it to capture the complex relationships and patterns in the data, resulting in more accurate stock price predictions. The careful selection of these factors ensures that the LSTM model is trained on meaningful inputs, enhancing its ability to forecast stock prices and potentially assisting in making informed investment decisions.

The stock market is influenced by crude oil prices, which reflect economic conditions. Existing studies present mixed findings, with some reporting positive associations and others indicating negative associations. Park and Ratti [4] discovered a negative influence on composite stock market returns but a positive influence on energy stock market returns. Filis and Chatziantoniou [5] found a negative influence on the stock markets of importing countries but a positive influence on the stock markets of exporting countries. Narayan and Sharma [6] identified that the influence depends on the scale of the enterprise. Bjornland [7] observed a positive influence on crude oil production stock prices and a negative influence on transportation stock prices. Lu and Chen [8] investigated the influence of WTI crude oil prices on stock prices of

*Corresponding Author.

transportation companies across eight countries and found that sea transportation is less affected due to increased crude oil transportation volume. Mohanty et al. [9] revealed a negative impact on US airlines during the financial crisis, while Shaeri et al. [10] identified a stronger impact on airlines. Although there is limited literature on the effect of crude oil prices on aviation stocks, Kristjanpoller and Concha [11] reported a negative relationship between oil prices and aviation stocks. The researchers utilized statistical analysis techniques to examine the impact of oil on stock prices in their academic paper. They employed various datasets and models, such as three-factor and two-factor asset pricing models, GARCH model, the standard market model, the event study methodology, and CAPM with GARCH models. However, no prior research has employed LSTM (Long Short-Term Memory) based on crude oil prices to predict airline stock prices. Predicting airline stock prices using LSTM based on crude oil prices hold significant value due to the influence of oil prices on airline operational costs, industry dynamics, risk management, investor confidence, and operational optimization. Accurate predictions can assist airlines in planning and budgeting for fuel costs, making informed strategic decisions, managing risk exposure, attracting investors, and optimizing operations, potentially impacting profitability and stock prices.

As a result, this study focuses on oil prices while considering the characteristics of airlines. From this perspective, there is a need to identify an accurate method for predicting airline stock prices using machine learning algorithms. Since stock prices do not exhibit seasonality, machine learning models are applicable and beneficial. Therefore, the study has implemented a long short-term memory (LSTM) model, which is one of the popular machine learning algorithms used in previous studies. However, previous studies simply input data into the model without considering data frequency or sample size. Data with different frequencies possess distinct structures, and employing simplistic machine learning algorithms may result in errors such as overfitting due to their complex methods. Thus, in this study, we investigate whether the information concealed in the economic and technological determinants of oil can accurately predict airline stock prices.

II. Related Work

Crude oil is a significant resource, and its price change is closely linked to the stock market, which serves as an indicator of economic development [12][13][14]. While some studies suggest a positive correlation between oil and stocks, other studies also report a negative correlation. Firstly, Park and Ratti [4] emphasize that although the crude oil price hurts the overall stock market returns, it has a positive impact on the energy stock market returns. Filis and Chatziantoniou [5] also conclude that an increase in crude oil prices negatively affects the stock market of crude oil-importing countries, but positively affects the stock market of crude oil-exporting countries. Secondly, Narayan and Sharma [15] analyze the stock market prices of 560 companies across 14 industries listed on the New York Stock Exchange (NYSE), and find that the impact of oil prices on different companies depends on their scale. Bjornland [7] concludes that an increase in crude oil prices positively affects the stock prices of crude oil production companies, but negatively affects the stock prices of transportation companies.

In particular, Lu and Chen [8] conducted a study on the effects of WTI crude oil prices on the stock prices of 160 transportation companies in eight countries. The findings revealed that sea transportation companies were able to mitigate the negative impact of rising crude oil prices by increasing their crude oil transportation volume, while air transportation companies were more vulnerable to crude oil price volatility. Mohanty et al. [9] examined the impact of WTI crude oil prices on six industries in the US, including airlines, gambling, hotels, recreational services, restaurants & bars, and travel & tourism. They discovered that crude oil prices had a significant negative effect on the stock price returns of airlines, with a stronger impact observed during the 2008-2009 financial crisis. Similarly, Shaeri et al. [10] found that the impact of crude oil price risk on airlines was more pronounced compared to other industries. Despite limited research on the relationship between crude oil prices and aviation stocks, Kristjanpoller and Concha [11] demonstrated a negative association between oil prices and aviation stocks. Specifically, they analyzed the impact of fuel price fluctuations on aviation stocks associated with the International Air Transport Association (IATA) and found a positive influence of oil prices on aviation stocks.

