Online-Journals.org

iTDAF IETI Transactions on
Data Analysis and Forecasting

iTDAF | eISSN: 2959-0442 | Vol. 1 No. 1 (2023) | OPEN ACCESS

https://doi.org/10.3991/itdaf.v1i1.32683

PAPER

Does the Sentiment Index Help Predict Crude Oil Prices?

Jin Shang, Tamotsu Nakamura, Shigeyuki Hamori(✉)

Graduate School of Economics, Kobe University, Kobe, Japan

hamori@econ.kobe-u.ac.jp

ABSTRACT

The price fluctuations in the crude oil market remarkably influence the global economy since crude oil is an essential source of energy and plays a determinant role in most industrial sectors. The tremendous development of social media has generated many applications that can be used for sentiment analysis to improve the prediction of crude oil prices. Many researchers have also used technical indicators to predict oil prices. This study integrated several machine learning approaches—random forest, support vector machine, and long short-term memory—with a dynamic expanding moving window and fixed moving window to forecast one-period-ahead West Texas Intermediate (WTI) spot prices. We assessed the forecasting performance of these models using the root mean squared error and then compared prediction accuracy among the sentiment indicator, the technical indicator, and the lagged values of WTI spot prices using the Diebold–Mariano test. The forecasting simulation and empirical results show that the sentiment indicator outperforms the technical indicator and lagged prices data set when predicting WTI spot prices using machine learning methods. In addition, this work examined that using the sentiment indicator provides better prediction performance than using the benchmark time-series analysis model ARIMA.

KEYWORDS

crude oil price, machine learning, sentiment index, random forest, support vector machine, long short-term memory

1 INTRODUCTION

The price fluctuations of important energy resources, which are fundamental inputs for many production and consumption activities, have an enormous influence on the global economy. Appropriate responses to energy price fluctuations are critical to the world economy, especially for important limited energy resources, such as crude oil. Numerous researchers [10], [11], [14], [15] have been studying crude oil for many years, and they have found that fluctuations in crude oil prices have a significant impact on macroeconomics.

Shang, J., Nakamura, T., Hamori, S. (2023). Does the Sentiment Index Help Predict Crude Oil Prices? *IETI Transactions on Data Analysis and Forecasting (iTDAF)*, 1(1), pp. 54–65. https://doi.org/10.3991/itdaf.v1i1.32683

Article submitted 2022-05-24. Resubmitted 2022-06-10. Final acceptance 2022-06-11. Final version published as submitted by the authors.

Many researchers have attempted to predict crude oil prices using diverse approaches [7], [18], [20], [27], [28], [29]. Rapidly developed machine learning approaches have attracted significant interest and have been applied to time-series forecasting. Wang & Wang [25] utilized the recurrent neural network (RNN) to predict crude oil prices. Luo et al. [17] proposed a new approach based on convolutional neural network models and showed it to be fairly accurate in estimating short-term crude oil futures prices. Many empirical results [6], [19], [24], [26] have shown the outstanding efficiency and accuracy of the machine-learning-based forecasting model. For instance, Moshiri and Foroutan [19] compared the linear models (the generalized autoregressive conditional heteroscedasticity models and autoregressive moving average models) to the nonlinear neural network models and found that the neural network models are superior to the linear models as they can forecast more precisely.

Amid the rapid development of Internet technology, social media have grown significantly, with networking services, users, and digital data rapidly increasing. As a result, big data obtained through social media can be integrated into the decision-making process. Such big data are deemed to be good sources of real-time information because of their high frequency and low acquisition costs. Therefore, it is now possible to utilize machine-learning approaches and big data obtained from social media to achieve breakthroughs in the time-series analysis of the energy market. Specifically, sentiment analysis can computationally recognize and classify opinions from collected text and then identify the polarity of the emotions toward a particular topic, which can be assessed as positive, neutral, or negative.

