Journal of Quality Measurement and Analysis *JQMA* 21(2) 2025, 53-67
e-ISSN: 2600-8602 *http://www.ukm.my/jqma*
https://doi.org/10.17576/jqma.2102.2025.05

EXPLORING THE EFFECTIVE TECHNICAL INDICATORS OF LSTM FOR WTI PRICE IN DIFFERENT REGIMES

(*Meneroka Petunjuk Teknikal Berkesan LSTM untuk Harga WTI dalam Rejim Berbeza*)

ZHUQIN LIANG & MOHD TAHIR ISMAIL*

ABSTRACT

This study aims to explore the effectiveness of technical indicators in predicting WTI crude oil prices and investigate their applicability in different oil price regimes. By collecting monthly WTI data, we analyzed the closing prices using the Markov switching regression (MS-Regression) modeling approach. The study reveals that the market conditions can be divided into two distinct periods: low oil price regime and high oil price regime. For each of these periods, we employed LSTM regression models with the 30 least correlated technical indicators as features to predict the closing prices for the subsequent period. The MSE, RMSE, MAE, and MAPE were calculated for the test sets of both regimes. For Regime 1, the results were 43.409, 6.588, 4.745, and 10.299%, respectively. For Regime 2, the corresponding values were 120.872, 10.994, 8.521, and 9.371%, respectively. The results demonstrate that the models perform more accurately in predicting price fluctuations during the high oil price regime. Furthermore, through feature importance analysis, we identified the effectiveness of `SMA_20`, `MIDPOINT_14`, and `BBANDS_middle_20_2` indicators in both low and high oil price regimes. Additionally, `BBANDS_upper_20_2` and `CCI_14` exhibit better predictive capabilities during the low oil price regime, while `NATR_14` and `RSI_14` prove more effective during the high oil price regime. These findings contribute to a better understanding of the role of technical indicators in predicting WTI crude oil prices in different oil price regimes.

Keywords: WTI; MS-Regression; technical indicators; LSTM; feature importance

ABSTRAK

Kajian ini bertujuan untuk meneroka keberkesanan penunjuk teknikal dalam meramalkan harga minyak mentah WTI dan menyiasat kebolehgunaannya dalam rejim harga minyak yang berbeza. Dengan mengumpul data WTI bulanan, kami menganalisis harga penutup menggunakan pendekatan pemodelan regresi pertukaran Markov (MS-Regression). Kajian ini mendedahkan bahawa keadaan pasaran boleh dibahagikan kepada dua tempoh berbeza: rejim harga minyak rendah dan rejim harga minyak tinggi. Bagi setiap tempoh ini, kami menggunakan model regresi LSTM dengan 30 petunjuk teknikal yang paling kurang berkorelasi sebagai ciri untuk meramalkan harga penutup untuk tempoh berikutnya. MSE, RMSE, MAE, dan MAPE telah dikira untuk set ujian kedua-dua rejim. Bagi Rejim 1, keputusannya adalah 43.409, 6.588, 4.745, dan 10.299%, masing-masing. Manakala bagi Rejim 2, nilai yang sepadan ialah 120.872, 10.994, 8.521, dan 9.371%, masing-masing. Keputusan menunjukkan bahawa model berprestasi lebih tepat dalam meramal turun naik harga semasa rejim pada harga minyak yang tinggi. Tambahan pula, melalui analisis kepentingan ciri, kami mengenal pasti keberkesanan penunjuk `SMA_20`, `MIDPOINT_14` dan `BBANDS_upper_20_2` dalam kedua-dua rejim harga minyak rendah dan tinggi. Selain itu, `BBANDS_upper_20_2` dan `CCI_14` mempamerkan keupayaan ramalan yang lebih baik semasa rejim harga minyak rendah, manakala `NATR_14` dan `RSI_14` terbukti lebih berkesan semasa rejim harga minyak yang tinggi. Penemuan ini menyumbang kepada pemahaman yang lebih baik tentang peranan penunjuk teknikal dalam meramalkan harga minyak mentah WTI dalam rejim harga minyak yang berbeza.

Kata kunci: WTI; Regresi MS; penunjuk teknikal; LSTM; kepentingan ciri

1. Introduction

Crude oil, often dubbed "black gold," reigns as the lifeblood of the global energy landscape. Its significance is unparalleled, fueling industries, transportation, and economies worldwide. As the primary source of gasoline, diesel, and jet fuel, it powers vehicles, planes, and ships, enabling global commerce and travel. Additionally, crude oil serves as a cornerstone for various products, from plastics to pharmaceuticals. However, its dominance comes with challenges, including environmental concerns and geopolitical tensions over reserves and distribution. Nevertheless, the world continues to pivot around this vital resource, shaping economies, policies, and daily lives.

Crude oil plays a pivotal role in the global economy, with its price fluctuations impacting various sectors and influencing economic decision-making. From Mann and Sephton (2016), we can infer that the West Texas Intermediate (WTI) crude oil has emerged as a key benchmark due to its significance in the global oil market. However, the WTI exhibits diverse trends and price levels under different economic backgrounds and world dynamics. Understanding and predicting these variations is crucial for informed decision-making and risk management. Therefore, this study focuses on collecting WTI crude oil futures monthly data and employing advanced analytical techniques to explore the existence of different economic regimes and their implications on WTI crude oil prices.

Previous studies have primarily focused on enhancing prediction accuracy by improving model performance, with relatively limited attention given to the effectiveness of model inputs. Additionally, whether the effectiveness of inputs varies across different economic regimes remains an open question worthy of investigation. Proposing a method to evaluate the effectiveness of various inputs across different economic regimes would significantly improve future researchers' ability to enhance the accuracy of time series forecasting while effectively reducing computational costs.

The concept of economic regimes refers to distinct periods characterized by unique economic conditions, such as expansion, recession, or stability. These regimes can significantly influence the behavior and performance of various financial assets, including crude oil. Exploring the existence of multiple regimes in the WTI crude oil market is crucial for comprehending the underlying economic factors driving price fluctuations. Based on the fitting of oil price and other economic indicators using Markov switching regression as outlined in Phoong *et al.* (2020), it can be observed that oil price exhibits different behaviors across various regimes. Consequently, this study employs the Markov Switching Regression approach with the aim of differentiating the time series data into two distinct regimes. This differentiation enables a deeper understanding of the economic backgrounds associated with WTI crude oil prices.