Researchers conducted a statistical analysis methodology to investigate the impact of oil on stock prices. Mohanty and Nandha (2011) utilized a three-factor asset pricing model to analyze monthly data from July 1992 to December 2008. Mohanty et al. [9] employed a two-factor model to analyze industry returns from September 1983 to August 2011. Narayan and Sharma [15] used the GARCH Model to analyze data from January 2000 to December 2008 for 560 US firms listed on the New York Stock Exchange. Nandha and Brooks [16] analyzed the Transport Indices of thirty-eight countries and WTI Crude Oil, as well as all companies in the S&P Transportation industry index from January 1986 to July 2008, using the Standard Market Model and Event Study Methodology. Kristjanpoller and Concha [11] analyzed data from 56 airlines for the period of 2008–1 to 2013–10, employing the CAPM and GARCH Models.

However, utilizing LSTM (Long Short-Term Memory) to predict airline stock prices based on crude oil prices can minimize similarity in academic papers when considering the following points. Firstly, crude oil prices significantly impact airline operational costs, as jet fuel constitutes a significant portion of expenses. Accurate prediction of crude oil prices can enable airlines to plan and budget their fuel costs more effectively, ultimately affecting profitability and stock prices. Secondly, the airline industry is highly competitive and sensitive to external factors, including changes in crude oil prices. Reliable predictions of crude oil prices can provide airlines with valuable insights into industry dynamics, facilitating informed strategic decision-making and operational adjustments, which in turn can impact stock prices as investors closely monitor industry trends. Thirdly, fluctuating crude oil prices introduce risks to airlines' financial performance and stock prices. LSTM-based predictions of crude oil prices can aid airlines in managing risk exposure by developing risk mitigation strategies such as hedging or adjusting pricing strategies, to minimize the impact of volatile oil prices on stock prices. Fourthly, accurate predictions of airline stock prices based on crude oil prices can enhance investor confidence. Factors such as fuel costs and industry dynamics are often considered by investors when making investment decisions in the airline sector. Reliable predictions of stock prices can attract and retain investors, leading to increased market demand for airline stocks and potentially driving up stock prices. Lastly, timely and accurate predictions of crude oil prices can help airlines optimize their operations for better efficiency. To mitigate the potential impact of rising oil prices, airlines can implement proactive strategies such as optimizing flight routes, adjusting ticket prices, and negotiating fuel contracts. By employing LSTM to predict airline stock prices using crude oil prices, valuable insights can be gained into fuel costs, industry dynamics, risk management, investor confidence, and operational efficiency. These insights can inform decision-making, optimize airline operations, and potentially influence stock prices, making it a valuable tool for financial forecasting in the airline industry.

III. METHODOLOGY

A. Data Collection

This study utilized oil and American Airlines (AA) stock price data from FinanceDataReader. The stock of American Airlines represents the airline industry stock price for the following reasons: Industry leader: American Airlines is one of the prominent airlines in the United States, with a wide network of routes including international air transportation connecting the U.S. and other countries. Moreover, it has a strong competitive position in the U.S. air transportation market, making it a representative indicator of the trend in airline stock prices. Market size: The aviation industry is a significant part of the global economy, closely tied to economic activities in countries around the world. Therefore, airline stock prices can serve as a representative indicator of the market size concerning global economic conditions and expectations for economic growth. Industry trends: The aviation industry is highly sensitive to fluctuations in the economy and various factors such as oil prices, exchange rate fluctuations, government regulations, and international geopolitical situations can impact airline stock prices. The stock of American Airlines reflects these industry trends, representing the movements in airline stock prices. Investor interest: Airline stocks garner significant attention from investors due to the industry's unique characteristics, competitive landscape, technological advancements, and profitability prospects, all of which influence investors' decisions. Therefore, the stock of American Airlines can serve as a representative indicator of the overall trends in the aviation industry and economic conditions, reflecting the factors that represent airline stock prices.