Concerning sentiment analysis, many researchers have used Twitter tweets to analyze the sentiments of the public during a specified time range because Twitter data are easily accessible and tremendous volumes of tweets are available for any specific day. For instance, one of the most famous studies [1] on stock market prediction used Twitter to conduct a sentiment analysis of the public emotion. The authors used Google Profile of the Mood States and Opinion Finder to analyze the tweets collected through the Twitter API. Moreover, they classified public opinion on Twitter into six different moods, such as happiness and anger. Deeney et al. [6] revealed that the sentiment indices developed for crude oil affected both WTI and Brent futures prices from 2002 to 2013. The empirical results in Li et al. [16] indicated the significant predictive advantage of sentiment analysis in forecasting oil price trends using news release data. Furthermore, online data, including news releases and social media networks such as Twitter could help to forecast oil price trends as the Internet and big data technologies continue to develop rapidly [16].

Meanwhile, many investors forecast the movement direction of prices by the technical indicator and some basic time-series analysis models such as the Autoregressive Integrated Moving Average (ARIMA) model. However, few prior studies have examined whether models using sentiment analysis data sets can outperform those using technical indicators and basic time-series analysis models.

Therefore, the first objective of this study is to compare the forecasting performance of the sentiment indicator (SI) obtained by sentiment analysis and the technical indicators using the machine learning approach. Second, this work also aims to verify whether the predictive power of the sentiment indicator data set can beat the basic time-series analysis model, i.e., ARIMA. Third, given that, the lagged values of explanatory variables are deemed to be highly correlated with price crashes according to the variable importance [5], we explored the prediction power of the lagged values of WTI spot prices and compared their forecasting performance to that of the sentiment indicator and the technical indicator.

In this study, we integrated the machine learning approaches with the dynamic expanding moving window (EMW) and fixed moving window (FMW) to forecast West Texas Intermediate (WTI) spot prices. Specifically, this study applied the RF, SVM, and LSTM approaches and utilized the dynamic techniques to obtain the dynamic changing parameter of each model to predict the crude oil prices of one-period-ahead (because the sentiment is more of real-time changing, we set the horizon of forecasting as one to check the predictive power of the sentiment indicator). Therein, we used some parameter-tuning techniques (e.g., grid-search) to select the best hyper-parameters for each model. Then, the statistically significant difference in predictive accuracies of different datasets (namely, lagged prices, technical indicators, sentiment indicators) are evaluated by the modified Diebold–Mariano (DM) test.

The main findings of this investigation are as follows. First, employing the technical indicator did not improve prediction performance relative to using lagged prices. Second, utilizing the sentiment indicator displayed better predictive performance than using lagged prices and was also better than utilizing the technical indicator in some cases. In addition, our results also show that machine learning models using sentiment index data sets have superior performance in crude oil price prediction than the fundamental time series analysis models ARIMA.

To the best of our knowledge, regarding crude oil price forecasting, this study is the first to compare the predictive performance of the sentiment indicators to that of technical indicators and lagged prices. Our results demonstrate the superior predictive power of the sentiment indicator, which should provide important new insights to investors.

The remainder of this paper is organized as follows. Section 2 provides a detailed description of the data, the methodologies, and the model evaluation measures applied in this paper. Section 3 presents the empirical results. Then, we analyze these results and evaluate the forecasting performance of the different datasets with different machine learning approaches and investigate the validity of the sentiment index, technical indicators, and lagged prices. Finally, Section 4 concludes this study.

2 DATA AND METHODOLOGY

2.1 Data

We collected WTI spot price data from Bloomberg, covering the period from January 6, 2019, to December 27, 2020. After cleaning the data, we obtained 614 daily data points. Then, based on the WTI spot prices, we calculated ten popular technical indicators, including 7-period and 21-period moving average, exponential moving average, 26-day exponential weighted moving average, 12-day exponential weighted moving average, 20-day standard deviation, Bollinger bands (upper and lower), moving average convergence divergence (MACD), and the relative strength index (RSI). We then utilized the Daily News Sentiment Index, which is collected from the Federal Reserve Bank of San Francisco, as the sentiment indicator. The Daily News Sentiment Index, calculated using the methodology developed by Shapiro et al. [22] and introduced by Buckman et al. [4], is a measure of economic sentiment based on economics-related news articles from 16 major U.S. newspapers.

We built three datasets to perform forecasting; as we sought to compare the predictive performance between the sentiment indicator, the technical indicator, and the lagged values, we set 3 different datasets to perform forecasting. Detailed information on the variables is provided in Table 1.