To establish a robust predictive model for WTI crude oil prices under different regimes, this research integrates the use of technical indicators as features in conjunction with the powerful Long Short-Term Memory (LSTM) regression model, it can be seen in Kakade *et al.* (2023), Karasu and Altan (2022) and Lee *et al.* (2021). Technical indicators provide valuable insights into market trends, volatility, and potential price movements. By incorporating these indicators as features, the LSTM model can capture complex patterns and dependencies in the data, enabling more accurate price predictions. Moreover, by employing feature importance analysis, this study aims to identify the key technical indicators that exert a significant influence on WTI crude oil prices under each regime, thereby enhancing our understanding of the underlying dynamics of the market. The findings of this research can potentially contribute to better price forecasting and decision-making in the oil market, offering meaningful references for traders and policymakers alike.

2. Literature Review

In the research endeavor to forecast WTI prices, identifying suitable independent variables for inclusion in predictive models is crucial. However, selecting these variables necessitates employing diverse models to ascertain their effectiveness in price prediction. This particular study focuses on examining the effectiveness of technical indicators. Yin and Yang (2016) has already demonstrated, through Ordinary Least Squares (OLS) regression, that technical indicators, when utilized as independent variables, can significantly enhance the accuracy of crude oil price predictions. Beyond linear models, deep learning models leveraging technical indicators have also been applied in financial markets. Lee *et al.* (2021) explores the predictive capabilities of LSTM models constructed using technical indicators for the stock market, achieving high accuracy. Furthermore, Benedetto *et al.* (2020) employs an entropy-based approach to study the relationship between crude oil volatility and prices, revealing that periods of high volatility introduce uncertainty into crude oil prices. Consequently, this suggests that technical indicators used to compute volatility can impact WTI price predictions, and the importance of such indicators warrants further examination.

Regarding the choice of models for testing effectiveness, based on prior research, we prioritize the use of LSTM models. Yin and Wang (2022) investigated the chaotic features and prediction of WTI crude oil prices using chaos theory and AI technology, highlighting the influence of noise on price forecasting. The study concludes that the EMD-LR-CHAOS model is the most accurate among the eight models tested. Additionally, Wu *et al.* (2019) integrates the LSTM model with the Ensemble Empirical Mode Decomposition (EEMD) method for predicting WTI prices, achieving superior predictive performance.

Zhang and Zhang (2015) focused on the movements of Brent and WTI crude oil price returns before and after the financial crisis. To analyze these movements, the researchers used a Markov regime-switching model that incorporated dynamic autoregressive coefficients. By utilizing this approach, they identified three distinct periods (the "sharply downward period," the "slightly downward period," and the "sharply upward period"). Gong *et al.* (2021) also employed a five-variable Markov switching VAR model to analyze the impact of different oil shocks on oil prices and identify driving factors. The empirical results indicate that oil inventory and speculative demand have significant effects on oil price fluctuations, highlighting the complex dynamics of multiple factors influencing oil prices under different regime conditions. Qian *et al.* (2022) investigated the predictability of geopolitical risk (GPR) on oil market volatility using an autoregressive Markov-regime switching model. The results demonstrate that high GPR is associated with heightened fluctuations in the oil market, and GPR exhibits strong predictive power for oil price volatility, particularly during recessions and geopolitical risk threats.

Zhang *et al.* (2023) proposes a novel approach incorporating nonlinear components of predictors to enhance crude oil return predictability, showing significant improvement both in and out-of-sample. These nonlinear diffusion indices better explain crude oil supply and demand dynamics, surviving robustness checks and extensions. Moreover, LSTM can handle nonlinear data. LSTM is widely employed for oil price prediction, Vo *et al.* (2020) proposes an efficient model, BOP-BL, based on Bi-LSTM for accurately predicting the fluctuating Brent oil prices. Experimental results show that BOP-BL outperforms state-of-the-art models in forecasting oil prices. Moreover, technical indicators are also applied to the LSTM model, Karasu and Altan (2022) proposed a new crude oil price prediction model that incorporates LSTM, technical indicators, and the CHGSO technique. The model effectively deals with the chaotic and nonlinear dynamics of both WTI and Brent crude oil time series. The selection of features is optimized using the CHGSO algorithm based on the logistic chaotic map, resulting in accurate price estimation. Lu *et al.* (2021) introduced a new research framework for selecting core influence factors and forecasting crude oil prices. It utilizes various methods like GLMNET, spike-slab lasso, and BMA for factor selection, and LSTM for price forecasting. The empirical results show that the variable selection-LSTM method outperforms benchmark methods in both level and directional accuracy of forecasting. While LSTM models have been widely utilized across numerous studies, achieving commendable results, there is a notable gap in the exploration of their efficacy across varying economic conditions. This research aims to delve into the performance and determining factors of LSTM models within diverse economic environments, with a specific focus on their capability to predict crude oil prices. Our study is dedicated to thoroughly examining how different economic contexts influence the effectiveness of LSTM models, highlighting the importance of understanding these dynamics in enhancing predictive accuracy.

3. Methodology

3.1. *Data collection*

We gather the monthly open price, high price, low price, close price and volume data of West Texas Intermediate (WTI) from Investing.com `https://cn.investing.com/` to constitute our focal subjects of study. Our research utilizes monthly data for WTI crude oil prices, as we infer from Balcilar *et al.* (2015) that monthly oil price fluctuations exert a certain impact on stock prices. Consequently, we deem monthly oil price data to possess greater economic significance. As the WTI serves as a key indicator for identifying economic cycles, we sought data with a higher signal-to-noise ratio. By focusing on monthly data, we aimed to mitigate the influence of short-term fluctuations and noise, enabling a more precise capture of long-term trends in asset prices. Additionally, Baumeister *et al.* (2015) demonstrates that while high-frequency data can predict fluctuations in oil prices, it does not significantly impact the actual monthly oil prices. Furthermore, the utilization of high-frequency data can substantially reduce computational costs.

3.2. *MS-Regression*

Markov Switching Regression (MS-Regression) is a powerful statistical technique tailored for analyzing time series data characterized by heterogeneous statistical properties across different economic regimes. Unlike traditional regression models that assume constant parameters over time, MS-Regression acknowledges the dynamic nature of economic systems by incorporating the principles of Markov chain theory.