In this study, oil price data for about seven years from January 4, 2016, to April 14, 2023, were collected through FinanceDataReader. The collected data is a total of 1,833 days of AA stock price data. The price data consists of six categories: Date, Open, High, Low, Close, Volume, and Change (price is based on dollars). Data is stored every 24 hours, so it was judged to be most suitable for short-term price prediction (24 hours later) to be conducted in this study. Fig. 1 shows the change in oil prices over time of the data. Fig. 2 shows the change in the stock price of AA over time of the data.

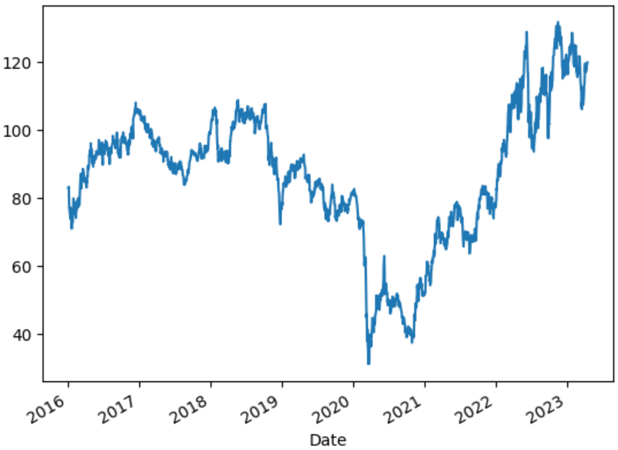

Fig. 1. Oil price distribution.

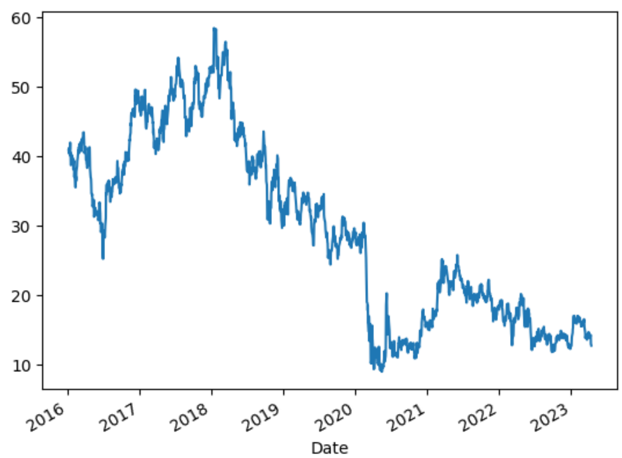

Fig. 2. AA stock price distribution.

B. Data Preprocessing

In this study, data pre-processing was performed using Python as follows. Pre-processing is a total of 4 steps, 1) converting raw data into a data frame, 2) data refinement, 3) data normalization, and 4) data division. First, the sequence length of data consisting of daily was set to 30, and it was reconstructed into a data frame matrix in units of 30 days. Second, in the data refinement process, null values or data marked as 0 among the data were replaced with the mean value to handle missing values. Third, all input values were converted into values between 0 and 1 through the data normalization process. Finally, data segmentation was performed. The entire collected Ethereum data was divided in a ratio of 7:3, and 7 parts of data (training data) were used for model learning, and 3 parts of data (test data) were used for testing. However, since the data is composed of time series, considering the order, the test data was selected as the most recent data. In addition, the hyper-parameters were divided into training and verification parts at a ratio of 8:2 to the 7 parts of the training data split for optimization.

C. Price Prediction with LSTM Model

Recurrent neural network (RNN) can reflect the sequence-related characteristics of financial time-series data, but it has the problem of gradient disappearance or gradient explosion. Also, its mining of historical information for financial time-series data is very limited. LSTM is a special RNN that can well handle the long-term dependencies of time-series data [17]. Therefore, the LSTM model is an improved RNN model, to some extent.