Table 1. Datasets used to predict WTI spot prices

	Containing Variables	Number of Variables
Dataset-SI	Today's price + Sentiment indicator	2
Dataset-TI	Today's price + Technical indicators	11
Dataset-lag	Today's price + lag1~lag5*	6

Notes: *Lagged prices from one to five periods; "lag1" denotes the one-period lagged prices of crude oil; "lag2" denotes the two-period lagged prices of crude oil; "lag3" denotes the three-period lagged prices of crude oil; "lag4" denotes the four-period lagged prices of crude oil; "lag5" denotes the five-period lagged prices of crude oil.

2.2 Methodology

This study applied the RF, SVM, and LSTM approaches combined with the EMW and FMW to estimate the optimal parameters of these models. The study also used trained models with dynamic changing parameters to predict one-period-ahead crude oil prices and evaluated the prediction performance of these models with the remaining test datasets. In the following, we will show the combination of methods in the form of an underlined (_) link between the machine learning model abbreviation (i.e., RF, SVM, and LSTM) and the type of abbreviation of the moving window (i.e., EMW and FMW), e.g., the RF_EMW represents the random forest model combining with the expanding moving window technique.

Random forest (RF). The RF methodology [3] is an ensemble machine learning technique. The RF algorithm combines multiple decision trees to facilitate forecasting performance. This methodology can prevent the overfitting problem when more trees are added to the forest, and it can improve the prediction performance because each tree is grown from the primal sample through bootstrap resampling and each tree is extended from the randomly selected feature.

Support vector machine (SVM). The SVM approach was proposed by Vapnik and Lerner [23]. Boser et al. [2] proposed a creative approach for generating nonlinear classifiers utilizing kernel functions to obtain maximum-margin hyperplanes. The fundamental principle of SVM regression is to minimize the error between the predicted value and actual value and find the hyperplane that maximizes the margin (distance) between two hyperplanes (decision boundary) within the tolerance of margin and make the hyperplane as flat as possible. The parameters for the regularization, margin, and tolerance of margin are determined using the grid search method.

Long short-term memory (LSTM). The LSTM algorithm was introduced by Hochreiter and Schmidhuber [13]. As a representative deep learning model, LSTM has an external loop structure similar to that of the RNN, as well as an internal circulation structure in its characteristic memory cells. Three types of gates are associated with self-connected recurrent weights in each memory cell to ensure that the signal can be transferred through several time steps to avoid gradient explosion or gradient descent. As an extension of the RNN, additional pieces of information can be used, which is similar to a memory in the LSTM unit for each time step.

These gates ensure that the network determines the remembering parts of the network in the next iteration as well as forgetting parts of it.

We adopted the grid search method to tune the batch size and the number of epochs, the optimization algorithm, learning rate, neuron activation function, and the number of neurons in the hidden layer. As a result, we defined the LSTM using 50 neurons in the first hidden layer and one neuron in the output layer. We set the input shape as a one-time step using the variables listed in Table 1. We employed the mean squared error (MSE) loss function, and the networks were trained using Adam's adaptive stochastic gradient descent optimizer.

Expanding moving window (EMW) and fixed moving window (FMW). This study used two patterns of moving window techniques for predicting one-period-ahead, the EMW and FMW, to investigate whether a difference in prediction performance exists when historical data are excluded.

The moving window statistics proceed iteratively with the prediction while the EMW or FMW size is extended or shifted by a one-time step in every iteration.

With respect to EMW, initially, the first window size was set to 307, the same as the validation data length (there are 307 observations from January 1, 2020, to December 27, 2020); when iterating the model fitting, the window size would add one period. For example, the first window was taken from January 6, 2019, to December 31, 2019, and was used to estimate January 2, 2020. Therefore, the framework utilized the dataset from period 1 to 307 to train the model, used the trained model to forecast period 308, then utilized the extended training dataset from period 1 to 308 to train the model again, and used the updated model to predict period 309. This process was iterated until the last period of the time series. The expanding-length window analysis was run 307 times for each model.

In terms of the FMW, the window size was determined to be 307. For instance, the first window, from January 6, 2019, to December 31, 2019, was used to estimate January 2, 2020. Therefore, the model used the dataset from period 1 to 307 to train the model and used this trained model to forecast period 308 and used the dataset from period 2 to 309 to train the model and utilized the updated model to predict period 310. This process was iterated until the last period of the time series. In total, the FMW analysis was executed 307 times for each model.