Molnár *et al.* (2023) and Uzoma and Florence (2016) both provide detailed explanations of the principles and implementation of the MS-Regression. In Molnár *et al.* (2023), the MS-Regression model is applied to GDP data to fit and identify recessionary and expansionary periods. Uzoma and Florence (2016) utilizes MS-Regression to study the structure of the Nigerian stock index prices and finds evidence of regime switching in stock market return series. Both studies implement the model using Python's statsmodels package.

Markov chain theory posits that the evolution of a system can be modeled as a sequence of states, where the probability of transitioning from one state to another depends solely on the current state. In the context of MS-Regression, these states represent distinct economic regimes, each with its own set of statistical characteristics such as mean, variance, and other parameters.

The core idea behind MS-Regression is to model the time series data as switching between these latent states, with the transitions governed by probabilistic factors. This allows the model to adapt to changes in the underlying economic environment, capturing the complex interactions and structural shifts that occur over time.

To implement MS-Regression, the model parameters are estimated using advanced statistical techniques such as maximum likelihood estimation. Additionally, the latent states of the Markov chain are identified through algorithms designed to infer the most likely sequence of states given the observed data.

3.3. *LSTM*

The LSTM regression algorithm, Long Short-Term Memory regression, is a sophisticated technique widely employed in the realm of time series analysis and prediction. Initially introduced by researchers such as Yao and Wang (2021) and Sunny *et al.* (2020), LSTM models are a subset of recurrent neural networks (RNNs), engineered with a specialized architecture to effectively capture and utilize long-range dependencies and intricate temporal patterns present in sequential data.

Unlike traditional feedforward neural networks, LSTM models are equipped with memory cells, which enable them to retain and selectively store information from past observations over extended periods. This unique capability empowers LSTM models to learn from historical data while mitigating the vanishing or exploding gradient problems often encountered in traditional RNNs, especially when dealing with long sequences.

In Chen *et al.* (2022), it is introduced that the LSTM network exhibits the characteristics of functioning as a model for dynamical systems, possessing the capability to dynamically correlate the inputs and outputs within such systems. By incorporating memory cells and gate mechanisms, LSTM models can effectively process and extract relevant features from historical observations, facilitating accurate predictions even when confronted with prolonged time lags and intricate temporal dependencies.

In our specific application, we utilize LSTM regression to predict future prices of WTI crude oil, a pivotal commodity in the global financial markets. To accomplish this, we curate a set of technical indicators that encapsulate various aspects of market behavior and price movements. By training the LSTM model on historical data encompassing these indicators, we aim to uncover the latent patterns and dynamics governing WTI crude oil prices.

Through this analysis, we seek to provide valuable insights into the underlying drivers of price movements in the crude oil market, thereby empowering stakeholders with enhanced forecasting capabilities and informed decision-making strategies. Ultimately, leveraging LSTM regression in time series forecasting endeavors holds the promise of unlocking new avenues for predictive analytics and optimizing trading performance in dynamic financial markets.

3.4. *Technical indicator*

In this study, we utilized the functions from the 'Ta-Lib' package in Python to calculate various technical indicators for monthly WTI crude oil data. 'TA-Lib' is a versatile open-source tool widely utilized in financial markets, it is introduced in Uras and Ortu (2021). With over 150 indicators and functions, it aids in analyzing historical market data to identify trends and patterns. Its features include trend analysis, momentum, and volatility indicators, making it invaluable for traders and analysts seeking to make informed decisions. The user-friendly interface and widespread adoption of 'TA-Lib' position it as a primary resource for individuals engaged in quantitative finance or algorithmic trading, elevating comprehension of market dynamics and refining trading strategies. Specifically, we computed the indicators in Table 1.

Table 1: Indicators

Indicator	**Definition**
ADX_14	14-period Average Directional Index, measuring the strength of a trend.
ADXR_14	14-period Average Directional Movement Rating, smoothing the ADX indicator.
AROONOSC_14	14-period Aroon Oscillator, measuring the difference between Aroon-Up and Aroon-Down.
ATR_14	14-period Average True Range, measuring market volatility.
ATR_20	20-period Average True Range.
BBANDS_lower_20_2	Lower Bollinger Band for 20 periods period with 2 standard deviations below the middle band.
BBANDS_middle_20_2	Middle Bollinger Band or 20-period Simple Moving Average.
BBANDS_upper_20_2	Upper Bollinger Band for 20 periods period with 2 standard deviations above the middle band.

Table 1 (Continued)

CCI_14	14-period Commodity Channel Index, identifying cyclical trends in a security's price.
CCI_20	20-period Commodity Channel Index, identifying cyclical trends in a security's price.
CMO_14	14-period Chande Momentum Oscillator, measuring momentum on a percentage basis.
DX_14	14-period Directional Movement Index, measuring the strength of a price movement.
EMA_10	Exponential Moving Average over the past 10 periods, giving more weight to recent prices.
EMA_20	20-period Exponential Moving Average.
EMA_50	50-period Exponential Moving Average.
HT_TRENDLINE	Hilbert Transform Trendline, indicating the long-term trend with a digital signal processing algorithm using 20-period data.
MACD	Moving Average Convergence Divergence, showing the relationship between two moving averages of a security's price.
MACDEXT	Moving Average Convergence Divergence Extension, similar to the MACD but allows for different settings for the moving averages.
MIDPOINT_14	14-period midpoint of the price.
MIDPRICE_14	14-period midprice, calculated as the sum of the highest high and the lowest low over a given period, divided by two.
MOM_10	10-period Momentum indicator which measures the rate of price change.
NATR_14	14-period Normalized Average True Range, providing the ATR expressed as a percentage of the close.
RSI_14	14-period Relative Strength Index, measuring the magnitude of recent price changes to evaluate overbought or oversold conditions.
SMA_20	20-period Simple Moving Average, calculating the average of a selected range of prices.
SMA_50	50-period Simple Moving Average.
STOCH_slow_d	Slow stochastic oscillator D-line, a moving average of the K-line.
STOCH_slow_k	Slow stochastic oscillator K-line, showing the position of the current price in relation to its high-low range over a set number of periods.
TRANGE	True Range, the greatest of the following: current high less the current low, the absolute value of the current high less the previous close, or the absolute value of the current low less the previous close.
TRIX_14	14-period Triple Exponential Moving Average, showing the percent rate of change in a triple exponentially smoothed moving average.
ULTOSC_7_14_28	Ultimate Oscillator that combines short, intermediate, and long-term moving averages into one number.
WILLR_14	14-period Williams %R, measuring the level of the close relative to the highest high for the period.
WMA_10	10-period Weighted Moving Average, putting more weight on recent data.