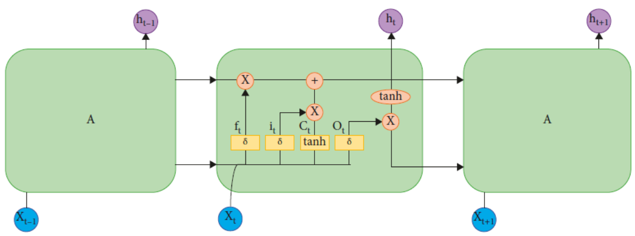

Fig. 3. LSTM network structure.

Fig. 3 shows the network structure of the LSTM. The basic unit of the LSTM model is a memory block, which includes a memory cell and three gate structures that control the state of the memory cell, forget gate, input gate, and output gate. To be specific, the forget gate decides to forget the useless historical information from the memory cell state, the input gate decides the influence of the current input data on the memory cell state, and the output gate decides the output information.

Firstly, the data that must be erased from the cell is identified based on the forget gate (ft) of the (1) as outlined below:

$$ft = \sigma(bf + Wf\,xt + Ufht - 1) \quad (1)$$

The sigmoid activation function σ is used to determine the amount of information retained. The xt refers to the current input vector, while ht represents the hidden layer vector. bt, xt, and xt are the bias, input weight, and loop weight of the forget gate, respectively, in the neural network architecture.

Subsequently, the information state is updated within the cell. The external input gate (it) is regulated by a sigmoid activation function as described in (2):

$$gt = \sigma(bg + Wg\,xt + Ught - 1) \quad (2)$$

Meanwhile, the update of the cell state (Ct) is carried out by equation (3), which utilizes Ct−1 as the input.

$$Ct = ft * Ct - 1 + gt * \tanh(bc + Wcxt + Ucht - 1) \quad (3)$$

The state of the memory cell at time t is represented by Ct.

Lastly, the output gate (Ot) of (4) regulates the information output in the following manner:

$$ht = (Ot)\tanh(Ct) \quad (4)$$

$$Ot = \sigma(bo + Wo\,xt + Uoht - 1) \quad (5)$$

The model's accuracy was assessed using Mean Squared Error (MSE), which is a common loss function for regression tasks in neural network models. MSE calculates the average of the squared differences between the predicted values and the actual values. This study chose to use MSE for several reasons. Firstly, MSE is a differentiable function, which makes it suitable for updating the model's weights using the backpropagation algorithm. Secondly, MSE accounts for the magnitude of the error, as larger errors are squared, emphasizing their impact and aiding in reducing prediction errors during model training. Thirdly, MSE is robust to outliers, making it suitable for evaluating model performance even in the presence of data outliers. The calculation of MSE was performed as shown in Equation (6), where yk represents the predicted value and tk represents the actual value.

$$MSE = \frac{1}{n}\sum_k (yk - tk)\,2 \quad (6)$$

where the variable "n" denotes the data size.

D. Train and Test the Model

In this study, after constructing the model, the hyperparameter optimization process of the LSTM model was performed to optimize performance. The primary process explored the number of neurons and the number of times of learning (Epoch). The two layers inside the LSTM model consist of n number of neurons. In this study, we experimented by changing the number of neurons to 16, 32, 64, 128, and 200. In addition, the number of times of learning was experimented by changing to 10, 30, 50, and 100. If the number of training is too small, the model does not train well, and if the number of training is too large, overfitting problems occur. In conclusion, when the number of neurons is 32 and the number of learning is 50, the verification data prediction results are the best.

As a secondary process, the optimal window size and activation function were optimized. The second process was carried out with the number of neurons and the number of learning selected in the first fixed at 32 and 50, respectively. The window size refers to the size of the previous data set used by the model in the process of learning, and the analysis was conducted while reducing it in the order of 10, 7, 5, and 3. An activation function means a function that converts an input value into an output value in the two hidden layers inside the LSTM. Recurrent activation is a commonly used time.