2.3 Model evaluation measures

Root mean squared error (RMSE). The RMSE is frequently used as a measure of the differences between the values (sample or population) predicted and the actual values observed. Generally, the RMSE is calculated as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \widehat{x}_i)^2}{N}} \tag{1}$$

where N is the number of non-missing data points, x_i is the actual observation time series, and $\widehat{x}_i$ is the estimated time series.

Modified Diebold–Mariano test. The DM-test was originally introduced by Diebold and Mariano [8]. In empirical analyses, when there are two or more time series forecasting models, it is often challenging to predict which model is more accurate or whether they are equally suitable. This test determines whether the null hypothesis (i.e., that the competing model's forecasting power is equivalent to that of the base model) is statistically true. We assume the actual values $\{y_t; t = 1, \ldots T\}$; two forecasts $\{\hat{y}_{1t}; t = 1, \ldots T\}$, $\{\hat{y}_{2t}; t = 1, \ldots T\}$; and forecast error ε_{it}, as follows:

$$\varepsilon_{it} = \hat{y}_{it} - y, i = 1, 2 \tag{2}$$

where ε_{it} denotes the forecast error and the loss function, $g(\varepsilon_{it})$, which is defined by the following function:

$$g(\varepsilon_{it}) = (\varepsilon_{it})^2 \tag{3}$$

Then, the loss differential, d_t, is expressed as follows:

$$d_t = g(\varepsilon_{1t}) - g(\varepsilon_{2t}) \tag{4}$$

Correspondingly, the statistic for the DM-test is expressed using the following formula:

$$DM = \frac{\bar{d}}{\sqrt{s/N}} \tag{5}$$

where $\bar{d}, s, and\, N$ denote the mean loss differential, the variation of d_t, and the number of data points, respectively.

The null hypothesis is $H_0 : \mathbb{E}\left[d_t\right] = 0, \forall t$, meaning that the two forecast models have equivalent forecasting performance. The alternative hypothesis is $H_1 : \mathbb{E}\left[d_t\right] \neq 0, \forall t$, representing the difference in accuracy between these two forecasts. Under the null hypothesis, the statistics for the DM-test are asymptotically $N(0,1)$ normally distributed. The null hypothesis is rejected when $\left\lceil DM \right\rceil > 1.96$.

Harvey et al. [12] proposed a modified DM-test, which they suggested was more suitable for small samples. The statistics for the modified DM-test are expressed as follows:

$$DM^* = \sqrt{\left[n + 1 - 2h + h(h-1)\right]n^{-1}}\, DM \tag{6}$$

where h represents the horizon, and DM refers to the original DM statistic. Here, we predicted one-period-ahead; hence, $h = 1$. Thus,

$$DM^* = \sqrt{(n-1)n^{-1}}\, DM \tag{7}$$

3 EMPIRICAL RESULTS

3.1 Prediction results

The results shown in Table 2 indicate that the lowest RMSE appears in the LSTM model with the fixed moving window. For most of the simulations, the EMW displays better performance than the FMW at the RMSE results level, which indicates that the past values (up to one year before) of the selected variables correlated with current prices. From the vertical perspective, the LSTMs generally demonstrate better predictions than the SVMs, and the SVMs are superior to the RFs. From the horizontal perspective, except for SVM_FMW, the dataset-SI surpasses the dataset-TI, and, except for RF_EMW and RF_FMW, utilizing the dataset-TI did not produce better predictions than utilizing the dataset lag.

Table 2. Results of three different datasets

	Dataset-Lag RMSE	Dataset TI RMSE	Dataset-SI RMSE
RF_EMW	2.2020	2.1470	1.9614
RF_FMW	2.2906	2.2129	2.1226
SVM_EMW	1.7329	1.8044	1.5054
SVM_FMW	1.8027	1.9044	1.5712
LSTM_EMW	1.5942	1.7314	1.4850
LSTM_FMW	1.6412	1.9761	1.4843

Notes: RF_EMW denotes the random forest model with expanding moving window; RF_FMW denotes the random forest model with fixed moving window; SVM_EMW denotes the support vector machine model with expanding moving window; SVM_FMW denotes the support vector machine model with fixed moving window; LSTM_EMW denotes the long short-term memory model with expanding moving window; LSTM_FMW denotes the long short-term memory model with fixed moving window.