We computed the correlation coefficient matrix, in absolute values, for all the technical indicators mentioned above. Next, we selected the k features with the lowest correlation from the correlation coefficient matrix as the input variables for our LSTM model. This method assists in data analysis and machine learning tasks by selecting features with lower correlation to the target variable, thereby reducing redundancy and enhancing the model's effectiveness and interpretability.

3.5. *Features importance analysis*

Feature importance analysis is a technique in machine learning used to assess the contribution of individual features towards model performance. It aims to determine the impact and importance of each feature on the model's predictive capability. In this approach, the feature importance analysis involves iteratively excluding each feature from the input data (In this case, setting the corresponding feature values to 0 in the test set) and evaluating the model's performance under different feature configurations, this method is also introduced in Zheng *et al.* (2017). By comparing the model's predictive accuracy or loss metric between the modified and original feature sets, the relative importance of each feature can be determined. When defining the loss metric, we elected to utilize the Mean Squared Error (MSE). This choice was grounded in the fact that our model training employed MSE as the loss function, and our definition of feature importance was also predicated on this loss function. Additionally, MSE has been adopted as the loss function for LSTM-based oil price prediction in both Xiao *et al.* (2021) and Wu *et al.* (2019). Consequently, we referenced their methodologies in making our decision. The importance of a single feature is calculated as the ratio of the MSE obtained when the feature is excluded to the MSE obtained when all features are included.

3.6. *Hybrid model*

In our hybrid model, we first model the monthly WTI data using MS-regression, resulting in two regimes. Then, we model the data of each regime using LSTM, where the features are selected technical indicators of each period's data, and the target is the closing price data of the next period's WTI. Subsequently, we employ Feature Importance Analysis to analyze the significance of each feature in this well-trained model. The flowchart is shown in Figure 1.

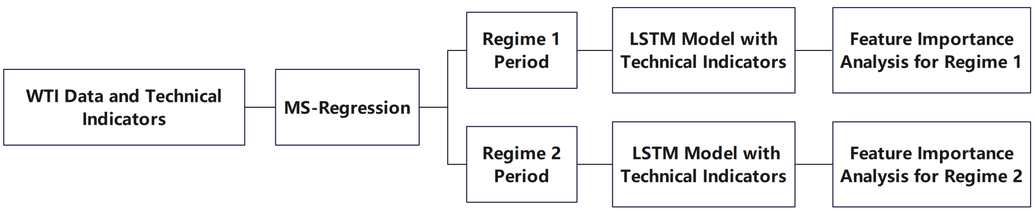

Figure 1: Flowchart of the hybrid model

4. Empirical Result and Analysis

4.1. *Data description*

We gathered a comprehensive dataset consisting of 493 monthly data for WTI, spanning from April 1, 1983, to April 1, 2024. Following this, we computed the technical indicators for the WTI. Subsequently, we selected a subset of data consisting of 256 monthly data, specifically from January 1, 2003, to April 1, 2024, encompassing approximately 21 years, to be used for modeling purposes. We computed the mean, variance, skewness, and kurtosis of this time series, resulting in a mean of 68.63, a variance of 535.69, a skewness of 0.24, and a kurtosis of -0.59. Moreover, the p-value of the Shapiro-Wilk test is 0.004, indicating that the returns data is unlikely to follow a normal distribution. The plot of the WTI closing price is shown in Figure 2.