The sigmoid function was used as it is, and the gate function was compared and analyzed with four functions, tanh, relu, linear, and softmax. As a result of the secondary analysis, when the window size was 7 and the gate function was composed of a combination of the tanh function, the verification data prediction result was the best. Additional hyperparameter selection work was carried out through the first and second optimization processes. The dropout ratio means the ratio of randomly disconnecting some of all the lines connected between neurons to prevent overfitting and was designated as 0.25 in this study. Batch size refers to the amount that is passed to the next network after learning by dividing the entire data. The larger the size, the more computer memory is used, so 2 was designated in this study. The sequence length and output dimension were configured in units of 30 days. And the optimizer used Adam, and the loss function used MSE (Mean Squared Error).

IV. RESULTS

In this paper, normalized closing price data was trained for 50 epochs. As a result of learning, the loss value converged close to 0 as shown in Fig. 4.

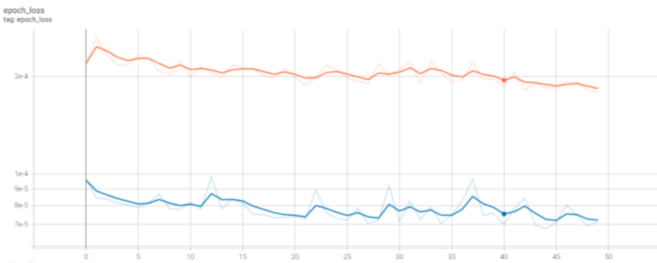

Fig. 4. Learning loss value.

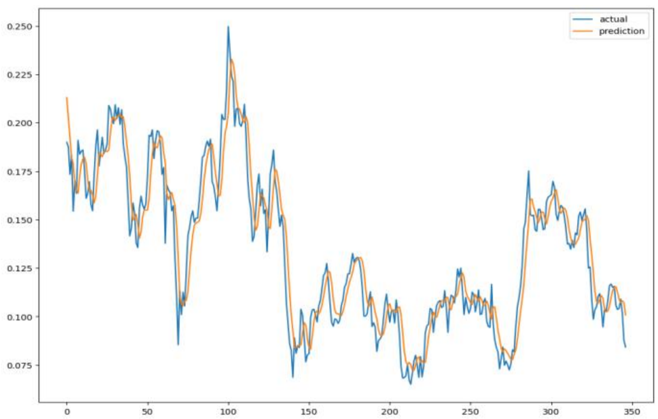

Fig. 5. Prediction results.

There was a case where the validation-set loss soared temporarily due to a sudden rise or fall in the stock price, but it generally converged to 0. The results of testing with test data using the trained model are shown in Fig. 5. Inverse normalization was performed on the test results before output. The MSE measured by the accuracy of the model shows a result of 0.0004. As a result of testing with the trained model, it was confirmed that the trend of the actual value was followed with high accuracy, as shown in Fig. 5.

V. CONCLUSION

The selection of influential factors for stock price prediction using LSTM is crucial as it allows for capturing relevant information that can impact stock prices. Economic indicators, financial statements, market sentiment, and news are significant factors that can affect stock prices. Including these relevant factors in the LSTM model enhances its ability to capture complex relationships and patterns in the data, leading to more accurate stock price predictions. Crude oil prices have been found to affect stock markets, but studies show mixed associations with some reporting positive and others negative effects. Previous studies have used statistical analysis methods to study the impact of oil prices on stock prices, but there is a gap in research predicting airline stock prices using LSTM based on crude oil prices. Accurate predictions can help airlines plan and budget fuel costs, make strategic decisions, manage risk exposure, and optimize operations. Therefore, this study focuses on accurately predicting airline stock prices using machine learning algorithms, particularly LSTM, considering the impact of oil prices on airline operations and industry dynamics. Different frequencies of data have different structures, and simple copy machine learning algorithms may have errors such as overfitting. Hence, this study investigates whether the information hidden in the economic and technological determinants of oil can accurately predict airline stock prices using LSTM. Based on the necessity of this study, this study developed an LSTM model that predicts the price of airline stock prices through oil prices. As a result of learning, the loss value converged close to zero. The MSE value for AA stock closing price predictions resulted in 0.00049.