These results suggest that the predictive power of the sentiment index is greater than that of the technical indicator and the lagged prices, providing additional supporting evidence for the assertion of Ni et al. [21] that the price fluctuation is more sensitive to the intraday sentiment. In addition, from the longitudinal results in Table 2, the RMSE results of the LSTM in general demonstrate excellent predictive capability, which is consistent with the findings of the previous studies conducted by Luo et al. [17] and Wang & Wang [25] and provides further empirical evidence from another perspective. LSTM is based on the recurrent normal neural network (RNN), which is a type of neural network for processing sequentially varying data. LSTM is superior in time series prediction since it can store and update information through some gates (e.g., sigmoid function and pointwise multiplication) to regulate the flow of information to decide which to forget and which to remember, solving the problems of gradient vanishing or gradient explosion and missing important information short-term memory, which are inherent to RNNs. On the contrary, the predictive accuracy of random forest is somewhat weaker than other models because random forest models ignore two important features when implementing predictions—namely, the internal time trend and the interdependence between variables.

3.2 Diebold–Mariano test results

Tables 3 and 4 show the Diebold–Mariano test results of comparing the predictive accuracy of the different datasets with different models.

Table 3. DM-test results of dataset-TI

Dataset-TI	vs. Dataset-Lag	
	DM-Test[1]	*p* Value
RF_EMW	–0.3506	0.726
RF_FMW	–0.4601	0.646
SVM_EMW	1.5578	0.120
SVM_FMW	1.8163	0.070*
LSTM_EMW	2.7812	0.006***
LSTM_FMW	2.8267	0.005***

[1]DM test indicates the modified Diebold–Mariano test statistic.
Notes: RF_EMW denotes the random forest model with expanding moving window; RF_FMW denotes the random forest model with fixed moving window; SVM_EMW denotes the support vector machine model with expanding moving window; SVM_FMW denotes the support vector machine model with fixed moving window; LSTM_EMW denotes the long short-term memory model with expanding moving window; LSTM_FMW denotes the long short-term memory model with fixed moving window. "***", "**" and "*" represent statistical significance at 1%, 5%, and 10%, respectively.

Table 4. DM test results of dataset SI

Dataset-SI	vs. Dataset-Lag		vs. Dataset-TI	
	DM Test[1]	*p* Value	DM Test[1]	*p* Value
RF_EMW	–1.6588	0.098*	–1.5567	0.121
RF_FMW	–1.0163	0.310	–0.6291	0.530
SVM_EMW	–2.2551	0.025**	–3.2246	0.001***
SVM_FMW	–1.8048	0.072*	–2.5691	0.011**
LSTM_EMW	–3.7211	0.000***	–4.5023	0.000***
LSTM_FMW	–2.8264	0.005***	–4.1962	0.000***

[1]DM-test indicates the modified Diebold–Mariano test statistic.
Notes: RF_EMW denotes the random forest model with expanding moving window; RF_FMW denotes the random forest model with fixed moving window; SVM_EMW denotes the support vector machine model with expanding moving window; SVM_FMW denotes the support vector machine model with fixed moving window; LSTM_EMW denotes the long short-term memory model with expanding moving window; LSTM_FMW denotes the long short-term memory model with fixed moving window. "***", "**" and "*" represent statistical significance at 1%, 5%, and 10%, respectively.

Tables 3 and 4 present the DM-test results of comparing the predictive accuracy of different datasets with different models. We set $g(\varepsilon_{1t})$ as the target model, which is the model before "vs.", and set $g(\varepsilon_{2t})$ as the contrast model, which is the model after "vs.". Thus, the numerator is (target-base). Therefore, if the DM test statistic is negative, that means the target model has a smaller variance than the base model, and its prediction performance is better than that of the base model.

First, comparing the dataset-TI with the dataset-lag (see Table 3) shows that the DM-test statistic results are negative only for RF_EMW and RF_FMW, which suggests that the dataset-TI is not better than the dataset-lag, verifying the forecasting results displayed in Table 2. Nevertheless, concerning the *p* values of the DM-test statistic, the significant statistical difference between the dataset-TI and the dataset-lag holds only for SVM_FMW, LSTM_EMW, and LSTM_FMW, which indicates that the lagged values of the prices are preferable to the technical indicator when predicting WTI spot prices utilizing SVM_FMW, LSTM_EMW, and LSTM_FMW.