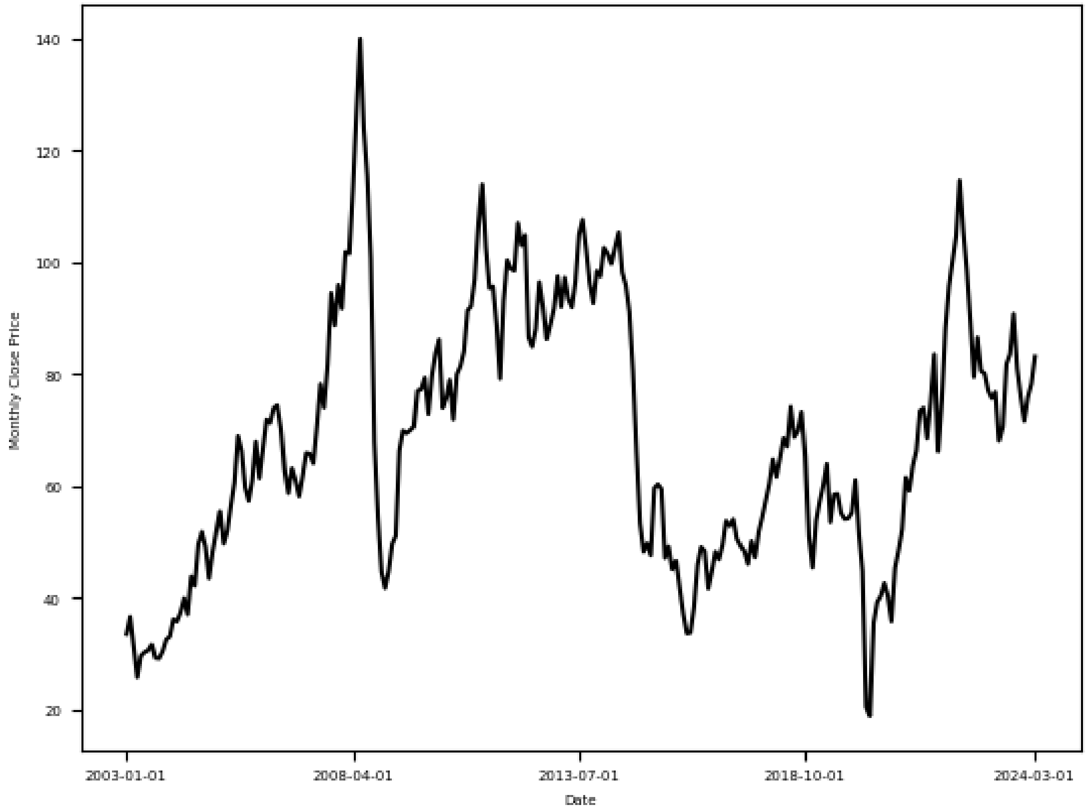

Figure 2: Line plot of WTI monthly close price

4.2. *Result of MS-Regression*

We fitted a time series of WTI close prices using the 'MarkovRegression' function from the 'statsmodels' library in Python, the implement process is also introduced in Das *et al.* (2022). In the selection of model hyperparameters, we explored alternative settings for the trend parameter and considered the possibility of distinct variances across different regimes. However, empirical findings revealed a lack of convergence in our data. Furthermore, given the monthly nature of our dataset, an excessive number of regimes led to insufficient periods within individual regimes. Ultimately, we opted for a model configuration incorporating a constant trend term and uniform variance across regimes, with the number of regimes set to two. The goodness-of-fit of the model is shown in Table 2.

Table 2: Model goodness-of-fit of MS-Regression

Metric	Value
Log-Likelihood	-1039.348
AIC	2088.696
BIC	2106.422
HQIC	2095.826

Parameters of the model:

$$\begin{aligned}
\textbf{Regime 1:}\quad & \mu_1 = 51.926 \quad (\text{Standard Error} = 1.170,\ \text{Z-value} = 44.382,\ \text{p-value} < 0.001) \\
\textbf{Regime 2:}\quad & \mu_2 = 90.327 \quad (\text{Standard Error} = 1.350,\ \text{Z-value} = 66.912,\ \text{p-value} < 0.001) \\
\textbf{Variance:}\quad & \sigma^2 = 170.135 \quad (\text{Standard Error} = 15.206,\ \text{Z-value} = 11.188,\ \text{p-value} < 0.001)
\end{aligned}$$

Regime Transition Probabilities:

$$\begin{aligned}
& p[1 \to 1] \quad \text{(Probability of staying in Regime 1)} \\
& p[2 \to 2] \quad \text{(Probability of staying in Regime 2)} \\
& p[1 \to 2] \quad \text{(Probability of transitioning from Regime 1 to Regime 2)} \\
& p[2 \to 1] \quad \text{(Probability of transitioning from Regime 2 to Regime 1)}
\end{aligned}$$

$$\begin{bmatrix} p[1 \to 1] & p[1 \to 2] \\ p[2 \to 1] & p[2 \to 2] \end{bmatrix} = \begin{bmatrix} 0.982 & 0.018 \\ 0.022 & 0.978 \end{bmatrix}$$

Based on the estimation results of the estimated parameters μ_1, μ_2, and σ, the result suggests that all of the estimated parameters are statistically significant, as evidenced by their low p-values and high Z-values. Therefore, we can have confidence in the accuracy of these parameter estimates within the context of the model. We can infer that WTI closing prices are primarily characterized by two distinct regimes: a low oil price regime and a high oil price regime. Furthermore, the transition probability from the low oil price regime to the high oil price regime is relatively low, indicating a reduced likelihood of transitioning from a period of low oil prices to high oil prices. Similarly, the probability of transitioning from the high oil price regime to the low oil price regime is also relatively low.

Then we get the smoothed marginal probabilities for regime 1 and regime 2 of the model, and they are shown in Figure 3. Based on the smoothed marginal probability results of the Markov Switching Regression Model described above, if, within this period, the smoothed marginal probability of regime 1 is higher than that of regime 2, we classify this period into regime 1. Conversely, it is classified into regime 2. We can identify the corresponding periods for the high oil price and low oil price, and it is shown in Table 3.

Over the past two decades, the WTI price has encountered numerous significant events in its historical trajectory. To elucidate why the WTI crude oil price remained at high (low) levels during these periods, we analyzed the corresponding economic backgrounds and historical events. Our objective was to provide supportive evidence for our model's findings and conclusions based on the prevailing oil price environment at that time. Table 4 is a brief background to explain the result of the model.

Table 3: WTI price regimes and periods

Regime	**Start Date**	**End Date**
Low Price	January 1, 2003	July 1, 2007
High Price	August 1, 2007	September 1, 2008
Low Price	October 1, 2008	September 1, 2009
High Price	October 1, 2009	October 1, 2014
Low Price	November 1, 2014	August 1, 2021
High Price	September 1, 2021	April 1, 2024