The significance of this study is as follows. First, it is meaningful in that it can present indicators such as more sophisticated predictions and risk management to airline companies. In addition, many researchers are attempting to develop new algorithms for predicting airline stock prices, and studies have been conducted with various types of statistical models. However, in this study, we developed a model with improved performance by constructing optimal hyperparameters based on LSTM, an existing algorithm. Oil price as our selected feature can compensate for the poor performance of a simple model and its limitations on overfitting. This result is meaningful in that the prediction accuracy of this study is superior to the results of previous studies that attempted to predict airline stock prices based on a simple machine-learning model.

REFERENCES

[1] J. Kaur, and K. Dharni, "Data mining–based stock price prediction using hybridization of technical and fundamental analysis", Data Technologies and Applications, Vol. ahead-of-print No. ahead-of-print., 2023, https://doi.org/10.1108/DTA-04-2022-0142.

[2] Q. Jia, Y. Zhu, R. Xu, Y. Zhang, and Y. Zhao, "Making the hospital smart: using a deep long short-term memory model to predict hospital performance metrics", Industrial Management & Data Systems, vol. 122 no. 10, 2022, pp. 2151-2174.

[3] J. Wang, Q. Cheng, and Y. Dong, "An XGBoost-based multivariate deep learning framework for stock index futures price forecasting", Kybernetes, Vol. ahead-of-print No. ahead-of-print.,2022, https://doi.org/10.1108/K-12-2021-1289.

[4] J. W. Park, and R. A. Ratti, "Oil price shocks and stock markets in the U.S. and 13 European countries," Energy Economics vol. 30, no. 5, 2008, pp. 2587–2608.

[5] G. Filis, and I., Chatziantoniou, "Financial and monetary policy responses to oil price shocks: evidence from oil-importing and oil-exporting countries," Review of Quantitative Finance and Accounting, vol. 42, no. 4, 2014, pp. 709–729.

[6] P. K. Narayan, and S. S. Sharma, "New evidence on oil price and firm returns," Journal of Banking and Finance, vol. 35, no. 12, 2011, pp. 3253–3262.

[7] C. H. Bjornland, "Oil price shocks and stock market booms in an oil exporting country," Scottish Journal of Political Economy, vol. 2, no. 5, 2009, pp. 232–254.

[8] J. R. Lu, and C. C. Chen, "Effect of oil price risk on systematic risk from transportation services industry evidence," The Service Industries Journal, vol. 30, no. 11, 2010, pp. 1853–1870.

[9] S. Mohanty, M. Nandha, E. Habis, and E. Juhabi, "Oil price risk exposure: the case of the U.S.," Travel and Leisure Industry, Energy Economy, vol. 41, 2014, pp. 117–124.

[10] K. Shaeri, C. Adaoglu, and S. T. Katircioglu, "Oil price risk exposure: a comparison of financial and non-financial subsectors," Energy, vol. 109, 2016, 712–723.

[11] D. W. Kristjanpoller, and D. Concha, "Impact of fuel price fluctuations on airline stock returns," Applied Energy, vol. 178, 2016, pp. 496–504.

[12] B. N. Huang, M. J. Hwang, and H. P. Peng, "The asymmetry of the impact of oil price shocks on economic activities: an application of the multivariate threshold model," Energy Economics, vol. 27, no. 3, 2005, pp. 455–476.

[13] S. A. Basher, and P. Sadorsky, "Oil price risk and emerging stock markets," Global Finance Journal, vol. 17, no. 2, 2006, pp. 224–251.

[14] J. I. Miller, and R. A. Ratti, "Crude oil and shock markets: stability, instability, and bubbles," Energy Economics, vol. 31, no. 4, 2009, pp. 559–568.

[15] P. K. Narayan, and S. S. Sharma, "Firm return volatility and economic gains: the role of oil prices," Economic Modelling, vol. 38, 2014, pp. 142–151.

[16] M. Nandha, and R. Brooks, Oil prices and transport sector returns an international analysis, Quant. Finance Account, vol. 33, no. 4, 2009, pp. 393–409.

[17] H. Ouyang, K. Huang, and H. Yan, "Prediction of financial time series based on LSTM neural network," Chinese Journal of Management Science," vol. 28, no. 4, 2020, pp. 27–35.