Second, the empirical results for dataset-SI are as follows. Comparing the dataset-SI with dataset-lag (see Table 3) shows that the DM-test statistic results are all negative, which suggests that the prediction performance of dataset-SI is superior to that of dataset-lag, which provides additional evidence in support of the forecasting simulation results shown in Table 2. Furthermore, the *p* values of the DM-test statistic (except for the RF with the FMW model) show that the differences in the predictive accuracy of these two sets of forecasts have obvious statistical significance, which implies that the sentiment indicator is more effective and has the greater predictive ability for predicting WTI sport prices than past prices from one up to five periods. Comparing the dataset-SI with the dataset-TI (see Table 4) shows that the DM-test statistic results are all negative, which suggests that the dataset-SI outperformed the dataset-TI, which is consistent with the prediction results demonstrated in Table 2. Notwithstanding, regarding the *p* values of the DM-test statistic, the significant statistical difference between the dataset-SI and the dataset-TI makes sense of SVM_EMW, SVM_FMW, LSTM_EMW, and LSTM_FMW, which indicates that the sentiment indicator is preferable to the technical indicator when predicting WTI spot prices utilizing SVM_EMW, SVM_FMW, and LSTM_EMW.

Table 5. DM-test results of dataset-SI comparing with ARIMA

Dataset-SI	vs. ARIMA	
	DM-Test[1]	*P* Value
RF_EMW	–1.8519	0.0650*
RF_FMW	–0.3009	0.7637
SVM_EMW	–3.0842	0.0022***
SVM_FMW	–2.8036	0.0054***
LSTM_EMW	–3.1696	0.0017***
LSTM_FMW	–3.1464	0.0018***

[1]DM-test indicates the modified Diebold–Mariano test statistic.
Notes: RF_EMW denotes the random forest model with expanding moving window; RF_FMW denotes the random forest model with fixed moving window; SVM_EMW denotes the support vector machine model with expanding moving window; SVM_FMW denotes the support vector machine model with fixed moving window; LSTM_EMW denotes the long short-term memory model with expanding moving window; LSTM_FMW denotes the long short-term memory model with fixed moving window. "***", "**" and "*" represent statistical significance at 1%, 5%, and 10%, respectively.

Table 5 shows the results of the machine learning predictions compared with those of the traditional ARIMA model. As is clear from Table 5, the predictive power of the sentiment index data set combined with the machine learning models is, by and large, remarkably superior to the basic time series analysis model ARIMA, and the predictive advantage of the SI data set is statistically significant, as indicated by the *p* value results of the DM.

To summarize, the dataset utilizing the sentiment indicator outperformed the dataset utilizing the lagged price values in forecasting WTI spot prices, and this was verified to have statistical significance, except for RF_FMW. At the same time, the sentiment indicator was also found to be preferable to the technical indicator when

SVM_EMW, SVM_FMW, and LSTM_EMW are being applied. On the other hand, the empirical results also provided evidence that the neural network model generally outperformed the machine learning models. Lastly, the Sentiment Index data set is generally significantly better than the benchmark model, ARIMA.

4 CONCLUSIONS

This study used the sentiment indicator, the technical indicator, and lagged values of WTI spot prices and utilized integrated machine learning approaches incorporating dynamic EMW and FMW techniques to forecast WTI spot prices. Specifically, this study applied the dynamic integrated RF, SVM, and LSTM approaches to predict the one-period-ahead oil prices and evaluate the forecasting performance via RMSE. We also used the modified DM-test statistic to investigate the statistically significant differences among these three different datasets.

As proposed by Gupta and Pandey [9], we adopted the LSTM approach to predict crude oil prices, but our approach used the sentiment index combined with the moving window techniques. As a result, our result showed superior forecasting performance. In addition, our study utilized the same predictor variable as the sentiment indicator. This result supported the findings reported by Li et al. [16], which used sentiment analysis to predict the trends of the crude oil prices. According to our results, it must be pointed out that as an extension of Li et al.'s study [16], the sentiment indicator also facilitates predicting the level of the crude oil prices.