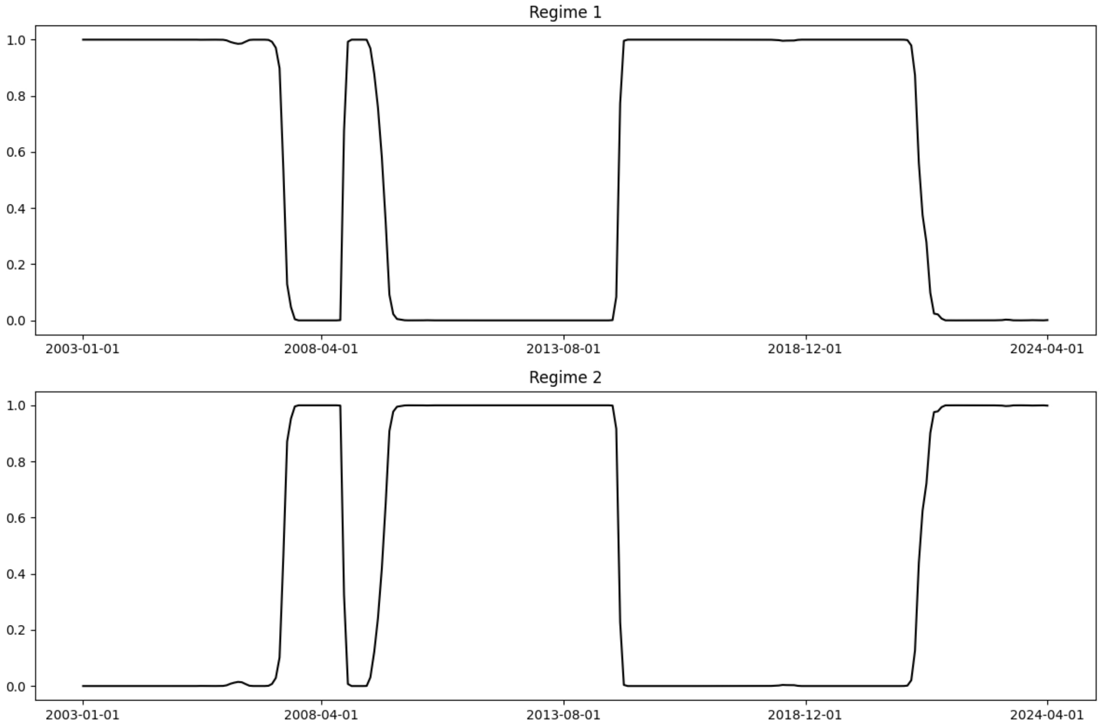

Figure 3: Smoothed marginal probabilities

4.3. *Result of LSTM*

Before training the LSTM model, we calculated the WTI crude oil technical indicators for each period following the method outlined in section 3.4. To avoid multicollinearity, we then eliminated indicators with high correlation, ultimately retaining 30 technical indicators for model training. Based on the Pearson correlation coefficients between indicators, we found that `EMA_10`, `EMA_20`, and `WMA_10` have relatively high correlation. The main reason is that these three indicators all calculate the weighted average price of WTI closing prices, and their calculation methods are quite similar. Therefore, it is relatively easy to obtain a high correlation among them, so we decided to remove them.

The decision to select 30 technical indicators was based on our experimental findings that using a smaller number of indicators led to models with severe overgeneralization issues, where the variance in predictions across different periods was minimal. It was only by adjusting the feature count to 30 that we could maintain the stability of the prediction outcomes. Subsequently, we divided the original dataset into two subsets based on regime 1 and regime 2. For each dataset, we randomly selected 80% of the data as the training set and the remaining 20% as the test set.

In constructing the LSTM model, we leverage the Python 'PyTorch' library. 'PyTorch' is a powerful open-source machine learning framework, renowned for its dynamic computation graph and seamless integration with Python. With its intuitive interface, it empowers researchers and developers to build and train neural networks effortlessly. PyTorch's flexibility enables rapid prototyping and experimentation, making it a favorite among the AI community. Its rich ecosystem provides modules for computer vision, natural language processing, and reinforcement learning, fostering innovation in diverse domains.

We specify the dimensions of input features, the size of the hidden layer, the dimensions of output features, and the number of LSTM layers to construct a network architecture capable of handling tasks such as time series prediction. The model processes input data comprising 30

Table 4: WTI price periods and background

Period	Background
2003-2007	Global oil production increased, particularly from non-OPEC countries such as Russia and the United States. There were no major conflicts or tensions in oil-producing regions during this time that could disrupt the global oil supply.
2007-2008	Global economic growth and increased demand from emerging markets led to a rise in oil demand. Concerns over supply disruptions were raised due to factors such as the Iran nuclear issue and tensions in the Middle East.
2008-2009	The financial crisis resulted in global economic recession and a decrease in demand, putting downward pressure on oil prices.
2009-2014	As the economy gradually recovered, especially in emerging markets, the oil demand increased, driving up prices.
2014-2021	Global oil supply was relatively abundant, especially with the rapid growth of shale oil production in the United States, leading to an oversupply. Global economic growth slowed down, resulting in decreased demand and downward pressure on oil prices.
2021-2024	The conflict between Russia and Ukraine resulted in a significant increase in energy prices. The conflict between the Palestinians and Israel has also led to a sharp rise in crude oil prices.

features (30 technical indicators), producing an output feature quantity of 1, which predicts the closing price of WTI.

We evaluated the selected parameters for Karasu and Altan (2022) and Li and Cao (2018), and ultimately decided to opt for these parameter choices. The size of the LSTM hidden layer is set to 100, with the number of LSTM layers set to 1. During the model's forward propagation process, we first initialize the hidden and cell states for the LSTM layer, then augment the input data with an additional time-step dimension to meet the input requirements of the LSTM layer. After the LSTM layer processes the sequential data, the output from the last time step is passed to a linear layer, which maps the LSTM output to the final prediction dimension.

Additionally, we configure the model's training iterations to 50 epochs, select the Mean Squared Error (MSE) as the loss function, and opt for the Adam Optimization Algorithm as the optimizer. Subsequently, we compute the model's predictive outcomes as illustrated in the Figure 4 and Table 5.

Figure 4 presents a comparison between the actual and predicted values for the two regimes. As the test set was randomly sampled, the line plots appear somewhat discontinuous. From the prediction errors of LSTM, it can be observed that the MSE, RMSE, and MAE of regime 2 are significantly higher compared to regime 1. This is mainly because regime 2 belongs to a period of high oil prices, where the volatility of oil prices is also relatively high. Consequently, the difficulty of accurately predicting numerical values increases, resulting in higher errors. However, we can observe from the MAPE that the MAPE of regime 2 is lower than that of regime 1, indicating that the model performs better in terms of predicting percentages during the high-price period of oil. Therefore, we infer that technical indicators are more effective in predicting price fluctuations during periods of high oil prices.

Table 5: Result of LSTM

	MSE	RMSE	MAE	MAPE
Regime 1	43.409	6.588	4.745	10.299%
Regime 2	120.872	10.994	8.521	9.371%

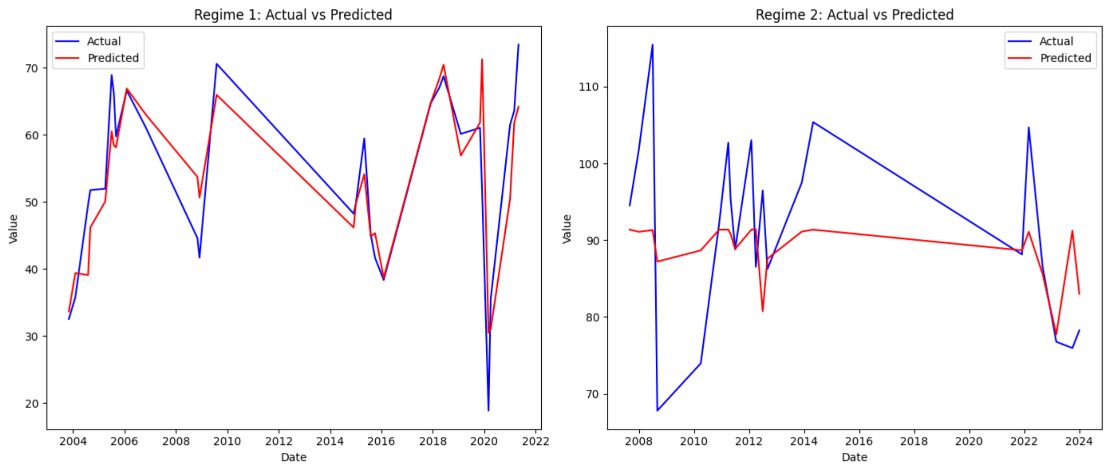

Figure 4: Comparison between actual and prediction

4.4. *Result of feature importance analysis*

By the approach outlined in section 3.5, we calculated the feature importance of the 30 technical indicators used for model training. The results are presented in the Table 6.

From this, we can see that during the period of low oil prices (regime 1), the most effective technical indicators are `BBANDS_upper_20_2` (3.707977), `CCI_14` (2.064058), `SMA_20` (2.085919), `MIDPOINT_14` (1.977182), and `BBANDS_middle_20_2` (1.869811). During the period of high oil prices (regime 2), the most effective technical indicators are `NATR_14` (1.218142), `MIDPOINT_14` (1.185067), `BBANDS_middle_20_2` (1.102351), `RSI_14` (1.092259), and `SMA_50` (1.085840). Therefore, we can conclude that `SMA_20`, `MIDPOINT_14`, and `BBANDS_middle_20_2` are effective indicators for predicting oil prices in any period, while `BBANDS_upper_20_2` and `CCI_14` are more suitable for the period of low oil prices, `NATR_14` and `RSI_14` are more suitable for the period of high price. Among the indicators, all of them are derived using data spanning 14 to 20 periods. Indicators calculated over 50 periods were not selected, allowing us to infer that the cycles influencing oil prices are predominantly of short to medium length. The indicators `SMA_20`, `MIDPOINT_14`, and `BBANDS_middle_20_2` are all technical indicators similar to the calculation of historical weighted average prices. Therefore, we can infer that regardless of whether it is a period of high oil prices or low oil prices, the future trend of oil prices is significantly influenced by the average prices of short to medium periods in the past. `BBANDS_upper_20_2` represents the upper pressure line of the Bollinger Bands, typically defined as the pressure level where an upward trend reverses. When the price reaches `BBANDS_upper_20_2`, the WTI price has significantly deviated from the average. `CCI_14` measures the deviation of prices from the average over the past 14 periods. Therefore, we can infer that when oil prices are low, the future prices of oil are more influenced by the deviation of short period and medium period prices from the average. `NATR_14` is a standardized indicator derived from considering the difference between the highest and lowest prices within each period, thus calculating the average true range. `RSI_14` is an indicator that measures the velocity and magnitude of price changes based on price fluctuations over a certain period. Both indicators share the commonality of considering historical price volatility over the past 14 periods. Therefore, we believe that during periods of high oil prices, the future trend of oil prices is likely to be influenced by the magnitude of price fluctuations in the short period and the medium period.

Table 6: Feature importance

Feature	Regime 1	Regime 2
ADX_14	0.938	1.034
ADXR_14	0.875	1.020
AROONOSC_14	0.928	0.999
ATR_14	1.531	1.005
ATR_20	1.034	0.997
BBANDS_lower_20_2	1.311	0.988
BBANDS_middle_20_2	1.870	1.102
BBANDS_upper_20_2	3.708	1.052
CCI_14	2.064	1.041
CCI_20	1.118	1.006
CMO_14	1.531	1.005
DX_14	0.950	1.073
EMA_50	1.472	1.029
HT_TRENDLINE	1.388	1.055
MACD	1.103	1.043
MACDEXT	1.087	1.034
MIDPOINT_14	1.977	1.185
MIDPRICE_14	1.579	1.036
MOM_10	1.115	1.002
NATR_14	0.989	1.218
RSI_14	0.954	1.092
ROC_12	1.349	0.985
SMA_20	2.086	0.978
SMA_50	1.007	1.086
STOCH_slow_d	0.945	1.012
STOCH_slow_k	0.933	1.035
TRIX_14	1.012	1.015
TRANGE	1.010	1.002
ULTOSC_7_14_28	0.872	0.989
WILLR_14	1.625	1.079

5. Conclusion

In this study, we delved into the intricate dynamics of crude oil, a vital energy source globally, and elucidated the pivotal strategic significance of understanding the trends in crude oil prices. Our research aimed to explore the primary factors influencing oil prices, acknowledging that these influences may vary across different economic cycles and historical contexts. To achieve this, we employed a two-step modeling approach, utilizing MS-Regression followed by LSTM regression, and incorporated a selection of common technical indicators as features.

Initially, we applied MS-Regression to segment historical market conditions into two distinct regimes: a low oil price regime and a high oil price regime. Our analysis revealed that this segmentation facilitated a more reasonable convergence of the models. Subsequently, LSTM regression models were developed separately for each regime, leveraging technical indicators to predict future closing prices. Notably, our findings indicated superior performance in predicting price fluctuations during periods of high oil prices.

During the training of LSTM models, we observed that a set of at least 30 technical indicators was necessary to achieve satisfactory performance, as a smaller set tended to result in overgeneralization on the test dataset. Consequently, we retained a substantial number of technically correlated indicators in our analysis. Furthermore, through feature importance analysis, we identified key indicators for predicting WTI closing prices in different oil price regimes.

Our results demonstrated that indicators such as `SMA_20`, `MIDPOINT_14`, and `BBANDS_middle_20_2` were effective across all periods, while others, like `BBANDS_upper_20_2` and `CCI_14`, showed greater efficacy in low oil price periods, and `NATR_14` and `RSI_14` were more suitable for high-price periods. This suggests that past short and medium-term price movements influence oil price trends, with deviations from moving averages exerting more influence during low oil price periods and volatility having a stronger impact during high oil price periods.

In the upcoming research, we will delve into incorporating additional technical indicators and monthly economic data (such as GDP, the US dollar index, and prices of other commodities) as time series features for LSTM modeling. Our objective is to explore whether we can identify more significant factors influencing WTI crude oil prices. Alternatively, we will consider alternative regression algorithms. In the realm of deep learning, we are aware of alternatives to LSTM, such as LSTM + CNN or Transformer models. Should these alternative models demonstrate superior predictive performance, they might unveil more suitable factors affecting WTI prices.