Moreover, we compared the difference in prediction performance between utilizing the sentiment indicator and utilizing the technical indicators.

The main findings of this study are as follows. First, the optimal prediction performance was observed in the LSTM_FMW approaches and in the dataset, including the sentiment indicator. Second, for most of our prediction simulations, the EMW demonstrated lower RMSE results than the FMW. Third, utilizing the sentiment indicator led to better forecasting than utilizing the technical indicators, as well as using lagged prices; the obvious statistically significant difference has been tested and verified by the modified DM-test in most of the cases. Lastly, the machine learning models using the Sentiment Index data set acquired by sentiment analysis outperformed the benchmark model, ARIMA.

The contributions of this study are summarized as follows. First, this study is the first to compare the different predictive performances between datasets (i.e., sentiment indicator dataset, technical indicator dataset, and lagged prices values dataset) when predicting WTI spot prices; the results confirm the superiority of the sentiment indicator. This result could provide guidance for investors involved in the commodities market. Second, this study is the first to combine the LSTM model with the dynamic moving window technique to predict crude oil prices. The empirical results show that LSTM_EMW and LSTM_FMW outperformed the other approaches in most cases. This study's combining of the LSTM with the moving window technique and its selection of the sentiment analysis indicator provide news insights that could help improve the prediction accuracy of crude oil spot prices and thus help investors.

Despite these findings, this study has limitations. For instance, the time range of its dataset is two years, which is usually considered somewhat short for feeding the machine learning models.

In future research, first, we intend to extend this study's time span to prove the rationality and robustness of its finding on the sentiment indicator's superiority.

Second, we plan to empirically investigate the difference between the tuned dynamic changing parameter models and unturned dynamic changing parameter models. Finally, it would be fruitful to apply this study's approaches to Brent crude oil prices and to increase the forecasting horizon.

5 ACKNOWLEDGMENT

We are grateful to two anonymous reviewers for their helpful comments and suggestions. This research was supported by JSPS KAKENHI Grant Number 22K01424.

6 REFERENCES

[1] Bollen J, Mao H, Zeng X (2011) Twitter mood predicts the stock market. J Comput Sci 2:1–8. https://doi.org/10.1016/j.jocs.2010.12.007
[2] Boser BE, Guyon IM, Vapnik VN (1992) A training algorithm for optimal margin classifiers. ACM 5:144–52. https://doi.org/10.1145/130385.130401
[3] Breiman L (2001) Random forests. Mach Learn 45:5–32. https://doi.org/10.1023/A:1010933404324
[4] Buckman SR, Shapiro AH, Sudhof M, Wilson DJ (2020) News sentiment in the time of COVID-19. Federal Reserve Bank of San Francisco 08:1–05. https://doi.org/10.24148/wp2020-24
[5] Chatzis PS, Siakoulis V, Petropoulos A, Stavroulakis E, Vlachogiannakis N (2018) Forecasting stock market crisis events using deep and statistical machine learning techniques. Expert Syst Appl 112:353–371. https://doi.org/10.1016/j.eswa.2018.06.032
[6] Deeney P, Cummins M, Dowling M, Bermingham A (2015) Sentiment in oil markets. Int Rev Financ Anal 39:79–185. https://doi.org/10.1016/j.irfa.2015.01.005
[7] De Souza e Silva A, Sutko D (2011) Theorizing locative technologies through philosophies of the virtual. Commun Theory 21:23–42. https://doi.org/10.1111/j.1468-2885.2010.01374.x
[8] Diebold FX, Mariano RS (1995) Comparing predictive accuracy. J Bus Econ Stat 13: 253–63. https://doi.org/10.2307/1392185
[9] Gupta V, Pandey A (2018) Crude oil price prediction using LSTM networks. International Journal of Computer and Information Engineering 12:226–230.
[10] Hamilton JD (1983) Oil and the macroeconomy since World War II. J Political Econ 91: 228–248. https://doi.org/10.1086/261140
[11] Hamilton JD (1996) This is what happened to the oil price–macroeconomy relationship. J Monet Econ 38:215–220. https://doi.org/10.1016/S0304-3932(96)01282-2
[12] Harvey D, Leybourne S, Newbold P (1997) Testing the equality of prediction mean squared errors. Int J Forecasting 13:281–91. https://doi.org/10.1016/S0169-2070(96)00719-4
[13] Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9: 173–17805. https://doi.org/10.1162/neco.1997.9.8.1735
[14] Hooker MA (1996) What happened to the oil price–macroeconomy relationship? J Monet Econ 38:195–213. https://doi.org/10.1016/S0304-3932(96)01281-0
[15] Kilian L (2009) Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market. Am Econ Rev 99:1053–1069. https://doi.org/10.1257/aer.99.3.1053
[16] Li J, Xu Z, Yu L, Tang L (2016) Forecasting oil price trends with sentiment of online news articles. Procedia Comput Sci 91:1081–1087. https://doi.org/10.1016/j.procs.2016.07.157