This study offers practical implications for key stakeholders by enhancing decision-making and strategic planning. For policymakers, the model provides valuable insights for designing energy policies, managing economic risks, and addressing market volatility. Traders can leverage the identified technical indicators to refine their strategies, anticipate market trends, optimize portfolios, and mitigate risks. Analysts can use the model to improve forecasting accuracy, enhance investment reports, and identify market inefficiencies. Furthermore, the findings contribute to a broader understanding of macroeconomic stability, highlighting the role of reliable oil price predictions in influencing inflation, exchange rates, and economic growth. These practical applications underscore the study's relevance and potential impact across various domains.

References

Balcilar M., Gupta R. & Miller S.M. 2015. Regime switching model of US crude oil and stock market prices: 1859 to 2013. *Energy Economics* **49**: 317–327.

Baumeister C., Guérin P. & Kilian L. 2015. Do high-frequency financial data help forecast oil prices? The MIDAS touch at work. *International Journal of Forecasting* **31**(2): 238–252.

Benedetto F., Mastroeni L., Quaresima G. & Vellucci P. 2020. Does OVX affect WTI and Brent oil spot variance? Evidence from an entropy analysis. *Energy Economics* **89**: 104815.

Chen R., Jin X., Laima S., Huang Y. & Li H. 2022. Intelligent modeling of nonlinear dynamical systems by machine learning. *International Journal of Non-Linear Mechanics* **142**: 103984.

Das S.R., Ostrov D.N., Casanova A., Radhakrishnan A. & Srivastav D. 2022. Optimal goals-based investment strategies for switching between bull and bear markets. *The Journal of Wealth Management* **24**(4): 8–36.

Gong X., Guan K., Chen L., Liu T. & Fu C. 2021. What drives oil prices?—A Markov switching VAR approach. *Resources Policy* **74**: 102316.

Kakade K.A., Ghate K.S., Jaiswal R.K. & Jaiswal R. 2023. A novel approach to forecast crude oil prices using machine learning and technical indicators. *Journal of Advances in Information Technology* **14**(2): 302–310.

Karasu S. & Altan A. 2022. Crude oil time series prediction model based on LSTM network with chaotic Henry gas solubility optimization. *Energy* **242**: 122964.

Lee M.C., Chang J.W., Hung J.C. & Chen B.L. 2021. Exploring the effectiveness of deep neural networks with technical analysis applied to stock market prediction. *Computer Science and Information Systems* **18**(2): 401–418.

Li Y. & Cao H. 2018. Prediction for tourism flow based on LSTM neural network. *Procedia Computer Science* **129**: 277–283.

Lu Q., Sun S., Duan H. & Wang S. 2021. Analysis and forecasting of crude oil price based on the variable selection-LSTM integrated model. *Energy Informatics* **4**(Suppl 2): 47.

Mann J. & Sephton P. 2016. Global relationships across crude oil benchmarks. *Journal of Commodity Markets* **2**(1): 1–5.

Molnár A., Vasa L. & Csiszárik-Kocsír Á. 2023. Detecting business cycles for Hungarian leading and coincident indicators with a Markov switching dynamic model to improve sustainability in economic growth. *Decision Making: Applications in Management and Engineering* **6**(1): 744–773.

Phoong S.W., Phoong S.Y. & Phoong K.H. 2020. Analysis of structural changes in financial datasets using the breakpoint test and the Markov switching model. *Symmetry* **12**(3): 401.

Qian L., Zeng Q. & Li T. 2022. Geopolitical risk and oil price volatility: Evidence from Markov-switching model. *International Review of Economics & Finance* **81**: 29–38.

Sunny M.A.I., Maswood M.M.S. & Alharbi A.G. 2020. Deep learning-based stock price prediction using LSTM and bi-directional LSTM model. *2020 2nd Novel Intelligent and Leading Emerging Sciences Conference (NILES)*, pp. 87–92.

Uras N. & Ortu M. 2021. Investigation of blockchain cryptocurrencies' price movements through deep learning: A comparative analysis. *2021 IEEE International Conference on Software Analysis, Evolution and Reengineering (SANER)*, pp. 715–722.

Uzoma U.E. & Florence A.U. 2016. Application of Markov-switching regression model on economic variables. *Journal of Statistical and Econometric Methods* **5**(2): 17–30.

Vo A.H., Nguyen T. & Le T. 2020. Brent oil price prediction using bi-LSTM network. *Intelligent Automation & Soft Computing* **26**(6): 1307–1317.

Wu Y.X., Wu Q.B. & Zhu J.Q. 2019. Improved EEMD-based crude oil price forecasting using LSTM networks. *Physica A: Statistical Mechanics and its Applications* **516**: 114–124.

Xiao W., Xu C., Liu H. & Liu X. 2021. A hybrid LSTM-based ensemble learning approach for China coastal bulk coal freight index prediction. *Journal of Advanced Transportation* **2021**: 5573650.

Yao T. & Wang Z. 2021. Crude oil price prediction based on LSTM network and GM (1, 1) model. *Grey Systems: Theory and Application* **11**(1): 80–94.

Yin L. & Yang Q. 2016. Predicting the oil prices: Do technical indicators help? *Energy Economics* **56**: 338–350.

Yin T. & Wang Y. 2022. Predicting the price of WTI crude oil futures using artificial intelligence model with chaos. *Fuel* **316**: 122523.

Zhang Y., He M., Wen D. & Wang Y. 2023. Forecasting crude oil price returns: Can nonlinearity help? *Energy* **262**: 125589.

Zhang Y.J. & Zhang L. 2015. Interpreting the crude oil price movements: Evidence from the Markov regime switching model. *Applied Energy* **143**: 96–109.

Zheng H., Yuan J. & Chen L. 2017. Short-term load forecasting using EMD-LSTM neural networks with a Xgboost algorithm for feature importance evaluation. *Energies* **10**(8): 1168.

School of Mathematical Sciences
Universiti Sains Malaysia
11800 Penang
MALAYSIA
*E-mail: liangzhuqin@student.usm.my, m.tahir@usm.my**

Received: 2 December 2024
Accepted: 26 February 2025

*Corresponding author