[17] Luo Z, Cai XJ, Tanaka K, Takiguchi T, Kinkyo T, Hamori S (2019) Can we forecast daily oil futures prices? Experimental evidence from convolutional neural networks. Journal of Risk and Financial Management 12:1–13. https://doi.org/10.3390/jrfm12010009
[18] Merino A, Albacete R (2010) Econometric modelling for short-term oil price forecasting. OPEC Energy Rev 34:25–41. https://doi.org/10.1111/j.1753-0237.2010.00171.x
[19] Moshiri S, Foroutan F (2006) Forecasting nonlinear crude oil futures prices. Energy J 27:81–95. https://doi.org/10.5547/ISSN0195-6574-EJ-Vol27-No4-4
[20] Naser H (2016) Estimating and forecasting the real prices of crude oil: A data rich model using a dynamic model averaging (DMA) approach. Energy Econ 56:75–87. https://doi.org/10.1016/j.eneco.2016.02.017
[21] Ni Y, Su ZY, Wang WR, Ying YH (2019) A novel stock evaluation index based on public opinion analysis. 2018 International Conference on Identification, Information and Knowledge in the Internet of Things, IIKI 2018. Procedia Computer Science 147: 581–587. https://doi.org/10.1016/j.procs.2019.01.212
[22] Shapiro AH, Sudhof M, Wilson DJ (2020) Measuring news sentiment. Federal Reserve Bank of San Francisco Working Paper 2017-01. https://doi.org/10.24148/wp2017-01
[23] Vapnik V, Lerner A (1963) Pattern recognition using generalized portrait method. Autom Remote Control 24:774–80.
[24] Wang D, Huang J (2005) Neural network-based adaptive dynamic surface control for a class of uncertain nonlinear systems in strict-feedback form. IEEE Trans Neural Netw Learn Syst 16:195–202. https://doi.org/10.1109/TNN.2004.839354
[25] Wang J, Wang J (2016) Forecasting energy market indices with recurrent neural networks: Case study of crude oil price fluctuations. Energy 102:365–374. https://doi.org/10.1016/j.energy.2016.02.098
[26] Wang S, Yu L, Lai KK (2005) Crude oil price forecasting with TEI@I methodology. J Syst Sci Complex 18:145–166.
[27] Wen F, Gong X, Cai X (2016) Forecasting the volatility of crude oil futures using HAR-type models with structural breaks. Energy Econ 59:400–413. https://doi.org/10.1016/j.eneco.2016.07.014
[28] Xie W, Yu L, Xu S, Wang S (2006) A new method for crude oil price forecasting based on support vector machines. Comput Sci ICCS 3994:444–451. https://doi.org/10.1007/11758549_63
[29] Ye M, Zyren J, Shore J (2006) Forecasting short-run crude oil price using high and low-inventory variables. Energy Policy 34:2736–2743. https://doi.org/10.1016/j.enpol.2005.03.017

7 AUTHORS

Jin Shang is a graduate student at the Graduate School of Economics, Kobe University, 2-1, Rokkodai, Nada-Ku, Kobe 657-8501, Japan (email: susanfeir@yahoo.co.jp).

Tamotsu Nakamura is a Professor at the Graduate School of Economics, Kobe University, 2-1, Rokkodai, Nada-Ku, Kobe 657-8501, Japan (email: nakamura@econ.kobe-u.ac.jp).

Shigeyuki Hamori is a Professor at the Graduate School of Economics, Kobe University, 2-1, Rokkodai, Nada-Ku, Kobe 657-8501, Japan (hamori@econ.kobe-u.ac.jp).